DRUG DISCOVERY AND STRUCTURAL STUDIES ON *Mycobacterium*

tuberculosis PROTEINS RELATED TO DRUG RESISTANCE AND PERSISTENCE

A Dissertation

by

GULCIN GULTEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,   James C. Sacchettini
Committee Members,   David H. Russell
                        Marcetta Y. Darensbourg
                        Margaret E. Glasner
Head of Department,   David H. Russell

May 2014

Major Subject: Chemistry

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis (TB), which is one of the leading infectious diseases worldwide. The current therapy for drug-sensitive TB is complex and lasts for at least six months. Improper use of antibiotics during this regimen has led to the emergence of drug resistance, which represents a grave threat to human health. This problem is further exacerbated by the ability of the bacterium to persist in the host in a non-replicating state despite the use of antibiotics. The majority of antibiotics currently used to treat tuberculosis only affect replicating bacteria. Therefore, it is critical to develop new antitubercular drugs that can shorten the current therapy while maintaining activity against persistent bacteria as well as the drug-resistant strains. In this dissertation, structural and drug discovery studies on *Mtb* proteins related to drug resistance and persistence are presented.

InhA, the enoyl-ACP-reductase enzyme of the mycolic acid biosynthesis pathway, is the molecular target of the antitubercular prodrugs isoniazid and ethionamide, and it is one of the best validated targets for *Mtb* drug discovery. A target-based high throughput screening and a structure-based drug design were performed to identify potent activation-free InhA inhibitors that were effective against drug-resistant *Mtb* strains. The molecular basis of InhA inhibition by these inhibitors was revealed by X-ray crystallography. In addition, the mode of action for ethionamide was revealed by using a cell-based activation system and X-ray crystallography. Furthermore, it was demonstrated that InhA is the clinically relevant primary target of isoniazid. The

regulation of InhA function was also studied, which revealed that phosphorylation of InhA occurs at its C-terminal. Phosphomimetic mutants showed that phosphorylation decreases InhA activity by decreasing the affinity toward cofactor NADH. The results of these studies are presented in Chapters II, III, IV, and V.

It is essential to understand the physiology of the bacterium to target the persistent state of *Mtb* infection. In Chapter VI, I report our studies on CarD, an essential *Mtb* transcription regulator that is required for persistent infection. The structure of the CarD/RNAP complex was determined by X-ray crystallography and the CarD-DNA interactions were investigated.

DEDICATION

This dissertation is dedicated to my family, Ayse, Kadir and Gokhan Gulten, who have always been a great support and source of encouragement for me all these years. Thank you for your constant support, unconditional love, and patience. I consider myself very fortunate to have you as family.

ACKNOWLEDGEMENTS

whom I learned a lot. Especially, I would like to thank my mentor, Dr. Feng Wang, for his time and for teaching me the necessary techniques to pursue my research, and Romano Miojevic for his help in work and non-work related issues.

Finally, I would like to thank the Turkish community in College Station and my dear roommates and friends for providing family away from family. I appreciate the support they provided throughout these years. Last but not the least; the greatest appreciation goes to my dearest parents, brother, and close relatives, without whom I would not be where I am now. I feel the deepest gratitude to God for making me who I am and giving me all the beautiful things in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION: FURTHERING OUR UNDERSTANDING OF DRUG

SUSCEPTIBILITY, DRUG RESISTANCE AND PERSISTENCE OF *Mycobacterium*

tuberculosis

Background: ***Mycobacterium tuberculosis* and TB**

Tuberculosis and* Mtb *pathogenesis

Tuberculosis (TB) is an infectious disease that is as old as the history of humanity (Wirth et al., 2008). In humans, TB is caused mainly by the airborne pathogen *Mycobacterium tuberculosis* (*Mtb*). The presence of the *Mtb* DNA in Egyptian mummies dating back to 2000 B.C. indicates that humankind has been in constant battle with this pathogen since ancient times (Zink et al., 2003).

TB is transmitted by inhalation of aerosol droplets containing the active *Mtb* that are directly expectorated from an individual with the acute disease. The pulmonary TB infection develops in lungs, where the bacteria are recognized and phagocytosed by the alveolar macrophages. These macrophages produce pro-inflammatory cytokines and chemokines, such as tumor necrosis factor (TNF)-α and interleukin-1 beta (IL-1β), to recruit more leukocytes to the infection site (Sakamoto, 2012). Neutrophils and monocytes phagocytose more bacteria together with the dendritic cells, which migrate to the lymph nodes to present mycobacterial antigens to lymphocytes and to activate the specific T cells (Silva Miranda et al., 2012). After phagocytosis, phagosomal maturation

1

is achieved by activation of macrophages with interferon-gamma (INF-γ) that is released by the T cells (Fenton et al., 1997). This results in the production of reactive-nitrogen (RNI) and reactive-oxygen (ROI) intermediates, acidification of the phagosome, and fusion of the phagosome with the lysosome (Schluger and Rom, 1998). However, *Mtb* can arrest the phagosome at an early stage of maturation and prevent acidification and the phagolysosome formation through various mechanisms such as blocking the proton ATPase pumps on the phagosomal membrane (Kaufmann, 2001; Stokes and Waddell, 2009). The pH of the unmatured mycobacterial phagosomes remains around 6.2, whereas the pH of the phagolysosomes remains around 5.5 or lower (Russell, 2011). The ability of *Mtb* to regulate the fusion of the phagosome with the lysosome is also seen in other pathogenic *mycobacterium* species such as *M. bovis*, *M. leprae* and *M. avium* (Russell, 2005).

Macrophages and the other immune cells gathered together through cytokine and chemokine signaling form the granuloma, which is the characteristic pathological signature of the tuberculosis disease. Inside the granuloma, macrophages that differentiate to epithelioid or foamy macrophages fuse to form giant, multinucleated cells. They are also surrounded by lymphocytes and fibroblasts (Silva Miranda et al., 2012). The mature granuloma represents an equilibrium between the virulent mycobacteria and the host immune response (Bold and Ernst, 2009). Previously, it was thought that granuloma formation was only beneficial to the host by containing and restricting the mycobacteria at the site of infection; however, studies performed with a close relative of *Mtb*, *Mycobacterium marinum*, revealed that mycobacteria exploit the

2

formation of granuloma during the early stages in order to spread the infection by recruiting and infecting the newly arriving macrophages (Davis and Ramakrishnan, 2009). The granuloma also protects *Mtb* from cytokine producing lymphocytes and provides a lipid rich medium for mycobacterial survival (Silva Miranda et al., 2012). Over time, the center of the granuloma undergoes necrosis (cell death), resulting in a caseous appearance. The bacteria in the caseous center of the granuloma are thought to be in a non-replicating persistent state, but are not completely killed (Trucksis, 2000). The caseating granuloma releases the bacteria into the extracellular environment where they can multiply again. Partially activated macrophages at the periphery of the granuloma and the non-activated macrophages throughout the body can phagocytose the released bacilli; however, *Mtb* can still proliferate in these cells (Trucksis, 2000). If the immune system fails to keep the infection under control with activated macrophages, then the tissue damaging immune response will dominate and the disease progresses to liquefaction and cavity formation (Rodriguez-Herrera and Jordan-Salivia, 1999; Trucksis, 2000). When this happens, the active *Mtb* multiplies extensively, spreading within the lung and through the respiratory tract allowing it to spread to new individuals (Lyadova, 2012).

One factor that makes *Mtb* such a successful pathogen is the fact that even the fully activated mature macrophages cannot eradicate 100% of the resident *Mtb*. Persistent bacteria proceed to a dormant state with reduced metabolic activity and adapt to the nutrient and oxygen poor environment inside macrophages. *Mtb* can live in the host without causing an active disease for decades in this non-replicating chronic state

3

(called latent-infection), which is held in control by the immune response. Latent disease is characterized by a positive tuberculin skin test, but displays no radiological, clinical, or microbiological evidence for the active disease (Trucksis, 2000). In fact, only around 10% of the *Mtb* infected individuals develop an active disease, and approximately 90% of the immunocompetent individuals infected with *Mtb* develop the latent disease. In general, immunocompromised individuals such as newborns and people with human-immunodeficiency-virus (HIV) develop the active disease upon primary infection. This is thought to correlate with a high bacterial load, increased virulence, or one's immunodeficiency (Hopewell and Jasmer, 2005). However, any deficiency in the immunity can cause reactivation of the latent disease (Trucksis, 2000). In many cases, the initial infection is controlled by the host immune response and the active disease develops later as a result of reactivation (Kaufmann, 2001). Aside from the lungs, other areas such as the central nervous system, internal organs, and the genitourinary tract can also be infected by *Mtb* that is known as extra-pulmonary TB. Indeed, around 15% of reactivated latent TB cases occur at extra-pulmonary sites (Sakamoto, 2012).

TB treatment

The tubercle bacillus was discovered in 1882 by Robert Koch (Koch, 1982). Before the introduction of anti-tubercular drugs in the 1940's, the main treatment of tuberculosis involved isolating patients into sanatoriums where they could receive plenty of fresh air, sleep, wholesome food, and exercise. The Bacille Calmette Guerin (BCG) vaccine, which was developed from an attenuated bovine tuberculosis strain in 1921,

4

exhibited a protective effect against TB in young children, but failed to prevent primary infection with adult pulmonary TB and reactivation of the latent disease. Also, the efficacy of the vaccine varies from strain to strain and the culture medium, making it an ineffective TB vaccine (McShane et al., 2012; Venkataswamy et al., 2012). However, the BCG vaccine is still administered to infants at birth in regions where TB is endemic (Zumla et al., 2013b). Despite constant efforts and research, to date, there is still no approved vaccine that exhibits a preventive effect against adult pulmonary TB (Koff et al., 2013).

Many anti-tubercular drugs have been developed to help mitigate TB. The first anti-tubercular drug was streptomycin (STM). It was isolated from the bacterium *Streptomyces griseus* in 1944 and then immediately used in clinical trials. In 1946, para-aminosalicylic acid (PAS) was found to be effective against TB. The frequent emergence of STM resistance led to the development of multi-drug therapies. PAS and STM combination therapy demonstrated better potency and less drug resistance (Mitchison and Davies, 2012). In the field of tuberculosis research, the landmark discovery was that of the anti-tubercular drug isoniazid (INH), which has been in use as a first-line drug since 1952. The discovery of INH was based on the observation that nicotinamide exhibited activity against tubercle bacillus (Chorine, 1945). Soon after, other nicotinamide based drugs such as pyrazinamide (PZA), ethionamide (ETH), and prothionamide (PTH) were synthesized. In the following years, ethambutol (EMB) was discovered. Screening for antibiotics from soil bacteria led to the discovery of many anti-TB drugs including cycloserine, kanamycin (KAN), amikacin (AMK),

capreomycin, and finally, rifampicin (RIF), which were included in multi-drug regimens and are still in use today (Zhang, 2005) **(Table 1-1** and **Figure 1-1)**.

Since the 1950s, a tremendous amount of research has been done on TB to understand, diagnose, and eliminate this ancient disease. Specifically, in the last 10-15 years a number of promising compounds exhibited potency against *Mtb* in both *in-vitro* and *in-vivo* models (Mitchison and Davies, 2012). However, the current TB treatment against the acute infection still depends on the antibacterials discovered 50 years ago. The current TB regimen against drug-sensitive *Mtb* consists of two-phases and includes the use of four first-line anti-TB drugs (INH, EMB, PZA, and RIF) for a period of two months, followed by a combination of INH and RIF for an additional four months. This regimen is administered for pulmonary and most extra-pulmonary TB regardless of the HIV coinfection, and can achieve cure rates of >95% if implemented under the DOTS (directly observed treatment, short-course) strategy (Zumla et al., 2013a). INH has a high early bactericidal activity, which causes a significant decrease in the colony-forming units in the first few weeks of the therapy. While the actively replicating bacteria can be efficiently killed during the initial phase of the treatment, the therapy is followed for six months in order to eradicate the persistent *Mtb* (Zumla et al., 2013a). However, the long-duration of treatment brings issues such as patient adherence, drug toxicity, intolerance, and pharmacokinetic drug-drug interactions, especially with the HIV drugs for the HIV-coinfected individuals.

Table 1-1. TB drugs and their molecular targets.

Drug	Year introduced	Clinical use	Molecular Target / Effect
STM	1944	No	S12 and 16S rRNA of 30S ribosomal subunit / Inhibits protein synthesis
PAS	1946	Limited use for XDR-TB	Dihydrofolate reductase / Inhibits folate biosynthesis
INH	1952	1st line	Enoyl-acyl-carrier-protein reductase (InhA) / Inhibits mycolic acid synthesis
PZA	1954	1st line	S1 component of 30S ribosomal subunit / Inhibits translation and *trans*-translation, acidifies cytoplasm
EMB	1961	1st line	Arabinosyl transferases / Inhibits arabinogalactan biosynthesis
RIF	1963	1st line	RNA polymerase, beta subunit / Inhibits transcription
AMK	1972	2nd line	30S ribosomal subunit / Inhibits protein synthesis
KAN	1957	2nd line	30S ribosomal subunit / Inhibits protein synthesis
Capreomycin	1963	2nd line	Ribosome / Inhibits protein synthesis
CIP	1981	2nd line	DNA gyrase / Inhibits DNA supercoiling
MFX	1996	2nd line	DNA gyrase / Inhibits DNA supercoiling
ETH	1961	2nd line	Enoyl-acyl-carrier-protein reductase (InhA) / Inhibits mycolic acid synthesis
Cycloserine	1955	2nd line	D-alanine racemase and ligase / Inhibits peptidoglycan synthesis
CFZ	1950s	3rd line	genomic DNA, potassium transporters / Proposed to intercalate to DNA
LZD	1990s	3rd line	50S subunit of ribosome / Inhibits protein synthesis
AMX/CLV	1978	3rd line	DD-carboxypeptidase and β-lactamase / Inhibits peptidoglycan biosynthesis
PA-824 nitroimidazole	2000s	Phase II trials	ND /Inhibition of cell wall, mycolic acid synthesis, NO poisoning of cytochrome *c* oxidase
SQ109	2003	Phase II trials	MmpL3 / Inhibits mycolic acid biogenesis
Bedaquiline	2004	MDR-TB	ATP synthase / Decreases intracellular ATP levels

Isonizid (INH)

Pyrazinamide (PZA)

Ethambutol (EMB)

Rifampicin (RIF)

Ethionamide (ETH)

Amikacin (AMK)

Kanamycin (KAN)

Ciprofloxacin (CIP)

Moxifloxacin (MFX)

p-amino salicylic acid (PAS)

Streptomycin (STM)

Capreomycin
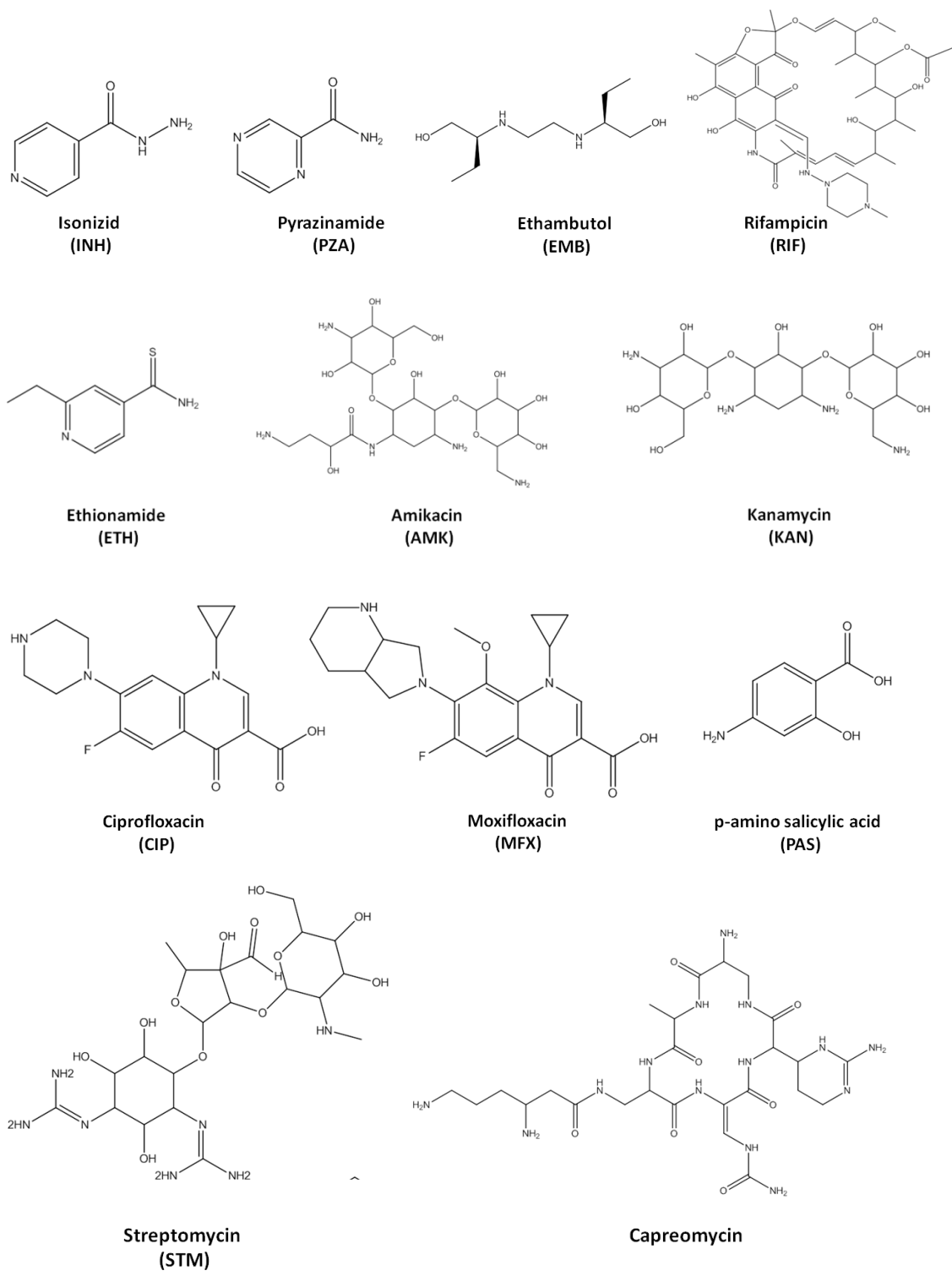
Figure 1-1. Chemical structures of the TB drugs mentioned in the text.

Clofazimine
(CFZ)

Bedaquiline

Cycloserine

Linezolid
(LZD)

Amoxicillin plus Clavulanate (AMX/CLV)

SQ109

PA-824 nitroimidazole
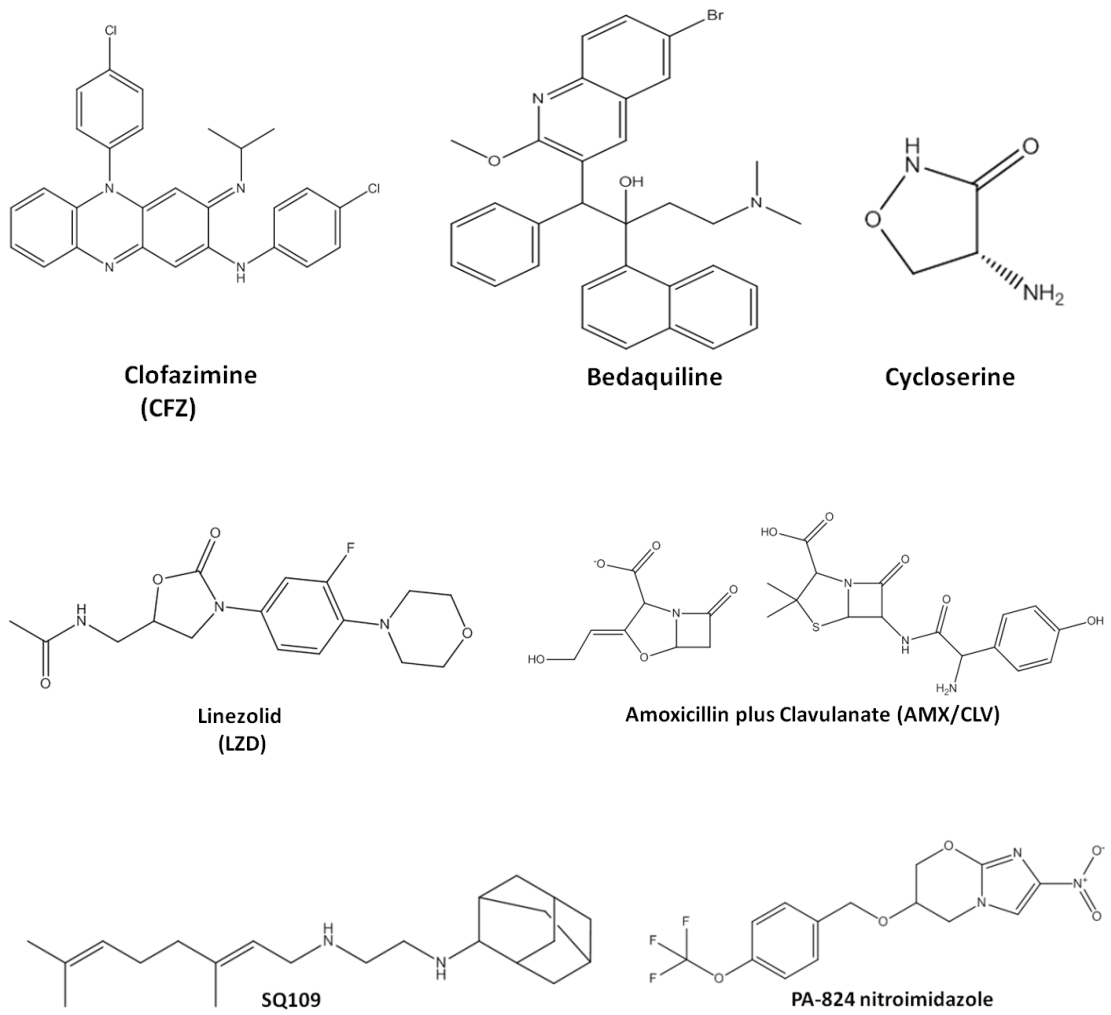
Figure 1-1 Continued.

This results in interruptions in the treatment and presents challenges to the current therapy for the drug-susceptible TB (Zumla et al., 2013a). Therefore, investment has been made in significant drug-discovery and development efforts that can shorten the therapy, with fewer drug-drug interactions that are more tolerable.

Problem of drug resistance and persistence

The improper use of antibiotics, together, with the incompletion of the therapy for drug-susceptible TB leads to the emergence of drug resistant (Multi-drug resistant (MDR) and extensively drug resistant (XDR)) tuberculosis. Furthermore, an individual who develops MDR-TB can transmit this form of the disease to others (Zumla et al., 2013b). MDR-TB strains are resistant to two or more drugs, usually to INH and RIF, and treatment of MDR-TB is administered for at least two years with second-line antibiotics such as amikacin (AMK), kanamycin (KAN), capreomycin, ciprofloxacin (CIP), moxifloxacin (MFX), ethionamide (ETH), and cycloserine **(Table 1-1)**. MDR-TB regimens include at least four second-line drugs administered under the DOTS strategy, which are determined based on factors such as previous treatment of the patient for MDR-TB, the presence of any underlying medical condition, drugs that were administered earlier to the patient, the adverse effects of the drugs, and the local resistance pattern of the disease (Zumla et al., 2013a; Zumla et al., 2013b).

XDR-TB are resistant to fluoroquinolones and at least one of the injectible second-line drugs: AMK, KAN, and capreomycin as well as INH and RIF (Campbell et al., 2011). Treatment of the XDR-TB is longer than the MDR-TB treatment and requires

the use of third-line antibiotics such as clofazimine (CFZ), linezolid (LZD), and amoxicillin plus clavulanate (AMX/CLV), which lack accessibility and are usually more toxic (Zumla et al., 2013a). While the MDR-TB cure rate is approximately 50% (Zhang, 2005), the mortality rate for XDR-TB is close to 100% for patients co-infected with HIV (Zumla et al., 2013b). About half a million new MDR-TB cases were detected in 2011, and about one-tenth of those cases exhibited XDR-TB (Abubakar et al., 2013). Most of the second- and third-line drugs are toxic, they have severe side effects and treatment of MDR- and XDR-TB is very costly compared to the standard regimen (Campbell et al., 2011).

The acquired drug resistance of *Mtb* occurs through chromosomal mutations (Ford et al., 2011). These mutations usually happen in either the target or the activator gene for the drug (Sacchettini et al., 2008). Mutations are also observed in the promoter or intergenic regions, altering the gene expression levels of proteins related to the drug mechanism or the efflux pump systems (Calgin et al., 2013; Muller et al., 2011). For example, INH and ETH are both prodrugs that need to be activated in the cell, and INH resistant (INH[R]) and ETH resistant (ETH[R]) strains mostly carry mutations in the activator gene of these prodrugs, *katG* and *ethA*, respectively. Typically, 20-80% of the INH[R] clinical isolates carry mutations in the *katG* gene. The most common *katG* mutation is the Ser315Thr mutation, which causes a deficiency in INH activation (Campbell et al., 2011; Suarez et al., 2009; Zhao et al., 2006). Similarly, 51% of ETH[R] isolates carry mutations in the *ethA* gene (Brossier et al., 2011; Morlock et al., 2003). On the other hand, 15-40 % of the INH[R] clinical isolates carry mutations on either the

target gene of this drug *inhA* (such as S94A mutation) or within the *inhA* promoter (Zhang et al., 2005b). The S94A mutation results in a decreased affinity of the enzyme for the activated form of the drug (INH-NAD(H) adduct) (Vilcheze et al., 2006), while the promoter related mutations cause overexpression of the InhA protein leading to resistance (Rouse et al., 1995). Since INH and ETH both target InhA, mutations or overexpression of InhA cause cross-resistance to INH and ETH. Other INH^R strains have also been reported with mutations in the *ndh* and *ahpC* genes encoding for NADH dehydrogenase II and alkyl hydroperoxidase, which modulate the INH activity, NAD metabolism (by increased NADH/NAD$^+$ ratios), and act as a marker of resistance, respectively (Miesel et al., 1998; Wilson and Collins, 1996). PZA is also a prodrug activated by the nicotinamidase/pyrazinamidase enzyme pncA, and PZA resistance is acquired through mutations in the *pncA* gene (Scorpio and Zhang, 1996). Resistance to RIF and fluoroquinoles occurs through mutations in the target genes *rpoB* and *gryA/B*, which encode for the RNAP β-subunit and DNA gyrase, respectively (Sandgren et al., 2009). RpoB mutations occur at the amino acid positions Asp435, His445, and Ser450 of the of RNAP β-subunit, altering the affinity of the enzyme for the drug (Gill and Garcia, 2011). The most common GyrA mutations are found at the amino acid positions Ala90 and Asp94, preventing the drugs from effectively binding the enzyme (Campbell et al., 2011).

TB drug resistance presents a major global health problem by threatening the efforts of TB control. There is an urgent need for discovery and development of new drugs that will be effective on the drug-resistant strains and that also shorten the duration

of the TB treatment. Another challenge of completely eradicating TB is the ability of the bacteria to survive in the host in a non-replicating persistent or chronic state of infection despite the use of antibiotics. *Mtb* has the ability to adapt to the acidic, lytic, and nutrient poor environment inside the activated macrophages, which leads to persistence (Monack et al., 2004). Furthermore, the local drug concentration inside the granulomas might not be adequate to cause bacterial death, or some population of the bacteria might adopt a psyhiological state that makes them less susceptible to the used drugs (Sacchettini et al., 2008). While some individuals can beat the primary infection by immune response without becoming sick, others are unable to eradicate the *Mtb*. Latent TB can reactivate after many years of the primary infection due to the dormant bacteria. It then leads to an active, progressive disease especially in immune-compromised people (Gomez and McKinney, 2004). Drug-resistance and persistence together lead to the development of chronic infection.

According to the World Health Organization (WHO), a new TB infection occurs every second and it is estimated that one-third of the world's population is infected with latent TB. In 2011, 1.4 million people died from TB, including 430,000 people co-infected with HIV (Zumla et al., 2013b). TB is one of the leading causes of mortality in developing countries (Zaman, 2010). The majority of the MDR-TB cases (~60%) are seen in Brazil, China, India, the Russian Federation, and South Africa. Of the new TB cases, 3.7% (~0.5 million people) are diagnosed as MDR-TB. However, this percentage rises up to 20% for previously reported cases (Abubakar et al., 2013). The WHO issued the DOTS strategy to identify, prevent, control, and cure TB. There are two main factors,

drug resistance and persistence, causing the resurgence of TB and making it such a difficult disease to eliminate (Sacchettini et al., 2008). These factors combine to make TB a serious global health threat.

Drug discovery efforts against Mtb

Determination of the genomic sequence of the virulent *Mtb* H37Rv strain in 1998 opened a new era for anti-tubercular drug discovery (Mitra, 2012). Among 4000 genes, around 40% were annotated with known functions, 44% were predicted with putative functions, and ~16% were designated as unique to this organism due to the lack of homology to the existing databases (Cole et al., 1998). Identification of the genes essential for the growth of the bacteria in various conditions suggested new drug targets for the development of novel anti-bacterial agents (Sassetti et al., 2003). In order to decipher the *Mtb* biology to facilitate drug discovery, the Tuberculosis Structural Genomics Consortium (TBSGC) was founded in 2000, and aimed to determine the structures of functionally important proteins and potential drug targets for *Mtb* (Chim et al., 2011; Ioerger and Sacchettini, 2009). To date, more than 750 protein crystal structures were determined for *Mtb* and deposited in the Protein Data Bank (PDB, http://www.rcsb.org), and ~250 of these structures were contributed by TBSGC members (Chim et al., 2011). Once the structure of a target protein is solved, it can be used to understand the molecular basis of cellular processes and macromolecular interactions (as will be demonstrated in Chapter VI of this dissertation), and further for

structure-based drug design studies (as will be demonstrated in Chapter V of this dissertation).

Three main strategies are followed toward drug development and to overcome drug-resistance against *Mtb*: **i)** identifying new drugs against the existing/validated drug targets, **ii)** identifying new drug targets that will be targeted by novel compounds, and **iii)** synergistic combination of new and existing drugs (Green and Garneau-Tsodikova, 2013). The known and validated targets include DNA gyrase and components of the cell wall, cell membrane, and mycolic acid biosynthesis. Target based high-throughput screening (HTS) through a biochemical assay and structure-based drug design are the two conventional methods for the identification of hit/lead compounds against a validated target. These two approaches for identification of inhibitors for the *Mtb* enoyl-reductase enzyme InhA will be demonstrated in chapters IV and V, respectively, of this dissertation. Conversely, identification and validation of new drug targets starts with the identification of an active compound against the organism in a whole-cell based HTS, followed up with additional experiments (forward chemical genetics, expression profiling, or chemical proteomics) to identify and verify the molecular target of the compound (Chan et al., 2010). The newly identified and promising drug targets for *Mtb* are the components of ATP homeostasis (such as the ATP synthase proton pump), the pantothenate synthesis pathway (such as pantothenate synthetase and pantothenate kinase), the enzymes in the shikimate pathway (such as dehydroquinase), the urease complex and its individual subunits, the chorismate-utilizing enzymes, the enzymes of

the glyoxylate shunt, and the drug transporters (such as the trehalose monomycolate transporter MmpL3) (Chim et al., 2011).

The recent developments in anti-TB drug discovery include the identification of new fluoroquinolones (for example, MFX, levofloxacin (LFX), gatifloxacin (GTX)), nitroimidazole compounds (PA-824 and OPC67683), and 1,2-ethylenediamine derivatives (SQ109) **(Table 1-1)**. Fluoroquinolones are a class of compounds effective against many bacterial species. They target the DNA gyrase and inhibit DNA supercoiling, which is especially important for DNA replication (Anand et al., 2011). They possess *in-vivo* and *in-vitro* bactericidal activity against drug-resistant *Mtb* and are included among the second-line TB drugs in the current tuberculosis therapy (Moadebi et al., 2007). PA824 and OPC67683, the 5-nitrorimidazole derivatives, are prodrugs with no cellular targets identified, and are potent against the drug-resistant and persistent bacteria. They are currently under Phase II and III clinical trials (Manjunatha et al., 2006; Singh et al., 2008). Bedaquiline is an ATP synthase inhibitor that is potent against drug-sensitive and resistant bacteria (Andries et al., 2005). It was approved by the US Food and Drug Administration (FDA) as an anti-TB drug in 2012. It is the only new drug in the last 50 years that has been approved for TB treatment; however, its use is restricted to the treatment of multi drug-resistant *Mtb* strains only.

Targeting the actively replicating drug-susceptible and drug-resistant *Mtb* through the inhibition of mycolic acid biosynthesis pathways

Mtb is a non-motile, rod-shaped, acid-fast, and aerobic bacterium belonging to the actinomycetales order of the actinobacteria class. *Mtb* is a slow growing organism with ~ 20 h of doubling time and contains a high G-C content (65%) in its genome (Cole et al., 1998). *Mtb* cells are impervious to gram staining or any other bacteriological stain due to the nature of the mycobacterial cell envelope. The mycobacterial cell envelope does not contain an outer membrane similar to the gram-negative bacteria, but has a unique cell wall structure that acts as an external permeability barrier. The mycobacterial cell wall is essential for mycobacterial viability, survival, and accounts for the intrinsic resistance of the bacteria to many chemical disinfectants, therapeutics and bactericidal agents (Brennan and Nikaido, 1995; Jarlier and Nikaido, 1994). In addition to the uniqueness of its structure and composition, the biosynthetic enzymes responsible for the synthesis of the characteristic features of the mycobacterial cell wall do not have homologs in mammalian systems. Therefore, the *Mtb* cell envelope is the preferred and promising source of molecular targets for drug discovery (Tomioka et al., 2008).

Mtb *cell wall*

The mycobacterial cell wall is composed of three covalently linked layers: peptidoglycan, arabinogalactan, and mycolic acids, which are often referred as the mycolyl-arabinogalactan-peptidoglycan complex (mAGP) **(Figure 1-2)**. The peptidoglycan layer encapsulates the cytoplasmic membrane, comprising of repeating

units of N-acetyl-β-D-glucosaminyl-(1→4)-N-glycolylmuramic acid crosslinked by L-alanyl-D-isoglutaminyl-meso-diaminopimelyl-D-alanine peptide chains (Schroeder et al., 2002). The arabinogalactan layer is composed of arabinan (D-Arabinofuranose) and galactan (alternating 5- and 6- linked β-D-Galactofuranose units) chains that are linked to the peptidoglycan through a phosphodiester linkage. Mycolic acids are attached to the branched hexaarabinofuranosyl termini of the arabinogalactan layer and constitute more than 40% of the weight of the cell wall (Schroeder et al., 2002). In addition to being covalently bound to the arabinogalactan layer, mycolic acids are also found in the cell wall in non-covalently bound forms such as trehalose 6,6'-dimycolate (TDM, cord factor). Lipoarabinomannan (LAM), phosphatidyllinositol mannosides (PIMs), trehalose monomycolates (TMMs), phenolic glycolipids (PGLs), and phthiocerol dimycocerosates (PDIMs) are the other extractable lipids present in the mycobacterial cell wall that play role in pathogenesis and virulence (Minnikin et al., 2002).
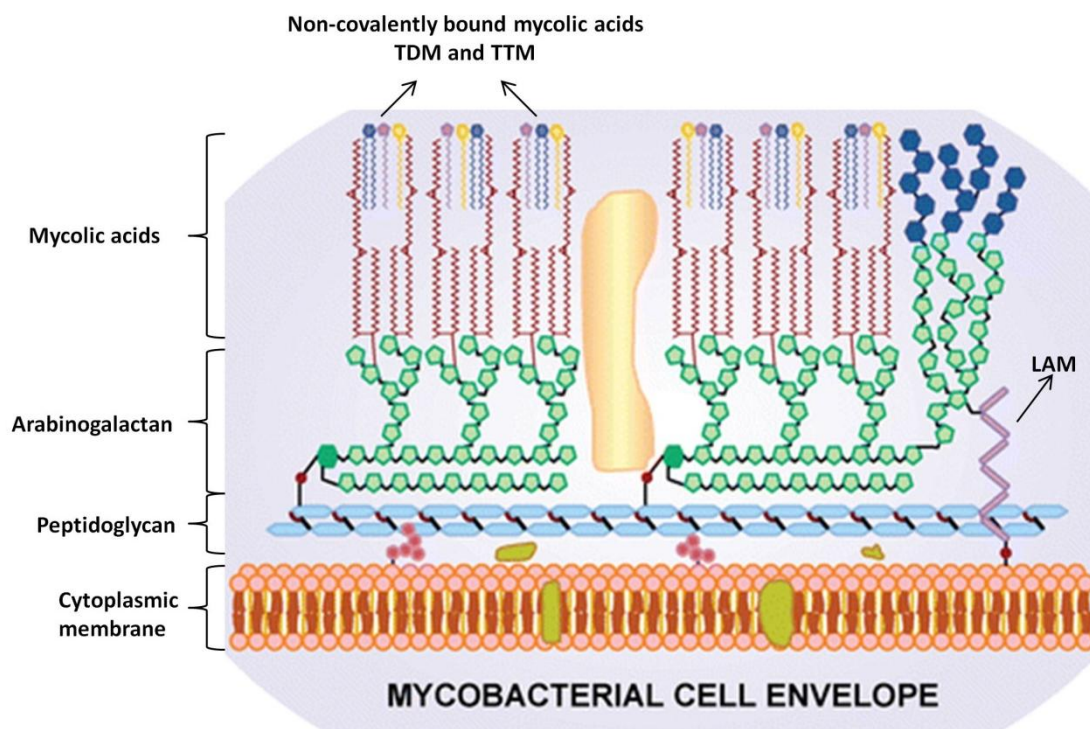
Figure 1-2. Schematic representation of the mycobacterial cell envelope. The figure is reproduced in part from Gokhale R.S. et al. (2007). Nat. Prod. Rep. *24*(2), 267-277, and reproduced with permission of © [2007] The Royal Society of Chemistry. http://dx.doi.org/10.1039/B616817P

Mycolic acid biosynthesis

The hallmark components of the mycobacterial cell wall are the mycolic acids, which are essential for mycobacterial survival and the key to its virulence (Dubnau et al., 2000). Mycolic acids are made of very long α-branched β-hydroxy fatty acids (C_{60}-C_{90}) and are divided into three different structural classes: the most abundant (~60%) α-mycolic acids, the next most abundant (~30%) methoxymycolic acids, and the least abundant (~10%) ketomycolic acids (Schroeder et al., 2002). The α-mycolates contain two *cis*-cyclopropanes in their mero chain, while the methoxymycolates and

19

ketomycolates both contain a *cis*-cyclopropane or an α-methyl *trans*-cyclopropane in the proximal position of the mero chain **(Figure 1-3)**. The methoxymycolates and ketomycolates differ by the presence of an α-methyl methyl-ether, or an α-methyl ketone moiety in the distal position of their mero chain **(Figure 1-3)**.
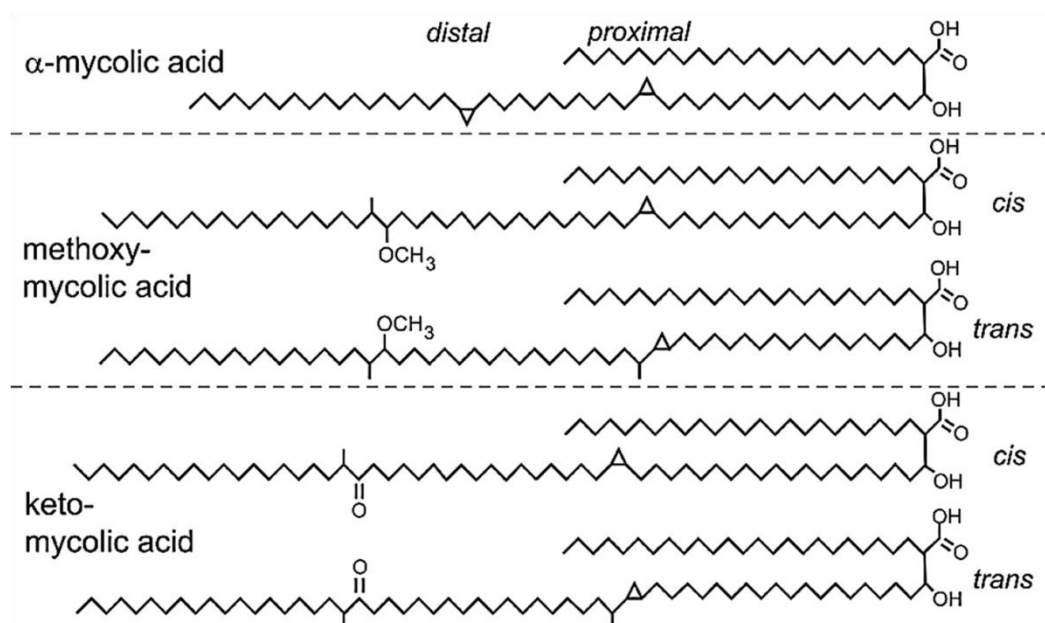
Figure 1-3. Chemical structures of the mycolic acids from *M. tuberculosis*. The figure is adapted from Takayama K et al. (2005). Clin. Microbiol. Rev. *18*, 81-101, and reproduced with permission of © [2005] American Society for Microbiology. http://dx.doi.org/10.1128/CMR.18.1.81-101.2005

In mycobacteria, the biosynthesis of mycolic acids is accomplished by two concerted type I (FAS-I) and type II (FAS-II) fatty acid biosynthetic pathways. The FAS-I system is also found in eukaryotes and is responsible for the *de novo* synthesis of

the medium-length (C_{16}-C_{26}) acyl-CoA chains. FAS-I functions as a single enzyme with multiple domains carrying discrete functions (Molle et al., 2010). In contrast, the FAS-II system is composed of multiple enzymes functioning successively. FAS-II elongates the FAS-I fatty acyl products leading to the production of meromycolates (C_{56}). Following the introduction of various functional groups to the meromycolate chain by cyclopropane synthases, the condensation of the FAS-I (C_{26}-acyl-CoA) and FAS-II (C_{56}-meromycolate) products by Pks13 (a polyketide synthase enzyme) yields the final product known as the mycolic acids (Bhatt et al., 2007b).

Specifically, the FAS-II system elongates the acyl-CoA products of FAS-I by using the substrate malonyl-AcpM, which is generated from malonyl-CoA and phosphopantothenylated holo-AcpM by the FabD enzyme (Takayama et al., 2005). In the first cycle of FAS-II, the malonyl-AcpM and the FAS-I acyl-CoA product are condensed to a β-ketoacyl-AcpM by the FabH enzyme **(Figure 1-4)**. The product of this condensation reaction is then reduced by the β-ketoacyl-AcpM reductase enzyme MabA. After that, the product is dehydrated by the β-hydroxyacyl-AcpM dehydratase enzyme HadABC, and finally reduced by the enoyl-AcpM reductase enzyme InhA **(Figure 1-4)** (Sacco et al., 2007). The resulting acyl-AcpM carries two more carbons when compared with the starting material, and it is fed into the next cycle as the substrate in the condensation reaction with malonyl-AcpM. However, this time the reaction is catalyzed by the KasA/KasB β-ketoacyl-AcpM synthases instead of FabH (Bhatt et al., 2007b). The successive iterations of the FAS-II cycle lead to the production of a very long chain (~C_{56}) of meromycolates.

21
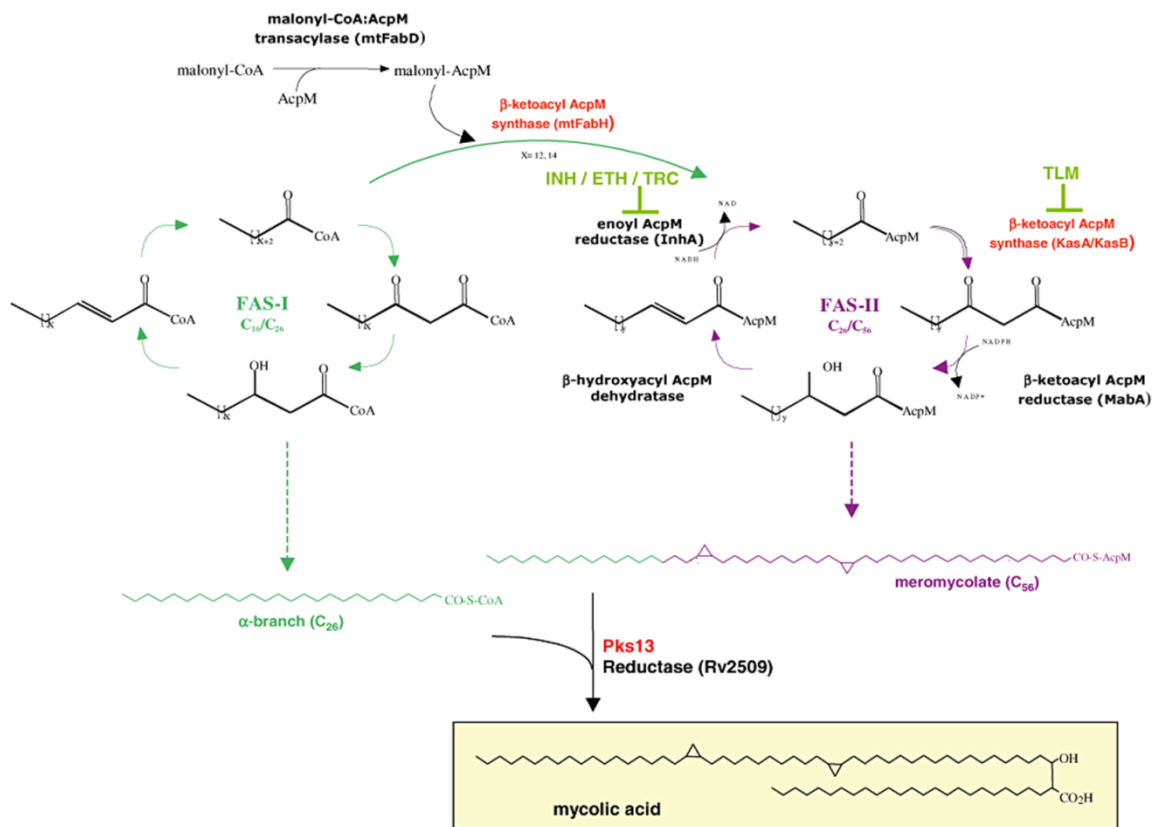
Figure 1-4. The FAS systems of mycobacteria. The FAS-I produces the medium-length (C_{16}-C_{26}) acyl-CoA chains and these precursors are used in FAS-II, producing the meromycolates (C_{56}). INH, ETH, TRC and TLM which inhibit InhA and KasA/B, respectively, are also shown above the enzymes they target. The figure is reproduced from Bhatt et al. (2007). Molecular Microbiology. *64*, 6, 1442–1454, with permission of © [2007] John Wiley and Sons. http://dx.doi.org/10.1111/j.1365-2958.2007.05761.x.

FAS-II system components as drug targets

Mycolic acids are essential for the integrity of the mycobacterial cell wall and play role in virulence and drug resistance (Singh et al., 2011). The enzymes in mycolic acid biosynthesis are also essential for the survival of mycobacteria (Bhatt et al., 2007b; Sacco et al., 2007). These enzymes represent excellent drug targets since no FAS-II system is present in humans (Kuo et al., 2003; Lu and Tonge, 2008; Payne et al., 2001). Components of the FAS-II system are conserved among many pathogens. Thus, selecting these enzymes as drug targets is also a useful strategy for broad spectrum antibiotic development (Lu and Tonge, 2008). Inhibitors that were designed to work against the enzymes of this pathway were experimentally shown to be suitable, potential antibacterial targets (Payne et al., 2001).

InhA

The enoyl reductase enzyme is one of the most studied enzymes as an antibacterial drug target in the FAS-II pathway. InhA is the *trans*-2-enoyl-ACP reductase and it catalyzes the last step in the elongation cycle of the FAS-II pathway while reducing the 2,3 double bond of *trans*-2-enoyl-ACP in an NADH dependent manner (Quemard et al., 1995). Recently, we have identified that InhA activity is regulated post-translationally by phosphorylation, and we elucidated the molecular basis of this regulation by X-ray crystallography (Molle et al., 2010). This study will be covered in Chapter III.

InhA has been identified as the primary cellular target of the anti-tubercular drugs isoniazid (INH), ethionamide (ETH), and also the anti-bacterial compound triclosan. The characterization of InhA as the molecular target of the frontline anti-tubercular drug INH and the second-line drug ETH was achieved after extensive genetic and biochemical analysis. Furthermore, there are several lines of evidence proving that InhA is the primary target of INH and ETH. First, the inactivation of *M. smegmatis* InhA and INH treatment results in similar inhibition of the mycolic acid biosynthesis, accumulation of saturated FAS-I end products, and similar morphological changes in the cell wall leading to cell lysis (Oliveira et al., 2007; Vilcheze et al., 2000). Next, overexpression of InhA in *M. smegmatis*, *M. bovis,* and *Mtb* causes 20- and 10-fold resistance to INH and ETH, respectively (Larsen et al., 2002). Also, clinical isolates resistant to INH contain mutations in the *inhA* gene and are cross-resistant to ETH (Morlock et al., 2003). Finally, the transfer of a single point mutation allele, S94A, of InhA to *Mtb* confers clinically relevant resistance levels to INH and ETH, and it inhibits mycolic acid biosynthesis (Vilcheze et al., 2006). Also, we have more recently demonstrated that InhA is the relevant molecular target of the antimycobacterial activity of INH, concluding the debates on the molecular target of this drug. This study will be discussed in Chapter II.

Even though INH and ETH both inhibit InhA activity, neither one interacts with InhA directly; they are prodrugs needing activation inside the cell. The INH activation is done by the catalase-peroxidase enzyme, KatG. KatG oxidizes INH to an acyl radical, which in turn forms a covalent adduct with NADH (Rozwarski et al., 1998). ETH

activation is done by a FAD containing monooxygenase, EthA, generating the ETH-NAD adduct (Vilcheze et al., 2005). Both the INH-NAD and the ETH-NAD adducts are tight binding inhibitors of InhA (K_i ~5 nM) (Wang et al., 2007). Noticeably, ~50 % of the clinical isolates of MDR and XDR *Mtb* strains, which are resistant to INH, carry mutations in the activator *katG* gene (Dessen et al., 1995; Kuo et al., 2003). The reported minimal inhibitory concentration (MIC) of INH against wild-type *Mtb* and the *katG S315T* strain is 1.5 μM vs. 200 μM (Kuo et al., 2003). In contrast, the majority of the ETH resistant strains carry mutations in the activator gene *ethA* and in the *ethR* gene, which encodes for a transcription repressor regulating the EthA expression negatively (Engohang-Ndong et al., 2004).

The determination of the crystal structure of InhA in complex with the INH-NAD adduct elucidated the mechanism of action of INH against InhA (Rozwarski et al., 1998). The INH-NAD adduct binds to the cofactor binding site of the enzyme adjacent to the fatty acyl substrate binding pocket, interacting with the active site residues **(Figure 1-5)**. Specifically, the pyridine ring of the isonicotinic acyl moiety makes aromatic stacking interactions with Phe149, hydrophobic interactions with the side chains of residues Gly192, Pro193, Leu218, Tyr158, and Trp222, and a water mediated interaction with Met155. When compared with the InhA:NADH structure, it is observed that the Phe149 side chain adopts a different conformation by flipping ~ 90° to interact with the pyridine ring of the INH-NAD adduct (Rozwarski et al., 1998). Rotation of the Phe149 ring opens up a new pocket deep down in the InhA active site and appeals as a promising position to explore for drug design **(Figure 1-5)**.
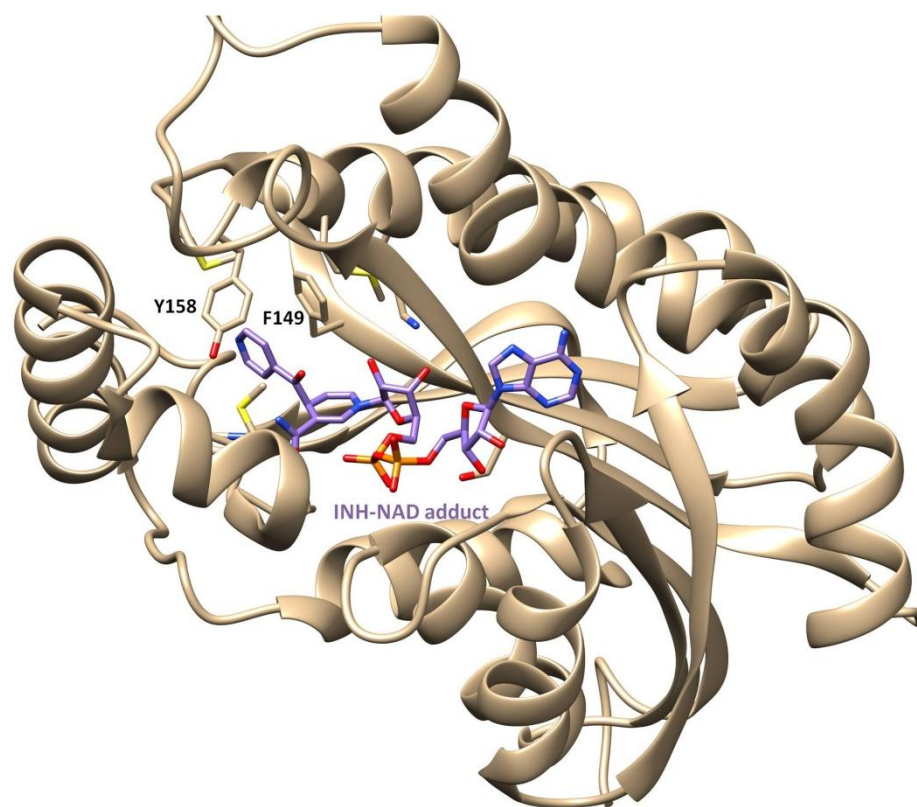
Figure 1-5. Binding mode of the INH-NAD adduct to the InhA active site (PDB ID: 1ZID). The INH-NAD adduct (purple sticks) binds to the cofactor binding pocket and interacts with the surrounding InhA residues such as Y158 and F149. F149 flips 90°compared to the InhA:NADH structure opening a new pocket, which is occupied by the pyridine ring of the adduct.

The mode of action of ETH against InhA is homologous to the INH mechanism of action. We determined the crystal structure of InhA in complex with the ETH-NAD adduct (Wang et al., 2007). Similarly to the INH-NAD adduct, the ETH-NAD adduct also binds to the cofactor binding site of the enzyme and the ethyl-isonicotinic-acyl moiety gets positioned to the new pocket accommodated by the flipping of the Phe149 side chain. The pyridine ring of the ethyl-isonicotinic-acyl group makes stacking

26

interactions with the Phe149 and is surrounded by the hydrophobic residues Tyr158, Met199, Trp222, Leu218, Met155, Met161, and Pro193 (Wang et al., 2007). Elucidation of the molecular mechanism of ETH action will be described in more detail in Chapter II.

InhA is the cellular target of INH and ETH as well as the well known antibacterial compound triclosan. Triclosan (TRC) exhibits a broad spectrum activity against various gram-positive and gram-negative bacteria including *P. falciparum*, *E. coli*, and *B. subtilis* by inhibiting the bacterial FAS-II enoyl-reductase enzymes (Fang et al., 2010; Freundlich et al., 2009; Kuo et al., 2003; Parikh et al., 2000). Unlike INH and ETH, triclosan is not a prodrug and does not require activation before it interacts with InhA. The reported IC_{50} (inhibitory concentration for 50% inhibition of enzymatic activity) for TRC against InhA is 1.1 μM; however, the MIC against *Mtb* is around 138 μM (40 μg/ml) (Freundlich et al., 2009). In addition, the bioavailability of TRC is quite low when compared with INH (Boyne et al., 2007), making it an attractive target for structure-based drug design to improve both the potency and the pharmacokinetic properties by chemical substitutions. Triclosan is composed of two phenyl rings (ring A and B) linked with an oxygen atom **(Figure 1-6)**. The A ring is substituted with a hydroxyl at the 1-position and a chlorine at the 5-position. The B ring has two chlorine substitutions at the 2'- and 4'- positions. Substitutions at these positions were explored to elucidate the structure-activity relationship (SAR). The hydroxyl of the A ring and the chlorines at the 2' and 4' positions of the B ring were found to be important for the potency of the compound (Freundlich et al., 2009). In contrast, modifications in the 5-Cl

position with hydrophobic groups yielded better TRC derivatives (Freundlich et al., 2009; Sullivan et al., 2006).
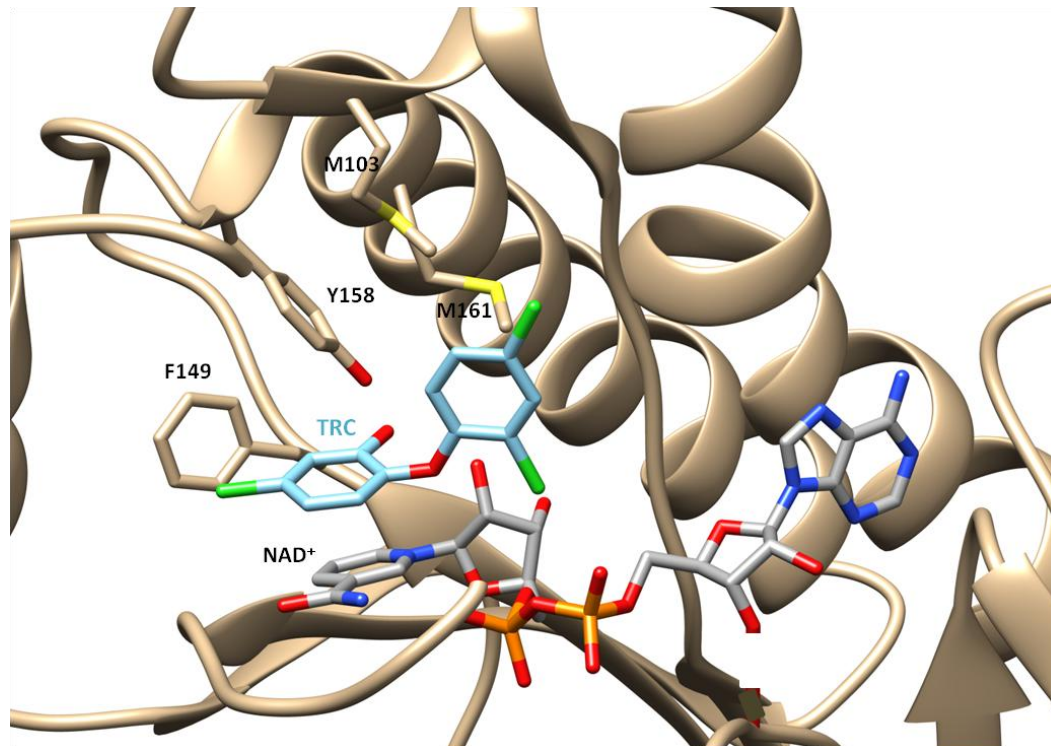
Figure 1-6. Binding mode of triclosan to the InhA active site (PDB ID: 1P45). Triclosan (TRC; blue sticks) interacts with NAD^+ (gray sticks) and the InhA active site residues. Y158, F149 and some other residues interacting with TRC are shown.

The mode of action of triclosan against InhA differs from INH and ETH. TRC is an uncompetitive inhibitor of InhA and requires the cofactor NAD^+ to be bound to the enzyme. TRC binds to the substrate binding pocket of InhA, stacking on top of the nicotinamide ring of NAD^+ with one of its aromatic rings (ring A) **(Figure 1-6)**. The

hydroxyl of this ring also makes H-bonding interactions with the 2'-hydroxyl of NAD$^+$ and the side chain of Tyr158. The rest of the TRC-InhA interactions are mediated through hydrophobic and van der Waals contacts between TRC and the residues Met103, Gly104, Met199, Phe149, Leu218, Ala157, Ile202, and Ile215 (Kuo et al., 2003). I will discuss our structure based drug design studies with triclosan analogs for the development of potent InhA inhibitors in Chapter V.

Since InhA is an established target of INH, ETH, and TRC, and most of the MDR-TB strains are resistant to INH due to the loss of drug activation by KatG, InhA has been the focus of many drug-discovery efforts by both target-based HTS and structure-based drug design approaches. Several classes of compounds such as indole-5-amides, pyrazole derivatives (Kuo et al., 2003), pyrrolidine carboxamides (He et al., 2006), arylamides (He et al., 2007), imidazopiperidines (Wall et al., 2007) and a natural product pyridomycin (Hartkoorn et al., 2012) were identified as direct InhA inhibitors with sub-micromolar IC$_{50}$ values. However, a majority of these compounds either showed no or very poor *in-vivo* activity. Some of these compounds even exhibited low membrane permeability against mycobacteria, limiting their use as anti-tubercular drugs (Lu et al., 2010). Therefore, more research is needed for identification of potent compounds that are active against InhA *in vitro*, and the drug-sensitive and drug-resistant *Mtb in vivo*. We undertook a target based HTS drug discovery approach in collaboration with GlaxoSmithKline to develop and characterize novel and activation free InhA inhibitors, which will be discussed in Chapter IV of this dissertation. We also

performed structure-based drug design using TRC as a template to develop InhA inhibitors, which will be covered in Chapter V.

In order to overcome the ETH resistance and lower the amount of the effective dose necessary to avoid toxic side effects, alternative approaches such as developing inhibitor compounds that will block the EthR-DNA interaction and then promote EthA production were followed. These approaches were based on the facts that increased EthA expression increases the ETH sensitivity, and an *ethR* null mutant BCG strain exhibits higher sensitivity to ETH than the parent strain (Belardinelli and Morbidoni, 2013). These studies resulted in identification of potent compounds that boosted the ETH activity both *in vitro* cultures and *in vivo* models (Flipo et al., 2012; Willand et al., 2009).

KasA/KasB

KasA and KasB both function as the β-ketoacyl-ACP synthase enzymes in the FAS-II elongation cycle. They catalyze the condensation reaction between malonyl-AcpM and the growing acyl chain (Slayden and Barry, 2002). The KasA enzyme has been shown to be essential for mycobacterial survival by both conditional depletion and transposon experiments (Bhatt et al., 2005; Sassetti et al., 2003). Thiolactomycin (TLM) is a natural product produced by both *Nocardia* and *Streptomyces* species and has been shown to inhibit bacterial Kas enzymes, including the mycobacterial KasA and KasB proteins (Brown et al., 2003). The antimycobacterial activity of TLM against *Mtb* is around 20 μg/ml (Kremer et al., 2000b). Overexpression of KasA and/or KasB in *M.*

bovis led to TLM resistance. However, it did not cause resistance to INH, ETH, isoxyl, or cerulenin, suggesting that the Kas enzymes are targeted by only TLM (Kremer et al., 2000b).

The crystal structure of *Mtb* KasA bound with TLM revealed the molecular basis of TLM interaction with KasA. TLM binds to the malonyl binding pocket of KasA forming two hydrogen bonds to the active site His311 and His345 residues **(Figure 1-7)**. The isoprenoid moiety of TLM points to a lipophilic pocket and intercalates between two peptide bonds, Ala279-Pro280 and Gly403-Phe404, further stabilizing the interaction (Luckner et al., 2009). Binding of TLM to the active site, introduces a small conformational change (~0.9 Å) on the loop comprised by residues Phe402-Gly406, which contains the gate keeper residue Phe404 for the adjacent acyl channel **(Figure 1-7)**.

To obtain more potent KasA inhibitors, several attempts were done on modification and optimization of the TLM scaffold, including adding more hydrophobic and longer substituents to the isoprenoid moiety. However, even subtle modifications at this position, disrupting the conjugated, planar structure, were not tolerated and the potency of the compound was lost (Kim et al., 2006; Kremer et al., 2000b). It was also proposed that the TLM molecule combined with a 40-50 carbon-long poly-ethylene-glycol (PEG) might exhibit a good affinity toward the enzyme while increasing the solubility and the stability of the compound (Luckner et al., 2009). TLM and Kas enzymes are considered to be unexploited targets, considering the lack of cross-resistance to other antibiotics and the susceptibility of the MDR strains to TLM (Kremer

et al., 2000b; Luckner et al., 2009; Singh et al., 2011). Therefore, they present potential

for anti-tubercular drug discovery.
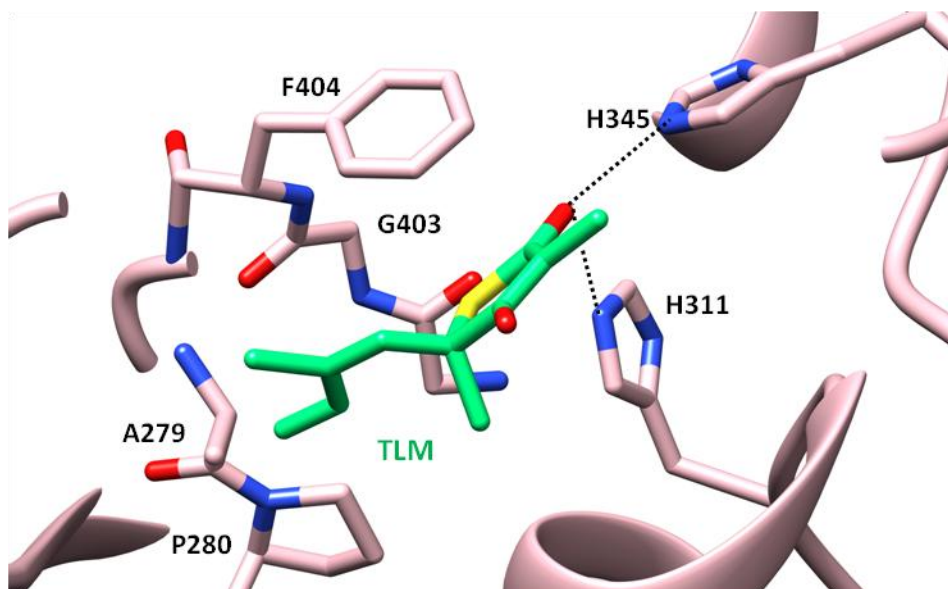
Figure 1-7. Binding mode of thiolactomycin to the KasA active site (PDB ID: 2WGE). Thiolactomycin (TLM; green sticks) interacts with the KasA active site residues His311 and His345. The isoprenoid moiety intercalates between the peptide bonds of Ala279-Pro280 and Gly403-Phe404. H-bonds are shown with dashed black lines.

HadABC and CMASs

Thiacetazone (TAC) is a bacteriostatic anti-tubercular drug that has been widely

used in combination with INH in Africa and South America (Davidson and Le, 1992).

The antimycobacterial activity of TAC against *Mtb* is 0.1 μg/ml (Alahari et al., 2007).

TAC is a prodrug known to be activated by EthA (Dover et al., 2007). The cyclopropane

mycolic acid synthases (CMASs), which modify the double bonds at specific sites on mycolic acid precursors to form cyclopropane rings, were shown to be inhibited by TAC. Examination of the mycolic acid profiles of the TAC treated mycobacteria showed a significant loss of cyclopropanation in mycolates (accumulation of uncycloproponated mycolates). Also, overexpression of CMASs (cmaA2, mmaA2, and pcaA) resulted in increased resistance to TAC (Alahari et al., 2007). In addition, the HadA, HadB, and HadC proteins, which form the HadAB and HadBC heterodimers and function as the β-hydroxyacyl-AcpM dehydratases in the FAS-II elongation cycle (Sacco et al., 2007), were recently proposed to be the molecular targets of thiacetazone (TAC). This was based on the experimental evidence that high levels of resistance to TAC were acquired upon overexpression of the *hadABC* operon and the isolated TAC-resistant mutants carried mutations in the *hadABC* genes (Belardinelli and Morbidoni, 2012). These results together suggest that TAC acts on the mycolic acid biosynthesis. Chemical analogues of TAC have been investigated against *M. avium*. Several of the analogues displayed better *in vitro* and *in vivo* activity than TAC (Bermudez et al., 2003). Structures of HadABC or CMASs in complex with TAC have not been determined yet. Determination of the molecular basis of TAC inhibition can aid the development of more potent inhibitors for these FAS-II system enzymes.

Understanding persistence for targeting non-replicating *Mtb*

Mtb is one of the most successful organisms at adapting to long-term residence in macrophage phagosomes, and this adaptation is essential for the survival, pathogenesis, persistence, and transmission of the bacterium (Schnappinger et al., 2003). This is what gives the bacterium its ability to persist in the host in a non-replicating state despite the use of antibiotics, where it will remain latent and then reactivate years later when the immune system wanes (Gupta and Chatterji, 2005; Honer zu Bentrup and Russell, 2001). "Persisters" are the bacteria that remain viable after exposure to antibiotics while also exhibiting resistance to a majority of available antibiotics (Hopewell and Jasmer, 2005). It is important to decipher the biology of persistent bacteria, and to understand the transition in and out of this state, to be able to target the latent infection (Bacon and Marsh, 2007). Most of the antibiotics currently used in the TB therapy are only potent against replicating bacteria; therefore, discovery of effective drugs against persistent bacteria is crucial. Furthermore, understanding the physiology of persistent *Mtb* will aid the development of new anti-TB agents (Sacchettini et al., 2008).

Transcriptional response of **Mtb** *in persistent state*

It is known that *Mtb* can prevent phagosomal maturation into phagolysosomes in order to avoid the hostile environment inside macrophages. Mature phagosomes expose the bacterium to the microbicidal agents including acids, lytic enzymes, oxygenated lipids, and reactive oxygen and nitrogen intermediates (ROI and RNI) such as the superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), the hydroxyl radical ($\cdot OH$), the

singlet oxygen ($^1O_2$) and nitric oxide (NO)) (Bartos et al., 2004). Also, the phagosomal environment is poor in iron and nutrients. To counter this oxidative, nitrosative, hypoxic, and carbohydrate poor phagosomal environment, *Mtb* undergoes a number of physiological changes that include decreased replication rate, minimal metabolic activity, switching to anaerobic respiration, and usage of fatty acids as the main carbon source (Russell, 2005). Expression profiling and transcriptome analysis suggested that *Mtb* induces fatty-acid degradation enzymes, DNA repair proteins, and iron scavenging siderophore production in macrophages. The genes associated with the persistent state fall into three main categories: respiratory enzymes, fatty acid catabolism and metabolic enzymes, and stress-related response proteins. The differential expression of genes ensures that the bacteria can develop the necessary survival mechanisms against varying stress.

The presence of the ROI and RNI, DNA damaging agents, and the lack of nutrients in the phagosome, induces a 'stress' response in bacteria. Under stress, mycobacteria, like all other bacterial species, proceeds to a non-replicating state by coordinately shutting-down the transcription of the genes for active metabolism, through a process called 'stringent response'. During stringent response, the genes for ribosomal proteins, ribosomal RNAs (rRNA), and cell-wall biosynthesis are downregulated (Godfrey et al., 2002). The stringent response is initiated and mediated by the synthesis of the hyper-phosphorylated guanine, ppGpp and pppGpp, referred together as (p)ppGpp, through a pyrophosphoryl transfer from ATP to GDP or GTP by the RelA enzyme (Dahl et al., 2003; Godfrey et al., 2002). The RNAP open complex is

35

destabilized by (p)ppGpp at selected promoters through direct binding to the RNAP secondary channel and coordinating to the active site Mg^{+2} atom of RNAP. The *relA* knockout *Mtb* strains deficient of (p)ppGpp production were found to be severely impaired in their ability to sustain persistent infection (Dahl et al., 2003), indicating the importance of this response mechanism for the long-term survival of *Mtb*. *Mtb* RelA is a bifunctional enzyme capable of both synthesis and hydrolysis of (p)ppGpp (Avarbock et al., 1999). In *E. coli*, synthesis and hydrolysis of (p)ppGpp is done by the RelA and SpoT enzymes, respectively (Murray and Bremer, 1996). However, it was discovered that (p)ppGpp production itself is not sufficient for the establishment of stringent response, and another protein encoded by the *dksA* gene was required in *E. coli* for this regulation (Paul et al., 2004a). *Mtb* does not have a DksA homolog by sequence analysis and no other component of the mycobacterial stringent response was identified. Recently, a functional homolog of *E. coli* DksA was identified in *Mtb* and annotated as CarD (Stallings et al., 2009). The CarD protein participates in the regulation of transcription under stress conditions; however, it has not been determined yet if it functions synergistically with (p)ppGpp in the same manner as DksA.

In the hypoxic environment of the phagosomes, *Mtb* can use nitrate instead of oxygen as a terminal electron acceptor for anaerobic respiration (Trivedi et al., 2012). Nitrate dependent respiration was also found recently to protect the non-replicating bacteria from acid stress and the reactive nitrogen intermediates (Tan et al., 2010). Therefore, the *Mtb* nitrate reductases, narGHJI, and the putative nitrate/nitrite transporter, narK2, are found to be highly upregulated in hypoxia (Wang et al., 2011).

36

The *narGHJI* and *narK2* genes are part of the DosR/S/T (Dos) dormancy regulon, which is a two-component signaling system composed of two sensor histidine kinases, DosS and DosT, and the cognate response regulator DosR (Chen et al., 2013). The Dos regulon regulates the expression of ~50 genes under hypoxic conditions and in response to nitric oxide, and it is especially important for the non-replicating bacteria (Hu and Coates, 2011). Other important members of this regulon include *hspX*, *acr*, *acg*, *fdxA*, and *pfkB* genes, which encode for a heat shock protein, an α-crystallin, a putative nitroreductase, ferredoxin, and phosphofructokinase, respectively (Voskuil et al., 2004; Zhang et al., 2012b).

The RNI production by macrophages were shown to be essential for host defense during both acute and persistent infection in murine models (Chan et al., 2005). As a response to the ROI and RNI, *Mtb* produces ROI scavenging enzymes catalase-peroxidase (KatG), superoxide dismutases (SodA and SodC), thioredoxin/thioredoxin reductase system (Trx-TrxR), truncated hemoglobins (trHbs), and the peroxidase and peroxynitrite reductase complex (AhpC, AhpD, SucB, and Lpd) (Voskuil et al., 2011). In order to maintain a reduced environment, *Mtb* utilizes mycothiol instead of glutathione, which is believed to act as an antioxidant buffer system under oxidative stress, and was also shown to be required for the ethionamide susceptibility of *Mtb* (Ung and Av-Gay, 2006; Vilcheze et al., 2008). In the presence of H_2O_2 and NO, in addition to the Dos regulon, the genes regulated by the transcriptional regulators SigH, SigE, IdeR, and FurA were also found to be highly induced (Voskuil et al., 2011). Among these, FurA regulates KatG, which is the only catalase-peroxidase enzyme of *Mtb*. KatG

degrades H_2O_2 to water and oxygen and is also involved in INH activation (Milano et al., 2001). IdeR regulates a variety of proteins including putative transporters, siderophore synthesis, and iron storage proteins (Rodriguez et al., 2002).

The granuloma is rich in fatty acids and the energy metabolism of persistent *Mtb* switches from carbohydrates to lipids inside the macrophages. Enzymes for the β-oxidation of fatty acids are utilized for the generation of acetyl-CoA that will be used in the tri-carboxylic acid (TCA) cycle (Kumar et al., 2011). *Mtb* replenishes the precursors required for gluconeogenesis by bypassing the two CO_2 generating steps of the TCA cycle through the glyoxylate shunt, which converts isocitrate to glyoxylate and malate with isocitrate lyase and malate synthase enzymes, respectively. Isocitrate lyase (Icl) is a key enzyme for mycobacterial virulence and survival in the persistent state that the *icl* gene is upregulated during chronic infection and an *icl* knockout *Mtb* strain cannot establish persistent infection in mice (McKinney et al., 2000). The glyoxylate shunt is present in prokaryotes, fungi, and plants but not in mammals, which makes the enzymes of this pathway attractive drug targets. While it is challenging to target Icl due to its small and highly polar active site, the other enzyme of the glyoxylate shunt, malate synthase, has been shown to be a druggable target (Krieger et al., 2012; Sharma et al., 2000). In addition to carbohydrate starvation, the phagosome environment is also poor in iron, which induces the upregulation of genes related to iron acquisition (mycobactin synthesis genes mbtA-L), iron transport, and iron storage (Reddy et al., 2012; Stokes and Waddell, 2009; Symeonidis and Marangos, 2012).

38

Overview of transcription

RNA polymerase (RNAP) is responsible for the DNA-directed synthesis of RNA, called transcription. While eukaryotes have multiple RNAPs, each functioning to synthesize a different class of RNA, bacteria have a single RNAP. The bacterial RNAPs are comprised of four distinct subunits, alpha (α_2), beta (β), beta prime (β') and omega (ω), forming the core enzyme. A fifth subunit, sigma factor (σ), reversibly associates with the RNAP, forming the holoenzyme during transcription initiation **(Figure 1-8)**. Currently, the crystal structures are available for the bacterial (*E.coli, Thermus thermophilus* and *Thermus aquaticus*) core RNAP (Zhang et al., 1999a), holoenzyme (Murakami, 2013; Murakami et al., 2002b; Vassylyev et al., 2002), holoenzyme with a "fork junction" DNA mimicking the open-complex (Murakami et al., 2002a), transcription initiation (Zhang et al., 2012a), and the elongation complexes (Vassylyev et al., 2007). The structure of the *Mtb* core/holo RNAP or any of its subunits has not been determined.

The architecture of the bacterial RNAPs is conserved among different species; however, it has been observed that the structural conservation among RNAPs is significantly greater than the sequence conservation (Darst, 2001). The overall structure of the bacterial RNAPs resembles a crab claw, where the β and β' subunits form the two pincers creating a deep channel (primary or active-site channel) with a width of 27 Å (Darst et al., 2002). The primary channel accommodates the double-stranded DNA and the DNA-RNA hybrid. The two α subunits are located at the opposite surface of the main cleft. They interact with the β and β' subunits through their N-terminal domains,

while they interact with the promoter DNA and transcription activators through their C-terminal domains (Geszvain and Landick, 2005). The ω subunit interacts with the carboxy terminal tail of the β' subunit and participates in RNAP assembly (Mathew and Chatterji, 2006). The RNAP molecule is approximately 150 Å long, 115 Å, tall and 110 Å wide in size (Darst, 2001). The top and bottom of the main channel is connected by a β' helical segment called the bridge helix. The active site is located at the base of the main channel formed between the pincers, where one or two Mg^{2+} ions are chelated by the three conserved aspartate residues of the β' subunit opposite to the bridge helix and several water molecules (Vassylyev et al., 2002). The active site is highly conserved among species and is equipped with the structural elements required for the catalysis and maintaining of the nucleic acid scaffolds. Comparison of the different RNAP structures revealed the conformational flexibility of the enzyme and identified the rigid core and mobile domains. While the α-N-terminal domains, ω subunit, β and β' regions around the active site form the rigid core modules, the β and β' pincer domains (protrusion (β1), lobe (β2), and β'-clamp) are mobile and can open/close the main channel, providing the grip on the dsDNA. The other mobile domain is the β-flap which covers the RNA exit channel (Geszvain and Landick, 2005). Conformational changes in these mobile domains are observed upon transition of RNAP from the core to the holo enzyme (Vassylyev et al., 2002).
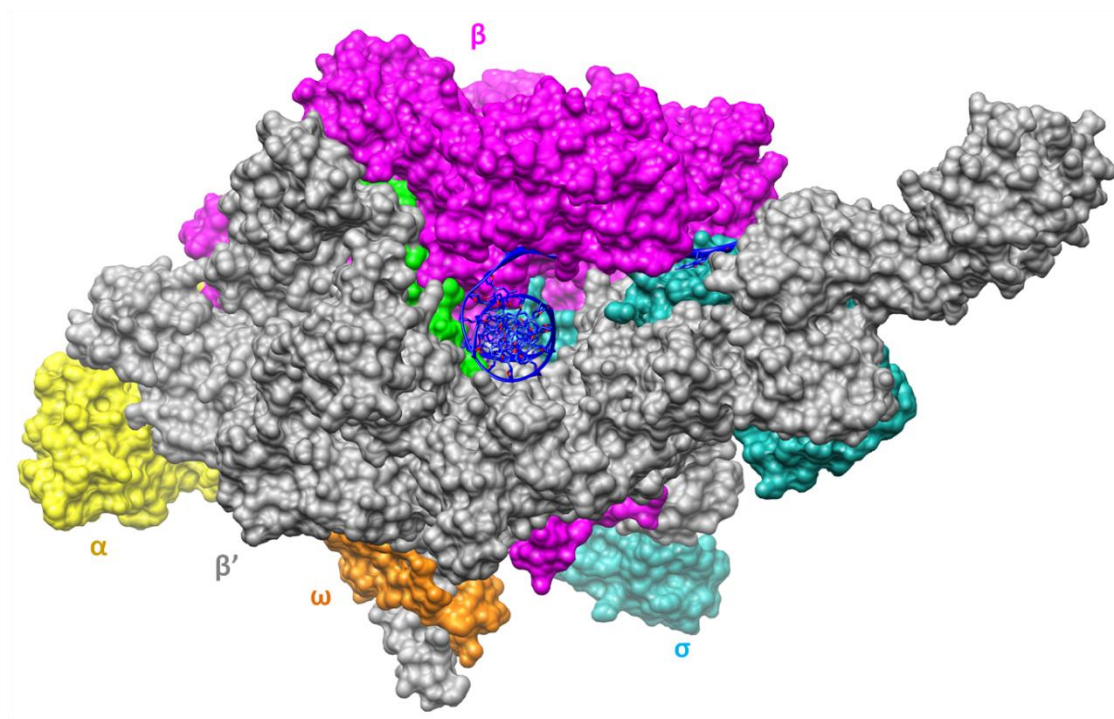
Figure 1-8. Structure of the bacterial RNAP. The crystal structure of the *Thermus thermophilus* transcription initiation complex is shown. Molecular surfaces for the α, β, β', ω and σ subunits are shown in yellow, magenta, gray, orange and cyan colors, respectively. The β'-bridge helix connecting the bottom and the top of the main channel is colored green. The template and non-template dsDNA are shown in ribbon and sticks in dark blue. The active site is located deep inside the main chain between the β and β' pincers, opposite to the bridge helix.

Transcription starts with binding of the holoenzyme to the specific promoter sites on the dsDNA that is selected by the σ-factor, forming the RNAP close-complex (RPc). Next, about 17 base pairs (bp) around the transcription start site get separated (melted) by the holocomplex, leading to the formation of the RNAP open-complex (RPo), and the RNA chain synthesis is initiated. Both the β (β1 and β2 lobes) and β' (β' clamp) subunits

interact with the downstream dsDNA in the initiation complex and participate in the formation and stabilization of the open-complex (Geszvain and Landick, 2005; Zhang et al., 2012a). Before processive RNA production starts, RNAP usually undergoes cycles of short (2-15 nucleotides) transcript formation and release known as the abortive transcription (Goldman et al., 2009). Once the nascent RNA chain becomes approximately ten nucleotides, RNAP can escape the promoter, the σ factor dissociates, and the core RNAP proceeds toward the chain elongation forming the transcription elongation complex (TEC). During elongation RNAP translocates both itself and the melted transcription bubble along the DNA template. The NTP substrates are fed to the active site through a secondary channel, which is lined within the β' subunit, forming a funnel from the surface of the enzyme to the active center (Darst, 2001). The newly synthesized RNA exits through the RNA-exit channel, which is covered by the β-flap and β'-lid regions. In prokaryotes, transcription termination occurs through one of two processes. It can occur when RNAP reaches a characteristic termination sequence/signal, which leads to the formation of a stem-loop RNA structure that will pause and release RNAP by disrupting the nucleic acid-protein interactions. Termination can also happen through the binding of Rho, which is a transcription termination factor, to the growing RNA chain, and pausing RNAP while unwinding the DNA-RNA hybrid (Darnell and Lodish, 2000).

Transcription regulation by RNAP interacting proteins

Gene regulation and transcription control is mainly mediated by the transcription factors and transcription regulators that interact with the RNAP and the promoter DNA sequences to alter the transcription activity. While the majority of the transcription factors bind to specific DNA sequences with their characteristic DNA binding domains and facilitate or block the RNAP-promoter DNA interactions (activators and repressors), some transcription regulators solely interact with RNAP, and some are involved in the structuring of the bacterial nucleoid (nucleoid-associated proteins). The *Mtb* genome encodes ~190 transcription regulators of which ~140 are uncharacterized/putative transcription factors, 13 are sigma factors, 11 are two-component systems, and 5 are unpaired response regulators (Manganelli et al., 2004). During the varying stages of infection, different transcription regulators are utilized to ensure differential gene expression as a response to the environmental conditions (Raman et al., 2004). Below, the *Mtb* RNAP-interacting transcription regulators that are important for the establishment of the persistent infection will be discussed.

<u>Sigma factors</u>

Sigma factors are indispensable components of the transcription machinery that provide promoter specificity. The *Mtb* genome encodes for 13 σ factors that all belong to the σ_{70} class, which is *E. coli*'s principal σ factor (Manganelli et al., 2004). Each mycobacterial sigma factor (sigA-M) has its own specificity and is expressed under different growth conditions, allowing transcription of a different subset of genes. The

primary σ factor, sigA (σ_A), is responsible for the transcription of housekeeping genes, it is essential for bacterial viability, and it is constantly expressed. The other sigma factors, sigB, sigC, sigD, sigE, sigF, sigH, and sigJ, were found to be induced upon various stress conditions, starvation, and within the macrophages. Therefore, they are associated with persistent infection and virulence (Smith et al., 2005). This is especially true for σ_D, σ_E, and σ_H, which regulate the genes related to the persistent infection. These genes include the *rel* operon, *icl*, and the genes encoding for fatty acid degradation enzymes (such as fadE24)*,* DNA repair proteins, thioredoxin and thioredoxin reductase (Rodrigue et al., 2006). *Mtb* strains with mutant σ_F, σ_B, and σ_J display attenuated virulence and are more sensitive to various stresses, suggesting their involvement in adaptation to the stationary phase (Rodrigue et al., 2006). The σ function is usually regulated by anti-sigma factors, which bind to σ and inhibit its interaction with RNAP and the promoter DNA. Four different anti-σ-factors were identified from the *Mtb* genome that regulate the functions of σ_E, σ_F, σ_H, and σ_L (Smith et al., 2005).

The σ_{70} related σ-factors contain up to four conserved regions (regions 1,2,3 and 4), which can further be divided into sub-regions (sub-regions 1.1-1.2, 2.1-2.4, 3.0-3.2 and 4.1-4.2), and classified into four different groups (groups 1,2,3 and 4) based on their structural and functional organizations (Geszvain and Landick, 2005). *Mtb* σ_A, σ_B, and σ_F belong to groups 1, 2, and 3, respectively, and contain at least three conserved regions (regions 2, 3, and 4). The rest of the ten mycobacterial σ-factors belong to group 4, and are also described as the extracytoplasmic function (ECF) σ-factors, containing only regions 2 and 4. The *Mtb* RNAP core or holoenzyme structure is not available; however,

44

the crystal structures of the RNAP holoenzyme from *Tth, E. coli* and *Taq* revealed the details of the sigma-RNAP interactions (Murakami, 2013; Murakami et al., 2002b; Vassylyev et al., 2002). It has also been observed from the crystal structures of initiation complexes that the sequence specific interactions between RNAP and the promoter DNA are provided by the conserved regions of the σ-subunit (Murakami et al., 2002a). The major contact interface between the core RNAP and σ occurs between the coiled-coil region of the β' clamp and the conserved σ sub-region 2.2 (Vassylyev et al., 2002). The other contacts are between the σ sub-region 3.1 and the β-subunit β1 protrusion domain, the σ sub-region 3.2 and the β' rudder, lid loops and the C-terminus of the β subunit, and the σ sub-regions 4.1 and 4.2 with the β flap and β' zinc finger domain. The σ-promoter DNA interactions occur through the sub-regions 2.4, 2.3, 3.0, and 4.2, which are involved in the -10 promoter recognition, transcription bubble melting, and the extended -10 element and the -35 promoter recognition, respectively (Rodrigue et al., 2006). The sub-region 4.2 was also shown to interact with other transcription regulators such as WhiB3 (Steyn et al., 2002).

RbpA

RbpA is a 14 kDa protein found only in actinomycetes and was first identified in *Streptomyces coelicolor*. It was found that in *Mtb*, during the stationary phase under starvation, the expression of the *rbpA* gene was upregulated ~8 fold (Betts et al., 2002). In *M. Smegmatis*, RpbA was associated with increased tolerance and resistance to antibiotic rifampicin (Dey et al., 2011). This was thought to occur via three alternate

mechanisms such as the competition of RbpA with rifampicin for the binding site, inducing allosteric conformational changes in the rifampicin binding site while influencing the drug's binding affinity, and an indirect mechanism affecting the membrane permeability (Hu et al., 2012). Recently, it was identified that RbpA functions as a σ specific transcriptional activator and it regulates the access of the primary and stress related σ-factors, σ_A and σ_B but not σ_F, to the core RNAP during the stress response and stationary phase (Bortoluzzi et al., 2013; Hu et al., 2012)

The NMR solution structure of RbpA revealed a well-defined core domain comprised of four β-strands and highly flexible N- and C-terminal regions. It was shown that RbpA interacts with σ_A and σ_B σ2 domains through these unstructured N- and C-terminal regions (Bortoluzzi et al., 2013; Tabib-Salazar et al., 2013). Two different studies suggested that RbpA interacts with the RNAP β-subunit. However, one study proposed RbpA binds close to the RNAP active site where rifampicin binds, while the other study proposed the sandwich-barrel hybrid motif of the β-subunit on RNAP surface is the interaction interface (Dey et al., 2010; Hu et al., 2012). These studies suggest that RbpA stabilizes the σ binding to the core RNAP and facilitates open-complex formation in order to promote transcription. Being an essential transcription regulator and playing an important role in the σ cycle makes RbpA a promising target to study (Tabib-Salazar et al., 2013).

CarD

Microarray experiments to identify genes that are regulated differently under stress conditions resulted in identification of the *Rv3583c* gene of *Mtb*. *Rv3583c* encodes for a 17 kDa protein, which is a putative transcriptional factor that is annotated as CarD. CarD expression was found to be upregulated 4-8 fold when the bacteria was grown under oxidative stress, starvation, and in the presence of DNA damaging agents. CarD depleted strains failed to regulate the rRNA and ribosomal protein levels, pointing out its involvement in the stringent response mechanism (Stallings et al., 2009). The ability of *Mtb* CarD to complement the DksA function in a Δ*dksA E. coli* strain verified its involvement in rRNA and ribosomal protein transcription regulation and suggested that it was a functional homolog to *E. coli* DksA. *In-vivo* studies also demonstrated that absence of the CarD protein diminishes the persistent *Mtb* infection in mice. In addition, the CarD protein was essential both *in-vivo* and *in-vitro* conditions for mycobacterial growth under stress as well as under normal growth conditions. These findings all suggested that CarD has a pivotal role in the establishment of persistence. Understanding the biology of this protein can provide insights into *Mtb*'s adaptation to persistence and may offer new potential drug targets.

The *Mtb* CarD protein belongs to the CarD-TRCF family of proteins. The N-terminus of CarD proteins shows ~30% sequence identity to the Transcription Repair Coupling Factor (TRCF) RNAP interacting domain (RID). TRCF removes the stalled RNAP from a site of damaged DNA, recruits the DNA repair proteins, and interacts with the RNAP β-subunit β1 domain through this RID domain (Smith and Savery, 2005).

47

CarD homologs are found in all mycobacterial species and they are highly conserved (>95%) among actinobacteria. On the other hand, no homolog is present in humans. *Mtb* CarD has ~30% sequence identity and ~50% sequence similarity to the CarD protein from *Myxococcus xanthus,* which is the first identified protein from this family. *M .xanthus* CarD is involved in the carotenogenesis pathway, which is responsible for the carotenoid synthesis upon exposure to blue light. *M. xanthus* CarD is part of a complex regulatory network, and interacts with both RNAP and CarG with its N-terminal protein-protein interaction domain. It also interacts with the *carQRS* promoter DNA with its C-terminal DNA interaction domain (Nicolas et al., 1996; Penalver-Mellado et al., 2006). The *M. xanthus* CarD DNA binding domain resembles the eukaryotic high mobility group A-1 (HMGA1) proteins, which interact with the minor grove of AT rich DNA sequences through their 'RGRP' protein sequence motifs called 'AT hook's. In contrast, *Mtb* CarD does not have an AT hook motif and only exhibits homology to the N-terminal domain of *M. xanthus* CarD. Therefore, *Mtb* CarD is also classified as a CarD-N-terminal-like (CdnL) protein (Garcia-Moreno et al., 2010).

Similar to *M. xanthus* CarD, *Mtb* CarD was also reported to interact with the RNAP β-subunit (Stallings et al., 2009). Based on the homology models generated according to the homologous TRCF-RNAP interaction, it was predicted that CarD proteins interact with the N-terminus (approximately the first 120 residues) of the β-subunit. For *Mtb* CarD, this interaction is crucial for transcription regulation and is required for mycobacterial survival and adaptation to stress conditions (Weiss et al., 2012). Furthermore, as was observed for RbpA protein, the CarD-RNAP interaction also

affects *Mtb*'s susceptibility to rifampicin. Loss of the CarD-RNAP interaction increases the sensitivity of mycobacteria to rifampicin (Weiss et al., 2012).

Little is known about the mechanism of CarD regulation of RNAP. Recently, we have determined the crystal structure of *Mtb* CarD complexed with the RNAP β-lobes. The uncomplexed structure of the β-lobes was also solved to compare the conformational differences of RNAP in the bound and unbound states. We then investigated the CarD-DNA interaction (Gulten and Sacchettini, 2013). Another study was performed with a CarD homolog protein from *Thermus thermophilus* and suggested that CarD functions during transcription initiation as an activator by favoring interactions between RNAP and promoter DNA (Srivastava et al., 2013). Our structural studies of the *Mtb* CarD/RNAP complex and the insights gained through this complex structure will be discussed in the Chapter VI of this dissertation.

CHAPTER II

DECIPHERING THE MECHANISM OF DRUG ACTION FOR ISONIAZID AND

ETHIONAMIDE AGAINST TUBERCULOSIS AND LEPROSY[*]

Overview

Thioamide drugs, ethionamide (ETH) and prothionamide (PTH), are clinically

effective in the treatment of *Mycobacterium tuberculosis (Mtb)*, *Mycobacterium leprae*

(*M. leprae)*, and *Mycobacterium avium (M. avium)* complex infections. Although

generally considered second-line drugs for tuberculosis, their use has increased

considerably as the number of multidrug resistant and extensively drug resistant

tuberculosis cases continues to rise. Despite the widespread use of thioamide drugs to

treat tuberculosis and leprosy, their precise mechanisms of action remain unknown.

Mycobacterium tuberculosis enoyl-acyl-ACP reductase (InhA) has been demonstrated to

be the primary target of isoniazid (INH) and ETH. Recently, it was postulated that *M.*

tuberculosis dihydrofolate reductase (DHFR) is also a target of INH, based on the

findings that a 4R-INH-NADP adduct synthesized from INH by a nonenzymatic

approach showed strong inhibition of DHFR *in vitro*, and overexpression of *M.*

---

tuberculosis dfrA in *M. smegmatis* conferred a 2-fold increase of resistance to INH. In the present study, a plasmid expressing *M. tuberculosis dfrA* was transformed into *M. smegmatis* and *M. tuberculosis* strains, respectively. The transformant strains were tested for their resistance to INH. Compared to the wild-type strains, overexpression of *dfrA* in *M. smegmatis* and *M. tuberculosis* did not confer any resistance to INH based on the MIC values. Similar negative results were obtained with 14 other overexpressed proteins that have been proposed to bind some form of INH-NAD(P) adduct. An *Escherichia coli* cell-based system was designed that allowed coexpression of both *M. tuberculosis katG* and *dfrA* genes in the presence of INH. The DHFR protein isolated from the experimental sample was not found bound with any INH-NADP adduct by enzyme inhibition assay and mass spectroscopic analysis. We also used whole-genome sequencing to determine whether polymorphisms in *dfrA* could be detected in six INH-resistant clinical isolates known to lack mutations in *inhA* and *katG*, but no such mutations were found. The *dfrA* overexpression experiments, together with the biochemical and sequencing studies, conclusively demonstrate that DHFR is not a target relevant to the antitubercular activity of INH. In addition, using the cell-based activation method, we now have definitive evidence that both thioamides form covalent adducts with nicotinamide adenine dinucleotide (NAD) and that these adducts are tight-binding inhibitors of *M. tuberculosis* and *M. leprae* InhA. The crystal structures of the

inhibited *M. leprae* and *M. tuberculosis* InhA complexes provide the molecular details of target–drug interactions. The purified ETH-NAD and PTH-NAD adducts both showed nano-molar K_is against *M. tuberculosis* and *M. leprae* InhA. Knowledge of the precise

structures and mechanisms of action of these drugs provides insights into designing new

drugs that can overcome drug resistance.

Introduction

In 1952, isoniazid (INH) was discovered to have bactericidal activity against

Mycobacterium tuberculosis (Bernstein et al., 1952). Since then, it has been used as a

potent front-line drug against tuberculosis (Bloch, 1977). On the other hand, INH is not

effective against *Mycobacterium leprae*, the causative agent of leprosy, most likely due

to the dysfunction of the *M. leprae katG* gene. Thioamide drugs, ethionamide (ETH) and

pro-thionamide (PTH), were introduced in the 1960s, and they are structurally similar to

INH **(Figure 2-1)**. They have been widely used in the treatment of mycobacterial

infections caused by *Mycobacterium tuberculosis*, *M. leprae*, and *M. avium* complex

infections (Fajardo et al., 2006; Yajko et al., 1987). ETH and PTH are both bacteriocidal

and are essentially interchangeable in a chemotherapy regimen. They are the most

frequently used drugs for the treatment of drug-resistant tuberculosis and, therefore, are

becoming increasingly relevant as the number of multidrug-resistant and extensively

drug-resistant cases is increasing worldwide (CDC, 2006; Crofton et al., 1997).

Moreover, ETH and PTH are also used in a combined chemotherapy regimen with either

dapsone or rifampin to treat leprosy (Katoch et al., 1992). Although we have previously

speculated about the mechanism of action of ETH in *M. tuberculosis* based on an

analogy to isoniazid's (INH's) mode of action (Banerjee et al., 1994; Rozwarski et al.,

1998; Vilcheze et al., 2006), definitive biochemical evidence that ETH targets InhA has

not been forthcoming.
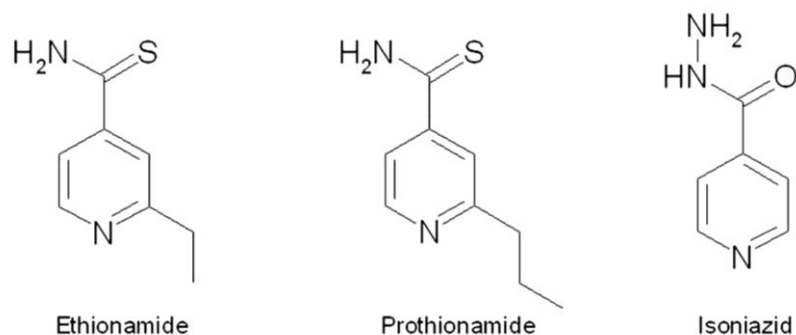
Figure 2-1. Chemical structure of ETH, PTH, and INH. Although these prodrugs have similar structures, INH is activated by a catalase-peroxidase, whereas ETH and PTH are activated by a flavin-dependent monooxygenase.

The mechanism of action of INH has been studied for more than 50 years. Through lipid profiling, INH was found to inhibit mycolic acid biosynthesis in *M. tuberculosis* (Winder et al., 1970). In addition, the INH-induced inhibition of mycolic acid biosynthesis was demonstrated to correlate with the bactericidal activity of INH (Takayama et al., 1972). Further analysis of the lipids of INH-treated *M. tuberculosis* indicated that the elongation of fatty acids beyond C26 was inhibited, which suggested that the target of INH is an enzyme in fatty acid elongation (Takayama et al., 1975). Similar to INH, ETH and PTH also inhibit mycolic acid biosynthesis (Takayama et al., 1972; Winder et al., 1971). INH is a prodrug that must first be activated by KatG, an

endogenous catalase/peroxidase (Zhang et al., 2005a). The mode of INH action remained unclear until an INH-NAD adduct was identified as the bound inhibitor in the active site of InhA, the enoyl-acyl ACP reductase involved in long-chain fatty acid biosynthesis, by protein crystallography (Rozwarski et al., 1998). It was hypothesized that KatG cleaves the hydrazide on INH to an isonicotinoyl radical, which then reacts with NAD to form an adduct that binds to and inhibits InhA (Timmins and Deretic, 2006). The crystal structure of InhA bound with the adduct indicates that an isonicotinoyl moiety was covalently attached to the 4-position of the nicotinamide ring of NAD cofactor in an S configuration. The chemical structure of the INH-NAD adduct was found to be consistent with the molecular weight obtained by the mass analysis (Rozwarski et al., 1998). Later studies demonstrated that INH-NAD adduct could be generated by a KatG-catalyzed oxidation in the presence of NAD (Lei et al., 2000; Wilming and Johnsson, 1999), which strongly inhibits InhA (K_i = 5 nM) to block mycolic acid biosynthesis (Rawat et al., 2003; Rozwarski et al., 1998; Vilcheze et al., 2006). Similar to INH, ETH and PTH are also prodrugs that require activation to exert antitubercular activity. KatG mutant strains resistant to INH are sensitive to ETH, indicating that ETH has a different activator (Fattorini et al., 1999; Morlock et al., 2003). Mutations of a gene designated *ethA* were repeatedly found in the clinical isolates resistant to ETH (DeBarber et al., 2000; Morlock et al., 2003). Like KatG, the overexpression of *ethA* in *M. smegmatis* resulted in substantially increased ETH sensitivity (Baulard et al., 2000). This evidence suggested that *ethA* is critical for the activation of ETH. *ethA* encodes a flavin monooxygenase found to catalyze the Baeyer-

Villiger reaction to detoxify aromatic and long-chain ketones (Fraaije et al., 2004). The enzyme is membrane associated and has a tendency to form large oligomers after purification (Fraaije et al., 2004; Vannelli et al., 2002). The monooxygenase activity of the purified EthA is very low (k_{cat} = 0.00045 s^{-1}), suggesting that the enzyme may require other proteins or cellular components to be completely functional (Fraaije et al., 2004). The active form of ETH has never been detected or isolated *in vitro*, although some inactive metabolites produced by the catalytic oxidation of ETH by EthA have been studied by TLC and HPLC (DeBarber et al., 2000).

Mutations within the protein-coding and promoter regions of *inhA* are frequently observed in clinical isolates resistant to INH (Guo et al., 2006; Telenti, 1998). An S94A mutation in InhA, which was originally identified in an INH-resistant *M. smegmatis* strain, was later found in three *M. tuberculosis* clinical isolates that conferred resistance to both INH and ETH (Banerjee et al., 1994; Morlock et al., 2003). The S94A allele of *inhA* has been transferred into *M. tuberculosis* by a specialized linkage transduction, which was sufficient to confer 5-fold resistance to INH (Vilcheze et al., 2006). Moreover, overexpression of *inhA* in *M. tuberculosis* was found to confer 10-fold resistance to INH (Larsen et al., 2002). Indeed, several *M. tuberculosis* clinical isolates resistant to INH contain mutations in the *inhA* gene, and all have been found to be cross-resistant to ETH (Lei et al., 2000). These genetic observations support that InhA is the primary target of INH.

Although genetic and biochemical studies have provided convincing evidence that InhA is the primary target of INH, other putative targets of INH have also been

proposed (Mdluli et al., 1998; Timmins et al., 2004). Previously, FAS-II elongation enzyme KasA was proposed to be the primary target of INH based on the mutations found in *kasA* of some clinical isolates and the formation of a complex between AcpM, KasA and INH (Mdluli et al., 1998). However, later it was demonstrated that InhA inhibition but not KasA inhibition induces the formation of this complex, and only InhA but not KasA gets inhibited by KatG activated INH *in vitro* (Kremer et al., 2003). Recently, 17 proteins other than InhA were identified from *M. tuberculosis* lysate that could tightly bind to an affinity matrix derived from INH-NADP or INH-NAD adducts by proteomic analysis (Argyrou et al., 2006a). Among these proteins, *M. tuberculosis* dihydrofolate reductase (DHFR) was shown to be strongly inhibited by an INH-NADP adduct *in vitro* ($K_{i\ \text{app}} = 1$ nM) in a separate study (Argyrou et al., 2006b). This INH-NADP adduct was synthesized by incubating INH and NADP in the presence of Mn(III) as a catalyst. The crystal structure of the complex indicated that an acyclic 4R INH-NADP adduct was selectively bound in the active site of DHFR. In addition, overexpression of *dfrA* in *M. smegmatis* caused a 2-fold increase of resistance to INH compared to the wild-type (Argyrou et al., 2006b). These observations were taken to suggest that *M. tuberculosis* DHFR is also a target of INH (Argyrou et al., 2006a; Argyrou et al., 2006b).

There are several observations which conflict with the proposal that DHFR might be a target of INH. First, the MIC difference (2-fold) between the *dfrA* overexpressed *M. smegmatis* strain and the wild-type strain is not significant compared to the error of this type of experiment. Second, *dfrA* has not been clearly shown to be essential to *M.*

tuberculosis. Although it is an essential gene for nucleotide biosynthesis in many other organisms (Gangjee et al., 2007; Kompis et al., 2005), disruption of *dfrA* by transposon mutagenesis was found not to attenuate *M. tuberculosis* infection in mice (Sassetti and Rubin, 2003). Third, the endogenous formation of the 4R INH-NADP adduct by KatG catalysis has not been demonstrated.

In this study, we elucidated the mechanism of drug action of ETH and PTH against *M. tuberculosis* and *M. leprae* using a cell based activation approach. To determine whether *M. tuberculosis* DHFR is a molecular target of INH, we investigated (i) whether overexpression of *dfrA* under relevant conditions confers resistance to INH in *M. smegmatis* and *M. tuberculosis*, (ii) whether the INH-NADP adduct is an activated INH product generated by KatG catalysis inside the cell, and (iii) whether mutations in *dfrA* could be observed in INH-resistant clinical isolates.

Results and discussion

Identification of the active form of ETH

To identify the active form of ETH, we and others have attempted to use purified EthA to activate ETH and inhibit InhA *in vitro* but have never been able to observe any InhA inhibition (unpublished data). This might be due to the fact that purified EthA forms large oligomers (200-600 kDa) and has a very low enzymatic activity *in vitro* (Fraaije et al., 2004). Purification of EthA is only possible in the presence of detergents suggesting that the protein is membrane associated even though no transmembrane helix

or hydrophobic segments that could contribute to membrane anchoring is recognized from the protein sequence (Fraaije et al., 2004). It is possible that other proteins or cellular components are required for the full EthA activity. Because *in vitro* activation of the drugs ETH and PTH has not been possible by either chemical or enzymatic approaches, we developed a cell-based activation method. In this system, recombinant *Mtb or M.leprae* EthA and InhA were co-overexpressed in the same *Escherichia coli* cell, and ETH or PTH was added to the culture to test whether the drugs would inhibit InhA upon activation. Although ETH and PTH are both potent drugs against *M. tuberculosis* (MIC = ~0.5 µg/ml) (Quemard et al., 1992), they do not affect *E. coli* growth, even at very high concentrations (100 µg/ml), which is primarily caused by the absence of an EthA homologue in *E. coli*.

InhA and EthA from *M. tuberculosis* were coexpressed in *E. coli* BL21 (DE3) in the presence of 100 µg/ml ETH. InhA was rapidly purified, and an *in vitro* enzyme assay was performed. InhA isolated from the experimental sample had <1% of the specific activity of InhA purified without the addition of ETH under the same assay condition. Mass analysis of denatured InhA from the experimental sample indicated the presence of a small molecule with a molecular weight of 798.2. This corresponds to the exact molecular weight of an ethyl-isonicotinic-acyl-NAD covalent adduct. Moreover, pure fractions of this small molecule showed strong inhibition to native InhA *in vitro* ($K_i = 7 \pm 5$ nM), which is as potent as the INH-NAD adduct, the active form of INH ($K_i = 5$ nM) (Vilcheze et al., 2006). When PTH was used in the same coexpression experiment, a compound with a molecular weight of 812.2 was identified that corresponds to the

exact weight of a propyl-isonicotinic-acyl-NAD adduct. This compound is also extremely potent against InhA *in vitro* (Ki = 2 ± 0.8 nM for *Mtb* InhA; Ki= 11 ± 6 nM for *M.leprae* InhA).

Other than ETH and PTH, thioamide drugs such as thiacetazone and isoxl have been shown to be activated by EthA EthA (DeBarber et al., 2000). The same cell-based method was applied to test thiacetazone. InhA and EthA were co-overexpressed in the presence of TAC and InhA was purified to test the enzymatic activity. The isolated InhA was not inhibited under the same assay condition. As expected, mass analysis did not show the existence of any tightly bound inhibitor. These results indicate that, unlike ETH or PTH, thiacetazone does not target InhA, even though all of these thioamides are activated by EthA in *M. tuberculosis*.

Structures of **Mtb** *and* **M.leprae** *InhA with ETH-NAD and PTH-NAD adducts*

We have determined the crystal structures of *M. tuberculosis* InhA in complex with the ETH-NAD and the PTH-NAD adducts, and *M. leprae* InhA in complex with the PTH-NAD adduct to 2.2, 2.5 Å and 2.1 Å resolution, respectively. *Mtb* and *M. leprae* InhA proteins are highly similar in sequence and overall structure (90% sequence identity; RMSD = 1.3 Å for $C_\alpha s$). The active site residues are also conserved. Unbiased electron density maps of each complex clearly indicated the presence of a modified NAD with an ethyl-isonicotinic-acyl or propyl-isonicotinic-acyl group covalently attached to the 4 position of the nicotinamide ring in a 4S configuration **(Figure 2-2A)**. The chemical structures of both inhibitors are consistent with the molecular weights

obtained by the mass analysis. Similar to the structure of InhA bound with the adduct INH-NAD (Rozwarski et al., 1998), the ethyl-isonicotinic-acyl, or the propyl-isonicotinic-acyl, moiety is found in a hydrophobic pocket that was formed by the rearrangement of the side chain of Phe[149] (**Figure 2-2B–D**). The ethyl-isonicotinic acyl or the propyl-isonicotinic-acyl group also forces the side chain of Phe[149] to rotate ~90°, forming an aromatic ring-stacking interaction with the pyridine ring **(Figure 2-3A)**. The pocket is predominantly lined by hydrophobic groups from the conserved side chains of Tyr[158], Phe[149], Met[199], Trp[222], Leu[218], Met[155], Met[161], and Pro[193], and is adjacent and partly overlapped with the fatty acyl substrate-binding site. Indeed, the atoms common to ETH-NAD, PTH-NAD, and INH-NAD are in nearly identical positions. The only difference is the extra ethyl or propyl group at the 2 position of the pyridine ring of ETH or PTH. The ethyl group contributes to the binding of ETH-NAD adduct by forming π-stacking interactions with the aromatic side chain of Tyr[158] at a distance of ~3.3 Å. It is also within van der Waal interaction distances with side chains of Leu[218] (3.3 Å) and Met[155] (3.2 Å). The hydrogen-bonding interactions between the phosphate group of the adduct and residues of the nucleotide-binding site are well conserved. Therefore, it is very likely that mutations, such as S94A, that decrease the binding of NAD(H) and the INH-NAD adduct would also weaken the binding of ETH-NAD and PTH-NAD **(Figure 2-4)**. This explains why the S94A mutant strain of *Mtb* is co-resistant to both INH and ETH.
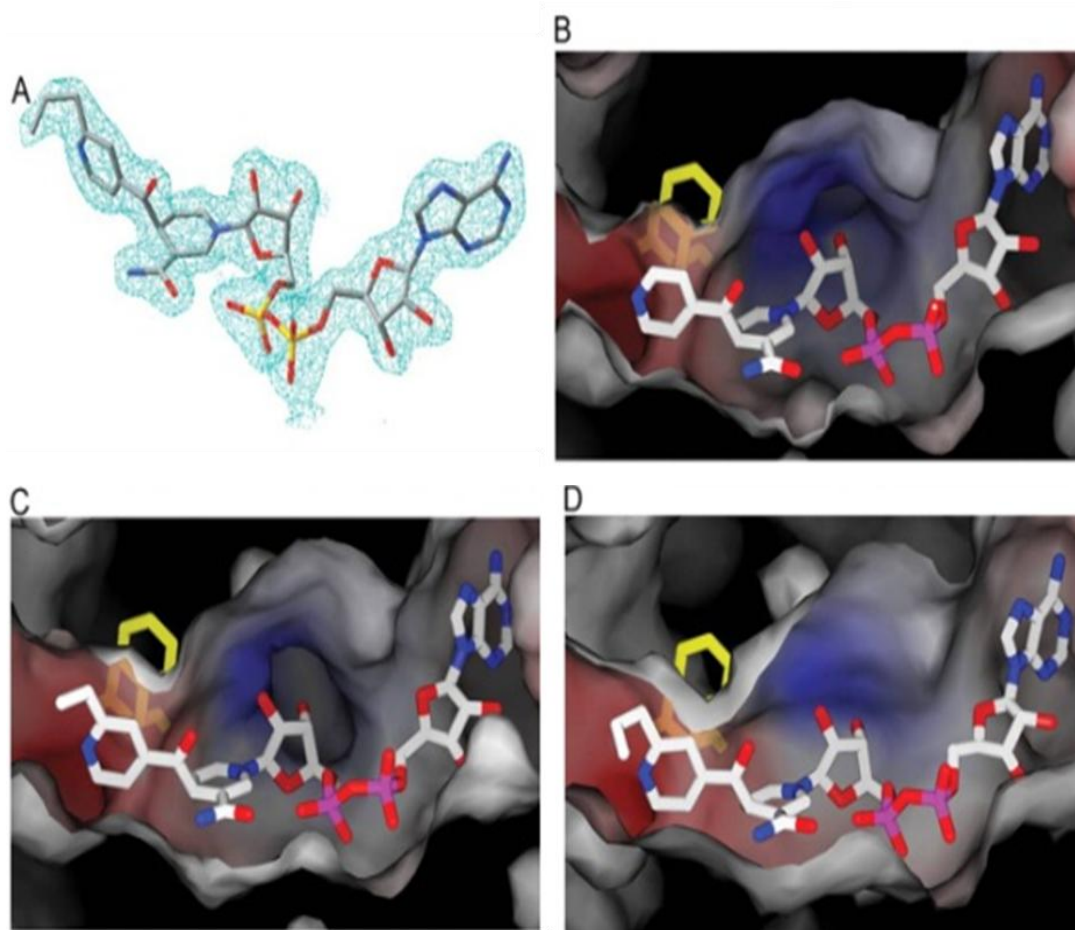
Figure 2-2. Active sites of the *M. tuberculosis* enoyl-acyl ACP reductases bound to inhibitors and the bound inhibitor. A) The crystal structure of PTH-NAD superimposed onto the simulated annealing omit electron density map contoured at 1 σ. Carbon atoms are gray, oxygen atoms are red, nitrogen atoms are blue, and phosphor atoms are orange. The 2-propyl-isonicotinic acyl group is covalently attached to the 4 position of the nicotinamide ring of NADH in a 4S configuration. **B)** Cross section through the surface of the InhA active site with bound INH-NAD. **C)** Cross section through the surface of the InhA active site with bound ETH-NAD showing that the 2-ethyl-isonicotinic acyl moiety protrudes into a hydrophobic binding pocket created by the rearrangement of the side chain of Phe[149] (shown behind the transparent surface), which is similar to INH-NAD. **D)** Cross section through the surface of the InhA active site with bound PTH-NAD, which has a similar binding mode to INH-NAD and ETH-NAD. The carbon atoms of the adduct inhibitors and Phe[149] are white and yellow, respectively.
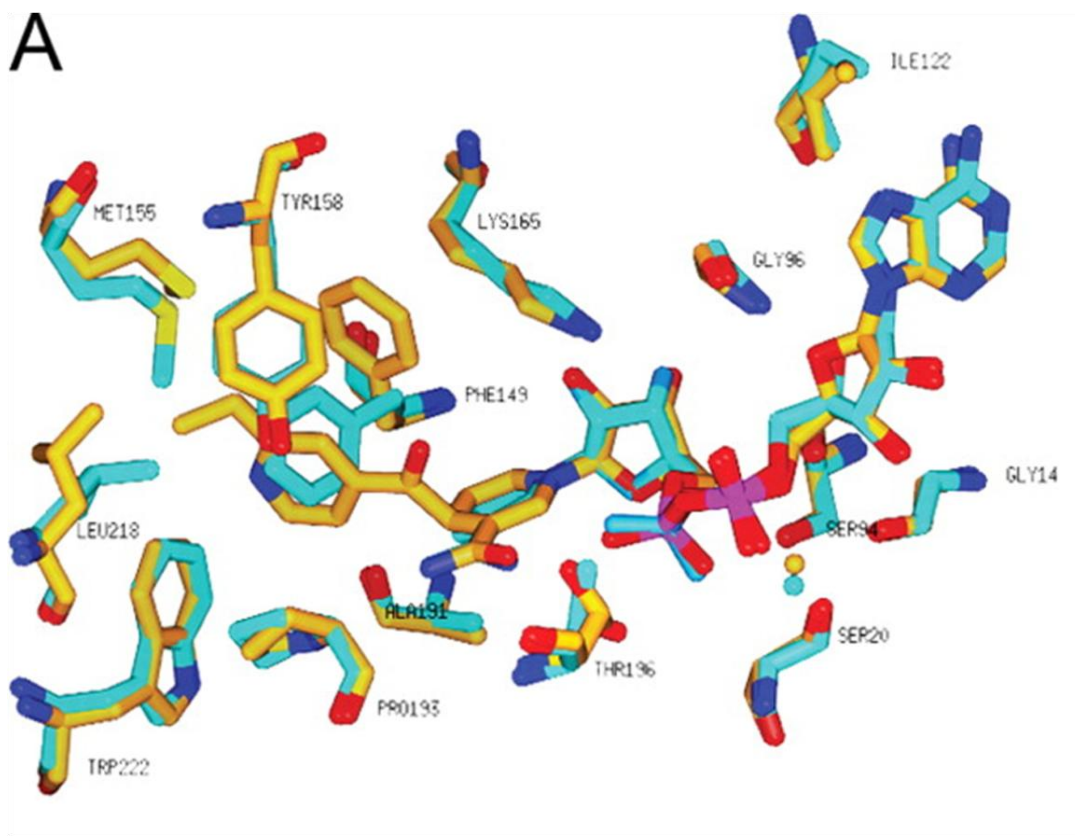
Figure 2-3. *M. tuberculosis* InhA with bound inhibitors. A) Superposition of active sites of the *M. tuberculosis* InhA:NADH structure and the InhA:ETH-NAD structure, showing the side chain of Phe[149] rotated 90° once the ETH-NAD adduct binds to the enzyme. The carbon atoms of residues and NADH in the InhA:NADH structure are cyan. The carbon atoms of residues and ETH-NAD in the InhA:ETH-NAD structure are gold. **B)** The stereo view of the active sites of the *M. leprae* InhA:PTH-NAD structure. The carbon atoms of residues and PTH-NAD adduct are gold and cyan, respectively. Other atoms are colored according to the atom type (red, oxygen atoms; blue, nitrogen atoms; yellow, sulfur atoms; and orange, phosphor atoms).
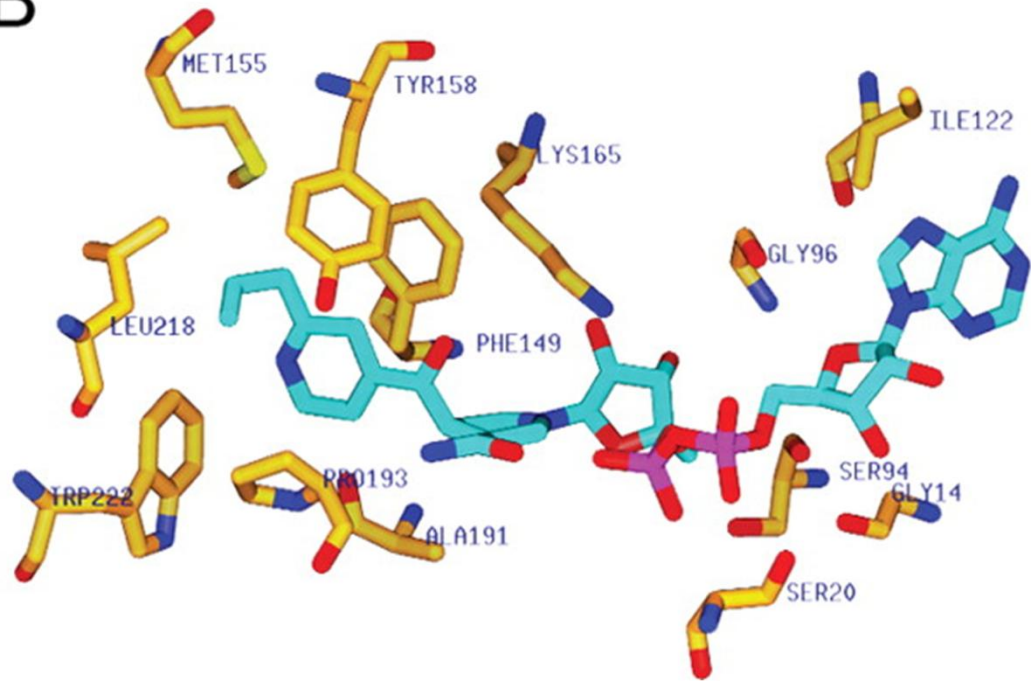
Figure 2-3 Continued**.**

The structures reveal that the active forms of ETH and PTH inhibit *M. leprae* InhA in a similar way to *M. tuberculosis* InhA. Although no clinical or experimental mutant of *M. leprae* InhA has been reported thus far, based on the binding mode of the PTH-NAD adduct, it is very likely that the mutations of InhA found in ETH-resistant *M. tuberculosis* mutant strains, such as S94A, would also confer resistance of *M. leprae* to ETH and PTH.
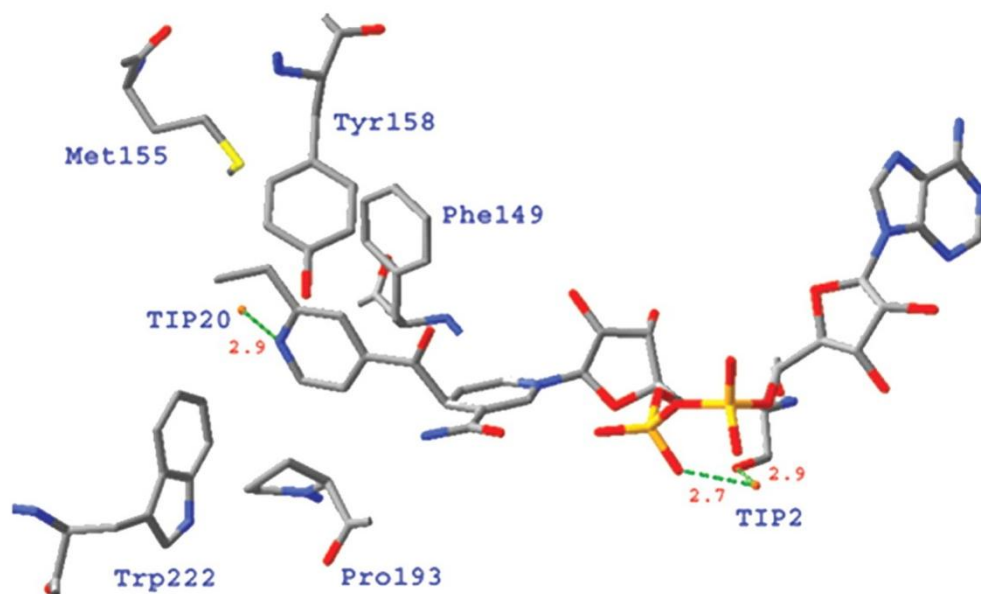
Figure 2-4. Selected interactions between ETH-NAD and the active site of InhA. A conserved water molecule, TIP20, forms a hydrogen bond interaction with the nitrogen atom of the 2-ethyl-isonicotinic acyl moiety of the inhibitor at a distance of 2.9 Å. The other water molecule, TIP2, is in the center of a hydrogen bonding network, which interacts with the oxygen atom of the phosphate group of the adduct and the hydroxyl group of Ser[94] at distances of 2.7 and 2.9 Å, respectively.

It has been proposed that a free radical metabolite intermediate could be generated through EthA oxidation of ETH, similar to the activation of INH. However, no active species that inhibit InhA were isolated *in vitro* (DeBarber et al., 2000), which suggests that an unknown cell component, either a protein or cell membrane, is required for the formation of the adduct by the free radical intermediate. We believe that the inactive metabolites isolated in previous attempts could result from side reactions and quenching of the free radical intermediate in solution. It is still not clear how the

64

thioamide is oxidized by EthA. Tokuyama et al. demonstrated that a thioamide could be used as a precursor of a synthon equivalent to an imidoyl radical in converting thioamides to corresponding indole derivatives. Bu_3SnH/Et_3B has been used as a free radical initiator in pure organic solvent (Tokuyama et al., 1999). Similarly, we postulate that ETH is converted to an imidoyl radical, and this imidoyl radical subsequently attacks NAD^+ to form an adduct, which is then converted to ethyl-isonicotinic-acyl-NAD adduct after hydrolysis to release the amine group. It is also possible that the imidoyl anion is the intermediate before forming the adduct with NAD **(Figure 2-5)**. However, based on the current evidence, we are not certain if this reaction is catalyzed by EthA alone or requires the involvement of additional enzymes.
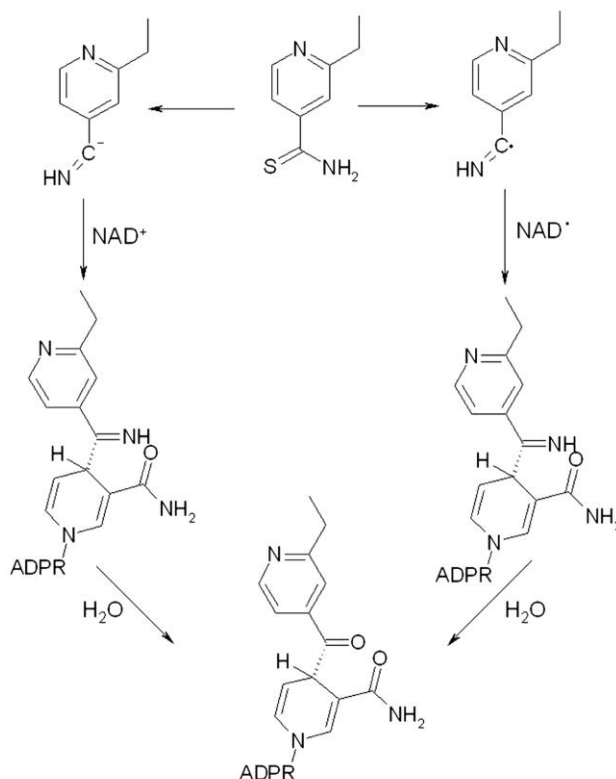
Figure 2-5. Possible reaction mechanisms for the activation of ETH and the formation of ETH-NAD. Two plausible mechanisms for the activation of ETH are shown. Either route will lead to the observed ETH-NAD adduct, retaining a tetrahedral carbon at position 4 of the nicotinamide ring.

INH-NADP adduct formation was not observed in an **E. coli-***based activation system coexpressing katG and dfrA.*

In the previous study, a synthetic INH-NADP adduct derived from INH demonstrated strong inhibition of *M. tuberculosis* DHFR *in vitro* ($K_{i\ app}$ = 1 nM) (Argyrou et al., 2006b). However, the INH-NADP adduct was synthesized by using an inorganic catalyst, Mn(III). Thus, the yield of adduct generated from this approach might

not truly reflect an enzyme-mediated process inside the cell. To better mimic the *in vivo* activation of INH, a cell-based activation system was designed to examine the KatG-catalyzed adduct formation and the inhibition of DHFR by the adduct. In this system, *katG* and *dfrA* were coexpressed in *E. coli* in the presence of INH to investigate whether the activated drug would inhibit DHFR. To construct this system, *katG* and *dfrA* were cotransformed into the *E. coli* BL21(DE3) strain and selected on 50 μg of kanamycin and carbenicillin/ml. The *E. coli* strain containing *katG* and *dfrA* genes was grown and induced in the presence and absence of INH, respectively. After the coexpression of both genes was confirmed by SDS-PAGE, recombinant KatG and DHFR proteins were readily purified.

Mass spectroscopy was used to determine whether DHFR purified from the experimental sample was bound with any inhibitor. Before the mass spectroscopic analysis, the purified DHFR was concentrated and then denatured, followed by filtration to separate small molecules from the denatured protein. The MALDI mass spectrum of the filtrate, which ranged between 200 and 1,200 Da, was carefully analyzed. We were not able to identify any compound that has a molecular mass corresponding to an INH-NADP adduct. The peaks shown on the spectrum most likely resulted from the cofactor NADP and some small fragments of the protein.

An enzyme assay was performed to determine the activity of purified DHFR. DHFR isolated from the experimental sample was found to be fully active (specific activity of 12 μmol mg^{-1} min^{-1}) compared to the enzyme purified from expression in the absence of INH. Since the acyclic 4R INH-NADP adduct is extremely potent against

67

DHFR *in vitro*, it would tightly bind to DHFR if the adduct is indeed generated by KatG catalysis inside the cell. However, both the activity assay result and the mass analysis indicated that no detectable amount of the INH-NADP adduct had bound to DHFR. The coexpressed KatG from the same experimental sample was purified and assayed for its activity *in vitro*. The specific catalase activity of isolated KatG was 17 mol mg^{-1} min^{-1}, a finding comparable to published data (21 mol mg^{-1} min^{-1}) (Lei et al., 2000), which confirmed that the lack of the INH-NADP adduct did not result from the absence of KatG activity. Therefore, the acyclic 4R INH-NADP adduct is not an activated INH product generated by KatG catalysis inside the *E. coli* cell-based system. Thus far, KatG is the only identified activator of INH, and it is very unlikely that INH could be activated by another unknown protein to form the INH-NADP adduct in *M. tuberculosis*. As a result, we conclude that the synthetic INH-NADP adduct is not biologically relevant to INH inhibition.

INH-NAD adduct was detected in the E. coli*-based activation system coexpressing katG and inhA*

It has been shown that KatG activates INH and catalyzes the formation of an INH-NAD adduct *in vitro* (Timmins and Deretic, 2006; Wilming and Johnsson, 1999). In order to demonstrate that this KatG-catalyzed INH-NAD adduct formation and its inhibition of InhA can be reproduced in the *E. coli*-based system, both *katG* and *inhA* were transformed into *E. coli* and coexpressed in the presence of INH. InhA was rapidly purified by a Ni-NTA affinity column, and an *in vitro* enzyme assay was performed.

InhA isolated from the experimental sample had 15% of the specific activity of InhA purified without the addition of INH under the same assay condition. The bound inhibitor was isolated by denaturing the InhA purified from the experimental sample, and a 1 μM concentration of the inhibitor led to complete inhibition of native InhA. The crystal structure of InhA in complex with the inhibitor was solved to 2.4 Å resolution. In the active site of InhA, an unbiased electron density map clearly indicated the presence of a modified NAD with an isonicotinic-acyl group covalently attached to the 4-position of the nicotinamide ring in a 4S configuration, which is consistent with the previously identified INH-NAD adduct (Rozwarski et al., 1998). This is the first time that the activation of INH by KatG and the formation of the INH-NAD adduct has been demonstrated in a whole-cell environment. This confirms that our *E. coli* cell-based system is capable of activating the prodrug INH.

Overexpression of dfrA does not confer resistance to INH

The most direct evidence to support that DHFR is a target of INH comes from the observation that overexpression of *dfrA* in *M. smegmatis* conferred a 2-fold increase in resistance to INH (Argyrou et al., 2006b). Therefore, we first reexamined whether overexpression of this protein using a strong promoter in *M. smegmatis* would lead to increased resistance to INH. Use of cosmids harboring a gene on a multicopy plasmid along with its native promoter might be a desired approach to overexpress putative targets. However, this can result in a failure to observe overexpression whether the expression of the protein is from a tightly regulated promoter. To eliminate the

69

promoter-specific effect, a novel T7-based expression system was developed for the efficient expression of proteins in mycobacteria. In brief, an *M. smegmatis* strain (mc^{2}4517) was generated by integrating a plasmid expressing the T7 polymerase under the acetamidase-inducible promoter into the genome at the mycobacteriophage L5 attachment site. *M. tuberculosis* genes were cloned under the T7 promoter on a separate episomal plasmid pYUB1062. Protein expression was induced by the addition of 0.2% acetamide. To ensure overexpression of *dfrA* and *inhA*, protein lysates were analyzed on SDS gels, and a high level of expression was observed within 6 h of induction as a distinct band corresponding to the expected size compared to the uninduced control **(Figure 2-6)**. The strains overexpressing *dfrA* (mc^{2}5097) and *inhA* (mc^{2}5089) were then tested for their susceptibility to INH and ETH and the results obtained are summarized in **Table 2-1**. The MIC for the *inhA* overexpression strain mc^{2}5089 increased more than 60-fold to 300 μg/ml for INH, compared to the wild-type strain. In contrast to InhA, we did not observe any increase in MIC for mc^{2}5097, the strain overexpressing DHFR, which had been reported to confer 2-fold resistance to INH when induced in *M. smegmatis* with an acetamidase promoter (Argyrou et al., 2006b). Other than DHFR and InhA, 16 additional proteins from *M. tuberculosis* were identified in a previous study (Argyrou et al., 2006a) that bound to an INH-NAD or INH- NADP adduct *in vitro*. We applied the same T7 promoter system to successfully overexpress 14 of these genes in *M. smegmatis*, but none of them could confer resistance to INH or ETH **(Table 2-1)**. Since this analysis was done in liquid culture, we also tested the ability of the strains to grow on plates with two times the MIC for INH. As shown in **Figure 2-7**, the *M. smeg-*

70
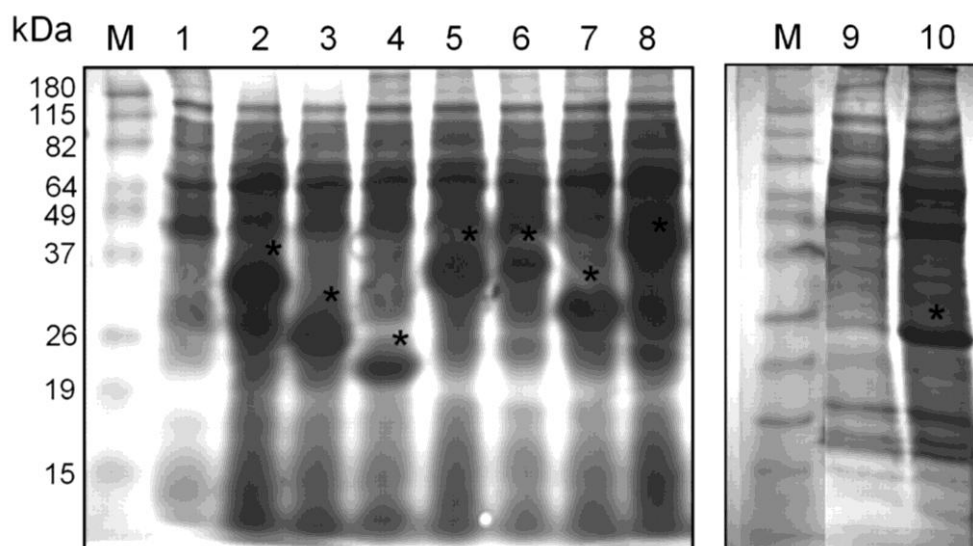
Figure 2-6. Overexpression of the proteins in *M. smegmatis* confirmed by SDS-PAGE analysis. Lanes: M, marker; 1, vector; 2, Rv0155 (PntAA, 40 kDa); 3, Rv1484 (InhA, 29.5 kDa); 4, Rv2763c (DfrA, 17.4 kDa); 5, Rv1996 (34.9 kDa); 6, Rv2623 (32.7 kDa); 7, Rv2766c (FabG5, 30.8 kDa); 8, Rv2858 (AldC, 50 kDa); 9, uninduced Rv2763c (DfrA); 10, induced Rv2763c (DfrA).

Table 2-1. MICs determined in 7H9 broth by serial dilution method after overexpressing the INH-NAD binding proteins from *M. tuberculosis* H37Rv in *M. smegmatis*.

Overexpressed protein	MIC (μg/ml)[a]	
	INH	ETH
None (mc²4517 parental strain)	4.7	9.4
Rv1484 (InhA)	300	150
Rv2763c (DfrA)	4.7	9.4
Rv3248c (SahH)	4.7	9.4
Rv0753c (MmsA)	4.7	9.4
Rv1187 (RocA)	4.7	9.4
Rv0155 (PntAA)	4.7	9.4
Rv2623 (universal stress protein)	4.7	9.4
Rv1996 (universal stress protein)	4.7	9.4
Rv0468 (FadB2)	4.7	9.4
Rv2691 (CeoB/TrkA)	4.7	9.4
Rv0091 (Mtn/Sah)	NA	NA
Rv2858c (AldC)	4.7	9.4
Rv1059 (unknown)	4.7	9.4
Rv0926c (unknown)	4.7	9.4
Rv3777 (probable oxidoreductase)	4.7	9.4
Rv2971 (probable oxidoreductase)	NA	NA
Rv2766 (FabG5)	4.7	9.4
Rv2671 (RibD)	4.7	9.4

[a] NA, not available (overexpression was not successful).

matis strain containing the overexpressed DHFR (mc^{2}5072) failed to grow on an INH-containing plate, while the strain overexpressing InhA (mc^{2}5071) showed complete resistance. To find out whether DHFR was indeed expressed, the strain containing the overexpressed DHFR was plated on medium containing the known DHFR inhibitor trimethoprim. Strain mc^{2}5072 (DHFR overexpresser) was able to grow on medium containing 50 or 150 μg of trimethoprim/ml. In contrast, the parental strain mc^{2}5069 or the strain overexpressing InhA (mc^{2}5071) failed to grow on trimethoprim-containing plates. One possible explanation for the difference between our findings and those reported previously (Argyrou et al., 2006b) might be that the growth inhibition experiment in the previous study was carried out at 30°C, which is not optimal for growth, whereas the growth inhibition experiments in our study were performed at 37°C. Our observation that the overexpression of *M. tuberculosis* InhA but not DHFR confers resistance to INH and ETH in *M. smegmatis* using this new expression system is consistent with InhA being the primary target of INH and ETH. Even though DHFR has been shown to bind an INH-NADP adduct *in vitro*, this phenomenon is not sufficient to confer INH or ETH resistance in mycobacterial cells.
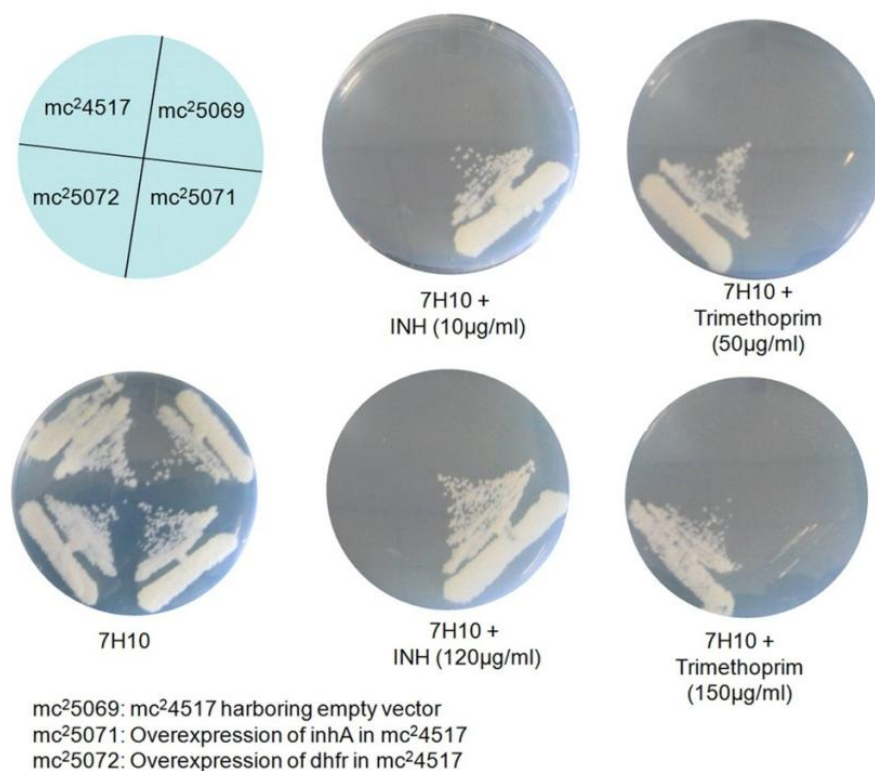
Figure 2-7. Growth of the wild-type *M. smegmatis* strain and transformed strains overexpressing DHFR or InhA on media containing the known DHFR inhibitor trimethoprim (50 or 150 µg/ml) or INH (10 or 120 µg/ml).

A second overexpression experiment was conducted using the H37Rv strain of *M. tuberculosis*, which is more disease relevant, compared to *M. smegmatis*. The *M. tuberculosis* H37Rv strain was transformed with the pMV261::*dfrA* plasmid. The MIC of INH for the *M. tuberculosis* strain overexpressing *dfrA* is identical to that of the wild-type H37Rv strain (0.06 µg/ml). On plates with three times the MIC for INH, no growth of the *M. tuberculosis* strain overexpressing *dfrA* was observed. In contrast, the *M. tuberculosis* strain overexpressing *inhA* survived **(Figure 2-8)**, which is consistent with

74

the previous report that the MIC for *M. tuberculosis* overexpressing *inhA* (1.0 µg/ml) increased 20-fold compared to the wild-type H37Rv strain (MIC 0.06 µg/ml) (Larsen et al., 2002). Therefore, similar to what we observed in our *M. smegmatis* experiment, overexpression of *dfrA* did not prevent growth inhibition of *M. tuberculosis* by INH, which genetically demonstrates that *M. tuberculosis* DHFR is not a relevant target of INH.

Absence of polymorphisms in dfrA in INH-resistant clinical isolates

As an alternative approach to determining whether DHFR might be a relevant target of INH, we used whole-genome sequencing to determine whether mutations in *dfrA* could be observed in INH-resistant clinical isolates. If DHFR were a target of INH, then it might be expected that mutations might occur in *dfrA* in INH-resistant clinical isolates lacking mutations in *inhA* or *katG*. Although mutations in *inhA* or *katG* and their promoters account for most cases of resistance to INH, the remaining 10 to 25% of cases must have mutations in other genes relevant to INH toxicity (Zhang et al., 1993).
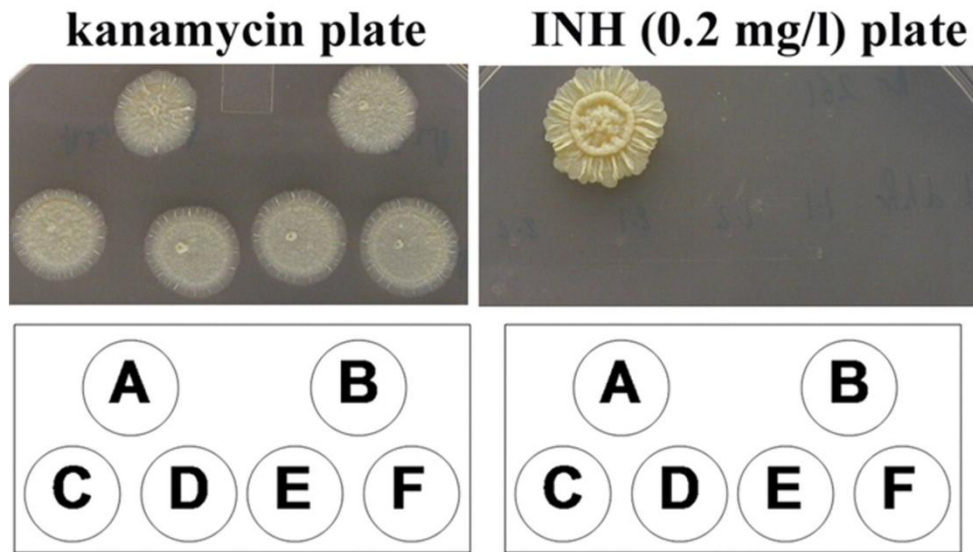
Figure 2-8. Growth of *M. tuberculosis* **H37Rv strains overexpressing DHFR or InhA on media containing INH (0.2 μg/ml) or kanamycin (100 μg/ml).** Spots: A, pMV261::*inhA* (InhA overexpresser); B, pMV261 (plasmid only); C to F, four independent transformants of pMV261::*dfrA* (DHFR overexpresser).

We sequenced the genomes of six clinical isolates that were INH resistant but known not to have relevant mutations in *inhA*, *katG*, or their promoters. The six strains were sequenced by using an Illumina Genome Analyzer II (see Materials and Methods). The depth of coverage (i.e., the number of reads overlapping each site, averaged over the whole genome) was $36.9x$ to $49.4x$ for each strain, providing high confidence in the final base calls. This method has been used successfully to identify known and novel mutations associated with resistance to INH, rifampin, pyrazinamide, kanamycin, ofloxacin, streptomycin, etc., in many other strains sequenced in our lab.

Coding regions for the 17 putative INH-NADH-binding proteins identified

earlier (Argyrou et al., 2006a) were searched for polymorphisms. The 100-bp regions upstream of the operon containing each gene were also searched for possible SNPs in the promoter region that might affect expression level. None of the six clinical isolates harbored any mutations in *dfrA* or its promoter **(Table 2-2)**. Furthermore, few nonsynonymous mutations were found in any of the other 16 genes identified as potential proteins that can bind an INH-NAD(P)H adduct. Strain 5358 showed the most mutations, due to its evolutionary distance from the others (5358 belongs to the *M. bovis* family), but most of the mutations in strain 5358 were also observed in the genome for INH-sensitive *M. bovis* BCG, implying that they are not responsible for INH resistance. The only sites where there are potentially relevant differences (i.e., nonsynonymous and not in BCG) are a 931-bp deletion of *mtn* in 5400, *aldC*(T21A) in 5297, the a-64c promoter mutation upstream of Rv1059, the 1-bp frameshift mutation in Rv1059 in 5297, and Rv3777:P101A in 5297.

For completeness, other genes traditionally associated with INH resistance were also searched, including *katG*, *iniC*, *ndh*, *ahpC*, etc. (Hazbon et al., 2006), but no known mutations responsible for INH resistance could be identified **(Table 2-3)**. Two strains showed mutations in *katG* (F129S in strain 5400 and G285V in strain 5324), and one strain has a mutation in *efpA* (Q513R in strain 5324). None of these mutations has been previously reported in INH-resistant strains. The *katG* F129S mutation is located at the dimer interface, and the *katG* G285V mutation is located at the mouth of the active site entrance, so their effect on INH activation is unclear. Mutations in *mshA*, which catalyzes the first step in mycothiol biosynthesis, have also been linked with INH and

Table 2-2. Mutations in genes identified as potential INH-NAD(P) binders[a]

	5071	5072	Mutation in strain 5297	5324	5358	5400	*M.bovis* BCG
Protein							
Rv2763c/*dfrA*	None	None	None	None	None	None	None
Rv3248c/*sahH*	None	None	None	None	None	None	None
Rv0753c/*mmsA*	None	None	None	None	None	None	None
Rv1187/*rocA*	None	None	None	None	None	None	None
Rv0155/*pntAa*	A274A	None	None	None	None	None	None
Rv2623	None	None	None	None	None	None	None
Rv1996	None	None	None	None	None	None	None
Rv0468/*fadB2*	None	None	None	None	L87L	None	L87L
Rv2691/*ceoB*	T117A	T117A	T117A	T117A	T117A	T117A	T117A
Rv0091/*mtn*	None	A155P	None	Q174Q	None	931-bp deletion	None
Rv2858c/*aldC*	None	None	T21A	None	P77P	None	P77P
Rv1059	a-64c[b]	1-bp deletion in P99	None	None	None	None	None
Rv3777	None	None	P101	None	None	None	None
Rv3777	V160A	V160A	V160A	V160A	V160A	V160A	V160A
Rv3777	L63L	L63L	L63L	L63L	L63L	L63L	L63L
Rv0926c	None	None	None	None	None	None	None
Rv2766c/*fabG5*	None	None	None	None	None	None	None
Rv2671/*ribD*	None	None	None	None	None	None	None
Rv2971	None	None	None	None	N152H	None	N152H

[a]Argyrou et al. (Argyrou et al., 2006a). Mutations are indicated relative to the amino acid in the H37Rv reference sequence. Mutations that are shared with *M. bovis* BCG are assumed not to cause isoniazid resistance. [b]Single nucleotide polymorphism in promoter region.

Table 2-3. Mutations in genes traditionally associated with INH resistance[a]

Protein	5071	5072	5297	5324	5358	5400	*M.bovis* BCG
					Mutation(s) in strain:		
Rv1484/*inhA*	None	None	None	None	None	None	None
Rv1908c/*katG*	None	None	None	G285V	R463L	F129S	R463L
Rv1909c/*furA*	None	None	None	None	A46V	None	A46V
Rv1854c/*ndh*	None	None	None	None	G313R	None	G313R
Rv0342/*iniA*	None	None	None	None	N88S, H481Q	None	N88S, H481Q
Rv0341/*iniB*	None	None	None	None	None	None	None
Rv0343/*iniC*	None	None	None	None	None	None	None
Rv3139/*fadE24*	None	None	None	None	None	None	None
Rv2245/*kasA*	None	None	None	None	None	None	None
Rv2246/*kasB*	None	None	None	None	None	None	None
Rv2428/*ahpC*	None	None	None	None	None	None	None
Rv2242	None	None	None	None	A363T	None	A363T
Rv0340	None	None	None	None	None	None	None
Rv1592c	I322V	I322V	I322V	I322V	I322V	I322V	I322V
Rv1772	None	None	None	None	Deletion	None	Deletion
Rv2846c/*efpA*	None	None	None	Q513R	T15R	None	T15R
Rv0486/*mshA*	N111S	N111S	None	N111S	None	N111S	None

[a] Mutations are indicated relative to the amino acid in the H37Rv reference sequence. Synonymous mutations are excluded.

ETH resistance, but the N111S allele appears in the Erdman strain, which is as susceptible to these drugs as H37Rv (Vilcheze et al., 2008). Although the true cause of INH resistance in these six strains remains undetermined, the lack of mutations in *dfrA* and the 16 other putative INH-NAD(P)-binding proteins suggests that they are unlikely to be targets of INH. We also have sequenced the *dfrA* gene in six additional *M. tuberculosis* strains known to be INH-resistant but with no known relevant mutations. None of these had a mutation in the *dfrA* gene. Consistent with our findings, a recent large-scale screening of 127 INH-resistant *M. tuberculosis* isolates from Singapore by PCR sequencing also revealed no mutations within *dfrA* (Ho et al., 2009).

Experimental methods

Cloning, expression, and purification

The *M. tuberculosis dfrA*, *katG* and *inhA* genes were cloned as previously described (Argyrou et al., 2006b; Dessen et al., 1995). *M. tuberculosis ethA* was cloned from genomic DNA (National Institutes of Health contract N01-AI-75320; Colorado State University). The amplified product was inserted into pET28b using the NdeI and NotI restriction sites. *M. leprae ethA* and *inhA* were cloned from genomic DNA. The amplified product of *M. leprae ethA* was inserted into pET15b using the NdeI and BamHI restriction sites. *M. leprae inhA* was inserted into pET30b using the NdeI and HindIII restriction sites.

The plasmids of *M. tuberculosis katG* and *dfrA, inhA and katG, inhA* and *ethA,*

and M.leprae inhA and *ethA* were singly and doubly transformed into *E. coli* BL21(DE3) (EMD Bioscience, catalog no. 69387-3). The strains containing both plasmids were cultured in LB-Miller medium containing 50 μg/ml kanamycin and 50 *μ*g/ml carbenicillin at 37°C until OD_{600} reached 0.5. Expression of both genes was induced for 20 h at 16°C by addition of 1 mM isopropyl β-D-thiogalactopyranoside. At the same time of induction, 100 μg/ml of INH or ETH or PTH was also added to the culture. The same protocol was used for the strains containing just the *dfrA* or *inhA* plasmids.

Recombinant *M. tuberculosis* DHFR and KatG was purified according to a previously described method (Argyrou et al., 2006b; Zhao et al., 2006). Recombinant InhA was purified according to the method described in Chapters III and IV. The coexpression and purification of *M. leprae ethA* and *inhA* were conducted using protocols similar to those used for the *M. tuberculosis* enzymes.

Mass spectroscopy analysis for adduct identification

InhA purified from the experimental strain containing both *inhA* and *ethA* genes was concentrated and heated for 40 s at 100°C. After the heat treatment, ETH-NAD or PTH-NAD was separated from denatured enzymes by filtration, using a centricon device (cutoff size = 30 kD). The concentration of ETH-NAD and PTH-NAD was determined by its absorbance at 260 and 326 nM (Lei et al., 2000). The molecular weight of both adducts was determined by matrix-assisted laser desorption/ionization (MALDI) performed on an ABI Voyager-DE STR (AME Bioscience): ETH-NAD, calculated weight = 797.2 and found weight = 797.3 (negative mode), and calculated weight =

799.2 and found weight = 799.2 (positive mode); PTH-NAD, calculated weight = 811.2 and found weight = 811.3 (negative mode). Mass spectrometry analysis for DHFR was performed in a similar way. Purified DHFR was heated for 60 s at 100°C. After the heat treatment, denatured enzyme was separated by filtration, using a Centricon (cutoff, 3 kDa). The filtrate was used for MALDI experiment.

DHFR, KatG and InhA enzymatic activity assays

All assays were carried out on a Cary 100 Bio Spectrophotometer at 25°C. DHFR assays are performed by monitoring the oxidation of NADPH and reduction of dihydrofolate (DHF) at 340 nm. Reactions were initiated by adding DHFR (10 nM) to assay mixtures containing NADPH (10 μM), DHF (4.5 μM), and phosphate buffer (pH 7.5, 50 mM). The KatG activity was assayed as previously described (Zhao et al., 2006). InhA activity was monitored by oxidation of NADH at 340 nm. Reactions were initiated by adding 50 μM of substrate dodecenoyl-CoA to assay mixtures containing 1 nM InhA, 100 μM NADH, and 3–2,000 nM of adduct inhibitors. The IC_{50} was determined from the dose-response plot of enzyme fractional activity as a function of inhibitor concentration. K_i was obtained by dividing the IC_{50} value by $1 + [S_1]/K_{m1} + [S_2]/K_{m2}$, where $[S_1]$ and $[S_2]$ are the concentrations of dodecenoyl-CoA and NADH, and K_{m1} and K_{m2} are their Michaelis constants.

Crystallization of InhA in complex with the INH-NAD, ETH-NAD and PTH-NAD adducts

Crystallization was accomplished by the hanging-drop vapor diffusion method. *Mtb or M. leprae* InhA in complex with INH-NAD, ETH-NAD or PTH-NAD was cocrystallized in hanging droplets containing 2 µl of protein solution at 10 mg/ml and with 2 µl of buffer (12% 2-methyl-2,4-pentanediol, 4% dimethyl sulfoxide, 0.1 M HEPES, and 0.025 M sodium citrate) at 16°C against 1 ml of the same buffer. Diamond or cubic shaped protein crystals appeared 4 days later.

Data collection and processing

Data were collected at 121 K using cryoprotection solution containing reservoir solution with an additional 30% MPD. Crystals of *M. tuberculosis* InhA:INH-NAD, InhA:ETH-NAD and *M. leprae* InhA:PTH-NAD diffracted x-rays to 2.4, 2.2 and 1.8 Å using the beamline 23-ID at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Diffraction data were collected from a single crystal with 1° oscillation widths for a range of 120°. Crystals of *M. tuberculosis* InhA:PTH-NAD were diffracted to 2.5 Å using a Raxis image plate detector coupled to a Rigaku x-ray generator using a copper rotating anode (CuK$_\alpha$, λ = 1.54 Å). The data were integrated and reduced using HKL-2000 (HKL Research, Inc.) (Otwinowski and Minor, 1997).

Structure determination and model refinement

Crystals produced from InhA in complex with ETH-NAD were isomorphous to

those of the native enzyme. Initial phases were obtained by molecular replacement using the apo-InhA structure (1ENY) and refined with CNS software (Brunger et al., 1998). F_o – F_c and $2F_o$ – F_c electron density maps were calculated, and an additional density resembling the inhibitor was found. The ligand was fit into the additional density, and the whole model was rebuilt using XtalView (McRee, 1999). During the final cycles of the refinement, water molecules were added into peaks above 3 σ of the F_o − F_c electron density maps that were within hydrogen-bonding distances from the appropriate protein atoms.

T7-based expression system for mycobacteria

Construction of M. smegmatis *expression strain, mc²4517*

pT7Pol26 was digested with BamHI to excise the RNA polymerase gene of bacteriophage T7. The 2.7-kb BamHI DNA fragment was ligated to BamHI-digested pSD26 plasmid, and the orientation was confirmed by digestion with the appropriate restriction enzymes. The DNA fragment of an acetamidase promoter fused to T7 polymerase was then cloned into pMV306, to generate pYUB1232. The expression strain mc²4517 was obtained after electroporating pYUB1232 into *M. smegmatis* mc²155.

Construction of expression plasmids

pYUB1062, the parental expression plasmid, was constructed by combining the

NarI/SspI DNA fragment (948 bp) from pET30a plasmid and the NarI/EcoRV DNA fragment (4,026 bp) of pMV206. Shuttle vector pYUB1062 contains an *oriE* for replication in *E. coli*, *oriM* for replication in mycobacteria, a hygromycin cassette for antibiotic selection, a T7 promoter/terminator, and a histidine tag region for the purification of overproduced protein. The various *M. tuberculosis* genes were cloned into pYUB1062 using NdeI or NcoI as the upstream site and an appropriate down-stream site.

Expression of INH-NAD- or INH-NADP-binding proteins from M. tuberculosis

M. smegmatis mc^{2}4517 cells were transformed with plasmids expressing INH-NAD- or INH-NADP-binding protein of *M. tuberculosis*. A single bacterial colony was grown in all cases at 37°C to an A_{600} of 0.6 to 0.8 in Middlebrook 7H9 broth medium supplemented with 0.2% Tween 80, 0.05% glycerol, kanamycin (20 μg/ml), and hygromycin B (100 μg/ml). Cell cultures were diluted 1:100 in 50 ml of the same liquid medium and cells were grown to mid-exponential phase (A_{600} 0.5 to 0.6) at 37°C at 120 rpm before induction with acetamide. The bacterial cultures were divided into equal volumes, and 0.2% (wt/vol) acetamide was added to the one of the aliquots (the other aliquot was used as a control). Expression of protein was determined after 6 h of induction by analyzing the cell-free protein extract by 10 to 20% SDS-PAGE.

MIC determination

M. smegmatis cells expressing the *M. tuberculosis* proteins were grown to an OD

of 0.6 and then induced with 0.2% acetamide for 6 h until the OD reached 1.0. The cells were then diluted by 10^5-fold and incubated in 200 μl of 7H9 medium containing INH (0, 0.6, 1.2, 4.7, 9.4, 18.6, 37.5, 75, 150, and 300 μg/ml) in a 96-well plate. The plates were incubated at 37°C and observed after 72 h. The MIC was determined as the concentration of drug at which no visible growth was observed. MICs for *M. tuberculosis* strains were determined by using an MTT assay (Martin et al., 2005). The concentrations of INH tested were 0, 0.015, 0.03, 0.06, 0.12, 0.24, 0.48, 0.96, 1.92, and 3.84 μg/ml.

Sequencing of clinical isolates

Six INH-resistant clinical isolates (strains 5071, 5072, 5297, 5324, 5358, and 5400) were obtained for whole-genome sequencing. The clinical isolates were collected from Mexico between 1991 and 1996 by the Laboratory of Clinical Microbiology, Instituto Nacional de Ciencias Me´dicasy Nutricio´n Salvador Zubira´n, Mexico City, under an IRB-approved protocol. Five of the six strains were genotyped as belonging to SNP cluster group SCG_3b, whereas strain 5358 belongs to the *M. bovis* family based on spoligotyping. MICs for INH ranged from 0.2 to 8.0 μg/ml compared to 0.03 μg/ml for H37Rv.

Genomic DNA was extracted by using a CTAB (cetyltrimethylammonium bromide) protocol as previously described (Larsen et al., 2007). The DNA library was constructed by using a genomic DNA sample preparation kit (Illumina). The sample was first fragmented by using a nebulization technique. Then, the double-stranded DNA

fragments comprised of 3 or 5 overhangs were converted into blunt ends, using T4 DNA polymerase and Klenow enzyme. Klenow 3 -to-5 exonuclease minus was used to add an "A" base to the phosphorylated 3 blunt end of the DNA fragments so that the fragments could be ligated to the adaptors, which have a single "T" base overhang at their 3 end. The ligated DNA was then size selected on a 2% agarose gel. DNA fragments of 300 bp were excised from the preparative portion of the gel. DNA was then recovered by using a Qiagen gel extraction kit and PCR amplified to produce the final DNA library. A total of 5 pmol of DNA from each strain was loaded onto a different lane of the sequencing chip (eight lanes total), and the clusters were generated on the cluster generation station of a Genome Analyzer II using the Illumina Cluster generation kit. Bacteriophage X174 DNA was used as a control.

The sequencing reaction was run for 36 cycles (tagging, imaging, and cleavage of one terminal base at a time), and four images of each tile on the chip were taken in different wavelengths for exciting each base-specific fluorophore. Image analysis and base-calling were done by using v0.3 of the Illumina GA Pipeline software.

The 36 bp reads that were generated for each strain were mapped (aligned) against H37Rv as a reference sequence (or *M. bovis* in the case of strain 5358, which was the closest reference strain by spoligotyping). Apparent differences (at sites where the consensus base from overlapping reads differed from the expected base in the reference sequence), along with sites where coverage was low or observed bases were heterogeneous, were identified, and local contig-building was used to resolve them into SNPs versus indels, based on alignment to the corresponding region in the reference

genome.

Additional unpublished results

Coexpression experiments for InhA-EthA, Dhfr-KatG and Dhfr-EthA

Initial coexpression trials for InhA and EthA or Dhfr and KatG were performed using a duet expression vector. The corresponding genes were successively cloned into the multiple cloning sites of the pETDuet (Novagen) vector. Each gene sequence was verified by DNA sequencing. Individual colonies harboring the InhA-EthA, Dhfr-KatG, or Dhfr-EthA pETDuet plasmids were inoculated into Luria broth (LB) medium. Coexpression was tested in both small (5 ml) and large scale (2 L) cultures. The cells were grown till the log phase and were induced with IPTG at 18 °C and 37 °C. Samples were aliquoted from each flask and subjected to SDS-PAGE analysis. However, it was found that coexpression from a single plasmid did not yield either a good amount of protein expression or soluble protein **(Figure 2-9B)**. Therefore, coexpression experiments were performed by co-transformation of two individual plasmids carrying each gene into the same *E. coli* cell. *Mtb* and *M.leprae* InhA and EthA, or *Mtb* Dhfr and KatG were cloned into individual expression vectors with different antibiotic selection markers (pET15b with carbenecillin marker and pET28b or pET30b (Novagen) with kanamycin marker), which were simultaneously cotransformed to *E. coli* BL21(DE3) cells. This approach yielded sufficient amount of soluble InhA, Dhfr and KatG proteins required for the subsequent analysis and experiments.
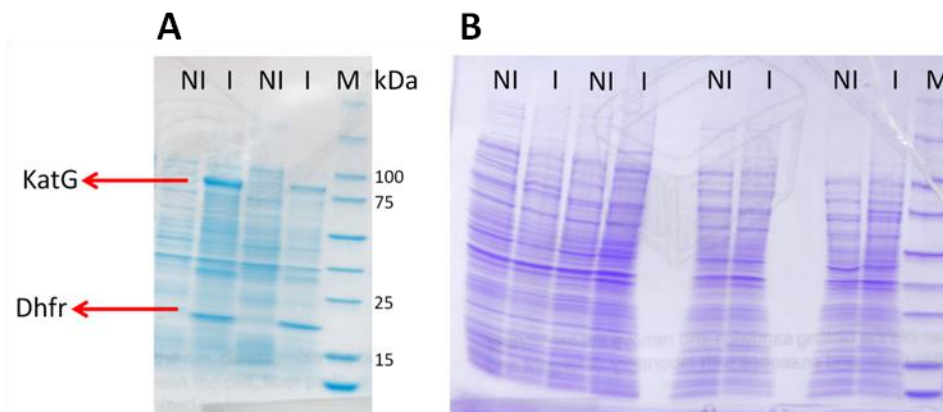
Figure 2-9. Recombinant coexpression of *Mtb* Dhfr and KatG proteins in *E.coli*. A) Plasmids carrying Dhfr and KatG genes were co-transformed to *E.coli* BL21(DE3) cells. Two individual colonies carrying both plasmids were inoculated into LB medium and protein expression was induced by 1 mM IPTG. Over-expression of both proteins (Dhfr-17.5 kDa and KatG- 80 kDa) was clearly visible in the SDS-PAGE gel. **B)** Dhfr and KatG genes were cloned into pETDuet vector and the presence of both inserts were verified by DNA sequencing. Four individual colonies were inoculated into LB medium and expression was induced by 1 mM IPTG. As seen from the SDS-PAGE gel, there was no over-expression for either protein. For both panels, NI represents before IPTG induction (non-induced), I is after IPTG induction (induced).

We have shown that both INH and ETH target InhA by forming structurally similar adducts with NAD^+. Since it was proposed that Dhfr was inhibited by the INH-NADP adduct, we tested if the same hypothesis can hold for ETH, which is a structural analog of INH. Similar to the Dhfr/KatG coexpression experiment, Dhfr and EthA were coexpressed in the presence of 100 μg/ml ETH, to determine if the cell activated ETH can inhibit Dhfr. However, despite the optimization of induction time, temperature and the expression host cell line, no soluble Dhfr protein was obtained when it was coexpressed with EthA **(Figure 2-10)**. It is very likely that the ETH-NAD adduct will

89

also not target Dhfr, as we have shown that the INH-NAD adduct does not bind to Dhfr *in vivo* (see Results and Discussion for details).
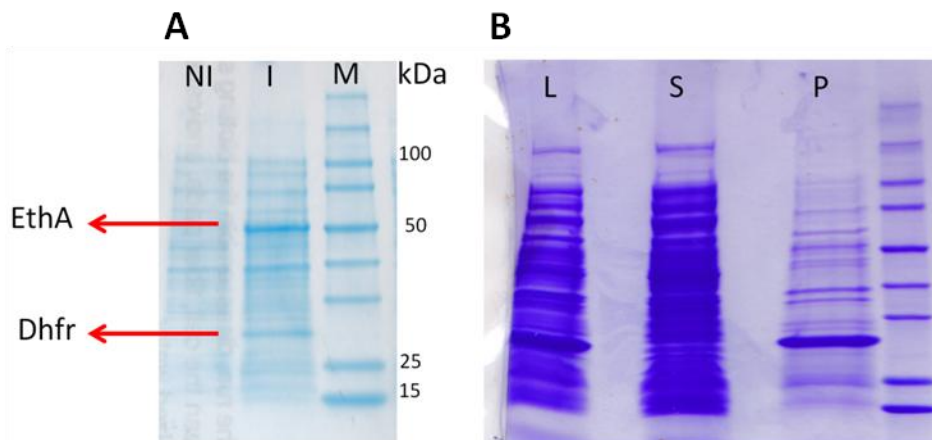
Figure 2-10. Recombinant coexpression of *Mtb* Dhfr and EthA proteins in *E.coli*. A) The plasmids carrying the *dhfr* and *ethA* genes were co-transformed to *E. coli* BL21(DE3) cells. A single colony carrying both plasmids was inoculated into 5 ml LB medium and protein expression was induced by 1 mM IPTG. Over-expression of both proteins (Dhfr- 17.5 kDa and EthA- 55 kDa) was clearly visible in the SDS-PAGE gel. **B)** Dhfr and EthA were overexpressed in the presence of ETH. Cells were lysed by French press and spinned at 16K rpm. The lysate (L), supernatant (S) and pellet (P) were run on a SDS-PAGE gel. As concluded from the gel, the majority of the Dhfr protein was in pellet (insoluble).

Inhibition of InhA activity by the cell-activated ETH

After coexpression of *Mtb* InhA and EthA in the presence of 100 μg/ml ETH, InhA was purified and the enzymatic activity was monitored. As a control experiment, either InhA was singly expressed in the presence of ETH or InhA and EthA were coexpressed in the absence of ETH. The InhA protein isolated from the experimental

sample (InhA+EthA+ETH) displayed very low (<1%) enzymatic activity **(Figure 2-11)**.

In contrast, the control experiments, where InhA and EthA were coexpressed in the absence of ETH, or InhA was singly expressed in the presence of ETH, showed that InhA activity was retained in these situations **(Figure 2-11)**. This demonstrated that ETH was a pro-drug and it could not inhibit InhA without activation by EthA. In addition, InhA inhibition was solely due to the *in vivo* activated ETH molecule (ETH-NAD adduct), and it was not an artifact caused by coexpression or the experimental design.
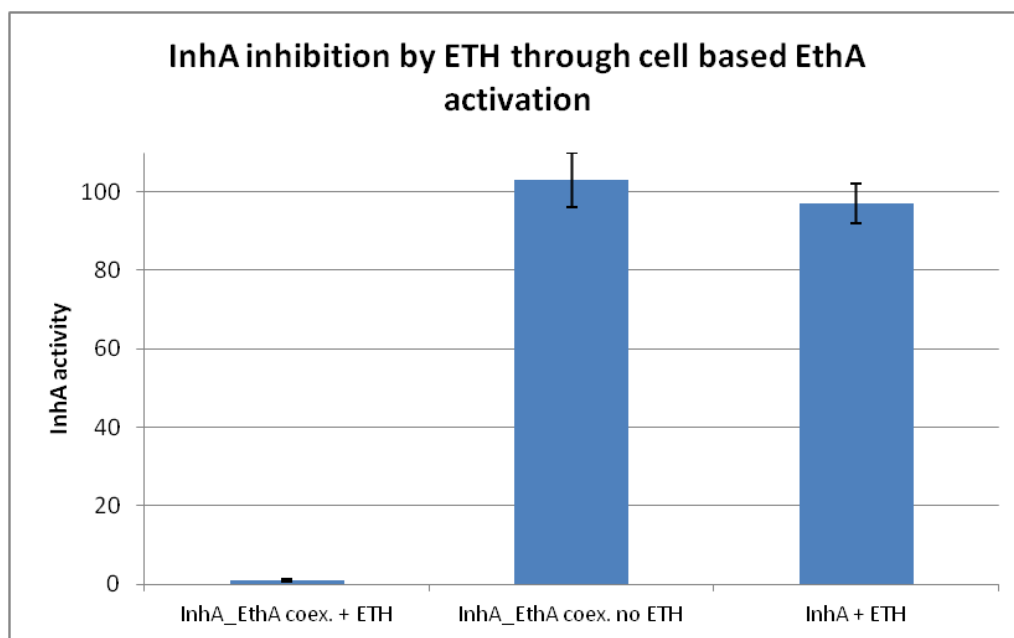
Figure 2-11. InhA inhibition after coexpression with EthA in the presence of ETH. InhA activity was inhibited only by the *in vivo* activated ETH.

Summary

Isoniazid (INH) and ethionamide (ETH) have been used for the treatment of *M. tuberculosis* and *M. leprae* infections, the causative agents of tuberculosis and leprosy diseases, respectively. Both INH and ETH are prodrugs requiring activation inside the cell. The mechanism of action for INH has been established. INH is activated by the catalase-peroxidase KatG enzyme and forms an adduct with NAD^+ in the 4-position of the nicotinamide ring in an S configuration. In contrast, the active form of ETH has not been demonstrated before this study. The enoyl-reductase enzyme, InhA, has been shown to be the molecular target of both drugs. Recently, it was proposed that the dihydrofolate reductase enzyme, Dhfr, of *Mtb* was also a target of INH, based on the findings that a non-enzymatically generated 4R-INH-NADP adduct strongly inhibited Dhfr *in vitro*, and Dhfr overexpression in *M. Smegmatis* confered 2-fold resistance to INH. In this study, an *E. coli* cell-based activation system was developed, which allowed the coexpression of the activator and the target gene of a drug in the presence of the drug compound. To determine the active form of ETH, *Mtb* and *M. leprae* InhA and EthA proteins were co-overexpressed in the presence of ETH, and the active form of ETH was isolated and characterized by mass spectrometry and X-ray crystallography. It was demonstrated that ETH displayed a similar mode of action as INH, and the biologically active form of ETH was consisted of an ETH-NAD^+ adduct in a 4S configuration. Similarly, to test whether Dhfr and InhA were molecular targets of INH, Dhfr/KatG or InhA/KatG proteins were coexpressed in the presence of INH. Mass spectrometry and enzymatic assays verified that only InhA but not the Dhfr protein was bound with the

92

INH-NAD(P) adduct. In addition, it was verified that overexpression of the Dhfr protein in *M. Smegmatis* and *Mtb* did not confer resistance to INH, and INH resistant clinical isolates did not carry polymorphisms in the *dfrA* gene. These findings conclusively demonstrated that Dhfr was not a target relevant to the antitubercular action of INH.

CHAPTER III

REGULATION OF INHA ACTIVITY BY PHOSPHORYLATION[*]

Overview

The remarkable survival ability of *Mycobacterium tuberculosis* in infected hosts
is related to the presence of cell wall-associated mycolic acids. Despite their importance,
the mechanisms that modulate expression of these lipids in response to environmental
changes are unknown. Here we demonstrate that the enoyl-ACP reductase activity of
InhA, an essential enzyme of the mycolic acid biosynthetic pathway and the primary
target of the anti-tubercular drug isoniazid, is controlled via phosphorylation. Thr-266 is
the unique kinase phosphoacceptor, both *in vitro* and *in vivo*. The physiological
relevance of Thr-266 phosphorylation was demonstrated using inhA phosphoablative
(T266A) or phosphomimetic (T266D/E) mutants. Enoyl reductase activity was severely
impaired in the mimetic mutants *in vitro*, as a consequence of a reduced binding affinity
to NADH. Importantly, introduction of inhA_T266D/E failed to complement growth and
mycolic acid defects of an inhA-thermosensitive *Mycobacterium smegmatis* strain, in a
similar manner to what is observed following isoniazid treatment. This study suggests
that phosphorylation of InhA may represent an unusual mechanism that allows *M.*

tuberculosis to regulate its mycolic acid content, thus offering a new approach to future anti-tuberculosis drug development.

Introduction

Mycolic acids are the major and distinguishing molecular components of the *Mycobacterium tuberculosis* cell envelope. These very long α-alkyl β-hydroxyl fatty acids are found either unbound as esters of trehalose or glycerol that are extractable with organic solvents or attached to the terminal penta-arabinofuranosyl units of arabinogalactan, the polysaccharide that, together with peptidoglycan, forms the insoluble cell-wall skeleton (McNeil et al., 1991). These lipids play an important role in the formation of an outer membrane and in cell-wall impermeability, virulence, immune evasion (Bhatt et al., 2007a; Bhatt et al., 2007b; Dao et al., 2008; Dubnau et al., 2000; Glickman et al., 2000; Rao et al., 2006) and the distinctive acid-fast staining of *M. tuberculosis* (Bhatt et al., 2007a). Recent studies have also revealed free mycolic acids in *M. tuberculosis* biofilms (Ojha et al., 2008). Mycolic acid synthesis is tightly related to cell division, as evidenced by the fact that isoniazid (INH) and ethionamide (ETH), two antitubercular drugs, inhibit mycolic acid biosynthesis, resulting in cell lysis (Takayama et al., 1972). This metabolic pathway thus represents an important reservoir of targets for new drugs (Dover et al., 2008), which are more critical than ever since the emergence of multidrug-resistant and extremely drug-resistant strains of *M. tuberculosis*.

Biosynthesis of mycolic acids depends on two distinct fatty acid synthases: the eukaryotic-like type I (FAS-I) and the prokaryotic-like type II (FAS-II) enzymes. FAS-I

95

is a single polypeptide that performs *de novo* biosynthesis of medium-length acyl-CoAs (C_{16} and C_{24}–C_{26}) (Bloch, 1975; Zimhony et al., 2004). These are used as primers by the FAS-II system and are iteratively condensed with malonyl- Acyl Carrier Protein (ACP) in a reaction catalysed by mtFabH, the β-ketoacyl-ACP synthase III of *M. tuberculosis* (Brown et al., 2005; Choi et al., 2000; Scarsdale et al., 2001). During the second step of the elongation cycle, the resulting β-ketoacyl-ACP product is reduced by MabA, the NADPH-dependent β-ketoacyl reductase of *M. tuberculosis* (Banerjee et al., 1998; Marrakchi et al., 2002). The resulting β-hydroxyacyl-ACP is then dehydrated by a set of dehydratases, HadABC (Brown et al., 2007; Sacco et al., 2007), and finally reduced by the NADH-dependent 2-trans-enoyl-ACP reductase, InhA (Quemard et al., 1995). The succeeding steps of condensation of the elongating chain with malonyl-ACP units are performed by the β-ketoacyl-ACP synthases KasA and KasB (Bhatt et al., 2007a; Kremer et al., 2002; Schaeffer et al., 2001), leading to very long-chain meromycolyl-ACPs (up to C_{56}), which are the direct precursors of mycolic acids (Kremer et al., 2000a; Takayama et al., 2005).

InhA belongs to the family of short-chain dehydrogenases/reductases (Dessen et al., 1995; Quemard et al., 1995). Genetic studies have revealed that *inhA* is essential in mycobacteria (Vilcheze et al., 2000). Thermal inactivation of InhA in a *Mycobacterium smegmatis* strain carrying an *inhA*-thermosensitive allele resulted in inhibition of mycolic acid biosynthesis and cell lysis in a manner similar to that seen in INH-treated bacteria (Vilcheze et al., 2000). It is now clear that INH is activated by the catalase peroxidase KatG to form a hypothetical isonicotinoyl radical that binds to NAD. The

resulting INH-NAD adduct inhibits InhA (Rozwarski et al., 1998), leading to accumulation of long-chain fatty acids, inhibition of mycolic acid biosynthesis and ultimately cell death (Vilcheze et al., 2000). INH resistance also involves several genes but only mutations in *inhA* have a dominant phenotype, whereas all other mechanisms of resistance (*katG, ndh, msh, nat*) are recessive (Vilcheze and Jacobs, 2007). The *inhA* gene product was originally identified as a putative target for INH and ETH in *M. smegmatis* (Banerjee et al., 1994). An increasing body of evidence has since pointed to InhA as the primary target of the two drugs (Kremer et al., 2003; Larsen et al., 2002; Vilcheze et al., 2000; Wang et al., 2007), culminating in the demonstration that transfer of an *inhA* S94A mutant allele in *M. tuberculosis* was sufficient for conferring resistance to both INH and ETH and establishing InhA as the clinically relevant target of INH (Vilcheze et al., 2006). InhA is one of the best-validated targets for the development of antitubercular agents. Further studies on INH, as well as on other small-molecule inhibitors of InhA, hold significant promise for the delivery of novel anti-tubercular agents effective against drug-resistant *M. tuberculosis* (am Ende et al., 2008; Freundlich et al., 2009; He et al., 2007; He et al., 2006; Oliveira et al., 2007; Tonge et al., 2007; Vilcheze and Jacobs, 2007). This prompted us to seek new alternatives for InhA inactivation, ultimately leading to mycolic acid and cell growth inhibition. Indeed, mycolic acid biosynthesis and cell division are very likely to be related (Lacave et al., 1989; Lacave et al., 1987; Takayama et al., 1972). We hypothesized that the activity of InhA might be controlled by post-translational modification in *M. tuberculosis*. This idea was supported by (i) the recent demonstration that the keto-reductase activity of MabA is

97

regulated by Ser/Thr protein kinase (STPK)-dependent phosphorylation, providing the first information on the molecular mechanism(s) involved in mycolic acid regulation through phosphorylation of a FAS-II enzyme (Veyron-Churlet et al., 2010) and (ii) the fact that genes encoding the two FAS-II reductases, MabA and InhA, are in the same operon in the *M. tuberculosis* genome (Banerjee et al., 1998; Cole et al., 1998).

Many of the stimuli encountered by *M. tuberculosis* are transduced via sensor kinases in the membrane, allowing the pathogen to adapt to survive in hostile environments. In addition to the classical two-component systems, *M. tuberculosis* contains 11 eukaryotic-like STPKs (Av-Gay and Everett, 2000; Cole et al., 1998). There is now an increasing body of evidence suggesting that, in *M. tuberculosis*, many STPKs are involved in regulating metabolic processes, transport of metabolites, cell division or virulence (Molle and Kremer, 2010). Signalling through Ser/Thr phosphorylation has recently emerged as a key regulatory mechanism in pathogenic mycobacteria (Molle and Kremer, 2010; Wehenkel et al., 2008).

As a first step to decipher an original molecular mechanism for future drug development by specifically targeting InhA, this study was undertaken to determine whether InhA represents a new substrate of *M. tuberculosis* SPTKs and to investigate whether phosphorylation negatively regulates InhA activity and consequently mycolic acid biosynthesis and mycobacterial growth.

Results

InhA is phosphorylated **in vitro** *by multiple Ser/Thr kinases (STPK)*

By analogy with the mode of action of INH, we reasoned that inhibition of mycolic acid biosynthesis might occur *in vivo* via post-translational modifications of InhA capable of reducing the enoyl reductase (ENR) activity of the enzyme. This prompted us to examine whether InhA could be modified by phosphorylation, a reaction that changes the physicochemical properties of defined Ser or Thr residues by introducing negative charges, which can ultimately affect the overall activity of the protein. This was first investigated *in vitro* in the presence of purified STPKs (PknA to PknL). The kinase domains of several transmembrane kinases from *M. tuberculosis* were expressed as GST-tagged fusion proteins and purified from *Escherichia coli* as reported earlier (Molle et al., 2006). The purified kinases were incubated with InhA and $[\gamma\text{-}^{33}P]$-ATP, resolved by SDS-PAGE and their phosphorylation profiles analysed by autoradiography. The presence of intense radioactive signals indicated that InhA was phosphorylated by multiple kinases, including PknA, PknB, PknE and PknL (**Figure 3-1A**). No signal was observed in the presence of PknF, PknH or PknK, all of which displayed various autokinase activities as reported earlier (Molle et al., 2006). These results clearly indicate that InhA is a specific substrate and interacts with various STPKs *in vitro*, suggesting that this key protein of the mycolic acid biosynthetic pathway might be regulated in mycobacteria by multiple extracellular signals.
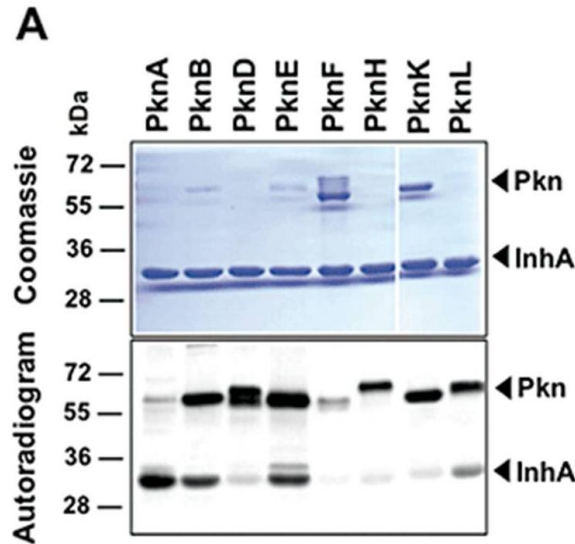
Figure 3-1. *M. tuberculosis* **InhA is phosphorylated** *in vitro* **by Ser/Thr kinases on residue Thr-266. A)** *In vitro* phosphorylation of InhA by multiple kinases. Eight recombinant STPKs (PknA to PknL) encoded by the *M. tuberculosis* genome were expressed and purified as GST fusions and incubated with purified His-tagged InhA and radiolabelled [γ-^{33}P]-ATP. The quantity between the STPKs varied from 0.2 to 1 μg in order to obtain the optimal autophosphorylation activity for each specific kinase. Samples were separated by SDS-PAGE, stained with Coomassie blue (upper panel) and visualized by autoradiography after overnight exposure to a film (lower panel). Upper bands reflect the autophosphorylation activity of each kinase and the lower bands correspond to the phosphorylation signal of InhA. **B)** Mass spectrometric analysis of PknB-phosphorylated InhA. MS/MS spectrum of the triply charged ion [M+3H]$^{3+}$ at m/z 1051.51 of peptide (241–269) (monoisotopic mass: 3151.52 Da). Unambiguous location of the phosphate group on Thr-266 was shown by observation of the 'y' C-terminal daughter ion series. Starting from the C-terminal residue, all 'y' ions lose phosphoric acid (-98 Da) after the Thr-266 phosphorylated residue. **C)** *In vitro* phosphorylation of the InhA_T266A mutant. Purified InhA_WT and InhA_T266A were incubated with PknB and [γ-^{33}P]-ATP. Samples were separated by SDS-PAGE, stained with Coomassie blue and visualized by autoradiography after overnight exposure to a film, as indicated.

B

T V C A L L S D W L P A T T G D I I...

y_{11}^Δ y_8^Δ y_5^Δ

$\Delta = - [H_3PO_4]$

..Y A D G G A H pT Q L L

pT$_{266}$ – peptide [241-269] from *in vitro*

Monoisotopic mass : 3151.52 Da ; m/z 1051.51 (+3)

C

Coomassie Autoradiogram

Figure 3-1 Continued.

InhA is phosphorylated on a unique threonine residue

Mass spectrometry was used to identify the number and nature of the phosphorylation sites on InhA. Such a method has been successfully used to elucidate the phosphorylation sites in a sequence-specific fashion for several *M. tuberculosis* STPK substrates (Barthe et al., 2009; Canova et al., 2009; Veyron-Churlet et al., 2009; Veyron-Churlet et al., 2010). InhA was incubated with unlabelled ATP in the presence of PknB (one of the most active kinases for InhA, **Figure 3-1A**) and subjected to mass spectrometric analysis after tryptic and chymotryptic digestion. Spectral identification and phosphorylation determination were achieved with the paragon algorithm from the ProteinPilot® 2.0 database-searching software (Applied Biosystems) using the phosphorylation emphasis criterion against a homemade database that included the sequences of InhA and derivatives. The phosphopeptides identified by the software were then validated by manual examination of the corresponding MS/MS spectra. Manual validations were performed based on neutral loss of H_3PO_4 from the precursor ion and the assignment of major fragment ions to b- and y-ion series or to the corresponding neutral loss of H_3PO_4 from these ions. The sequence coverage of the protein was 94% and phosphorylation occurred only on peptide (241–269) with an 80 Da mass increment from 3071.53 Da (theoretical MW) to 3151.52 Da (monoisotopic mass). The MS/MS spectra unambiguously confirmed the presence of a phosphate group on Thr-266 (**Figures 3-1B** and **A-1A**). Then, in order to prevent *in vitro* phosphorylation on Thr-266, this residue was changed to alanine by site-directed mutagenesis. The corresponding phosphoablative InhA_T266A mutant was expressed and purified as a

His-tagged protein in *E. coli* BL21(DE3)Star harbouring pETPhos_inhA_T266A. The resulting InhA_T266A mutant protein was purified and incubated with PknB in the presence of $[\gamma\text{-}^{33}P]$-ATP. Following separation by SDS-PAGE and analysis by autoradiography, total abrogation of the phosphorylation signal was observed compared with InhA_WT (**Figure 3-1C**). Similar results were obtained when InhA_T266A was incubated in the presence of either PknA, PknE or PknL (data not shown), indicating that Thr-266 represents the phosphoacceptor for all four kinases.

Taken together, these results indicate that Thr-266 represents the unique phosphorylation site in InhA and suggest that phosphorylation of this residue is likely to play a critical role in the regulation of InhA activity.

In vivo *phosphorylation of InhA*

To corroborate the *in vitro* results, it was necessary to confirm the phosphorylation state of the InhA protein *in vivo*. The *inhA* gene was cloned into the pMK1 mycobacterial expression vector under the control of the strong promoter *hsp60* (**Table A-1**). The resulting construct was used to transform *Mycobacterium bovis* BCG Pasteur in order to allow over-production of recombinant His-tagged InhA, which was purified and subjected to 2D-gel electrophoresis. As shown in Figure 3-2A, two spots, presumably corresponding to the non-phosphorylated form followed in the acidic direction by a mono-phosphorylated form of InhA, were clearly detected. Definitive identification and localization of Thr-266 as the unique phosphorylation site in InhA *in vivo* was achieved by mass spectrometric analysis using InhA purified from *M. bovis*

BCG cultures. LC-MS/MS identified an 80 Da mass increase corresponding to the peptide (241–269), supporting the conclusion that InhA is phosphorylated *in vivo* in *M. bovis* BCG **(Figures 3-2B and A-1B)**. Together, these results confirm that Thr-266 corresponds to the primary phosphorylation site both *in vitro* and *in vivo*.

ENR activity is strongly reduced for the InhA_T266D and InhA_T266E phosphomimetic mutants

Earlier studies showed that the acidic Asp or Glu amino acids qualitatively mimic the effect of phosphorylation with regard to functional activity. Following a strategy successfully used to demonstrate the role of phosphorylation on the condensing activity of mtFabH (Veyron-Churlet et al., 2009) and the β-ketoacyl-ACP reductase MabA (Veyron-Churlet et al., 2010), we first expressed and purified the phosphomimetics InhA_T266D and InhA_T266E. Next, we determined the *in vitro* ENR activity of these mimetics and compared their enzymatic activity with those of InhA_WT in the presence of increasing concentrations of trans-2-dodecenoyl-CoA (DD-CoA). Figure 3-3A clearly shows a strongly reduced activity of the two phosphomimetic mutants. The effect of mutation at T266 on the enzymatic activity was then calculated as the remaining per cent activity. The phosphoablative InhA_T266A protein showed very similar catalytic activity to the wild-type enzyme, indicating that this residue is not involved in catalysis **(Figure 3-3A, 3-3B,** and **Table 3-1)**.
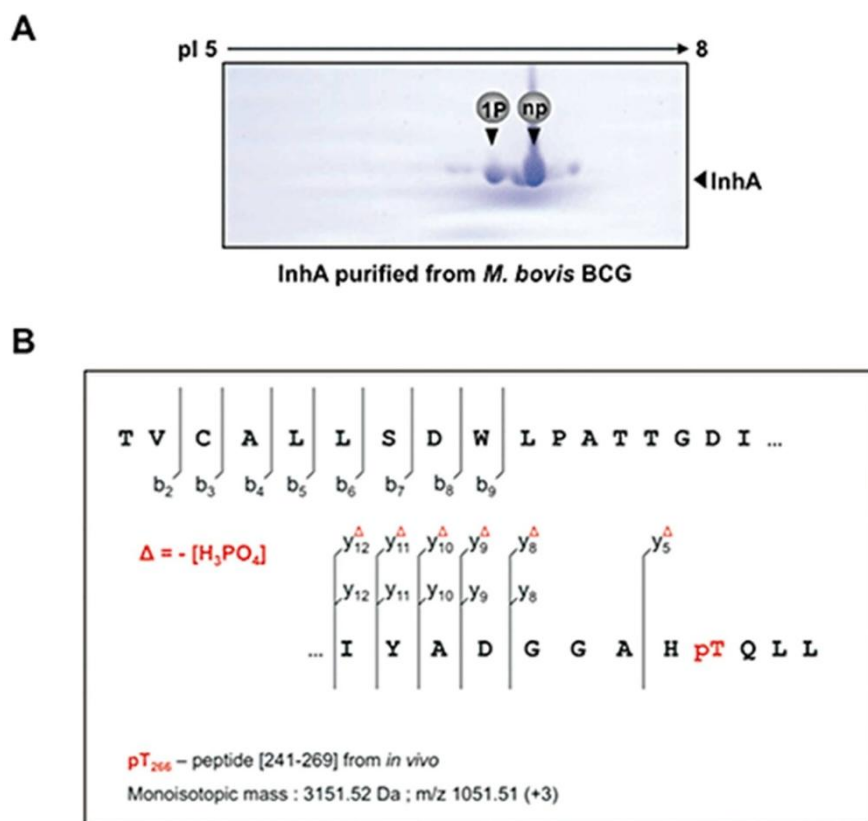
Figure 3-2. *In vivo* phosphorylation of InhA in mycobacteria. A) Phosphorylation profile of InhA purified from recombinant *M. bovis* BCG carrying the pMK1_*inhA*_*WT*. Bacteria were lysed and the soluble fraction was incubated with Ni-NTA agarose beads to purify the His-tagged InhA. The protein preparation was loaded on a 7 cm immobiline strip (Bio-Rad, pH 5–8) and electrophoresed on a Protean IEF Cell (Bio-Rad) for the first dimension and on a 10% SDS-PAGE for the second dimension, then stained with Coomassie blue. The grey circles represent the different phosphorylated isoforms, as indicated: np, non-phosphorylated; 1P; mono-phosphorylated. **B)** Mass spectrometric analysis of InhA purified *in vivo*. MS/MS spectrum of the triply charged ion $[M+3H]^{3+}$ at m/z 1051.51 of peptide (241–269) (monoisotopic mass: 3151.52 Da). The phosphate group on T266 was unambiguously located by observing the 'y' C-terminal daughter ion series. Starting from the C-terminal residue, all 'y' ions lose phosphoric acid (-98 Da) after the T266 phosphorylated residue.

InhA_T266E and InhA_T266D mutants showed only ~30% residual enzymatic activity. Subsequent enzymatic studies were focused on Asp and Glu mutants to determine the reason behind this significant loss of function. Initial velocity studies with Asp and Glu mutants suggested that both followed a random order reaction mechanism with preferred NADH binding prior to DD-CoA binding, which was reported previously for wild-type InhA (Parikh et al., 1999; Quemard et al., 1995) **(Figures 3-3A and 3-3C)**. This preferred-ordered pathway towards NADH binding resulted in sigmodial initial velocity curves as the NADH concentrations were varied at a fixed DD-CoA concentration **(Figure 3-3A, right panel)**. On the other hand, when the concentration of DD-CoA was varied at fixed concentrations of NADH, a pattern similar to substrate inhibition was observed **(Figure 3-3A, left panel)**. The steady-state kinetic parameters, $k$cat and Km, calculated for DD-CoA for InhA_WT, InhA_T266A, InhA_T266D and InhA_T266E mutants are presented in Table 3-1. Both Asp and Glu mutants exhibited Km values for DD-CoA very similar to those of the wild-type and T266A mutant enzymes but $k$cat and Vmax values for the wild type and Ala mutant showed three- to fourfold higher catalytic turnover and maximum velocity compared with the phosphomimetic mutants. $k$cat/Km, which defines the substrate specificity, is almost the same for all four enzymes. This suggests that the overall observed decrease in the activity (100% wild type versus 30% mutants) is not due to the deficiency in the DD-CoA substrate binding.
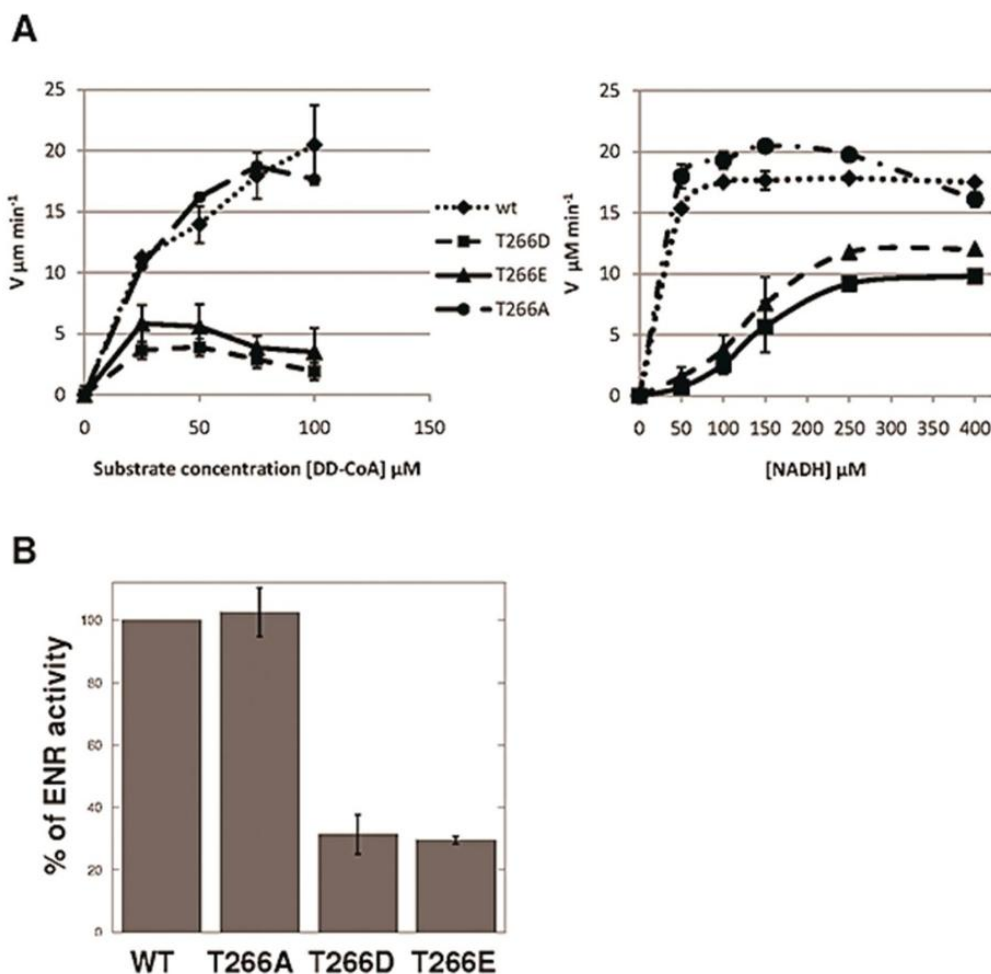
Figure 3-3. Enoyl reductase (ENR) activity of InhA_WT and mutant derivatives. InhA_WT and the single mutants carrying a Thr→Ala, a Thr→Asp or a Thr→Glu mutation at position 266 were purified from recombinant *E. coli*, dialysed and assayed for ENR activity. **A)** Enzymatic activity in the presence of 150 μM NADH and increasing concentrations of DD-CoA (left panel). Initial velocity for the wild-type and mutant enzymes measured as the NADH concentration was varied at a fixed DD-CoA concentration (50 μM) is presented in the right panel. **B)** Activity of the various InhA variants in the presence of 100 μM NADH and 50 μM DD-CoA. The activity of InhA_WT was arbitrarily set at 100%. Values are means of triplicates and are representative of three sets of experiments with independent protein preparations. Error bars represent the standard error. **C)** Proposed preferred pathway reaction mechanism for InhA_T266D and InhA_T266E enzymes. Both of the enzymes bind preferably to NADH first, followed by DD-CoA binding. Since the enzymes kinetically prefer NADH binding first, decreased affinity towards NADH can affect the overall *in vitro* activity of the mutant enzymes.

C
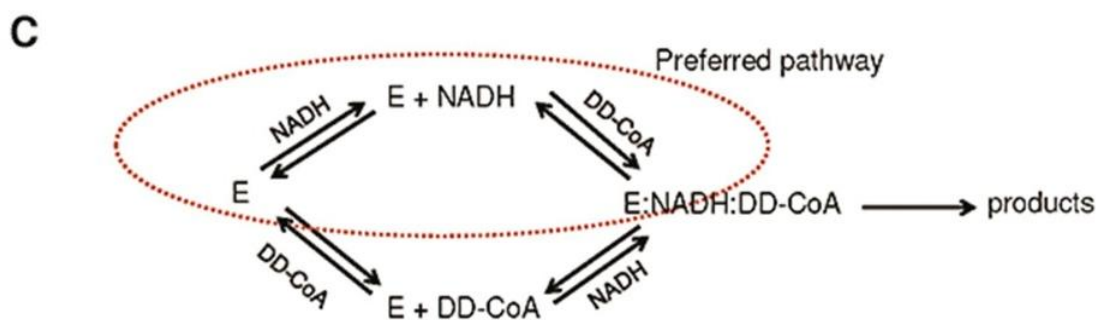
Figure 3-3 Continued.

Fluorescence experiments clearly indicated that both Asp and Glu mutants had three- to fourfold decreased binding affinity for NADH while InhA_T266A displayed approximately twofold reduction compared with InhA_WT as judged by the K_d values **(Table 3-1)**. This suggests that T266 is an important residue for NADH binding and mutation of this residue can affect the binding of the enzyme to NADH. However, the overall effect on enzymatic activity is not linearly related to the fold decrease in NADH binding affinity since T266A is catalytically as active as the wildtype enzyme. Because phosphomimetic mutant enzymes kinetically prefer NADH binding first, the overall activity for these mutant proteins were affected significantly by the decreased affinity towards NADH. It is therefore very likely that the change in the NADH binding affinity is responsible for the decreased activity of the phosphomimetic InhA mutants.

Table 3-1. Steady-state kinetic parameters K_m, k_{cat}, V_{max} and k_{cat}/K_m for DD-CoA, for wild-type and mutant InhA enzymes.

	Activity (%)	K_m (µM)	k_{cat} (min^{-1})	V_{max} (µMmin^{-1}mg^{-1})	k_{cat}/K_m (min^{-1} µM^{-1})	K_d (NADH) (µM)
InhA_WT	100	40.9 ± 15.4	320.4 ± 52.7	11.4 ± 1.9	9.9 ± 2.7	1.5 ± 0.2
InhA_T266A	102.6 ± 7.8	52.3 ± 3.6	427.4 ± 34.4	15.3 ± 1.2	8.2 ± 0.2	3.5 ± 0.4
InhA_T266D	31.4 ± 6.3	19.6 ± 5.6	87.3 ± 4.4	3.1 ± 0.1	6.0 ± 2.7	4.7 ± 0.6
InhA_T266E	29.5 ± 1.2	20.3 ± 4.9	149.4 ± 11.9	4.1 ± 0.8	6.9 ± 2.3	6.0 ± 0.6

Dissociation constants for NADH are presented. The values were determined in triplicate and the average numbers calculated from three different experiments are represented. For each InhA enzyme, k_{cat}/K_m was calculated from three independent experiments and the mean k_{cat}/K_m is presented. Values are mean ± standard error.

Local structural changes generated by Thr-266 mutations

Recombinant InhA_WT, InhA_T266D and InhA_T266E proteins were purified from *E. coli* and subjected to crystallographic studies. In the InhA_WT structure, the OG1 atom of T266 interacts with the NE2 atom of Q267 (3.08 Å) and carbonyl oxygen of G263 (2.74 Å) **(Figure A-2)**. For the T266D/E mutants, due to the loss of the Thr residue, this direct H-bonding interaction between residues T266 and Q267 is lost. Instead, a new water molecule (H$_2$O #9 and #5 for InhA_T266D and InhA_T266E mutants respectively) is introduced between residues 266 and 267 **(Figure A-2)**. Regarding the InhA_T266D structure, this makes H-bonding interactions with the NE atom of Q267 (3.01 Å), the carbonyl oxygen of G263 (2.74 Å) and the carbonyl oxygen of T254 from the neighboring subunit (2.77 Å). In InhA_T266E these interaction distances are 3.0 Å, 2.71 Å and 2.79 Å respectively. Notably, the direct interaction between the side-chains of residues 266 and 267 is completely abrogated in the mutant

structures. On the other hand, OD1 and OD2 atoms of D266 interact with NE2 and ND1 of H265 (3.07 Å and 3.45 Å respectively), causing a ~15 degree flip in the H265 ring. For E266, the distances are: OE1 of E266 with NE2 of H265 4.4 Å; ND1 of H265 3.04 Å; OE2 of E266 with ND1 H265 3.5 Å and NE2 of H265 3.8 Å. OE2 of E266 and OD2 of D266 also H-bonds with the carbonyl oxygen of T254 from the neighbor subunit (3.0 Å), which is not present in the wild-type InhA:NADH structure (3.7 Å) **(Figure A-2)**. In summary, mutation of Thr→Asp or Thr→Glu led to local structural changes limited to the surrounding of the mutated residue.

Growth defect of fast- and slow-growing mycobacteria overexpressing InhA phosphomimetic proteins

Mycobacterium smegmatis mc^{2}155 was transformed with pMK1 derivatives allowing constitutive expression of the different *inhA* alleles under the control of the strong *hsp60* promoter **(Table A-1)**: *inhA_WT*, the phosphoablative *inhA_T266A* or the phosphomimetic *inhA_T266D* and *inhA_T266E*. Transformed mycobacteria were selected on Middlebrook 7H10 plates. As shown in Figure 3-4A, which illustrates the morphology and size of *M. smegmatis* colonies after 4 days' incubation at 37°C, it is clear that overexpression of InhA_T266A did not impair growth of *M. smegmatis* as compared with the strain overproducing InhA_WT. In contrast, overexpression of either InhA_T266D or InhA_T266E was accompanied by a decrease in the growth rate compared with the strains overexpressing InhA_WT or InhA_T266A **(Figure 3-4A)**. Similar results were also observed in *M. bovis* BCG strains harbouring these plasmids

(Figure A-3). Taken together, these results indicate that overexpression of the phosphomimetic InhA alleles severely altered mycobacterial growth in both fast- and slow-growing mycobacterial species. Our results show that the impaired ENR activity of InhA_T266D or InhA_T266E is likely to be responsible for this growth defect, presumably by inhibiting FAS-II activity, either interacting with other FAS-II system enzymes or forming non-functional oligomers. Because InhA acts as a tetramer in mycobacteria, overproduction of InhA_T266D or InhA_T266E may lead to unproductive InhA multimers, which may coexist with functional InhA multimers.

The hypothesis that the impaired *in vitro* activity of InhA_T266D or InhA_T266E is linked to a different oligomerization state than InhA_WT can, however, be excluded since both mutant enzymes formed only tetramers in solution as demonstrated by gel-filtration chromatography (data not shown), and by the fact that we could crystallize and determine the structures of the tetrameric InhA_T266D and T266E mutants **(Figure A-4)**. Western blot analysis from *M. smegmatis* and *M. bovis* BCG cultures revealed both the endogenous and the recombinant InhA proteins in each strain, which can be clearly distinguished by the presence (recombinant InhA) or absence (endogenous InhA) of a His-tag **(Figures 3-4B and A-3)**.
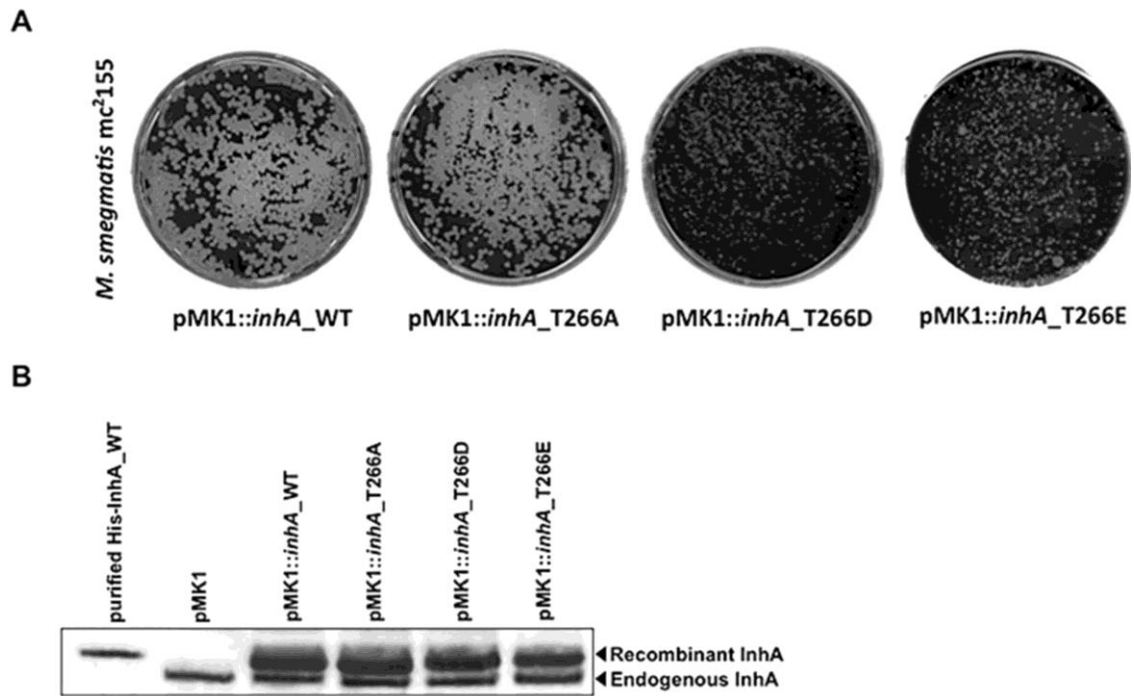
Figure 3-4. Overexpression of the various InhA variants and effect on growth in *M. smegmatis*. A) Electrocompetent *M. smegmatis* mc^{2}155 was transformed with the empty pMK1 construct, the pMK1_*inhA_WT*, pMK1_*inhA_T266A*, pMK1_*inhA_T266D* or pMK1_*inhA_T266E* to allow constitutive expression of the various *inhA* alleles under the control of the strong *hsp60* promoter. Transformed mycobacteria were plated and incubated at 37°C for 3 days. **B)** InhA expression levels in the InhA-overproducing *M. smegmatis* mc^{2}155 strains. Western blot analysis of *M. smegmatis* mc^{2}155 cultures overexpressing the phosphoablative *inhA*_T266A and phosphomimetic *inhA*_T266D and *inhA*_T266E alleles were harvested, resuspended in PBS and disrupted. Equal amounts of crude lysates (20 μg) were loaded onto a 4–12% acrylamide gel, subjected to electrophoresis and transferred onto a membrane for immunoblot analysis using rabbit anti-InhA antibodies. Endogenous and recombinant InhA proteins are indicated by arrowheads.

Although the recombinant InhA forms are produced in large excess compared with the endogenous InhA, the presence of significant amounts of endogenous protein may explain the partial growth defect in the phosphomimetic strains. The growth impairment also hampers further examination and evaluation of the possible link between growth defect and mycolic acid biosynthesis inhibition in the mycobacterial strains overproducing the phosphomimetic mutants. We then explored whether this could be demonstrated by generating isogenic strains carrying either the phosphoablative or the phosphomimetic *inhA* alleles in *M. tuberculosis*.

Transfer of T266D or T266E point mutations in InhA is lethal to M. tuberculosis

Using specialized linkage transduction (Vilcheze et al., 2006) we could transfer a single-point mutant allele (T266A) into *inhA* in *M. bovis* BCG and *M. tuberculosis* H37Rv (**Figure 3-5A**), indicating that introduction of the phosphoablative *inhA* allele was not lethal. In addition, the growth curves of *M. tuberculosis* bearing either the *inhA_WT* or the *inhA_T266A* allele were similar, indicating that introduction of the T266A mutation did not alter growth of *M. tuberculosis* (data not shown). In contrast, and despite several attempts, we could not transfer the T266D or T266E alleles (**Figure 3-5A**) into a wild-type genetic background, suggesting that these single mutations were lethal to *M. bovis* BCG and *M. tuberculosis* strains. To test this hypothesis, we performed allelic replacements using *M. tuberculosis inhA* merodiploid strains constructed by integrating a functional copy of *inhA* into the chromosome of *M. tuberculosis* H37Rv using the integrative pMV361 vector, which inserts at the *attB*

113

chromosomal site. Gene replacements were carried out using specialized linkage transduction as described above. Interestingly, only in the presence of a complementing copy of the wild-type *inhA* were we able to introduce the T266D or T266E mutation into both the *M. tuberculosis* H37Rv and the *M. tuberculosis* CDC1551 merodiploid strains, as shown by Southern blot **(Figure 3-5B)** and PCR analyses (data not shown). Sequence analysis confirmed the presence of the expected *inhA* alleles in the various strains (data not shown).

Taken together, these results indicate that a single T266D or T266E point mutation within InhA confers lethality on *M. tuberculosis*. They also strongly suggest that, in the absence of wild-type InhA, transfer of a phosphomimetic allele leads to cell death.
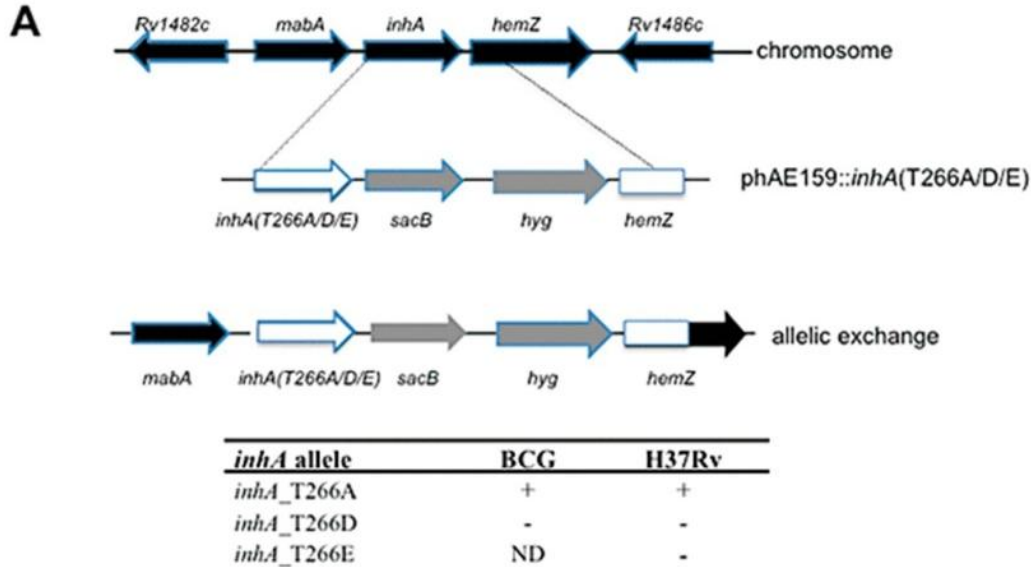
Figure 3-5. Construction and analysis of *M. tuberculosis inhA*_T266A, *inhA*_T266D and *inhA*_T266E isogenic strains. A) Schematic representation of the specialized transduction phage. A replicating shuttle phasmid phAE159 containing *inhA* carrying the T266A, T266D or T266E mutation, *sacB*, a *hyg* resistance cassette, and *hemZ* was used to transduce *M. bovis* BCG or *M. tuberculosis* H37Rv. When recombination occurs before the point mutation, this results in a recombinant strain carrying the T266A, T266D or T266E mutation. Hygromycin-resistant transductants are screened by PCR and the presence of the desired mutations is confirmed by sequencing. The results of introducing phosphoablative (T266A) or phosphomimetic (T266D, T266E) mutations into the chromosomal *inhA* gene in the various strains are indicated: +, isogenic strain obtained; -, no isogenic strain obtained; ND, not determined. **B)** Allelic exchange replacement in InhA merodiploid strains. Merodiploid strains of *M. tuberculosis* H37Rv (Rv::*inhA*) and CDC1551 (1551::*inhA*) were obtained following transformation with the integrative construct pMV361_*inhA*. These strains were used as receptor strains for specialized transduction, as described above. Hygromycin- and kanamycin-resistant strains were screened by PCR and sequencing. Southern blot analysis using the *inhA* probe of the corresponding BamHI restriction profiles of DNA from Rv::*inhA* + phAE159::*inhA*(T266E) (lane 1), Rv::*inhA* + phAE159::*inhA*(T266D) (lane 2), H37Rv + phAE159::*inhA*(T266E) (lane 3), H37Rv + phAE159::*inhA*(T266D) (lane 4), H37Rv + phAE159::*inhA*(T266A) (lane 5), H37Rv (lane 6), CDC1551 + phAE159::*inhA*(T266D) (lane 7), CDC1551 + phAE159::*inhA*(T266E) (lane 8), 1551::*inhA* (lane 9), 1551::*inhA* + phAE159::*inhA*(T266D) (lane 10) and 1551::*inhA* + phAE159::*inhA*(T266E) (lane 11). Sizes of the expected BamHI fragments are indicated.

B

BamHI   2947 bp   BamHI

mabA   *inhA*   *hemZ*   *Rv1486c*

M. tuberculosis
(H37Rv or CDC1551)

BamHI   6628 bp   BamHI

mabA inhA   *sacB*   *hyg*   *hemZ*   *Rv1486c*

M. tuberculosis
+ phAE159::*inhA(T266A/D/E)*

BamHI   3002 bp   BamHI

Phsp60 inhA   *kan*

pMV361::*inhA*

6.6 kb →

3 kb →

1   2   3   4   5   6   7   8   9   10   11

H37Rv   CDC1551
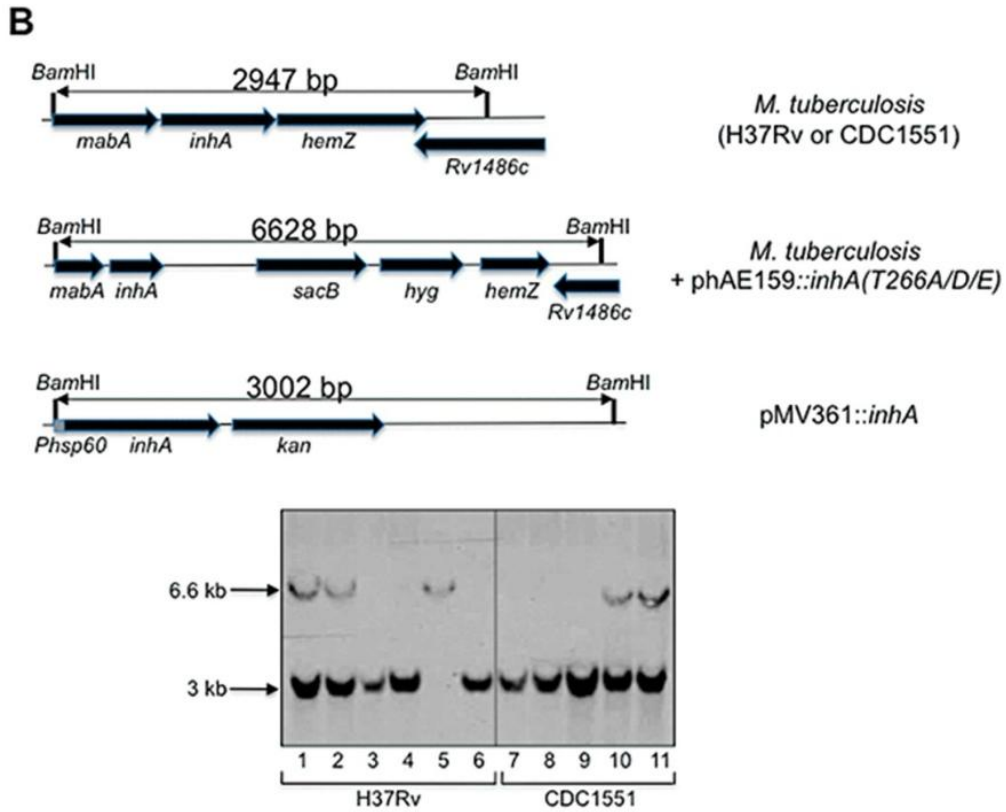
Figure 3-5 Continued**.**

Lack of ENR activity complementation with the InhA phosphomimetic mutants in
InhA-*thermosensitive mutant strain*

To investigate whether cell death occurring in a phosphomimetic strain is associated with inhibition of mycolic acid biosynthesis, we took advantage of using an *inhA*thermosensitive (Ts) mutant of *M. smegmatis* (m^{2}2359) carrying an *inhA_V238F* allele (Vilcheze et al., 2000). At a non-permissive temperature (42°C), thermal inactivation of InhA occurred, resulting in mycolic acid biosynthesis inhibition and

leading to a drastic morphological change and cell death in a manner similar to the result in INH treated cells (Vilcheze et al., 2000). *M. smegmatis* m^{2}2359 was therefore transformed with either the empty pMK1 vector or the pMK1 derivatives, allowing constitutive expression of the different *inhA* alleles, and selected for kanamycin resistance. Cultures were grown at the permissive temperature (30°C), then plated and grown at either 30°C or 42°C for 4–5 days. Figure 3-6A shows that cells transformed with constructs carrying either the *inhA_WT* or *inhA_T266A* alleles could grow on Middlebrook plates at the non-permissive temperature, indicating that functional complementation occurred. In sharp contrast, neither the *inhA* (T266D) nor the *inhA* (T266E) alleles restored growth at 42°C, indicating that introduction of these mutations was lethal in the absence of functionally active InhA **(Figure 3-6A)**. These phenotypes cannot be attributed to an eventual alteration of the expression levels of the various InhA mutants, as demonstrated by Western blot analysis **(Figure 3-6B)**. Comparable expression levels of the various InhA variants were found at either 30°C or 42°C. Moreover, due to the presence of an additional His-tag in the pMK1_InhA-derived constructs, it became possible to distinguish the recombinant variants (InhA_WT, T266A, T266D and T266E) from the endogenous protein (InhA_V238F) in the various strains. Importantly, following a temperature shift to 42°C, only the strains overexpressing InhA_WT or InhA_T266A could grow in broth medium, whereas those harbouring either the empty pMK1 plasmid or the pMK1-inhA_T266D/E constructs failed to grow **(Figure 3-6C)**. Overall, these results indicate that the T266D or T266E phosphomimetic proteins (even when present in large excess compared with the

117

endogenous protein) cannot complement thermal inactivation of InhA_V238F at the non-permissive temperature.

The usefulness of these strains prompted us to examine whether this lethal phenotype was related to *de novo* mycolic acid biosynthesis inhibition. This was achieved by shifting the temperature of the various strains to 42°C prior to labelling with [1-^{14}C]-acetate. After a further 3 h incubation at 42°C, cells were harvested and fatty acid methyl esters (FAMEs) and mycolic acid methyl esters (MAMEs) were extracted and separated by thin layer chromatography (TLC). As shown on the autoradiogram, and consistently with a previous study (Vilcheze et al., 2000), mycolic acid biosynthesis was almost completely abrogated in *M. smegmatis* mc^{2}2359 carrying the empty pMK1 at 42°C **(Figure 3-6D)**. In contrast, *M. smegmatis* mc^{2}2359 transformed with either pMK1_*inhA_WT* or pMK1_*inhA_T266A* displayed a wild-type mycolic acid profile at both 30°C and 42°C. Importantly, synthesis of mycolic acids in the phosphomimetic strains was severely impaired at the non-permissive temperature. Synthesis of α- and epoxy-mycolates was almost completely abrogated. The shorter-chain α′-mycolates, only present in this particular mycobacterial species, appeared less affected, as could be observed also when *M. smegmatis* cultures were treated with INH **(Figure 3-6D)**. This observation is consistent with previous studies reporting that α′-mycolates are less affected than the other mycolic acid subtypes after treatment with FAS-II-inhibitory drugs such as INH (Kremer *et al*., 2003) or thiolactomycin (Slayden et al., 1996).
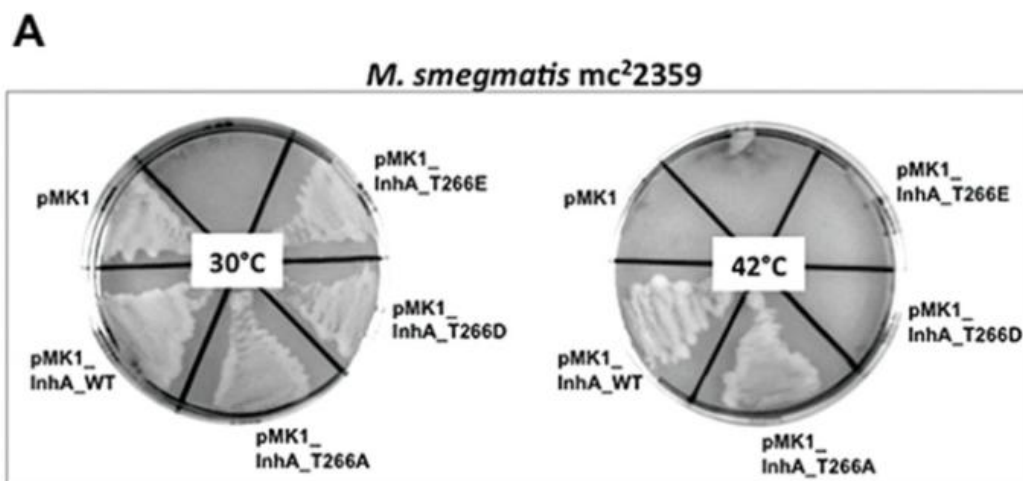
Figure 3-6. Functional complementation of *M. smegmatis* m²2359 with InhA_WT or InhA_T266A but not with phosphomimetics InhA_T266D or InhA_T266E. A) An *inhA(Ts)* mutant of *M. smegmatis* (mc²2359) resistant to INH was transformed with pMK1, pMK1_*inhA_WT*, pMK1_*inhA_T266A,* pMK1_*inhA_T266D* or pMK1_*inhA_T266E* and grown either at the permissive temperature (30°C) or at the non-permissive temperature (42°C). **B)** InhA expression levels. The various InhA levels of the parental strain and complemented mc²2359 strains were analysed by immunoblotting using rabbit anti-InhA antibodies. Cells were grown at the permissive temperature (30°C) or shifted at the non-permissive temperature (42°C), harvested, resuspended in PBS and disrupted. Equal amounts of crude lysates were loaded onto a 4–12% acrylamide gel, subjected to electrophoresis and transferred onto a membrane for immunoblot analysis. Endogenous (InhA_V238F) and recombinant InhA proteins are indicated by arrowheads. **C)** Growth curves of the strains described in (A) in Sauton broth medium, following temperature shift at 42°C. **D)** *De novo* biosynthesis inhibition of mycolic acids. Cultures of *M. smegmatis* mc²2359 harbouring pMK1, pMK1_*inhA_WT*, pMK1_*inhA_T266A,* pMK1_*inhA_T266D* or pMK1_*inhA_T266E* were grown at 30°C. The cultures were shifted to 42°C for 3 h and labelled by adding [1-^{14}C]-acetate (1 μCi ml^{-1}). After 3 h at 42°C, cultures were harvested and FAMEs and MAMEs were extracted and analysed by one-dimensional TLC using petroleum ether/acetone (95/5, v/v). Detection of the radiolabelled FAMEs and MAMEs was performed by autoradiography, exposing the TLC plates to X-ray films for 24 h. α, α′ and epoxy correspond to α-mycolates, α′-mycolates and epoxy-mycolates respectively.

Figure 3-6 Continued**.**

120
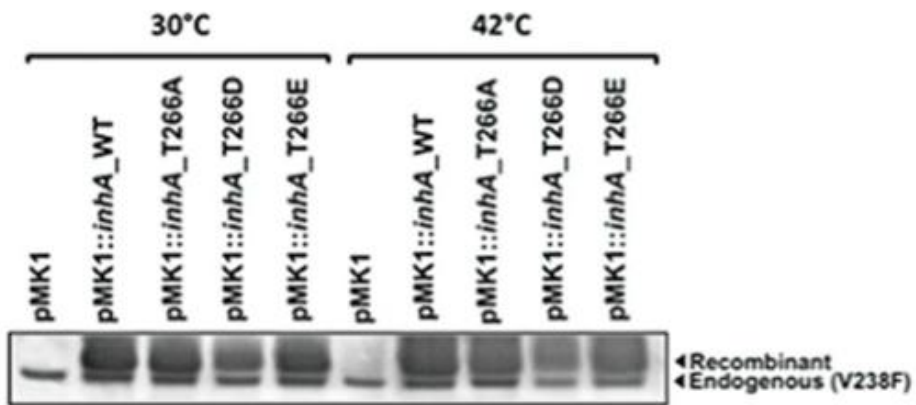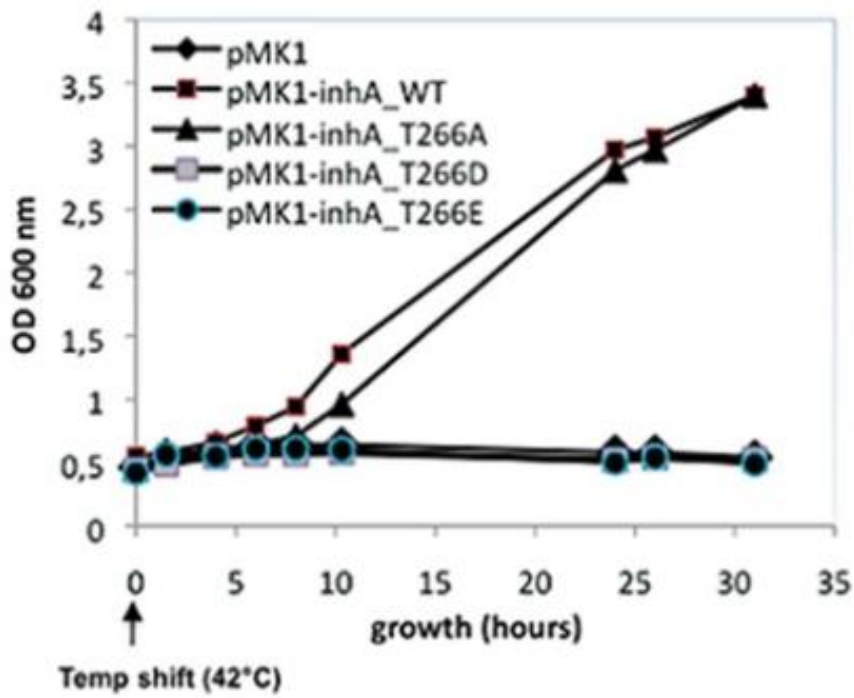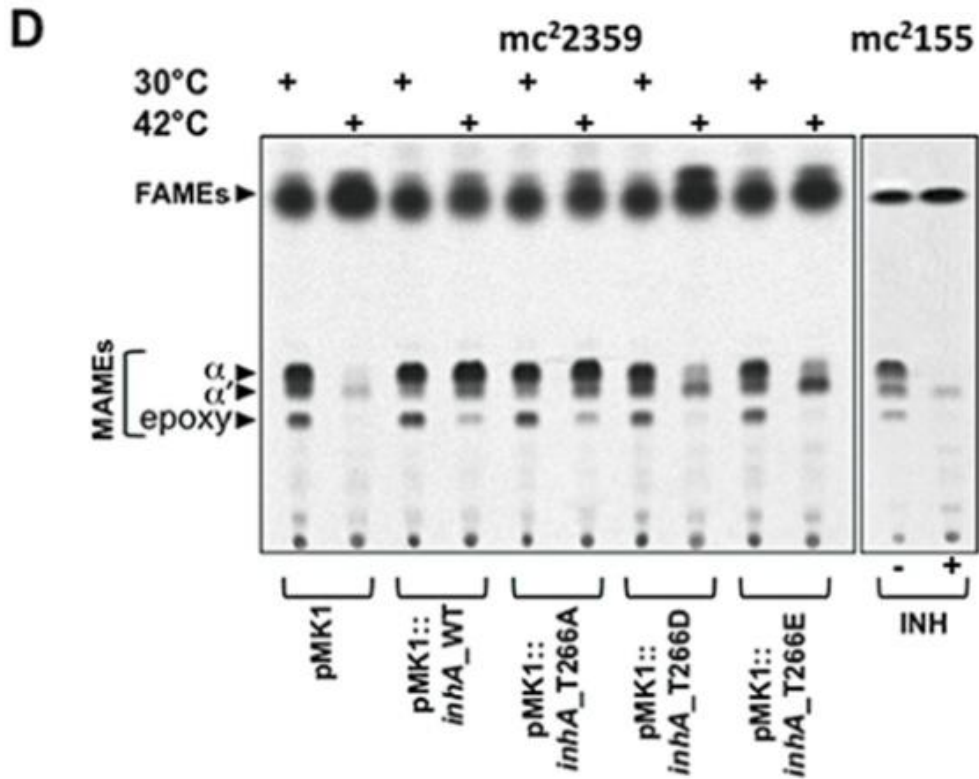
Figure 3-6 Continued**.**

Overall, these results clearly indicate that, in the absence of a functional endogenous protein, the T266D or T266E mutants cannot restore mycolic acid production, leading to mycobacterial growth inhibition. Importantly, this mycolic acid profile was reminiscent of that observed in INH-treated cultures (Vilcheze et al., 2000), leading to the conclusion that phosphomimetic mutants of InhA inhibit mycolic acid biosynthesis in a way similar to INH treatment.

Discussion

In this study, a combination of genetic, biochemical and structural approaches was used to provide, for the first time, evidence that phosphorylation of InhA negatively regulates *de novo* biosynthesis of mycolic acids in *M. tuberculosis*. We found that phosphorylation occurred both *in vitro* and *in vivo* on a unique residue, identified as Thr-266, although InhA was not identified in the recent phosphoproteomic study performed on *M. tuberculosis* (Prisic et al., 2010). Moreover, enzymatic analyses clearly indicated reduced ENR activity of InhA phosphomimetic mutant proteins, suggesting that phosphorylation may negatively regulate mycolic acid production. This hypothesis is supported by the following three facts: transfer of single T266D or T266E mutations into *M. tuberculosis* was lethal, unless an additional copy of *inhA* was present in a merodiplid strain; overexpression of InhA_T266D/E severely impaired growth in both fast- and slow-growing mycobacterial species; and T266D or T266E phosphomimetic proteins could not complement thermal inactivation of InhA_V238F at the non-permissive temperature in an *inhA*(*Ts*) mutant of *M. smegmatis*, leading to strong inhibition of α- and epoxy-mycolic acid production and growth arrest. It is noteworthy that, during review of our manuscript, an independent study has been accepted for publication (Khan et al., 2010), presenting very similar results and conclusions by demonstrating that (i) InhA was phosphorylated *in vivo* at Thr-266, (ii) phosphorylation of InhA resulted in decreased enzymatic activity, and (iii) an InhA_T266E mutant was unable to rescue a *M. smegmatis* conditional *inhA* gene replacement mutant. Therefore, from both studies it can be inferred that STPKdependent phosphorylation of InhA may represent an

important mechanism controlling mycolic acid biosynthesis and largely supports and extends the emerging theme that Ser/Thr phosphorylation plays a critical role in the regulation of cell-wall biosynthesis and cell division in mycobacteria (Molle and Kremer, 2010; Wehenkel et al., 2008).

One hypothesis could be STPK-dependent phosphorylation of InhA is used by *M. tuberculosis* during the nonreplicating dormant state, known to correlate with a slowdown of cell division, energy metabolism and lipid biosynthesis (Betts et al., 2002). However, further studies are required to investigate whether InhA (hyper)phosphorylation participates in the inhibition of mycolic acid metabolism during the 'persistent' state.

From a purely mechanistic perspective, our kinetic studies revealed that both InhA phosphomimetics followed a preferred-order pathway with NADH binding preceding DD-CoA binding. Both mutant proteins exhibited a three- to fourfold decreased binding affinity to NADH even though the mutation site was not at the cofactor binding pocket of InhA. Since there is a kinetic advantage when InhA binds to NADH first, it appears very likely that the overall activity for the mutant proteins was affected significantly by the decreased affinity towards NADH. T266 is very close to the C-terminal end of the protein (266 of 269 residues) and crystallographic studies indicated that mutation at this residue was not accompanied by significant differences in the secondary structure and did not alter the overall backbone conformation of the protein. Both T266E and T266D complexed with NAD^+/NADH crystallized in the same space group ($P6_222$) as the wildtype protein, with one molecule in the asymmetric unit

(Table A-3). As for InhA_WT, all mutant proteins formed tetramers in solution as judged by the gel-filtration chromatography profile (data not shown), indicating that mutation of the Thr-266 residue does not alter the oligomeric state of the protein. Moreover, the comparison of the tetrameric structures generated by crystallographic symmetry-related molecules for the wild type and Asp/Glu mutants revealed that the Rmsd value between the superimposed structures was 0.3 Å **(Figure A-4)**. Structural differences due to Thr→Asp or Thr→Glu replacements were limited to the surroundings of the mutated residue and introduced only subtle and local structural changes, perturbing the H-bonding network and introducing water molecules **(Figure A-2)**. The active site of InhA protein was ~15– 20 Å away from the mutation site in InhA_T266 **(Figure A-5)**. Although no significant structural difference or conformational change was observed at either the active site residues or the substrate/cofactor binding pockets between native and mutants structures, it remains possible that mutations cause changes in the network of interactions and affect the interaction and communication between the subunits of the functional tetramer. This in turn may affect NADH binding affinity and the overall enzymatic activity of the phosphomimetic InhA mutants. Based on our enzymatic analyses, it is noteworthy that NADH binding is affected even though the mutation site is ~16 Å away from the cofactor binding pocket.

InhA is one of the best-validated targets for antitubercular agents and several recent studies reported the development of new InhA inhibitors (am Ende et al., 2008; Freundlich et al., 2009; He et al., 2007; He et al., 2006; Oliveira et al., 2007; Tonge et al., 2007; Vilcheze and Jacobs, 2007). Our work suggests that displacement of the

unphosphorylated/phosphorylated InhA balance in favour of the phosphorylated isoform rapidly leads to mycolic acid cessation, as happens following INH treatment, and to mycobacterial growth inhibition. Therefore, an elegant hypothesis arising from the present work is that, by increasing the activity of the kinases, it may be possible to directly alter mycobacterial growth, opening new and original perspectives for future anti-tuberculosis drug development. Indeed, small molecules that modulate the activity of STPK may be of great therapeutic value in inhibiting *M. tuberculosis* growth. Bryostatin, a natural product synthesized by a marine bacterium, which activates eukaryotic intracellular STPKs (Hale et al., 2002), is one of these molecules. Interestingly, bryostatin acts directly on the *Bacillus subtilis* Ser/Thr kinase PrkC, which contains an extracellular domain able to bind to peptidoglycan fragments and this signals the bacteria to exit dormancy by stimulating germination (Shah et al., 2008). PrkC, like *M. tuberculosis* PknB, possesses the PASTA (Penicillin And Ser/Thr kinase Associated) domains, which are found in the extracellular portion of membrane associated STPKs and which have been proposed to bind to peptidoglycan and act as signalling molecules. In this context, bryostatin or other STPK-activating molecules, along with the recent structural determination of the *M. tuberculosis* PknB PASTA domains (Barthe et al., 2010), may provide new therapeutic strategies to be developed against tuberculosis. From our results, it can also be inferred that STPK-induced phosphorylation of InhA would be active on *M. tuberculosis* clinical isolates carrying the *inhA* alleles (I21, S94 or I194) that confer resistance to INH (Hazbon et al., 2006; Ramaswamy et al., 2003), since

phosphorylation occurs exclusively on Thr-266, a residue that has never been linked to INH resistance.

The present study provides a foundation for further investigation of a seemingly important functional linkage between STPKs and the FAS-II system. It also provides conceptual advances in our understanding of the mycolic acid metabolic adaptation and regulatory events exploited by *M. tuberculosis* to adapt its mycolic acid cell-wall content. Although very challenging, future studies should now help to identify extracellular cues sensed by the different kinases and leading to InhA phosphorylation. This will not only allow us to understand how *M. tuberculosis* senses its environment and mediates its response in a co-ordinated manner to regulate mycolic acid biosynthesis, but also to possibly link InhA phosphorylation to the establishment of the non-replicating persistent state. One can also anticipate that similar strategies involving STPK-dependent mechanisms will be found to be used by pathogenic mycobacteria to regulate expression of other cell-wall lipids/glycolipids to respond to the various signals encountered during infection or latency.

Experimental procedures

Bacterial strains, plasmid, phage and growth conditions

Strains used for cloning and expression of recombinant proteins were *E. coli* TOP10 (Invitrogen) and BL21(DE3)pLysS (Novagen) or BL21(DE3)Star (Novagen)

grown in LB medium at 37°C. Media were supplemented with ampicillin (100 µg ml^{-1}) or kanamycin (25 µg ml^{-1}), as required. Mycobacterial strains used were *M. smegmatis* mc^{2}155*, M. bovis* BCG Pasteur 1173P2 and *M. tuberculosis* H37Rv, Erdman and CDC1551. Mycobacteria were grown on Middlebrook 7H11 agar plates with OADC enrichment (Difco) or in Sauton's medium containing 0.05% tyloxapol (Sigma) supplemented with kanamycin (25 µg ml^{-1}) or hygromycin (50 µg ml^{-1}) when required. The shuttle phasmid phAE159 (Bardarov et al., 2002) and the integrative plasmid pMV361 (Stover et al., 1991) were reported earlier.

Cloning, expression and purification of recombinant InhA and mutant proteins

For *in vitro* kinase assays, the *inhA* gene was amplified by PCR using *M. tuberculosis* H37Rv genomic DNA as a template and a set of primers containing NdeI and NheI restriction sites **(Table A-2)**. This amplified product was digested with NdeI and NheI and ligated into pETPhos (Canova et al., 2008), generating pETPhos_*inhA_WT* **(Table A-1)**. Sitedirected mutagenesis was directly performed on this vector using inverse-PCR amplification with the self-complementary primers **(Table A-2)**. All constructs were verified by DNA sequencing. Recombinant InhA proteins were overexpressed in *E. coli* BL21(DE3) and purified as described (Veyron-Churlet et al., 2010). Fractions containing pure InhA proteins were pooled, dialysed when required and stored at -20°C until further use. For enzymatic or crystallographic studies, *inhA_T266A, inhA_T266D* and *inhA_T266E* were amplified from H37Rv genomic DNA using the *inhA*_forward primer (containing an NdeI site) with the appropriate reverse primer

(containing a HindIII site) **(Table A-2)**. The amplified products were digested with NdeI and HindIII and ligated into pET30b (Novagen). Mutation sites were verified by DNA sequencing. Expression and purification of InhA_WT (Quemard et al., 1995), InhA_T266A, InhA_T266D and InhA_T266E proteins were performed in *E. coli* BL21(DE3) as described (Freundlich et al., 2009).

*Construction of the integrative pMV361_**inhA** and cosmid p004_**inhA_hemZ***

inhA was amplified from H37Rv genomic DNA using primers inhAE1 and inhAH1 **(Table A-2)**. The PCR fragment was cut with EcoRI and HindIII and cloned into EcoRI–HindIII-cut pMV361 **(Table A-1)**, generating pMV361_*inhA*. To produce p004_*inhA_hemZ*, the *hemZ* gene was PCR-amplified from H37Rv genomic DNA using the RR and RL primers **(Table A-2)**. *inhA* was then PCR-amplified using upstream LL primers and downstream LR-266A, LR-266D or LR-266E primers **(Table A-2)**. PCR fragments were cut with Van91I, cloned into Van91I-cut p004 **(Table A-1)**. All constructs were sequenced before use.

Specialized transduction

The transducing mycobacteriophages were prepared as described (Bardarov *et al*., 2002). Briefly, the recombinant cosmids p004_*inhA_hemZ* were cut with PacI and ligated to the PacI-cut shuttle phasmid phAE159. The resulting phasmids were packaged *in vitro* (GigapackII, Strattagene) and transduced into *E. coli* HB101. Phasmid DNA was

electroporated into *M. smegmatis* mc2155 and the resulting transducing phages were amplified to obtain high-titre phage lysates. *M. bovis* BCG and *M. tuberculosis* H37Rv, Erdman and CDC1551 strains (100 ml) were grown to log phase (OD_{600} about 1.0), spun down, washed twice with mycobacteriophage (MP) buffer (50 mM Tris, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM $CaCl_2$; 100 ml) and finally resuspended in MP buffer (10 ml). The cell suspension (1 ml) was mixed with high-titre phage lysate (1010–1011 pfu ml^{-1}, 1 ml) and incubated at 37°C for 4 h. After centrifugation, the cell pellet was resuspended in Middlebrook 7H9 medium (0.4 ml) and plated onto Middlebrook 7H10 plates containing 75 µg ml^{-1} hygromycin. Plates were then incubated at 37°C for 4 weeks and transductants screened by PCR using the primer pairs TW361 and TW398 **(Table A-2)** to verify that the *inhA–hemZ* region had been disrupted and TW 361 and p004R **(Table A-2)** to check the integration of the *sacB–hyg* cassette between *inhA* and *hemZ*. Confirmation of allelic exchange was by Southern blotting on genomic DNA of HygR transductants digested with BamHI and probed with *inhA*.

In vitro *kinase assay*

In vitro phosphorylation was performed as described (Molle et al., 2003) with 4 µg of InhA in 20 µl of buffer P (25 mM Tris-HCl, pH 7.0; 1 mM DTT; 5 mM $MgCl_2$; 1 mM EDTA) with 200 µCi ml^{-1} [γ-^{33}P]-ATP corresponding to 65 nM (PerkinElmer, 3000 Ci $mmol^{-1}$), and 0.2–1.0 µg of kinase in order to obtain for each specific kinase its optimal autophosphorylation activity for 30 min at 37°C. Cloning, expression and

purification of the eight recombinant GST-tagged STPKs from *M. tuberculosis* were described previously (Molle et al., 2006).

Mass spectrometry analysis

Purified InhA_WT and mutant derivatives were subjected to *in vitro* phosphorylation by GST-tagged PknB as described above, except that [γ-^{33}P]-ATP was replaced with 5 mM unlabelled ATP. Purified InhA from *M. bovis* BCG cultures was subjected to mass spectrometry without further treatment. Subsequent mass spectrometric analyses were performed as previously reported (Fiuza et al., 2008).

Overexpression of InhA proteins in mycobacteria

inhA was amplified by PCR using *M. tuberculosis* H37Rv chromosomal DNA as template and a set of primers containing an NdeI and an EcoRI restriction site **(Table A-2)**. This amplified product was then digested by NdeI and EcoRI and ligated into the pMK1 expression vector **(Table A-1)**. The resulting construct was electroporated into *M. bovis* BCG and transformants were grown in broth medium and harvested for immunoblotting analysis as described previously (Kremer et al., 2003).

ENR assay

Enoyl reductase activity was monitored spectrophotometrically by following the change in absorption at 340 nm via the oxidation of NADH to NAD$^+$. All reactions were performed with a Cary 100 spectrophotometer at 25°C in 100 mM phosphate buffer pH

130

7.5 using 80 nM of either wild-type or mutant enzyme. Kinetic parameters K_m, k_{cat} and V_{max} for DD-CoA were determined at a fixed saturating concentration of NADH (150 μM) while varying the concentration of DD-CoA (25– 100 μM) using the equation $v = k_{cat}[E]_0[S]/(K_m + [S])$. The percentage of remaining activity was calculated by comparing the rate of the reaction of the mutant enzyme to the rate of the reaction of the wild-type enzyme in the presence of 50 μM DD-CoA and 100 μM NADH with wild-type enzyme activity set at 100%.

Fluorescence measurements

Binding constants of NADH to the wild-type, Asp and Glu mutant enzymes were obtained by measuring the protein fluorescence using a Cary Eclipse spectrofluorometer at 25°C. Excitation and emission wavelengths were 280 nm and 335 nm respectively. Emission was first monitored between 300 and 450 nm and 335 nm was selected as λ_{max}. Approximately 0.7–0.76 μM protein solutions in 100 mM Pipes buffer pH 7.0 were titrated with various concentrations of NADH (0–20 μM). The contribution of NADH fluorescence to final fluorescence intensity was checked with a control experiment by titrating NADH into buffer in the absence of protein. The fluorescence intensities at λ_{max} for a corresponding NADH concentration were corrected for dilution factors and NADH contributions before any calculations. At the concentrations used for InhA and NADH, the fluorescence intensity change due to inner filter effect was negligible. K_d values were calculated using the equation $\Delta F = \Delta F_{max} - K_d*(\Delta F/[NADH])$, where F represents fluorescence intensity (Li and Lin, 1996).

Complementation studies of the inhA(Ts) **M. smegmatis** *strain*

Cultures of *M. smegmatis* strains mc^{2}2359 (InhA(Ts), INHresistant) were grown at 30°C. Competent bacteria were prepared and transformed with pMK1, pMK1_i*nhA_WT*, pMK1_*inhA_T266A*, pMK1_*inhA_T266D* or pMK1_*inhA_T266E*. Clones selected on kanamycin were grown in Sauton medium at 30°C to mid-log phase and plated at 30°C or 42°C for 3–5 days.

Mycolic acid biosynthesis

Cultures of *M. smegmatis* mc^{2}2359 (InhA(Ts), INH-resistant) were grown to mid-log phase in Sauton medium at 30°C. The temperature was then shifted to 42°C for 3 h and cultures were labelled by adding [1-^{14}C]-acetate (1 μCi ml^{-1}). After a further 3 h incubation at 42°C, cells were harvested and FAMEs and MAMEs were extracted as reported (Kremer *et al.*, 2002). Equal amounts of counts were subjected to TLC using petroleum ether/acetone (95/5, v/v) and exposed overnight to a Kodak X-Omat film. As a control of mycolic biosynthesis inhibition, *M. smegmatis* mc^{2}155 (InhA_WT, INHsensitive) was treated with 50 μg ml^{-1} INH for 3 h and labeled by adding [1-^{14}C]-acetate (1 μCi ml^{-1}) for another 3 h at 37°C prior to mycolic acid extraction and analysis.

132

Additional unpublished results

ENR assays with in vitro *phosphorylated InhA*

In order to determine whether phosphorylation has any effect on InhA enzymatic activity, two serine-threonine kinases of *Mtb*, PknB and PknF, were used in an *in vitro* phosphorylation assay. As a negative control, two kinase-dead mutants of PknB (PknB^{K40M}) and PknF (PknF^{K41M}) were also included in the assays. The kinase domains were expressed in *E. coli* and purified to homogeneity as described previously (Molle et al., 2006). As seen from the *in vitro* phosphorylation experiments done with [γ-^{33}P]-ATP **(Figure 3-1A)**, InhA was phosphorylated by PknB but not by PknF. InhA and PknB/F were incubated at different molar ratios (1:20 or 1:10 kinase:InhA) at 37 °C in the presence of 1 mM ATP, 5 mM MgCl$_2$, and 1 mM DTT in phosphate buffer pH 7.5. Control experiments lacking either kinase, ATP or both were also set in the same way. 1 μl from this reaction mixture was aliquoted and InhA activity was monitored by the oxidation of NADH at 340 nm. However, the assay results were inconsistent and no difference in the InhA activity between the phosphorylation and control reactions was observed. Also, there was no correlation of InhA activity with the type of the kinase used in the assay; i.e., no difference was observed if the assay was performed with the wild-type or K40M/K41M mutant kinases. There may be a couple of reasons for the inconsistency in these experiments. The first issue was the stability of InhA in the phosphorylation reaction mixture. It was observed that InhA protein heavily precipitated after incubation at 37 °C, which meant that 1 μl aliquot from each reaction mixture did

not have equal amount of InhA protein to be used in the enzymatic assay. Also, since the efficiency of the phosphorylation reaction (phosphorylated/non-phosphorylated InhA ratio) was not known, it was difficult to compare the enzymatic activity from a mixture of phosphorylated and non-phosphorylated protein. Therefore, to overcome these issues, the phosphomimetic T266D and T266E mutants of InhA were generated. Even though aspartate and glutamate do not mimic the size of a phospho group perfectly, they do mimic the negative charge. Mimicking of serine and/or threonine phosphorylation by aspartate or glutamate has been widely used in the literature (Leger et al., 1997; Veyron-Churlet et al., 2009; Veyron-Churlet et al., 2010).

Intrinsic fluorescence quenching experiments

The affinity of the wild-type and mutant InhA proteins for NADH was determined by the intrinsic fluorescence quenching. InhA has four tryptophan residues, two of which (Trp222 and Trp230) lie close to the NADH binding pocket. It has been observed that InhA undergoes conformational changes upon transition from the apo form to the NADH bound form (Dias et al., 2007). Particularly, the substrate binding loop of the enzyme (residues 196-219) gets ordered and moves ~2 Å upon NADH binding (Dias et al., 2007). In addition, there are multiple tyrosine residues positioned close to the NADH binding pocket (especially the catalytic Tyr158), which may also contribute to the intrinsic protein fluorescence. Therefore, quenching of the intrinsic fluorescence by NADH titration reflects the NADH binding to the enzyme.

Initial excitation and emission scans showed that InhA had an excitation maximum at 280 nm. The emission was monitored between 300-450 nm and 335 nm was selected as the λ_{max}. At this emission wavelength, NADH did not contribute to the fluorescence intensity since the $\lambda_{excitation}$ and λ_{emis} for NADH were around 340 nm and 450 nm, respectively.

As the NADH concentration was increased, a decrease was observed in the fluorescence **(Figure 3-7A)**. When the wild-type and the T266D and T266E mutant proteins were compared, it was observed that the fluorescence quenching (the difference between F_{max} and $F_{observed}$; i.e., ΔF) was greater for the wild-type enzyme than the mutants **(Figures 3-7B and 3-7C)**. This indicated that NADH bound tighter to the wild-type enzyme. The K_d values for NADH binding were calculated from the fluorescence quenching data as described in the Experimental Procedures.

A
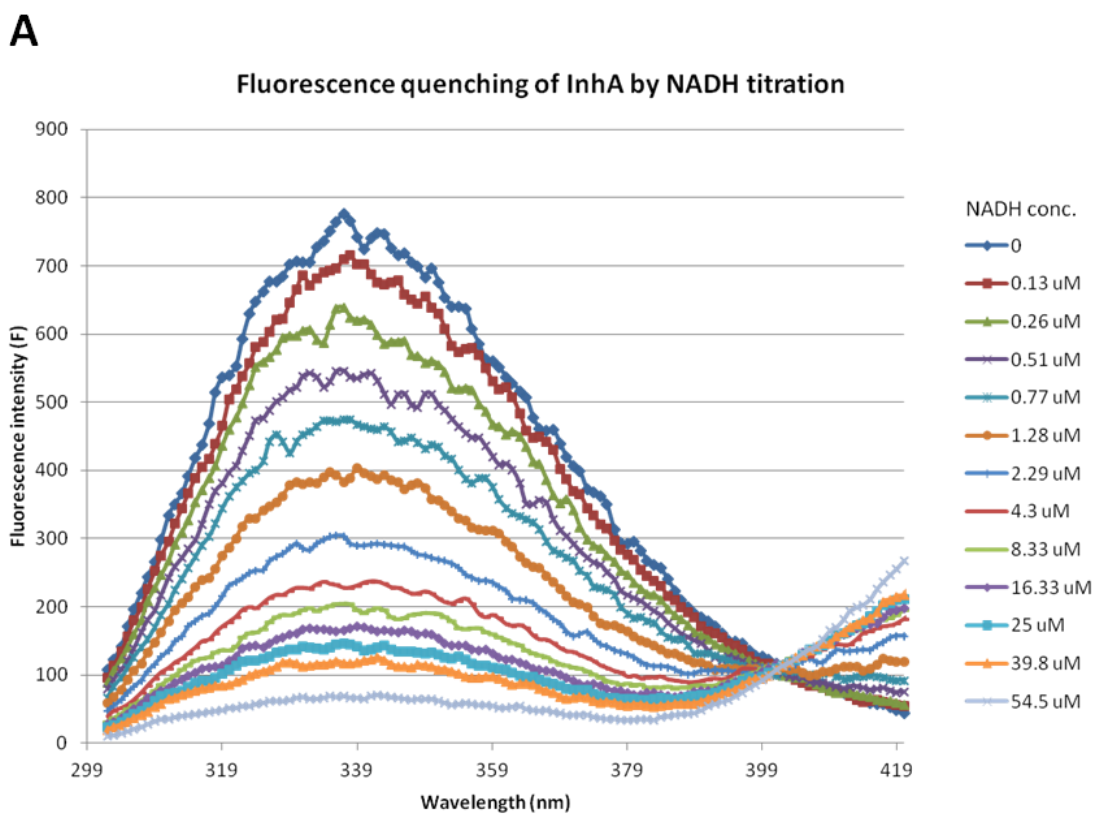
Figure 3-7. Intrinsic fluorescence quenching experiments for the wild-type and mutant InhA enzymes. A) Fluorescence quenching of wild-type InhA by NADH titration. The fluorescence intensity decreased as the concentration of NADH (0-54.5 μM) increased. Each titration is indicated with a different line color and the corresponding NADH concentrations are given on the right of the chart. **B and C)** Comparison of the fluorescence quenching of the wild-type and the T266 mutant InhA proteins. ΔF_{wt} was greater than ΔF_{T266A}, ΔF_{T266D} and ΔF_{T266E}, indicating that NADH bound tighter to the wild-type enzyme.
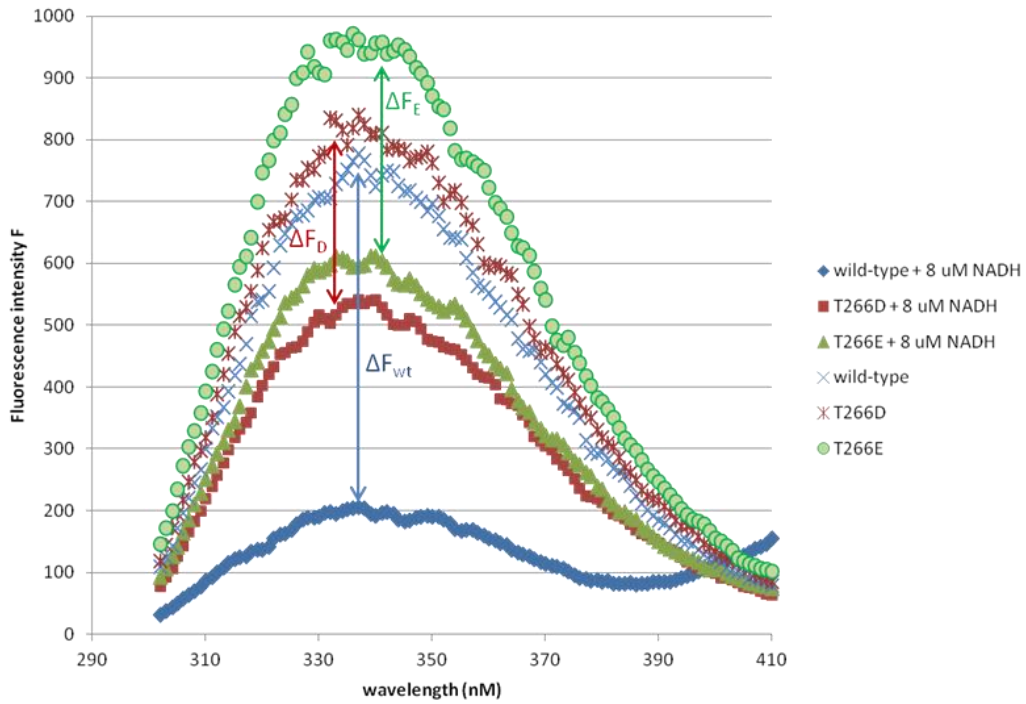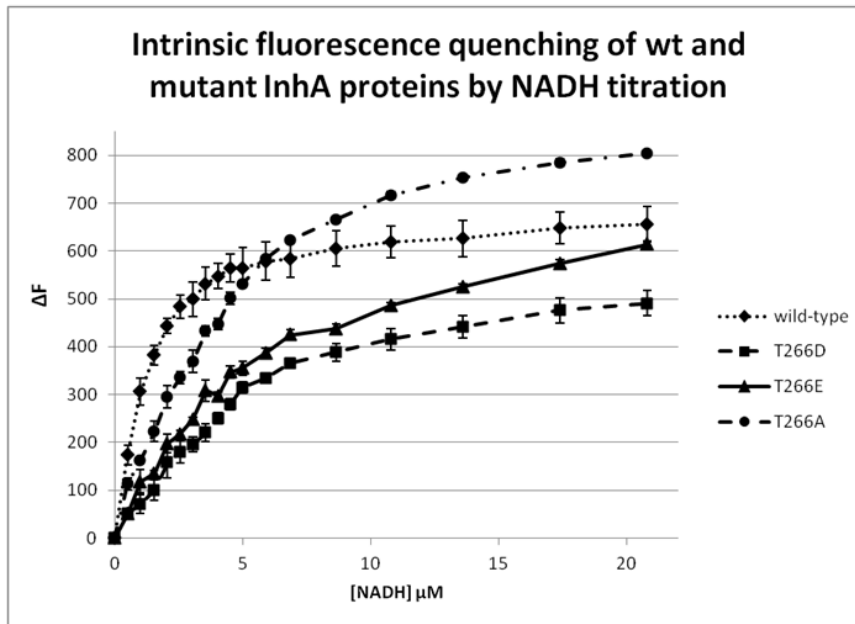
Figure 3-7 Continued.

137

Summary

InhA has a pivotal role in mycolic acid synthesis and it is essential for mycobacterial viability (Vilcheze et al., 2000). Being the molecular target of isoniazid and ethionamide, it still is one of the best-validated targets for antitubercular drug discovery. Therefore, understanding InhA regulation is quite important.

Phosphorylation/dephosphorylation is considered a central signal transduction mechanism that allows regulation of different cellular processes like growth, differentiation, mobility, and survival (Stock et al., 1989). It was recently shown that in *Mtb*, members of the FASII system, MabA, FabH, KasA, KasB and HadABC, were substrates of multiple serine-threonine kinases (STPK), and their functions were regulated by phosphorylation. While for the KasA, FabH, MabA, and HadABC enzymes phosphorylation decreased the enzymatic activity, for the KasB enzyme phosphorylation positively affected (increased) the activity (Molle and Kremer, 2010). In this study, it was demonstrated that InhA was also phosphorylated at a single threonine residue by the STPKs. Phosphorylation regulated the InhA function negatively. Kinetic studies and fluorescence experiments performed with the phosphomimetic T266D and T266E mutants showed that they retained only 30% of the enzymatic activity, which was due to their reduced affinity for NADH. The crystal structures of the wild-type and T266D/T266E mutants revealed the local structural changes around the mutated residue and indicated no significant conformational changes in the protein backbone and at the active site. InhA functions as a tetramer and it is possible that mutation of T266 alters the network of interactions between the different subunits. It was noteworthy that even

though the residue T266 lay ~16 Å away from the NADH binding site, mutation of this residue affected the cofactor binding and the enzyme activity significantly.

The effect of InhA phosphorylation on cell viability and growth was also demonstrated by genetic experiments. The *M. smegmatis* and *M. bovis* strains over-expressing the phosphomimetic InhA proteins displayed growth defects compared with the strains harboring the wild-type or the phosphoablative plasmids. The transfer of a point T266A mutation into the *inhA* gene in *Mtb* indicated that introduction of the phosphoablative allele was not lethal and did not have any effect on cell growth. In contrast, the transfer of T266D or T266E point mutations to the *inhA* gene was not feasible, suggesting that these alleles were lethal for *Mtb*. These results together implied that the presence of the phosphomimetic proteins became lethal for the cell because of the decreased InhA activity.

It is intriguing that all of the FASII enzymes are regulated by phosphorylation. Why does *Mtb* choose to phosphorylate all of the enzymes belonging to the same pathway? One suggested explanation is that altering the activity of these enzymes will lead to a tightly regulated system, which will allow adaptation to various growth conditions (Molle and Kremer, 2010). While phosphorylation of each individual FASII enzyme does not abolish the enzymatic activity completely, the cumulative effects of the partially reduced enzyme activities could result in a complete stop of the mycolic acid production.

CHAPTER IV

ACTIVATION-FREE INHIBITOR DISCOVERY STUDIES FOR INHA

Overview

The enoyl-reductase enzyme of the FASII system in *Mtb*, InhA, which is the biological target of the anti-tubercular drug isoniazid, has been shown to be essential for mycobacterial viability. Inhibition of InhA activity leads to the accumulation of FASI end products, which results in the disruption of mycolic acid biosynthesis, causing mycobacterial cell death. Considering that no FASII system is present in humans, InhA has been the focus of many drug discovery efforts since the 1990s. However, despite several reports on identification of potent InhA inhibitors, many of which were also effective *in vivo* against *Mtb*, INH remains the only first-line drug targeting InhA in the current TB treatment. The majority of the INH resistant *Mtb* clinical isolates were found to develop resistance through the mutations in the INH activator *katG* gene, which underlines the importance of the discovery of activation-free inhibitors.

In this work, we report novel InhA inhibitors that do not require activation by KatG. A target-based high-throughput screening (HTS) performed by GlaxoSmithKline (GSK) among a library of a million compounds resulted in the identification of a previously unexplored class of InhA inhibitors displaying good potency. These compounds effectively inhibited InhA enzymatic activity at nanomolar concentrations. The structural basis of InhA inhibition by this new class of GSK inhibitor compounds was determined by X-ray crystallography. The crystal structures of InhA complexed

with the GSK compounds shed light on the binding mode of these inhibitors, providing future assistance for further inhibitor development. Subsequent chemical optimizations yielded InhA inhibitor lead compounds, which exhibited potent activity against drug-sensitive and drug-resistant *Mtb* strains with good pharmacokinetic properties. They represent a promising class of inhibitors for future studies.

Introduction

Although the drugs currently used to treat tuberculosis (TB) were discovered more than 50 years ago, TB still accounts for 1.4 million deaths every year (Zumla et al., 2013b). While the majority of the TB cases are treatable with the current regimen of drugs, nevertheless, the rising incidences of multi-drug resistant and extensively drug resistant cases worldwide is threatening to nullify our advances in the fight against this disease. The current TB treatment regimen is long, complex and has several adverse effects, so lack of adherence is not unusual, leading to suboptimal responses (treatment failure and relapse), emergence of resistance, and continuous spread of the disease. In the last forty years there has been only one new drug approved for TB (Andries et al., 2005). Thus, there is a critical need for the development of drugs with shorter, simpler regimens as well as novel mechanisms of action that can be used for the treatment of the drug-resistant forms of the disease.

Both target-based and phenotypic screening approaches have been employed for the identification of new anti-tubercular drug leads (Kondreddi et al., 2013; Magnet et al., 2010; North et al., 2013). While a limited, but significant number of examples exist

141

for the latter (Rao et al., 2013; Remuinan et al., 2013), target-based approaches have encountered very limited success as previously demonstrated in the antibacterial field (Payne et al., 2007). Rather than invalidating the target-based approaches *per se*, this situation highlights the disconnect between the concepts like genetic validation of target essentiality and the amenability of that target to small molecule drug discovery. Confidence that inhibition of a single target will allow resolution of an infection cannot be fully supported by genetic validation of essentiality. A deeper understanding of the systems biology and the mechanisms underlying the antibiotic killing are important for the discovery of new antimicrobial therapies through the target-based approaches. Additionally, for reasons that are not always obvious, some targets are clearly more chemically tractable than others. For example, protein and cell wall synthesis and DNA gyrase have delivered multiple classes of published leads and marketed drugs, whereas there are no known inhibitors for many other essential gene products, despite a long history of antibacterial research (Kohanski et al., 2010; North et al., 2013; Payne et al., 2007). In the antitubercular field, only a very limited number of targets such as InhA, RpoB, DNA Gyrase, ATP synthase, and DprE1 have been shown to be targeted by potent bactericidal drugs or promising leads.

Isoniazid (INH) is a frontline anti-TB drug targeting InhA and it is an essential component of the TB treatment regimen. It has been administrated to more than 40 million people saving millions of lives. Its outstanding bactericidal effect rapidly blocks the bacterial transmission and significantly improves the patient's condition after a few weeks of treatment. Despite the seemingly simple structure of INH, its mode of action

has remained elusive for many years. INH penetrates the tubercle bacilli by passive diffusion and is activated by the bacterial anti-oxidant enzyme KatG to a range of reactive species including the isonicotinoyl radical, which forms an adduct with the nicotinamide adenine dinucleotide (NAD^+). This adduct inhibits the enoyl-ACP reductase enzyme encoded by the *inhA* gene leading to the block of mycolic acid biosynthesis. The dependency on KatG activation for the INH-mediated killing is also the source of the main clinical weakness associated with the use of INH, as between 40-95% of INH-resistant *Mtb* clinical isolates develop mutations in *katG,* leading to a decreased activation of INH to its active form. While mutations are also detected in clinical isolates within the *inhA* promoter region, these can be successfully treated in most instances by increasing the dose of isoniazid.

On the basis of the mode of action of INH, it has been proposed that direct inhibitors of *Mtb* enoyl-ACP reductase would retain the outstanding antitubercular-cidal profile of INH while overcoming most of the problems associated with its pro-drug nature, such as the resistance and toxicity issues. Also, compounds inhibiting InhA without requiring activation by KatG could be active under anaerobic conditions, given the fact that the catalase-mediated activation is suppressed by hypoxia. Novel classes of direct InhA inhibitors have been identified previously using high-throughput screening strategies. They include the indole-5-amides (Kuo et al., 2003), the pyrazole derivatives (Kuo et al., 2003), the pyrrolidine carboxamides (He et al., 2006), the arylamides (He et al., 2007), and the imidazopiperidines (Wall et al., 2007). Additionally, the natural product pyridomycin was found to target InhA (Hartkoorn et al., 2012). However, most

of these compounds showed a lack of correlation between the enzymatic inhibition and the whole cell activity, had moderate potencies or narrow selectivity windows, making them unsuitable for further progression as drug leads.
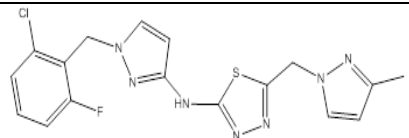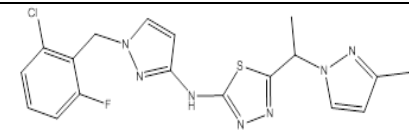
GlaxoSmithKline, under the sponsorship of the TB Alliance has carried out a high-througput screen against InhA using the GSK compound collection and has identified the thiadiazole series as the most promising small molecule anti-tubercular family. The structural basis of InhA inhibition by the thiadiazoles has been determined by X-ray crystallography. In the present study, we present the novel lead compound and its attractive anti-tubercular properties.

Results and discussion

Biochemical and antimicrobial characterization of the hit structure GSK826625A

Several screening campaigns were run against InhA with an overall 0.63% of true hits after the removal of the quenchers. The campaign covered a total of 1.9 million compounds (HTS details to be published elsewhere) unearthing GSK826625A (GSK625A) and GSK826613A (GSK613A) **(Table 4-1)** as the most attractive representative structures.

Table 4-1. Structures and activities of the hit compounds GSK613A and GSK625A.
Compounds were assessed for activity against *M. tuberculosis H37Rv* and against a broad-spectrum bacterial panel, cytotoxicity in HepG2 cells, clearance (*in vitro* and *in vivo*) and bioavailability (% F) in mice. o/e: over-expression

	GSK613	GSK625
InhA IC$_{50}$	8 nM	4 nM
MIC H37Rv	2 μM	1 μM
MIC shift o/e	> 4 fold	> 4 fold
Antibacterial panel	≥64 μg/mL	≥64 μg/mL
HepG2 cytotox	> 100 μM	> 100 μM
In vitro Cli mouse	11.9 mL/min.g	14.9 mL/min.g
In vitro Cli human	2.9 mL/min.g	2.8 mL/min.g
In vivo Cli mice	157mL/min/Kg	120 mL/min/Kg
% F	4 %	88 %

GSK613A (N-(1-(2-chloro-6-fluorobenzyl)-1H-pyrazol-3-yl)-5-(1-(3-methyl-1H-pyrazol-1-yl)ethyl)-1,3,4-thiadiazol-2-amine) was progressed to enzymatic mode of action studies. GSK613A was a selective inhibitor against *Mtb* enoyl-ACP reductase and it was inactive against human FAS. Steady state competition experiments showed that the hit structure competed with the fatty acid substrate for binding to the enzyme, but it

was uncompetive with NADH, consistent with inhibitors binding to a pre-formed enzyme:NADH complex **(Figure 4-1)**.

Best fit: NADH uncompetitive and DDCoA competitive

Compound	fixed substrate	Variable substrate	Model of inhibition	Kiu (uM)	Kic (uM)
GSK826613A	500 uM NADH	DDCoA	Competitive	-	0.040
GSK826613A	100 uM DDCoA	NADH	Uncompetitive	0.034	-
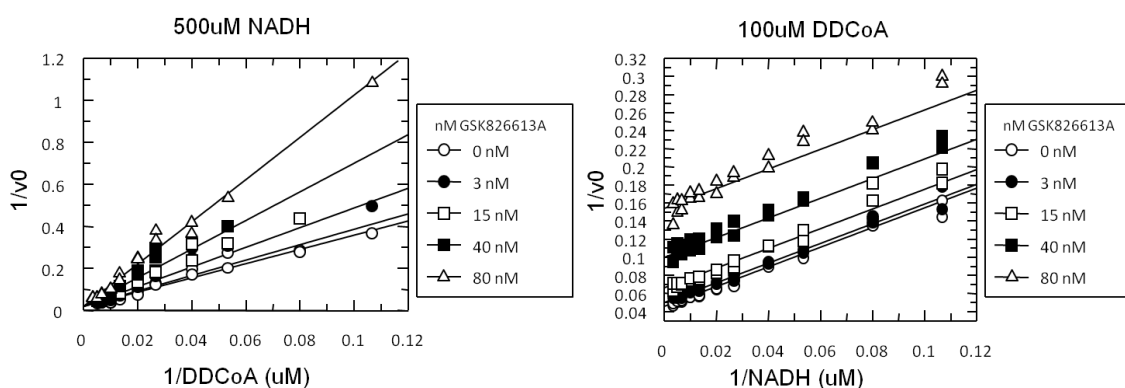
Figure 4-1.Inhibition of InhA by members of the thiadiazole family. Double reciprocal plots in the absence of inhibitor (○) and in the presence of 3 nM (●), 15 nM (□), 40 nM (■) and 80 nM (△) inhibitor compound. The patterns of inhibition observed is consistent with the compound binding to the InhA:NADH complex.

Whole cell mode of action confirmation in InhA overexpressor

In clinical isolates, over-expression of InhA translates into INH and ETH resistance, which confirms that the enoyl ACP-reductase is the main antitubercular target for both drugs (Vilcheze et al., 2006). Similarly, two *Mycobacterium smegmatis* and *M. bovis* BCG strains over-expressing the *M. smegmatis* or *Mtb* InhA protein were

constructed and the susceptibility of these strains for the compounds were tested in order to verify that the whole-cell activity of the inhibitors was mediated mainly or solely by the inhibition of InhA. A reproducible shift in minimum inhibitory concentration (MIC) values higher than 4-fold in strains overexpressing the target was considered as an indication that the antitubercular activity was mainly mediated by InhA inhibition. The thiadiazoles GSK625A and GSK613A met this criterion **(Table 4-1)**.

Structure of the InhA: GSK625A complex

InhA was co-crystallized with GSK625A and GSK2136490A (GSK490A) in the presence of the cofactor NAD^+. GSK490A is an analog of GSK625A, and lacks the halide substituted phenyl ring attached to the pyrazole ring **(Figure 4-2)**. GSK490A displayed ~6000 fold less inhibitory activity ($IC_{50} = 28$ μM) against InhA compared with the GSK625A ($IC_{50} = 4$ nM). The reason behind this potency difference was revealed by crystallography.

The enzyme:inhibitor complex crystallization was successful only in the presence of the cofactor NAD(H) in agreement with the uncompetitive character of these inhibitor compounds **(Figure 4-1)**. Crystals of the InhA:NAD^+:GSK490A ternary complex were determined to be in the $P6_222$ space group with one molecule in the asymmetric unit (ASU), whereas the crystals of the InhA:NAD^+:GSK625A complex belonged to the $P2_12_12_1$ space group with four molecules in the ASU. Both structures were solved by molecular replacement using the previously determined InhA:NADH structure as a model (PDB ID: 3OEW) (Molle et al., 2010), to a resolution of 2.0 Å and

2.9 Å, respectively. The structures were refined to R_{work} of 19% and 20%, and R_{free} of 21% and 23%, respectively, with excellent stereochemistry **(Table 4-2)**.

The crystal structures revealed that both compounds bound to the hydrophobic substrate binding pocket of InhA, which was surrounded by the residues Met103, Phe149, Tyr158, Ala198, Met199, Ile202, Leu207 and Ile215. Comparison of the InhA:GSK inhibitors and the InhA:NAD$^+$:C16-fatty acyl substrate analog structure (PDB ID: 1BVR) (Rozwarski et al., 1999) revealed that the aromatic rings of the GSK compounds occupied the same pocket as the thioester and the *trans* double bond portion of the C16-substrate analog **(Figure 4-2A)**, and they positioned proximate to NAD$^+$, which is in agreement with the competitive behavior of the inhibitors relative to the DDCoA substrate.

Table 4-2. Data collection and refinement statistics for the InhA:GSK625A and InhA:GSK490A structures. Highest resolution shells are given in parenthesis.

	InhA: GSK490A	InhA:GSK625A
Data collection		
Space group	$P6_222$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)		
α, β, γ (^{0})	97.9, 97.9, 139.4	90.2, 104.2, 190.8
	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Wavelength (Å)	1.542	1.542
Resolution (Å)	24.2-2.0 (2.01)	47.7-2.91 (2.91)
Completeness (%)	99.7 (99.9)	97.7 (95.5)
Redundancy	8.3 (8.0)	3.9 (3.8)
I/I_σ	21.8 (3.0)	13.9 (1.8)
R_{sym} (%)	2.8 (35.2)	6.1 (57.4)
Refinement		
Resolution	23.2-2.01	34.7- 2.91
No. reflections	26922	39444
R_{work}/ R_{free}	0.19/ 0.21	0.20/ 0.23
No. atoms		
Protein	2058	8264
Water	179	55
B-factors		
Protein	35.5	74.1
Water	42.8	54.9
R.m.s deviations		
Bond lengths (Å)	0.007	0.012
Bond angles ($^\circ$)	1.426	1.726
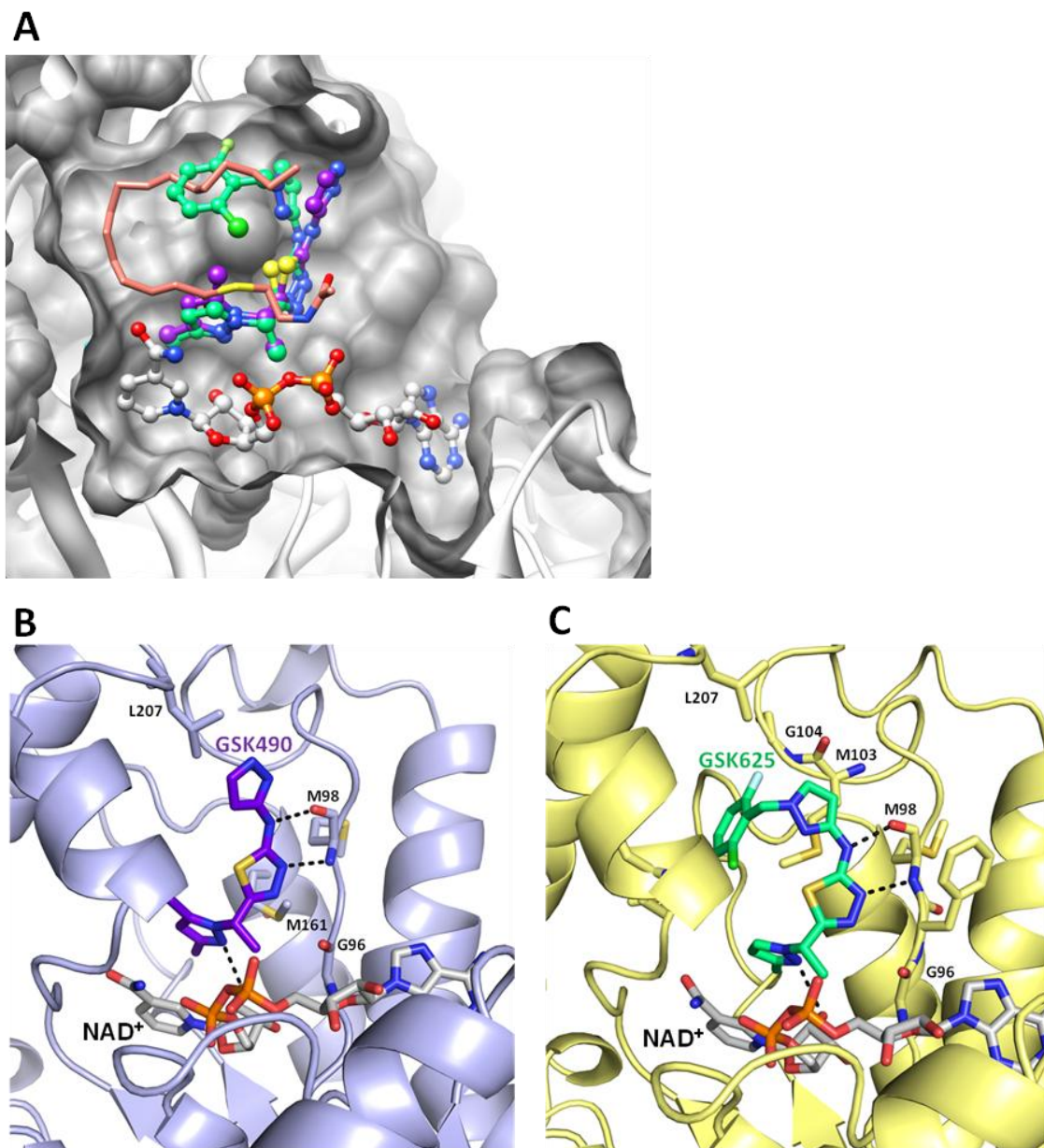
Figure 4-2. Binding mode of the thiadiazoles to the InhA active site. A) A cross-section through the surface of the active site of InhA. InhA:NAD$^+$:GSK structures (490 (purple) and 625 (green), respectively) are superposed with the InhA:NAD$^+$:C16-fatty acyl substrate analog (orange) structure. As seen from the figure, the GSK compounds occupied the substrate binding pocket**. B-C)** Key interactions between the GSK compounds 490 and 625 and the InhA active site. Compounds interacted with NAD$^+$ (gray) and M98 via H-bonds; with G96, F97, M103, F149, M161, I202, and L207 through hydrophobic and van der Waals interactions. GSK625 has additional hydrophobic and van der Waals interactions with G104, A157, Y158, A198, and M199 compared with GSK490, which results in more than 1000 fold better potency. Atom coloring is oxygen red; nitrogen blue; sulfur yellow; chlorine and fluoride green. H-bonds are shown with dashed black lines.

The specificity and potency of the thiadiazoles for InhA was provided by the interactions with both NAD^+ and the protein backbone. GSK625A and GSK490A interacted with the cofactor NAD(H) via π-π stacking interactions through their pyrazole rings (3.6 - 4 Å) **(Figure 4-2)**. In addition, the nitrogen of the pyrazole ring made a hydrogen bond to the 2' hydroxyl of the ribose moiety of NAD^+ (2.7 and 3.2 Å). The other key interactions were the direct H-bonding between (i) the nitrogen of the thiadiazole ring and the amide NH of Met98, and (ii) the nitrogen linking the pyrazole and the thiadiazole rings and the carbonyl oxygen of Met98 (2.4 & 2.9 Å, respectively) **(Figure 4-2)**. Met98 was the only residue that participated in a direct hydrogen bonding interaction with the thiadiazoles.

Hydrophobic and van der Waals interactions between the inhibitors and the side chains of the residues Gly96, Phe97, Met103, Phe149, Met161, Ile202, and Leu207 dominated the ligand-protein interactions for both compounds **(Figure 4-2)**. This was not surprising since the inhibitors occupied the hydrophobic substrate binding pocket. Compared with the GSK490A, the halide substituted aryl group of GSK625A made additional hydrophobic and van der Waals interactions with Gly104, Ala157, Tyr158, Ala198, and Met199. In particular, the fluoride atom of GSK625A interacted with Gly104. This can explain a shift of three orders of magnitude in the enzymatic potency of the two thiadiazoles (GSK490A IC_{50}=28 μM; GSK625A IC_{50}=4 nM).

Noticeably, neither GSK490A nor GSK625A caused a significant conformational change in the main chain backbone of InhA (rmsd is 0.5 Å), or perturbed the position or conformation of the active site residues. Exceptionally, in the InhA:GSK625A structure,

the substrate binding loop (especially the helix between the residues 208-226) moved 2.2 Å away from the active site (the distance between the C_αs of Glu209) compared with the InhA: NAD(H) structure, due to the presence of the halide substituted phenyl moiety of the inhibitor. Unlike the INH-NADH adduct, GSK compounds did not cause the flipping of the Phe149 side chain, thus they did not interact with the isonicotinic acid binding pocket. The catalytic residue Tyr158 also adopted the identical position as in the InhA:NAD(H) structure (PDB ID: 3OEW) (Molle et al., 2010). It is worth noting that Tyr158 did not make any hydrogen bonds with the GSK inhibitors and it participated in ligand binding only through van der Waals interactions (4.2-5.4 Å). In contrast to the previously identified InhA inhibitors such as pyrolidine carboxamides (He et al., 2006), Genzyme10850 and triclosan (Kuo et al., 2003), the GSK inhibitors did not interact with Tyr158 directly. Consequently, the GSK compounds are representatives of a new class of inhibitors, which do not need this conserved network of interaction with Tyr158 for their potency.

Recently, it was proposed that high affinity slow binding inhibition of InhA is related to the ordering of the substrate binding loop through the increased residence time of the inhibitor compounds on the enzyme (Lu and Tonge, 2008). Importantly, in both InhA:NAD$^+$:GSK490A and InhA:NAD$^+$:GSK625A structures, the substrate binding loop (residues 197-226) was ordered and well resolved in the crystal structure, as previously seen with the high-affinity slow onset FabI inhibitor structures. This implies that the compounds may also be slow, tight binding inhibitors, which generally enhances *in vivo* activity and increases their potential as antitubercular drugs.

The thiadiazoles are bactericidal

In order to test whether the thiadiazoles share the fast killing mode of action of INH, we determined the *in vitro* killing rates of GSK625A at 20X MIC in growth medium inoculated with approximately 10^6 bacteria, during seven days (**Figure 4-3A**). After the first two days of incubation, GSK625A behaved similarly to INH, killing more than 99% of the initial inoculum. Due to the high spontaneous mutation frequency to INH resistance, the number of bacteria increased after 4 days of culture in the presence of this drug. In contrast, the number of colony forming units kept decreasing down to the detection limit for the cultures treated with GSK625A. A very similar effect was also observed against *Mtb* growing inside the THP1 differentiated macrophages (**Figure 4-3B**), indicating the potent activity against the actively growing intracellular bacteria.
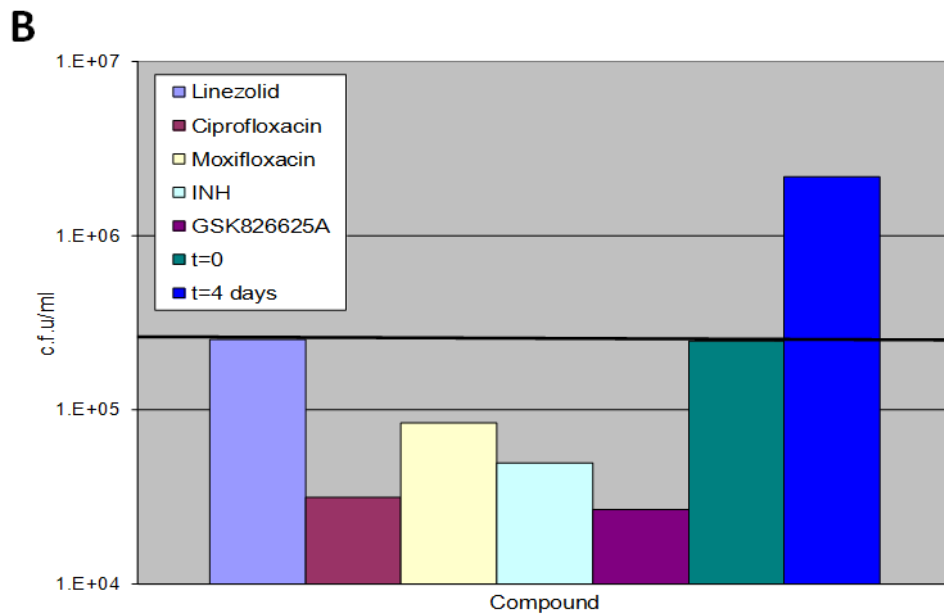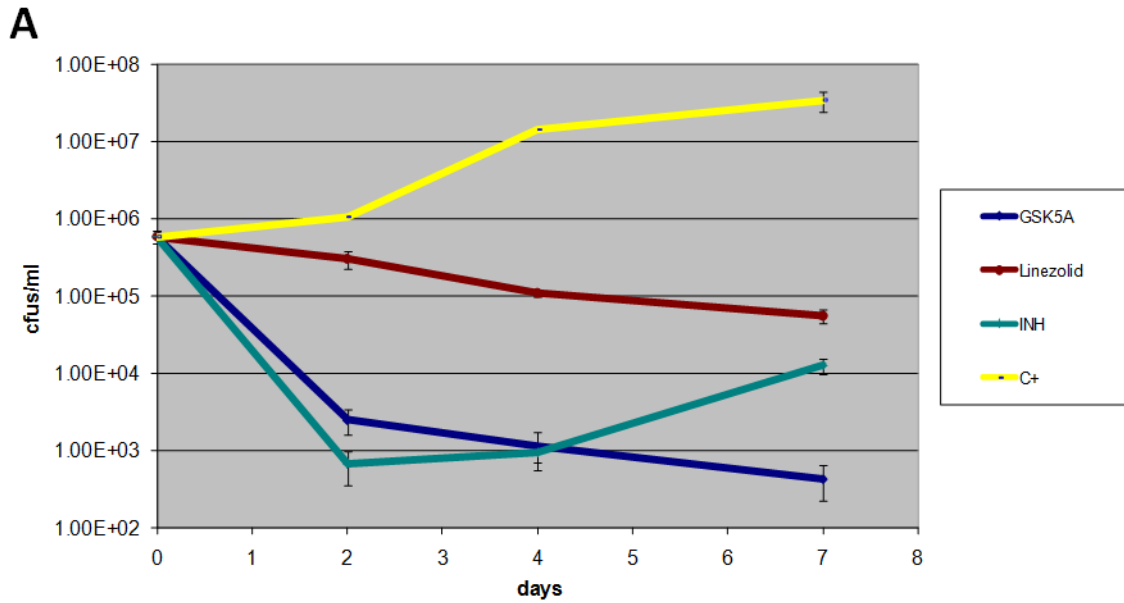
Figure 4-3. *In vitro* **activity of GSK625A. A)** Time killing kinetics of GSK625A in growth medium. **B)** Activity of GSK625A against *Mtb* growing inside THP1 macrophages.

GSK625A generates resistant mutants at lower frequency than INH and mutations

map in the InhA operon as established by targeted PCR

The *in vitro* rate of spontaneous resistance for GSK625A at 20X the MIC was $3.7x10^{-8}$ mutants/cfu, very similar to the frequency obtained with rifampicin and two orders of magnitude lower than that for isoniazid (Bergval et al., 2009). This rate is consistent with the presence of a single defined target. GSK625A resistant mutants exhibited a shift in MIC values by at least an order of magnitude and displayed no cross-resistance with other antitubercular compounds such as INH or RIF **(Table 4-3)**. Since the GSK compounds bound to the enzyme competitively with the DDCoA, it could be expected that mutations in the substrate binding site could affect the interaction of the thiadiazoles with the protein, while leaving the INH and ETH activity unaffected. The *inhA* alleles from single colonies of the seventeen mutants were amplified and sequenced. Four different point mutations were found in the protein coding regions from all mutants, producing a single amino acid change at either of two positions: Gly96 or Met103. These two amino acids map to the active site of the enzyme **(Figure 4-2)** and have not been reported previously to be involved in resistance to any InhA inhibitor.

Table 4-3. GSK625A resistant strains carry mutations in the *inhA* gene. Spontaneous resistant mutant colonies were isolated and sequenced resulting in a number of key InhA mutations being identified. All mutations gave rise to significant MIC shifts against GSK625A.

	H37Rv MIC	InhA operon mutations			
		G96V fold change	M103V fold change	M103I fold change	M103T fold change
GSK625A	1	>16	>16	>16	>16
INH	0.4	1	1	1	1
Rifampicin	0.002	1	1	1	1

In order to link the point mutations in *inhA* to the resistant phenotype, the drug sensitivity of the encoded enzyme and the bacteria expressing the mutant genes were analyzed. Wild-type InhA and the four different point mutants were over-expressed as untagged proteins in *E. coli* and purified. All mutant enzymes displayed *in vitro* V_{max} and K_m values similar to the wild type enzyme. A good correlation was observed between the IC_{50} and the MIC values **(Table 4-4)**. For the whole-cell analysis, the complete *inhA* operon (*fabG1-inhA-hemZ*) from the wild type H37Rv strain and from the resistant mutants were cloned in a multicopy plasmid and transformed into fast growing *M. smegmatis* mc^{2}155 and into slow growing *M. bovis* BCG and *M. tuberculosis* H37Rv. In all cases, the genetic transformation of the mutant operons conferred resistance to the thiadiazoles well above that obtained by expressing the wild type version from the multicopy plasmid **(Table 4-3)**.

Table 4-4. Activity of GSK625 against InhA purified mutant proteins and *Mtb* H37Rv carrying the mutated alleles.
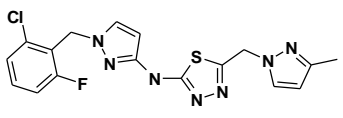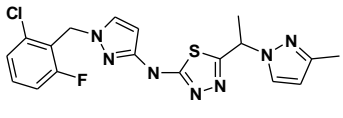
InhA allele	InhA IC50 (uM)	IC50 Ratio	MIC (uM)	MIC Ratio
Wild Type	0.0028	1	1	1
M103I	0.396	144	>125	>125
M103V	1.011	368	>125	>125
M103T	0.829	302	>125	>125
G96V	>1	>357	>125	>125

Activity of the thiadiazoles against sensitive and M(X)DR **Mtb** *clinical isolates*

Other InhA inhibitors, such as INH and ETH have been in use for decades and resistant *Mtb* strains have emerged without identifiable mutations in *inhA* or in any other studied gene (Vilcheze and Jacobs, 2007). It was therefore relevant to test a variety of clinical drug-resistant isolates for the possibility of pre-existing cross-resistance to the thiadiazole class. Two thiadiazole compounds were tested in a first instance against a group of INH-resistant clinical isolates with characterized mutations, either in KatG (the most frequent resistant class) or in nucleotide position 15 upstream of the transcription start site for InhA (promoter-up mutations). The compounds retained full activity against the first group of mutants, while a moderate shift (4-8 times) in MIC was observed against the group of strains over-expressing InhA, as described above for the laboratory constructs **(Table 4-5)**.

Table 4-5. **Activity of thiadiazole derivatives against INH resistant clinical isolates.**
CI: Clinical isolate; NT: not determined.

Compound/Strain		Wild Type H37Rv	-15 InhA CI1	CI2	CI3	KatG mutant CI4	CI5	CI6	CI7
GSK826613A	MIC (ug/ml)	2	>8	>8	>8	2	2	2	1
	Ratio	1	>4	>4	>4	1	1	1	0.5
GSK826625A	MIC (ug/ml)	1	8	8	NT	1	0.5	0.5	NT
	Ratio	1	8	8		1	0.5	0.5	
INH (ug/ml)	MIC (ug/ml)	0.25	2	2	2	64	32	32	32
	Ratio	1	8	8	8	256	128	128	128

GSK625A has been tested against recent clinical isolates of *Mtb* at two different TB reference centers (National Institute of Health Carlos III, Madrid and Hospital Vall d'Hebron, Barcelona) and two observations were noteworthy. The first one was that the MIC_{90} obtained with 100 recently isolated *Mtb* stains **(Table 4-6)** ranged from 0.3 to 3 μM, showing that the thiadiazole sensitivity of our laboratory H37Rv strain was representative for the majority of the *Mtb* strains circulating in Spain, which were therefore sensitive to the compound. The second was that GSK625A was active against the most resistant isolates present in the sample (13 XDR and 12 MDR), with MIC values similar to that of the laboratory strain H37Rv independently of the resistance profile of the clinical strains **(Tables 4-6 and 4-7)**.

Table 4-6. Activity of GSK625A against recently isolated clinical strains.

Strain	MIC(uM)	Strain	MIC(uM)	Strain	MIC(uM)	Strain	MIC(uM)
175	1	10	3	2177	0.3	2557	1
186	0.3	105	3	2194	1	2637 HSRZ	3
191	3	122	3	2195	1	2638 HSRZ	3
200	0.3	135	3	2199 HS	1	2644 H	9
205	0.3	143 HSERZ	1	2200	0.3	2645 HER	1
207	0.3	153	3	2206 HR	3	2901	1
209	0.3	163	3	2207 HR	3	2935	1
215	3	1901 H	1	2221 M	0.1	2936	1
224	0.3	1943	1	2225	0.3	2944	1
228	0.3	1946	1	2227	1	2945	3
236	9	1947	1	2229	1	2948	3
238	0.3	1969	1	2231	1	2980	1
241	1	1973	0.3	2234	1	2982	3
243	3	2014	0.3	2247 H	1	2983	0.3
248	1	2018	0.3	2271 HRSZ	0.3	3020	1
250	3	2019	3	2285 HERZ	3	3023	3
252	9	2020	1	2316 H	1	3035	3
256	1	2038	1	2471	1	3043	1
258	3	2047	0.3	2482	1	34	3
263	0.3	2049	0.3	2496	1	42 HSERZ	3
268	9	2140	0.3	2497	1	52	0.3
271	1	2152	1	2521	0.3	62	9
275	3	2157	0.3	2527	1	70	9
301	3	2166	0.3	2545	1	86	3
320	1	2169 HR	1	2547	1	91	1
						H37Rv	1

Resistences				Genotype	
H	INH			KatG (S315G)	
R	Rif			KatG (S315G); -15 InhA	
E	Etahmbutol			-15 InhA	
Z	Pirazinamyde			INH^R No KatG, No InhA	
S	Streptemycin				
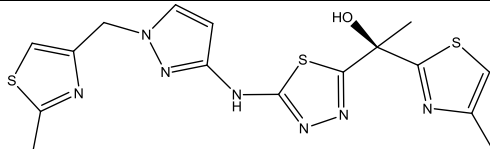
Table 4-7. Activity of GSK625A against *Mtb* strains resistant to different antitubercular compounds. The resistant profile of the tested isolates is shown.

Compound/Strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	H37Rv
Rifampicin	R	R	R	S	S	R	R	R	R	R	R	R	R	R	R	R	S	S	R	S	S	R	S	R	S	R	R	R	R	R	S
Isoniacid	R	R	S	S	S	R	R	R	S	R	R	R	R	R	R	R	S	R	R	R	S	R	S	R	R	R	R	R	R	R	S
Etambutol	R	R	S	S	S	R	R	R	S	R	R	R	R	R	R	S	R	R	S	S	ND	S	R	S	S	R	R	R	R	R	S
Streptomycin	R	R	S	S	S	R	R	S	R	R	S	S	R	R	R	R	S	R	R	R	ND	S	R	S	R	R	R	R	R	R	S
Pirazinamide	R	ND	R	S	S	R	R	R	S	R	S	S			R	R	S	R	S	S	S	R	S	R	S	R	R	S	R	R	S
Capreomycin	R	R	R	ND	S	R	ND	S	S	S	R	S	R	S	R	R	S	S	R	S	R	S	S	S	S	S	R	S	S	R	S
Kanamycin	R	R	R	R	S	R	R	S	S	R	S	R	S	R	R	R	S	S	R	S	S	S	S	R			ND	R	R	S	S
Amikacin																				S	S	S	S		S	R					S
Moxifloxacin																			S	S	R	R	R		R	S					S
Ciprofloxacin	R	R	R/S	R	R	R	R	R	R	R						R	R	R	S	R	S	R	R	R	R		R	R	R	R	S
Ofloxacino											R	R	R	R			R	R	R	S	R	R	R	R		S					S
Etionamide	S	S	S	S	S	S	S	S	S	S	S			R	S	R	S	S	R	ND	S	S	R	S	S	S	R	S	S	S	R	S
PAS											R			R		R	S	S	R	S	R	S		S	S						S
GSK826625A (uM)	3	9	1	1	3	3	3	9	1	1	1	3	27	3	3	1	0	3	3	1	1	27	1	3	3	3	27	3	9	3	3

Lead optimization and identification of GSK2505693A

Despite the interesting *in vitro* anti-tubercular profile stated in Table 4-1, the initial thiadiazole hits were affected by a number of compound development liabilities (drug metabolism and pharmacokinetics (DMPK) properties) that precluded their progression to the available *in vivo* acute murine TB efficacy model. Further medicinal chemistry efforts were hence dedicated to the identification of optimized compounds that were able to retain or even improve the anti-tubercular profile and simultaneously meet the good pharmacokinetic exposure values. These efforts led to the identification of GSK2505693A (GSK693A) as a new lead representative structure **(Table 4-8)**.

Table 4-8. Structure of the optimized lead GSK693A. The lead was assessed for activity against *M. tuberculosis H37Rv* both intra- and extracellularly and against a broad-spectrum bacterial panel, cytotoxicity in HepG2 cells, clearance (*in vitro* and *in vivo*) and bioavailability (% F) in mice.

	 GSK693A
InhA IC$_{50}$	7 nM
MIC H37Rv	0.2 µM
MIC intracellular	0.2 µM
Antibacterial panel	≥64 µg/mL
HepG2 cytotox	> 100 µM
In vitro Cli mouse	2.1 mL/min.g
In vitro Cli human	0.2 mL/min.g
In vivo Cli mice	157mL/min/Kg
% F (at 100 mg/Kg)	92 %

GSK693A improved the previous leads' *in vitro* potency values against H37Rv and was able to reach acceptable exposure levels after a single oral administration at 50 mg/kg. The cross resistance experiments with the previously isolated mutants showed that the observed TB activity for the new lead compound was target related (data not shown). GSK693A fulfilled the minimum requirements in terms of *in vitro* potency and the PK properties to be progressed to *in vivo* proof of concept efficacy studies in both acute and chronic TB murine infection models.

Acute and chronic **in vivo** *efficacy of GSK693A*

The *in vivo* acute efficacy of GSK693A was evaluated using a previously reported model (Rullas et al., 2010) at doses of 30, 100 and 300 mg/Kg, using Moxifloxacin as a positive control at 30 mg/Kg. GSK2505693A at 100 mg/Kg was shown to match the efficacy of moxifloxacin at 30 mg/Kg **(Figure 4-4A)**. The higher dose employed (300 mg/kg) produced a log cfu reduction in the lungs of infected mice equivalent to INH (data not shown). GSK693A was also progressed for evaluation in an established model of TB infection. In this model GSK693A (300 mg/Kg) was able to almost match the behaviour of isoniazid (25 mg/Kg) **(Figure 4-4B)**.

A
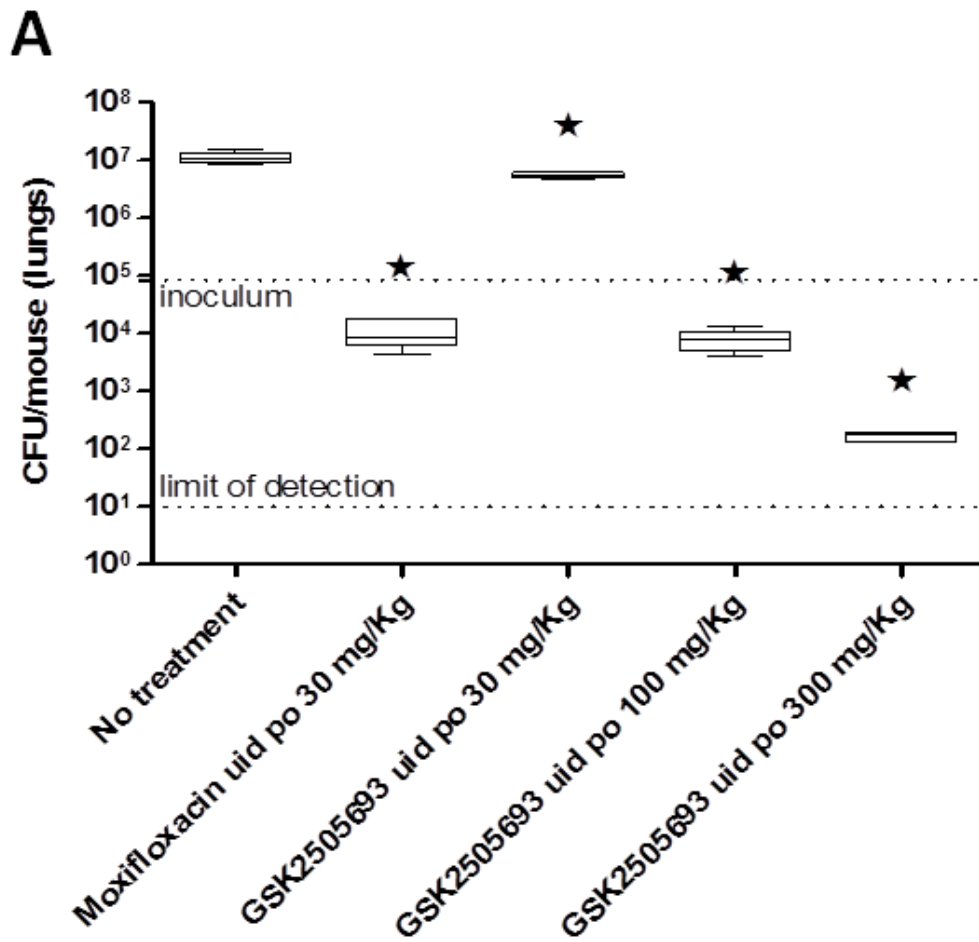
Figure 4-4. Dose-response studies of GSK693A. A) In an acute murine model of TB efficacy. **B**) In chronic, established murine model of TB infection.
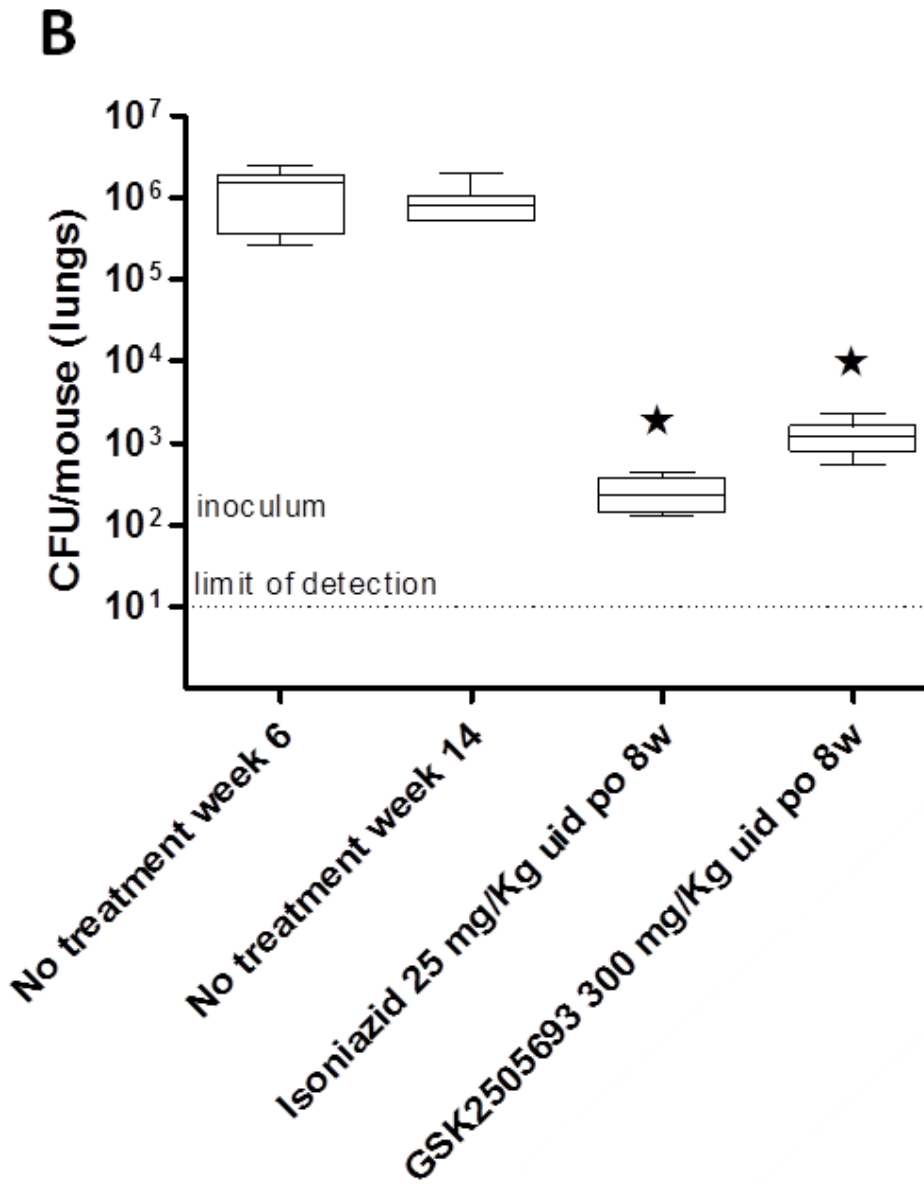
Figure 4-4 Continued.

Experimental procedures

Bacterial strains and culture

Mycobacterium smegmatis mc^{2}155 (Snapper et al., 1990) and *M. bovis* BCG Pasteur (Institut Pasteur) were grown at 37ºC in Middlebrook 7H9 broth (Difco) supplemented with 0.0025% Tween 80 and 10% albumin-dextrose-catalase (ADC) or on Middlebrook 7H10 plates supplemented with 10% oleic acid-albumin-dextrose-catalase (OADC). Cell-free extracts were done in 7H9 supplemented with 100 ml of 10X AS solution (5% albumin solution in salt: 10 mg albumin, 1.7 mg NaCl in 200ml water), 2.5 ml of 10% Tween 80 solution and 0.1% carbon substrate (acetamide, succinate, or glucose). *Escherichia coli* DH5α was grown in LB broth (LB).

Cloning and purification of InhA

The i*nhA* gene was amplified from *Mtb* genomic DNA by PCR and cloned into the pET15b vector (Novagen) using *NcoI* and *BamHI* restriction sites. The plasmid was transformed into BL21(DE3) *E.coli* cells for protein overexpression. Cells carrying InhA over-expression plasmid were cultured in terrific broth media together with 100 μg/ml carbenicillin as the antibiotic resistance selection marker at 37ºC with continuous shaking till the OD$_{600}$ reached 0.6-0.8. Cells were cooled down to 18ºC and induced with 0.5 mM IPTG for 16-18 hours, which gave ~20 mg of soluble protein per liter culture. The cells were harvested and resuspended for lysis in 50 mM PIPES buffer pH 6.8 (buffer A) including DNAse and protease inhibitor PMSF, lysed by French press and

were centrifuged at 15K rpm for 45 minutes. The supernatant was loaded on a blue-sepharose column which was pre-equilibrated with buffer A. The fractions containing the InhA protein were eluted from the column by using a gradient of buffer B (50 mM PIPES pH 6.8 and 2 M NaCl). These fractions were pooled and then loaded onto an hydrophobic octyl-sepharose column pre-equilibrated with buffer B. After consecutive wash steps with buffer A, the majority of the impurities were removed and the InhA protein was eluted from the column with water. The fractions containing InhA were pooled and supplied with 50 mM PIPES pH 6.8 buffer immediately. As the final purification step, the protein was concentrated and run through a Superdex S200 gel filtration column. The purity of the protein was checked by SDS-PAGE and activity was verified by enzymatic assay. The protein was concentrated to 10 mg/ml, stored in -80ºC and used for further biochemical and crystallographic studies.

InhA biochemical assays

InhA catalyzes the last step in the elongation cycle of the FAS-II pathway and reduces the 2,3 double bond of trans-2-enoyl-ACP in a NADH dependent manner (Quemard et al., 1995). High Throughput Screening (HTS) and Led Optimization (LO) biochemical assays were based in the oxidation of the cofactor in the presence of dodecenoyl-CoA (DDCoA).

InhA inhibition by file compounds in a high throughput format was assessed using a substrate-induced quenching (SIQ) assay as described previously (Vazquez et al., 2006). Reaction mixtures (3µl) containing 100 µM test compound, 150 µM DDCoA,

100μM NAD$^+$, 30μM NADH, 10 nM resorufin, 0.2% w/v pluronic acid F-127, 0.01% BSA, 1% DMSO and 1.25nM InhA in 30mM PIPES buffer at pH 6.8 were incubated in glass-bottom 1536-well plates for 1 hour at 20°C and 95% relative humidity. Plates were then read in a confocal microscope (4-Channel Reader, Evotec Technologies) using 1D-FIDA settings with beamscanner, 532nm excitation with a laser power of 150μW for 200ms/well. Confocal FLINT data from each screening plate were normalized against control wells containing reaction mixtures in the absence of test compounds (high controls) and in the absence of enzyme (low controls). Assay quality was monitored for each plate using the Z' factor (Zhang et al., 1999b) and the inhibition observed in additional control wells containing a known inhibitor at its IC$_{50}$. In order to identify and discard signal quenchers, each plate was pre-read before addition of the enzyme.

Enzymatic activity to support LO phase was measured fluorimetrically by following the NADH oxidation at λ_{exc}=340 nm and λ_{em}= 480 nm, using 50 μM NADH and 50 μM DDCoA as substrates. Dose-response experiments to determine IC$_{50}$ were performed using 5 nM InhA, percentage of remaining enzymatic activity (%AR) at different compound concentrations were calculated with the next formula [%AR = 100*((sample – control 2)/(control 1- control 2))] where sample is the enzymatic activity for each compound concentration, control 1 is the enzyme activity in absence of any compound and control 2 is NADH oxidation in absence of the enzyme. IC$_{50}$ were calculated fitting %AR to a 2 parameter equation [%AR = 100% /(1+ (compound conc/ IC$_{50}$)^s)], where s is a slope factor, using GraFit 5.0.12 software (Eritacus Software Ltd). All reactions were run in 30 mM PIPES buffer, pH 6.8, at 25 ºC.

Crystallization

Crystallization of the native protein and the protein-inhibitor complexes was done by using both hanging drop and sitting drop methods. For all of the crystallization trials, NAD^+ was used as the cofactor at 10 fold molar excess compared to the protein concentration. Protein-inhibitor complex crystals were obtained by co-crystallization, except for the compounds GSK133 and GSK915. The complex structures with GSK133 and GSK915 were obtained by soaking of the native protein crystals crystallized in 20% (v/v) Peg400, 5% (v/v) ethylene glycol, 0.2 M HEPES pH 7.4 in the presence of NAD^+. The crystals were soaked with 2-4 mM of inhibitor compound for 30 min-4 h. All of the inhibitor compounds were dissolved in 100% DMSO and 10-100 mM stock solutions were used in the experiments. For co-crystallization trials, briefly, 0.2 mM InhA, 2 mM NAD^+ and 1-2 mM inhibitor compound was incubated at room temperature for ~20 min and screened against the sparse matrix crystallization conditions. The crystallization conditions for the different InhA:GSK inhibitor co-crystals were as follows: GSK876 and GSK735 in 10-20% (v/v) 2-Methyl-2,4-pentanediol, 0.1 M HEPES pH 7.5, 4% (v/v) DMSO and 50 mM NaCitrate pH 6.5; GSK218 in 0.2 M NaAcetate trihydrate, 0.1 M Tris HCl pH 8.5, 30% (w/v) Peg4000; GSK713 in 0.1 M N-(2-acetamido) iminodiaceticacid pH 6.8, 6% (v/v) DMSO, 14-24% (w/v) Peg3350, 0.18 M NH4(OAc); GSK490 in 1.6 M tri-sodium citrate dehydrate pH 6.5; GSK625 in 1.1 M Sodium malonate, 0.1 M HEPES pH 7.0, 0.5% (v/v) Jeffamine ED-2001 pH 7.0.

Data collection and processing

The diffraction data were collected at 120 K under cryoproction by a Rigaku Raxis detector coupled to an X-ray generator with a rotating copper anode (λ= 1.541 Å) and at the beamlines 23ID and 19ID of Advanced Photon Source, Argonne National laboratory. Data was processed by HKL2000 (Otwinowski and Minor, 1997). InhA-GSK inhibitor complex crystal structures were solved by molecular replacement (MR) using the PDB entry 3OEW as a search model. The models were built by XtalView and Coot and were further refined by using Refmac in CCP4 package (Murshudov et al., 1997; Winn et al., 2011). Data processing statistics are given in Table 4-2.

Cloning and overexpression of InhA in mycobacteria

Double digestion with *BglII* and *BamHI* liberates *M. smegmatis* and *Mtb inhA* genes from plasmids pATB15 and pATB14, respectively. The 918bp and 917bp fragments were cloned into pATB45 that was previously linearized by *BamHI* digestion. Ligations were performed into *E. coli* DH5α cells. The *inhA* gene in the final plasmids was verified by sequencing. Electroporation of *M. smegmatis* mc^{2}155 and *M. bovis* BCG were carried out according to the method of Snapper et al. (Snapper et al., 1990) and the transformants were selected on Middlebrook 7H10 plates supplemented with 10% OADC and 50 µg/ml hygromycin. Verification of the transformants was made by colony PCR using the PuRe Taq Ready-To-Go PCR Beads kit (GE Healthcare) with the primers P1 (5'-AATCCAAAGTTCAAACGAGGGG-3') and P2 (5'-CCACCACCCGATAAGA GAAAGG-3').

Protein expression levels were followed by SDS-PAGE. *M. smegmatis* mc^{2}155 and *M. bovis* BCG cultures were grown in Middlebrook 7H9 broth with 0.0025% Tween 80 and 10% ADC to OD_{600nm}=0.6. The cells were harvested by centrifugation (5000 rpm for 5 min) and resuspended in 5 ml sterile distilled water. 0.5 ml of *M. smegmatis* mc^{2}155 or *M. bovis* BCG cells were used to inoculate 250 ml conical flasks containing 100 ml minimal medium with 1% acetamide, 0.1% succinate, or 0.1% succinate plus 1% acetamide (for *M.Smegmatis*) or with 0.1% acetate, 0.1% acetate plus 1% acetamide, 10% ADC or 10 % ADC plus 1% acetamide (for *M.bovis*). Cultures were incubated at 37ºC and harvested after 24 hours (*M. smegmatis*) or 7 days (*M. bovis*).

Cell-free extracts were prepared from both cultures using the Mini-BeadBeater (Biospec Products). 50 ml of bacterial cells were resuspended in 1 ml buffer of 50 mM HEPES/KOH pH 7.5, 10 mM $MgCl_2$, 60 mM NH_4Cl, 10% (v/v) glycerol, 5 mM 2-mercaptoethanol and mixed with 0.5 ml of 0.1mm sterile glass beads and shaken for 1 min three times, while incubating the samples on ice for 1 min between pulses. The supernatant was recovered after centrifugation. Cell-free extracts were analysed by SDS-PAGE and protein concentration was determined using the BCA protein assay (Pierce).

Isolation and characterization of* Mtb *H37Rv GSK625A resistant mutants

GSK625A-resistant *M. tuberculosis* H37Rv mutants were isolated by plating 10^8 CFU on Middlebrook 7H10 plates supplemented with 10% OADC containing the compound at 20µM (20X MIC). After 4 weeks of incubation at 37ºC, colonies were identified and selected through several passages in inhibitor-containing plates.

Single colonies were used to amplify and sequence the *inhA* alleles. DNAs of the 17 resistant mutants were extracted from liquid cultures grown in Middlebrook 7H9 broth supplemented with 0.0025% Tween 80 and 10% ADC. 1 ml of culture was centrifuged; the pellet was resuspended in 500 µl of distilled water and incubated at 90ºC for 1 h, and then filtrated by using 0.22µm filters. 5 µl of the supernatant was used as a source of genomic DNA for amplification of the *inhA* gene. PCRs of the entire gene were performed for each mutant by using the PuRe Taq Ready-To-Go PCR Beads kit. A 988 bp region was amplified with the primers Forward (5'-CAGCTTCCTGGCTTCC GAG-3') and Reverse (5'-TAACGTTCTCCAGGAACGG-3'). PCR products were purified and sequenced using the dRhodamine Terminator Cycle Sequencing Ready Reaction kit in an ABI Prism 310 automated DNA sequencer (Applied Biosystems).

Complementation experiments

The whole *fabG1-inhA-hemH* operon from wild type H37Rv and mutants carrying the four mutant alleles (M103V, G96V, M103I, M103T) were cloned in the shuttle vector pSUM36 (Ainsa et al., 1996). *M. tuberculosis* H37Rv was transformed with the five plasmids and transformants were selected using kanamycin, and verified by colony PCR using the PuRe Taq Ready-To-Go PCR Beads kit. The sequences of the primers were as follows: pSUM36_2, 5'-GTTGTGTGGAATTGTGAGCGG-3'; Inharev1, 5'-GAAACGCGATCGACGAGTCGG-3'. Positive transformants were used for minimum inhibitory concentration (MIC) determination experiments.

MIC determination against mycobacteria and clinical strains

The measurement of the MIC against *Mycobacterium* strains for each tested compound was performed in 96-well flat-bottom, polystyrene microtiter plates in a final volume of 200 μl. Ten two-fold drug dilutions in DMSO starting from 50 mM were performed. Drug solutions were added to Middlebrook 7H9 medium and Isoniazid (INH) (Sigma Aldrich) was used as a positive control with two-fold dilutions of INH starting at 160 μg/ml. The inoculum was standardized to approximately $1x10^7$ cfu/ml and diluted 100 fold in Middlebrook 7H9 broth. This inoculum (100 μl) was added to the entire plate except the G-12 and H-12 wells that were used as blank controls. All plates were placed in a sealed box to prevent drying of the peripheral wells and incubated at 37ºC without shaking for six days. Resazurin solution was prepared by dissolving one tablet of resazurin (VWR) in 30 ml of sterile phosphate buffered saline. Of this solution, 25 μl were added to each well. Fluorescence was measured (Spectramax M5 Molecular Devices, Exc. 530nm, Emis. 590 nm) after 48 hours to determine the MIC value.

The BACTEC MGIT 960 System (Becton Dickinson) was used for MIC determination in clinical isolates (Institute Carlos III and Hospital Val d'Hebron) following the manufacturer instructions.

Intracellular activity determination

Bacteria were grown in Middlebrook 7H9 broth supplemented with 10% ADC, 0.4% Glycerol and 0.05% Tween 80 until the mid-log phase. The bacteria were then harvested, washed with RPMI (Roswell Park Memorial Institute) medium with 10%

FCS (fetal calf serum) and resuspended in the same buffer. The suspension was dispersed by sonication, using a probe sonicator (Ultrawave), at power level moderate and pulse level 40 KHz; 2 rounds of 8 pulses (1st round 1min/pulse, 2nd round 30 secs/pulse). Between the pulses samples were placed in ice for 5 secs to avoid heat damage. Two centrifugation cycles at 1500 rpm 5 min recovering the supernatant were performed to collect the final bacterial suspension. Bacteria were quantified by measuring the absorbance at 600nm, where 0.1 OD corresponded to 3×10^8 bacteria.

PMA-differentiated THP-1 cells (PMA 20 ng/mL for 48h) were infected with *Mtb* H37Rv at a multiplicity of infection of one. Infection was performed in antibiotic free RPMI supplemented with 10% FCS. After adding bacteria, culture plates were incubated for 4 hours at 37°C with 5% CO_2. Non-engulfed bacteria were removed by washing 5 times with warm RPMI.

Master plates of compounds were prepared and added to each well (final DMSO concentration lower than 0.5%). After 5 days of incubation the cells were lysed in 100 µl of 0.06% SDS for 10 min at room temperature and dilutions were prepared. Lysate dilutions of 1:10 and 1:100 were each plated separately in duplicate sets on 7H10 agar plates supplemented with OADC. The plates were then allowed to dry and subsequently incubated in a humidified incubator at 37°C. Colonies were first counted on 18th day, and then again on the 21st day of the incubation.

General antimicrobial activity assay

Whole cell antimicrobial activity was determined by broth microdilution using the Clinical and Laboratory Standards Institute recommended procedure, Document M7-A7, "Methods for Dilution Susceptibility Tests for Bacteria that Grow Aerobically". Compounds have been evaluated against a panel of Gram-positive and Gram-negative organisms, including *Enterococcus faecium*, *Enterococcus faecalis*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Escherichia coli* and *Streptococcus pyogenes*. The MIC was determined as the lowest concentration of compound producing a >80 % reduction in fluorescence observed.

HepG2 cytotoxicity assay

HepG2 cells (HB-8065) were fed fresh medium (Essential Minimum Eagle Medium, EMEM, supplemented with 5% FCS and 2 mM L-glutamine) on the day before subculturing. On the day of plate seeding, 100 μL of 100,000 cells/mL was added to every well of a collagen coated, black clear bottom, 96-well microplate (Becton Dickinson) except in column 11, where 100μL of culture medium was dispensed. The plates were incubated for 24 h before the addition of 10 doses of 1:2 dilutions of test substances to achieve a final concentration of 0.5% DMSO. Plates were incubated for 48 hours at 37ºC under 5% CO_2 and 95% relative humidity. Then the culture medium was removed and 200μL of fresh culture medium with 50 μL of Resazurin solution were added. After 90 minutes, plates were removed from the incubator and were left at room temperature protected from light for 15 minutes to allow the fluorescence to stabilize.

174

Resazurine (BDH) was used to read out the viability of the cells. The fluorescence was measured at an excitation wavelength of 515 nm and an emission wavelength of 590 nm in a Microplate reader1420 Multilabel HTS counter Victor 2 (Wallac).

The fluorescence value of each well was corrected by subtracting the background value (average of column 11) from the absolute value. The percent of inhibition was calculated relatively to the DMSO control wells (average of column 12). For each compound, the average value of the duplicate samples was calculated and the curve was fitted to Sigmoidal dose-response (variable slope) nonlinear regression curve adjustment (GraphPadPrism 5 (GraphPad Software, Inc).) in order to calculate the IC_{50} (Tox_{50}).

Microsomal fraction stability assay

Pooled mouse, rat, dog and human liver microsomes were purchased from Xenotech. The microsomes (final protein concentration 0.5 mg/ml, 5 mM $MgCl_2$) and the test compound (final substrate concentration 0.5 µM; final DMSO concentration 0.5 %) in 0.1 M phosphate buffer pH 7.4 were pre-incubated at 37°C prior to the addition of NADPH (final concentration 1 mM) to initiate the reaction. The final incubation volume was 600 µl. Control samples were included for each tested compound, where 0.1 M phosphate buffer pH 7.4 was added instead of NADPH (minus NADPH). Midazolan was included as a control in every experiment. Each compound was incubated for 30 minutes and samples (90 µl) were aliquoted at 0, 5, 10, 20 and 30 minutes. The minus NADPH control was sampled at 0 and 30 minutes only. The reactions were stopped by the addition of 200 µl of acetonitrile:methanol (3:1) containing an internal standard,

175

followed by the centrifugation at 3700 rpm for 15 minutes at 4°C to precipitate the protein. Quantitative analysis was performed using specific LC-MS/MS conditions. The ln peak area ratio (compound peak area/internal standard peak area) was plotted against time and the gradient of the line was determined. Subsequently, half-life and intrinsic clearance were calculated using the equations below:

Elimination rate constant (k) = (-gradient)

Half life ($t_{1/2}$) (min) = 0.693/k

Intrinsic Clearance (CLint) (ml/min/g protein) = (V x 0.693) / $t_{1/2}$ where V is the incubation volume ml/g microsomal protein.

Killing kinetics and **in vivo** *assays*

Bacteria were grown at 37°C in 7H9 broth ADC Tyloxapol to mid-exponenetial phase and then diluted in 10ml fresh Middlebrook 7H9 to an $5x10^5$ cfus/ml. Incubation was continued after the addition of compounds at 20X the MIC. At specified time points, aliquots of cultures were withdrawn, serially diluted in 7H9 broth Tyloxapol and plated on solid culture medium. Plates were then incubated at 37°C and CFU were counted after 3 to 4 weeks.

Specific Pathogen-free 6-8-week-old female C57BL/6j mice (18 – 20 g) were obtained from Harlan (Harlan Interfauna Iberica, Spain). The experiments were performed at AAALAC-accredited GlaxoSmithKline Laboratory Animal Science animal facilities in Tres Cantos (Madrid, Spain). The mice were kept in air-conditioned facilities with fifteen air changes per hour. Room temperature and relative humidity were 22 ± 3

ºC and 40- 70%, respectively. The mice were accommodated in groups of up to five individuals in Tecniplast® type IV cages with autoclaved dust free corncob bedding (Panlab, Barcelona, Spain). The mice were maintained under a twelve hours light/dark period. Autoclaved tap water and γ-irradiated pelleted diet were provided ad libitum.

All the experiments were approved by the DDW Ethical Committee. The animal research complied with Spanish and European Union legislation on Animal Research and GlaxoSmithKline policy on the Care and Use of Animals.

Pharmacokinetic studies

Experimental compounds were administered by oral gavage at 50 mg/kg single dose at a volume of 20 ml/kg to n=5 mice. All mice received treatment in the fed state. 25 µl of peripheral total blood was taken from the lateral tail vein of each mouse at 15, 30 and 45 minutes, 1, 2, 4, 8 and 24 hours for the establishment of compound concentrations. LC-MS was used as the analytical method of choice to determine the concentration of the compound in blood. The non-compartmental data analysis (NCA) was performed with WinNonlin Phoenix 6.3 (Pharsight, Certara L.P) and supplementary analysis was performed with GraphPad Prism 5.

In vivo *acute efficacy assessment*

The mice were intratracheally infected with 100,000 CFU/mouse (*Mtb* H37Rv). The products were administered for eight consecutive days starting one day after infection. Lungs were harvested 24 hours after the last administration. For the chronic

assay, mice were infected with 100,000 CFU/mouse and the products administered daily

seven days a week for eight consecutive weeks starting six weeks after the infection. All

lung lobes were aseptically removed 24 hours after the last administration, homogenized

and frozen. Homogenates were plated in 10% OADC-7H11 medium for 14 days at 37ºC.

The homogenates from compound treated mice were incubated for 18 days at 37ºC in

plates supplemented with 0.4% (w,v) activated charcoal (Sigma Aldrich) to prevent the

effect of product carryover.
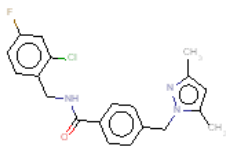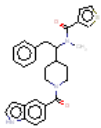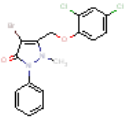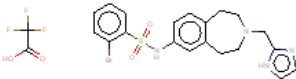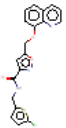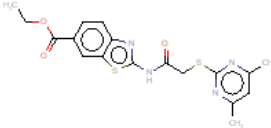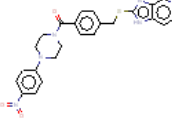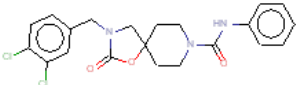
Additional results

HTS screening and lead selection

Several screening campaigns were run by GSK against InhA, which covered a

total of 1.9 million compounds. The initial average hit rate of 3.7% was reduced to

0.63% of true hits after the removal of false positives. Dose-response studies were

performed on the true hits using compound concentrations ranging from 1.7 nM to 100

μM, and several chemotypes showing IC_{50} less than 10 μM were found **(Table 4-9)**. All

of the identified inhibitors were progressed to the enzymatic mode of action studies

where non-specific enzyme inhibitors were removed. The new filtered InhA inhibitors

were selective against InhA. Steady state competition experiments showed that most of

the compounds competed with the fatty acid substrate for binding to the enzyme, but

they were uncompetive with NADH, which is consistent with inhibitors binding to a pre-

formed enzyme:NADH complex.

The antitubercular activity of the newly identified InhA inhibitors was tested against *M. tuberculosis* H37Rv in liquid medium **(Table 4-9)**. A robust correlation was observed between the enzymatic and whole cell activities for compounds from the same series. Thiadiazoles or arylbenzamides were found to be nanomolar inhibitors of InhA with MICs against H37Rv in the micromolar range, while for the other series, like the triazines, IC_{50} and MIC values were very similar. This contrast between the chemical series may reflect differences in cell penetration, intracellular accumulation or enzymatic mode of action, but it could also stem from off-target activities. When the physico-chemical profiles of the selected InhA inhibitors were evaluated and compared, the thiadiazole series emerged as the family with the best balanced scaffold **(Table 4-9)**. A complete profile of GSK826613 and GSK826625 thiadiazoles was built to ascertain their value as antitubercular leads. As described in the results and discussion section of this chapter, further development of these series yielded the identification of GSK2505693A as the potent lead compound that was effective against M(X)DR *Mtb* strains.

Table 4-9. InhA inhibitor hits identified from the HTS screen. Hits belonged to different chemical series and displayed good inhibitory activity ($IC_{50} < 10$ μM) against InhA.

SERIES NAME	HIT STRUCTURE	SMILES	InhA IC50 (nM)	H37Rv MIC (uM)
Arylbenzamides	SB713520		70	16
Piperidinyl benzamides	GSK2067714A		40	>62
Phenyl-pyrazolone	GR44249X		400	>125
Benzazepin sulfonamide	SB735701M		40	31
ISOXAZOLE CARBOXAMIDE	GSK2069070A		15	>62
BENZOTHIAZOLS	GW487324X		16	>125
PIPERAZINYL BENZAMIDES	GSK1400069A		10	>62
THIADIAZOLES	GSK826625A		2	1
SPIRO PIPERIDENES	GSK703100A		200	16

180

Crystal structures of the InhA:GSK compound complexes

      The co-crystal structures of the InhA:GSK compounds were determined by X-ray crystallography. The tested compounds belonged to different chemical series to understand the differences in the binding mode of these inhibitors. Compounds received from GSK were directly used for crystallographic studies without further chemical modifications.

      In total, 13 compounds were received and eight of these compounds, which had various degrees of inhibitory activities (IC_{50}: 2 nM to 28 μM), were successfully co-crystallized with InhA. The co-crystal structures were obtained for two compounds from the thiadiazole series (compounds 490 and 625; **Figure 4-2**), four compounds from the aryl benzamide series (compounds 218, 876, SB713 and 915; **Figure 4-5**), one compound from the benzazepin sulfonamide series (compound 735; **Figure 4-6** (co-crystal structure with compound 246 was determined by GSK)), and one compound from the benzothiazole series (compound 133; **Figure 4-6**).
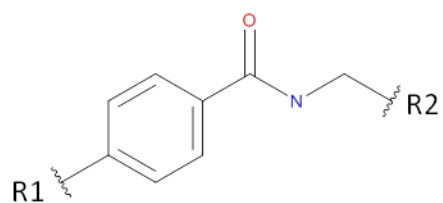
Cmpd#	R1	R2	IC_{50} (nM)
713*			70
876*			60
218*			30
890			50
915*			100
918			90

Figure 4-5. The aryl benzamide series of GSK inihibitors. The scaffold of this series and the corresponding substituents for each compound is shown. The compounds which were successfully co-crystallized with InhA are marked with an asterix (*).

182
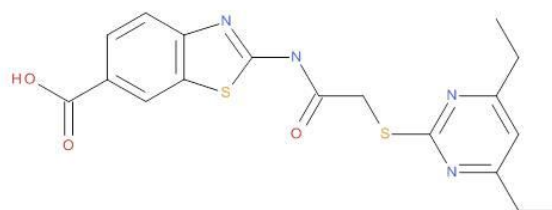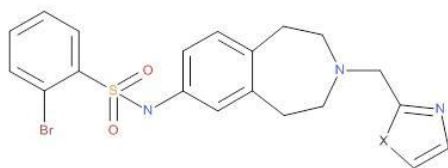
Cmpd #	X	IC$_{50}$ (nM)
735*	N	40
246*	S	180

Cmpd # 133* IC$_{50}$ = 20 nM

Figure 4-6. The benzazepin sulfonamide (735 and 246) and the benzothiazole (133) series of GSK compounds. The compounds which were successfully co-crystallized with InhA are marked with an asterix (*).

The co-crystal structures of InhA:GSK compounds revealed that all of the investigated compounds bound to the hydrophobic substrate binding pocket of InhA, which was surrounded by residues Met199, Leu207, Ile215, Leu218, Met103, Met155, Met161, Phe149, Tyr158, Ala198 and Ile202 **(Figure 4-7)**. As was observed for the thiadiazole series, compounds belonging to the aryl benzamide, the benzazepin sulfonamide or the benzothiazole series also formed the enzyme:inhibitor complex only in the presence of the cofactor NAD^{+}, verifying the uncompetitive character of the inhibitor compounds relative to NADH.

183
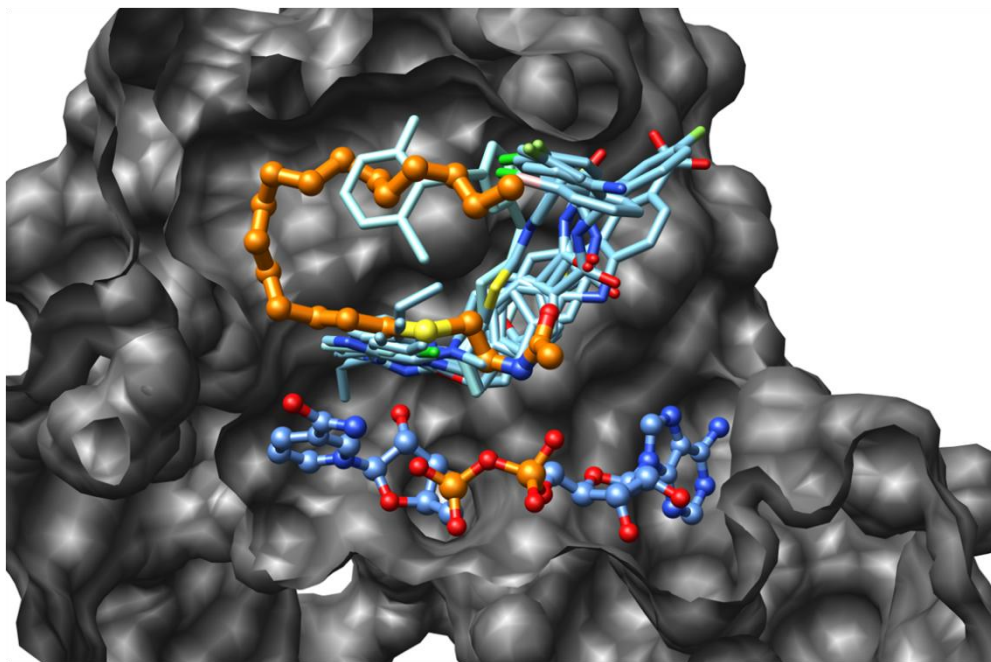
Figure 4-7. Binding mode of the GSK inhibitors to the InhA active site. Eight InhA:GSK co-crystal structures (light blue) are superimposed with NAD^+ (blue) and the C16-fatty acyl substrate analog (orange). A cross-section through the surface of the active site of InhA is shown. As seen from the figure, GSK compounds occupied the substrate binding pocket.

As a common scaffold, all of the investigated GSK inhibitors, including the thiadiazole series, had aromatic pyrazole/thiazole/pyrimidine/imidazole or pyridine rings, which interacted with the nicotinamide ring of NAD^+ via π-π stacking interactions (3.6 - 4 Å). In addition, the nitrogen of these aromatic rings made H-bonds to the 2' hydroxyl of the ribose moiety of NAD^+ (2.6-3 Å), except for the compound GSK218. For compound GSK218 this interaction was provided by the H-bonding of the linker methoxy oxygen atom between the 2-methylpyridyl and the phenyl rings to the 2' hydroxyl of the NAD^+ ribose **(Figure 4-8)**. This direct H-bonding interaction with NAD^+ was important for tight inhibitor binding, but it was not enough for potency. This

can be clearly seen by comparing the co-crystal structures of compounds GSK490 and GSK876. Both compounds interacted with NAD^+ through direct H-bonding with the nitrogen atom in their pyrazole and pyrimidine rings, but they displayed IC_{50} values of 28 μM and 60 nM.
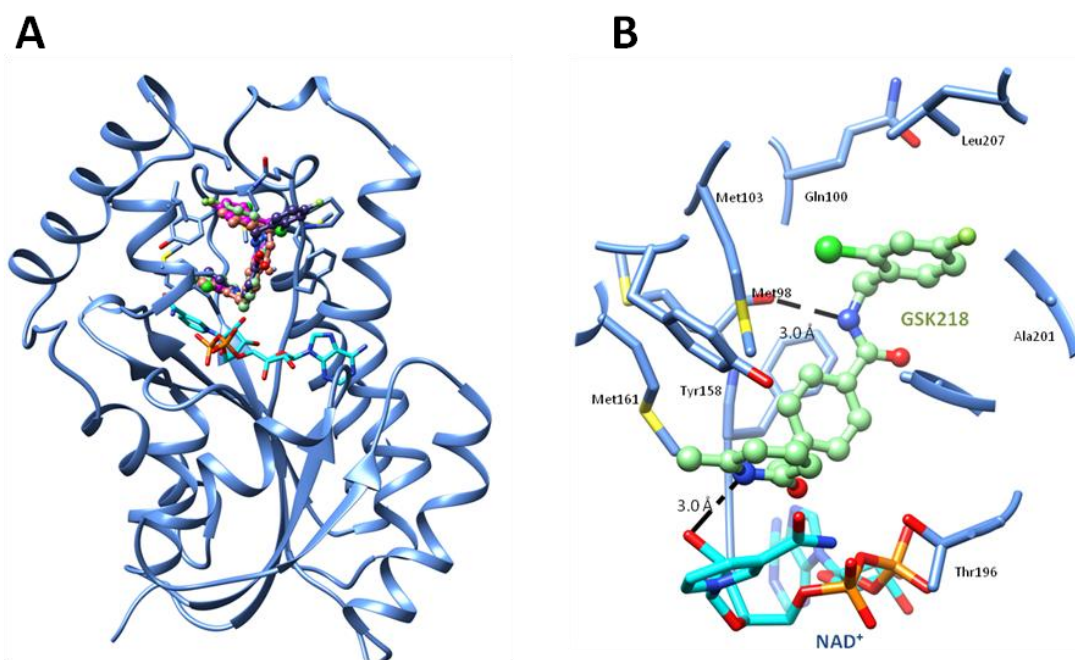
Figure 4-8. Binding mode of the aryl benzamide series to the InhA active site. A) Four InhA:GSK co-crystal structures are superimposed. The GSK compounds are shown in ball-and-stick in green (218), magenta (SB713), purple (876) and orange (915) colors. NAD^+ is shown in cyan sticks. **B)** Key interactions between the aryl benzamide compounds and the InhA active site. As a representative, interaction of the compound 218 (green) with the substrate binding pocket residues is shown. GSK218 interacts with NAD^+ (cyan) and M98 via H-bonds; with M199, M161, P99, Q100, F149, Y158 and F97 through hydrophobic and van der Waals interactions. For both panels atom coloring is: oxygen red; nitrogen blue; sulfur yellow; chlorine and fluoride green.

Based on the InhA:GSK inhibitor co-crystal structures, one of the key interactions identified between the inhibitor compounds and InhA was the direct H-bonding between the inhibitor compound and the carbonyl oxygen and/or the amide NH of the main chain Met98. The range of this interaction varied between 2.7 to 3.3 Å **(Figures 4-8 and 4-9)**. For compounds belonging to the thiadiazole series, these interactions were present between (i) the nitrogen of the thiadiazole ring and the amide NH of Met98, and (ii) the nitrogen linking the pyrazole and the thiadiazole rings and the carbonyl oxygen of Met98 **(Figure 4-2)**. In comparison, for compounds belonging to the aryl benzamide series, this interaction was provided by the amide nitrogen of the inhibitor and the carbonyl oxygen of Met98 (2.7-3.4 Å) **(Figure 4-8)**. For compounds 735 and 133 belonging to the benzazepin sulfonamide and the benzothiazole series, H-bonding was present between the sulfonamide nitrogen and the carbonyl oxygen of Met98 (3.0 Å), and the sulfur atom of the benzothiazole ring and the amide nitrogen of Met98 (3.3 Å) **(Figure 4-9)**.
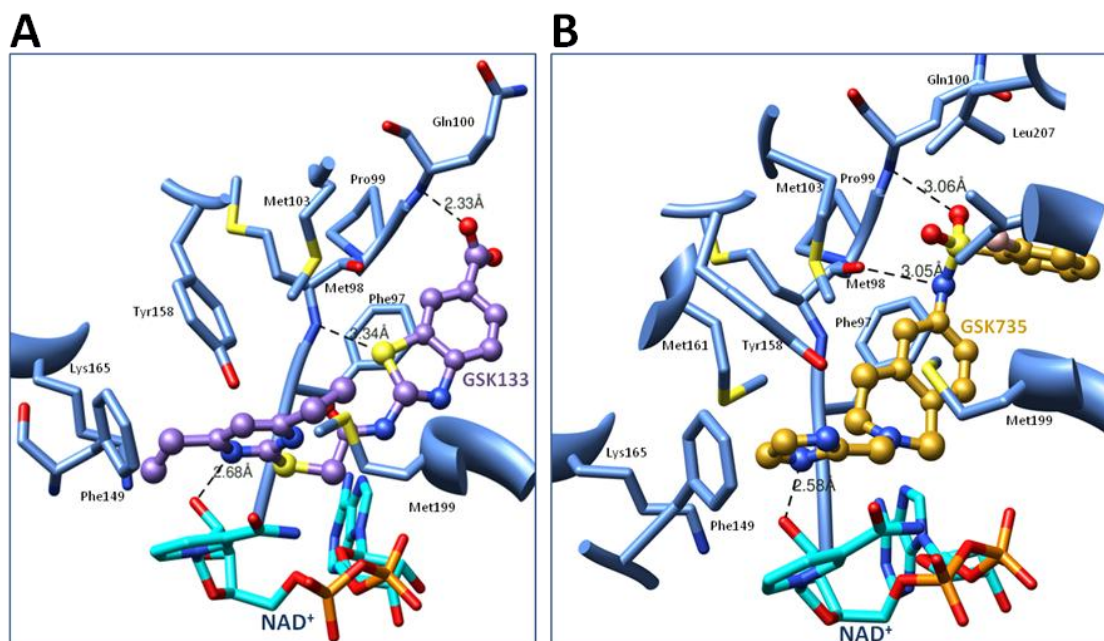
Figure 4-9. Binding mode of the benzothiazole and the benzazepin sulfonamide series to the InhA active site. Details of the InhA:GSK133 (**A**) and the InhA:GSK735 (**B**) interactions. Both GSK133 (purple) and GSK735 (gold) H-bond to NAD^+ (cyan), Met98 and Gln100, and make hydrophobic and van der Waals interactions with the surrounding residues. Compounds also interact with the nicotinamide ring of NAD^+ via π-π stacking.

Hydrophobic and van der Waals interactions between the inhibitors and the side chains of the active site residues Met161, Phe97, Leu207, Met199, Met103, Ala198, Thr196, Gly96, Pro99, Gln100, Phe149, Tyr158 and Ile202 dominated the ligand-protein interactions for all classes of the inhibitors (**Figures 4-8 and 4-9**). This was expected as the inhibitors occupied the hydrophobic substrate binding pocket of the enzyme. Moreover, compounds 133 and 735 were unique in their direct, strong H-bonding interaction with the amide NH of Gln100 through their carboxylate and sulfonyl groups, respectively (2.3 and 3.0 Å). Furthermore, compounds having an additional halide

substituted aryl group toward the end of the substrate binding pocket (compounds 218, 625, 876, 713, 915, and 735) formed additional hydrophobic and van der Waals interactions with the main chain residues (Gln100, Met98, Gly104, Leu207 and Ile202) compared with the compounds 490 and 133. It was observed that the compounds involved in multiple interactions with the protein residues were more potent against the enzyme. This difference can be clearly seen when the compounds 490 and 625 were compared.

As was observed with the thiadiazole series, inhibitors belonging to the other series also did not cause a significant conformational change in the main chain backbone of InhA. In all of the co-crystal structures, the substrate binding loop (residues 197-226) was ordered and present in the crystal structure. Similar to GSK826625, GSK218 and GSK713 also caused the movement of the substrate binding loop by 3.3 Å and 1.6 Å away from the active site compared with the InhA:NAD(H) structure. This movement was caused by the interaction of the substrate binding loop with the halide substituted benzene moiety of the inhibitor compounds. The shift of the substrate binding loop indicates the flexible nature of the substrate binding crevice, which is necessary to accommodate the varying chain length fatty acyl substrates. Excluding the substrate binding loop, the RMSD of the main chain C_αs between the InhA:NAD(H) and InhA:NAD$^+$:GSK compound structures was less than 0.5 Å. The catalytic residue Tyr158 did not make any H-bonds with the inhibitors; it participated in ligand binding only through van der Waals interactions (4.2-5.4 Å). In addition, the side chain of Phe149 adopted the identical position as in the InhA:NAD(H) structure. Unlike the INH-

188

NADH adduct, none of the GSK compounds caused the flipping of the Phe149 side chain, and so they did not interact with the isonicotinic acid binding pocket.

Summary

Drug discovery is a long and tedious process. It takes at least 10-15 years for a single drug to come to the market. There are many factors considered before a compound can be administered as a drug, such as *in vivo* activity, low frequency of drug resistance, low toxicity, oral availability, and affordability. For target-based approaches, the drug discovery process starts with the target selection, hit/lead identification, verification and optimization. However, a highly potent compound active *in vitro* against an enzyme does not guarantee that a similarly good activity will also be obtained *in vivo* against the organism. Even after compounds are optimized for *in vivo* activity, they may not exhibit good pharmacokinetic (PK) properties (bioavailability, solubility, Absorption, Distribution, Metabolization and Excretion (ADME)), which will block further progress into drug development. Therefore, it is very important to run and invest resources in drug discovery projects, especially for diseases such as tuberculosis, which is particularly difficult to eradicate due to the unique biology of its causative agent *Mtb*.

In this study, the collaboration with GSK and TB alliance resulted in the identification of compounds that exhibited potent antitubercular activity. These compounds were activation-free inhibitors of the InhA enzyme. The co-crystal structures of the InhA:GSK compounds revealed the details of the protein-inhibitor interactions and highlighted the substituents or chemical entities required for potency. These features

were considered in designing new analogs of the hits. The GSK inhibitors bound to the hydrophobic substrate channel of InhA, and achieved their specificity and potency against the enzyme by three types of interactions: (1) π-π stacking interaction with NAD^+, (2) direct H-bonding interaction with NAD^+ and the main chain backbone carbonyl oxygen and the amide NH (especially for Met98), (3) hydrophobic and van der Waals interactions with the side chain residues surrounding the substrate binding pocket, by the presence of hydrophobic moieties. It was observed that interaction of the compounds with the cofactor NAD(H) and the substrate binding loop was necessary for high potency.

As a result of this collaboration, we reached the objective of developing novel activation-free InhA inhibitors that are active against the INH resistant *Mtb* strains carrying mutations in the *katG* gene. The lead thiadiazole compounds represent a previously unexplored class of InhA inhibitors. However, the initial compounds of this series did not have optimum PK values and additional chemical optimization was needed. Further chemistry yielded the final compound GSK693 as a new lead representative structure. GSK693 exhibited good antitubercular activity *in vivo* as well as in murine models. It is a promising compound for further drug development studies.

CHAPTER V

STRUCTURE-BASED DRUG DESIGN STUDIES FOR INHA WITH TRICLOSAN

ANALOGS[*]

Overview

Triclosan has been previously shown to inhibit InhA, an essential enoyl acyl carrier protein reductase involved in mycolic acid biosynthesis, the inhibition of which leads to the lysis of *Mycobacterium tuberculosis*. Using a structure-based drug design approach, a series of 5-substituted triclosan derivatives was developed. Two groups of derivatives with alkyl and aryl substituents, respectively, were identified with dramatically enhanced potency against purified InhA. The most efficacious inhibitor displayed an IC_{50} value of 21 nM, which was 50-fold more potent than triclosan. X-ray crystal structures of InhA in complex with four triclosan derivatives revealed the structural basis for the inhibitory activity. Six selected triclosan derivatives were tested against isoniazid-sensitive and resistant strains of *M. tuberculosis*. Among those, the best inhibitor had an MIC value of 4.7 μgmL^{-1} (13 μM), which represents a tenfold improvement over the bacteriocidal activity of triclosan. A subset of these triclosan analogues was more potent than isoniazid against two isoniazid-resistant *M. tuberculosis*

strains, demonstrating the significant potential for structure-based design in the development of next generation antitubercular drugs.

Introduction

For nearly fifty years, isoniazid (INH) has been utilized as a frontline drug to treat tuberculosis (TB) (Vilcheze and Jacobs, 2007). INH is a key component of short course chemotherapy for TB treatment, which entails six months of daily administration, and is also the sole component for TB prophylaxis, which includes nine months of daily administration. The emergence of INH-resistant strains has compromised TB control programs worldwide (CDC, 2006; Crofton et al., 1997).

INH and ethionamide (ETH) are known to target the *Mycobacterium tuberculosis* enoyl acyl carrier protein (ACP) reductase (InhA) (Banerjee et al., 1994; Rozwarski et al., 1998; Vilcheze et al., 2006; Wang et al., 2007), validating it as an excellent antitubercular drug target. INH and ETH are prodrugs, and their activities are dependent on activation in *M. tuberculosis* by KatG, a catalase/peroxidase enzyme, and EthA, a flavin monooxygenase, respectively (DeBarber et al., 2000; Morlock et al., 2003). Upon activation, INH or ETH forms a covalent adduct with NAD cofactor, which inhibits InhA. As prodrugs, INH and ETH are highly specific and effective. However, mutations in activators *katG* and *ethA* have been linked to most of the clinical resistance in the diagnosed cases of drug-resistant TB (Hazbon et al., 2006; Morlock et al., 2003). Compounds that do not require activation and directly target InhA represent a promising approach to circumvent this resistance mechanism. The first report of such small-

molecule InhA inhibitors came from our laboratories in 2003, featuring triclosan and two members of the Genzyme compound library (Kuo et al., 2003). Triclosan, for example, inhibits InhA without the need for prior activation, although its use as an antitubercular may be limited by its less than optimal bioavailability (Wang et al., 2004). These small molecules, however, represent reasonable starting points for structure-based drug discovery efforts to afford effective InhA inhibitors. Aiding such an effort are several crystal structures, such as those of InhA:NADH (Dessen et al., 1995), InhA:NAD$^+$:triclosan (Kuo et al., 2003), InhA: NAD$^+$:Genz-10850 (Kuo et al., 2003), InhA:NAD$^+$:C16-substrate (Rozwarski et al., 1999), and InhA: INH-NAD (Rozwarski et al., 1998) that were accessible at the outset of these investigations. During the course of this work, Sullivan et al. reported the X-ray structures of two triclosan analogues with 5-alkyl (*n*-pentyl and *n*-octyl) moieties, lacking the two B-ring chlorines (Vilcheze et al., 2006). This collection of structures provided a precisely defined active site of InhA and a thorough understanding of the ligand–enzyme interactions that render potent enzyme inhibition.

One promising route for the design of potent InhA inhibitors involves triclosan, a commercially available compound that has been reported to inhibit the enoyl acyl ACP reductases (ENR) from several species, including *Plasmodium falciparum (*Surolia and Surolia, *2001), Escherichia coli (*Heath *et al., 1998), Bacillus subtilis (*Heath *et al., 2000), Brassica napus, (*Roujeinikova *et al., 1999) and Pseudomonas aeruginosa (*Hoang and Schweizer*, 1999).* The structures of ENR from *E. coli* (FabI) (Levy et al., 1999), *B. napus* (Roujeinikova et al., 1999), *P. falciparum* (PfENR) (Perozzo et al.,

2002), and *M. tuberculosis* (InhA) (Kuo et al., 2003) bound with triclosan have been characterized. In our laboratories, a series of 5-substituted triclosan derivatives was designed and synthesized in order to optimize the potency of triclosan in parallel against both purified PfENR (Freundlich et al., 2007) and InhA. This report discusses these efforts focused specifically on the SAR developed versus purified InhA and INH-sensitive and resistant *M. tuberculosis*.

Results and discussion

Our efforts began with a survey of the complex structures of InhA bound with INH–NAD (Rozwarski et al., 1998) and triclosan (Kuo et al., 2003), respectively. Two hydrophobic cavities capable of being filled were identified: the substrate-binding site and the pocket into which the isonicotinoyl group of INH–NAD protrudes. To date, all reported inhibitors of InhA occupied the hydrophobic cavity of the substrate-binding site, except the INH–NAD and ETH–NAD adducts. The isonicotinoyl moiety of the INH–NAD and ETH–NAD adducts were found in a hydrophobic pocket formed by movement of the side chain of Phe 149. The pocket was underneath the fatty acyl substrate-binding site **(Figure 5-1)** and, lined predominantly by hydrophobic groups from the side chains of Tyr 158, Phe 149, Met 199, Trp 222, Leu 218, Met 155, Met161, Gly 192, and Pro 193. This pocket may also serve as a portal to the external solvent to the left side of the active site **(Figure 5-1)**. The active forms of both INH and ETH occupied the same pocket and were extremely potent against InhA ($K_i =5$ nM and 7 nM, respectively) (Vilcheze et al., 2006; Wang et al., 2007), validating this cavity as a

suitable site to target with new inhibitors. Superimposition of the structures of InhA:INH–NAD and InhA:NAD$^+$:triclosan indicated that the chlorine atom at the 5-position of the triclosan A-ring was about 2 Å away from the binding pocket of isonicotinoyl moiety of the INH-NAD adduct and was in contact with Pro 193, Met 199, and Phe 149 through van der Waals interactions. Based on this structural information, we hypothesized that it may be possible to replace the 5-chloro with various moieties to occupy this isonicotinoyl binding pocket and, thus, increase the *in vitro* activity against InhA.

It should be noted that this strategy contrasts with that of Sullivan et al. to extend relatively long *n*-alkyl chains off what is essentially the triclosan 5-position (in molecules where the two B-ring chlorines have been excised) to mimic substrate analogue *trans*-2-hexadece-noyl-(*N*-acetylcysteamine)thioester, whose structure with InhA we reported in 1999 (Rozwarski et al., 1999).
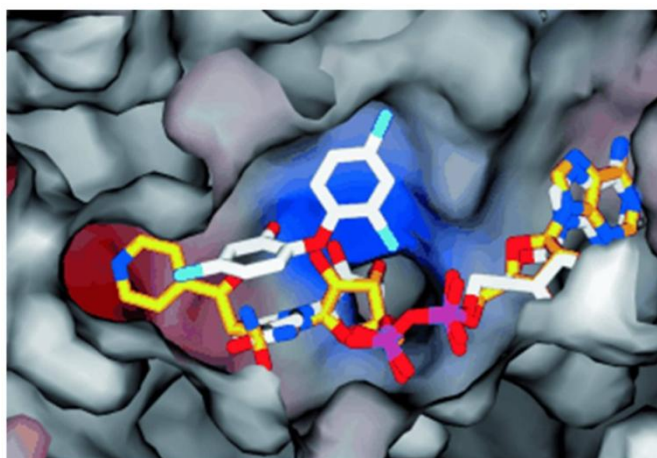
Figure 5-1. Cross-section through the surface of the InhA active site of super-imposed structures of InhA in complex with the INH–NAD adduct and triclosan. The carbon atoms of the INH–NAD adduct and triclosan are colored in gold and white, respectively. Other atoms are colored according to the atom type: oxygen, red; nitrogen, blue; chlorine, cyan; phosphorus, purple. This figure was produced using Spock (version V1.ob170) (Christopher, 1998).

A series of triclosan derivatives with modifications at the 5-position of triclosan was evaluated for their inhibition of purified InhA. These small molecules were prepared during the course of a concurrent program to investigate 5-substituted triclosan analogues as PfENR inhibitors (Freundlich et al., 2007). Inhibitors with hydrophobic substituents, such as alkyl groups (compounds 2 and 7) were much more potent than those with hydrophilic substituents (compounds 3, 4 and 5) **(Table 5-1)**. This result is consistent with our proposal that the 5-substituent of triclosan projects into a hydrophobic cavity of InhA. It is interesting to note the lack of activity of phenyl derivative 6, which may be explained by a potential steric clash with Phe 149 (as clarified below).

Following the initial modest activity of compounds 2 and 7, a series of 5-alkyl triclosan derivatives was examined for their inhibitory activity against InhA. Their potency against InhA appeared to increase with the chain length of the 5-alkyl groups **(Table 5-2)**. Highest inhibitory potencies were observed for compounds 10 and 12 with four-carbon chains. Sullivan observed the same trend with 5-alkyl substituted, des-chloro triclosan derivatives (Sullivan et al., 2006).

Table 5-1. *In vitro* **activities of select triclosan derivatives against** *M. tuberculosis* **InhA.** [a]

Compound	R	InhA IC$_{50}$ [nM]
triclosan	Cl	1100±180
2	Me	800±99
3	2H-tetrazol-5-yl	>10 000
4	COOH	>10 000
5	C(O)NH$_2$	>10 000
6	Ph	>10 000
7	CH$_2$(C$_6$H$_{11}$)	110±31

[a] Values reported as the mean ±standard error for at least three in-dependent measurements.

Table 5-2. *In vitro* **activities of select triclosan derivatives against** *M. tuberculosis* **InhA.**[a]
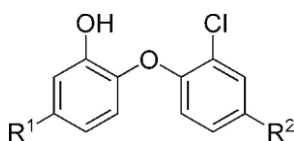
Compound	R^1	R^2	InhA IC$_{50}$ [nM]	
8	CH$_2$CH$_3$	Cl	120	±19
9	(CH$_2$)$_2$CH$_3$	Cl	91	±15
10	(CH$_2$)$_3$CH$_3$	Cl	55	±20
11	CH$_2$CH(CH$_3$)$_2$	Cl	96	±46
12	(CH$_2$)$_2$CH(CH$_3$)$_2$	Cl	63	±9
13	CH$_2$CH(CH$_3$)CH$_2$CH$_3$	Cl	130	±56
14	2-pyridyl	CN	>10 000	
15	3-pyridyl	Cl	>10 000	
16	4-pyridyl	CN	>10 000	
17	CH$_2$(2-pyridyl)	Cl	29	±11
18	CH$_2$(3-pyridyl)	Cl	42	±10
19	CH$_2$(4-pyridyl)	CN	75	±16
20	o-CH$_3$-Ph	Cl	1300	±77
21	o-CH$_3$-Ph	CN	>10 000	±110
22	m-CH$_3$-Ph	Cl	870	
23	p-F-Ph	Cl	>10 000	
24	CH$_2$Ph	Cl	51	±6
25	(CH$_2$)$_2$Ph	Cl	21	±8
26	(CH$_2$)$_3$Ph	Cl	50	±14

[a] Values reported as the mean ±standard error for at least three independent measurements.

It was found that the *in vitro* activity of those inhibitors was optimal at a carbon chain length of eight. Triclosan inhibitors show better *in vitro* activity than their des-chloro counterparts, with 5-substituents of the same carbon chain length, from the report by Sullivan et al. (ethyl, IC_{50} =120 ± 19 nM vs. 2000 ± 700 nM; *n*-butyl, IC_{50} =55 ± 20 nM vs. 80 ± 15 nM), suggesting that the two chlorine atoms on the B-ring contribute to the binding of the inhibitor to the enzyme. Molecular modeling suggests that this may be due to favorable van der Waals interactions of the 4'-Cl with Phe 97 and Met 103. These results imply that the hypothesis of Sullivan et al. (Sullivan et al., 2006) that the B-ring chlorines do not contribute to efficacy may not be correct.

The alkyl substituents investigated in previous studies were all linear in nature. In the present report, compounds 11, 12, and 13, with β- or γ-branched methyl substituents, failed to improve upon the potency of their straight-chain counterparts of the same chain length (compounds 9 and 10). Similarly, cyclohexylmethyl analogue 7 exhibited an IC_{50} value of 110 nM. The 2.8 Å crystal structure of InhA in complex with compound 7 **(Figure 5-2)** was solved to provide a structural basis for the activity difference amongst the alkyl substituents. The crystal belonged to I4$_1$22, a space group for InhA that has not been reported previously. It is worth noting that the substrate binding loop (residues 195–205) was ordered in the structure of InhA bound with compound 7, while it has been disordered in the structures of InhA bound with triclosan, and all other triclosan derivatives solved to date. The structure was readily superimposed on that of InhA in complex with 5-pentyl-2-phenoxy-phenol (Sullivan et al., 2006), which intriguingly features a somewhat strained C4-C5 portion of the pendant alkyl chain, and the C16

substrate analogue (Rozwarski et al., 1999) (overlapping the position of C4 to C7 of the U-shaped acyl chain). The 5-cyclohexylmethyl group formed predominantly hydrophobic interactions with the side chains of Phe 149, Ile 215, Leu 218, Met 155, Tyr 158, and Met 199. Based on this structure, we hypothesized that 5-substituents with a longer alkyl chain would create more extensive hydrophobic interactions in this pocket. Therefore, it was not surprising that the potency of the 5-alkyl triclosan analogues increased with the chain length. However, it is not so obvious from a structural point of view why our limited subset of methyl-branched inhibitors do not exhibit greater potency than the nonbranched inhibitors of the same chain length. A possible explanation could be steric clashes with one or more of the side chains of Phe 149, Tyr 158, and Met 199 with the branched methyl groups. Compared to the structure of InhA bound with the 5-pentylphenol derivative designed by Sullivan et al., the side chain of Met 199 on the substrate-binding loop flipped ~ 100° to form a hydrophobic interaction with the cyclohexyl group at a distance of 3.6 Å. In addition, the B-ring of compound 7 also rotated ~ 30° from its position in 5-pentyl-2-phenoxyphenol to allow the 4'-chloride to form a weak hydrogen-bonding interaction with the amide NH of Met 98 at a distance of 3.2 Å.
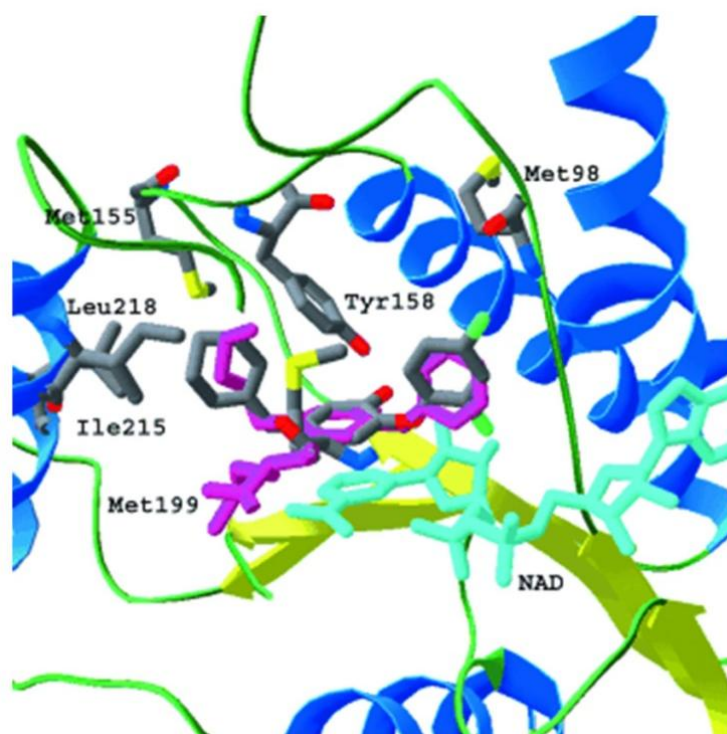
Figure 5-2. The superposition of crystal structures of InhA (ribbons and tubes, key residues in stick format) in complex with 5-pentyl-2-phenoxyphenol (Sullivan et al., 2006) (purple) and compound 7 (carbon atoms in gray) in the presence of NAD$^+$ (cyan). The B-ring of the 5-pentyl triclosan analogue was in a different orientation from 7, because of the lack of two chlorides. The cyclohexylmethyl group of compound 7 was in a similar position to the pentyl group of 5-pentyl triclosan analogue. This figure, in addition to Figure 3 and 4, was made using SwissPDB viewer (version 3.7) (Guex and Peitsch, 1997).

Analogues of phenyl 6 were examined to improve upon the surprisingly poor enzyme inhibition observed (**Table 5-2**). Pyridyl derivatives 14–16 and simple phenyl analogues 20–23 were also poor InhA inhibitors. Although the crystal structures for InhA bound with 5-phenyl or 5-pyridyl triclosan analogues have not been obtained to date, the inhibitors were modeled (McRee, 1999) into the active site based on the

structure of InhA:triclosan. A directly attached aryl group at the 5-position appeared to be too close to the side chain of Phe 149 (shortest distance was ~ 1.4 Å between the aryl substituent and the phenyl group of Phe 149), leading to steric clashes. The hydrophobic pocket of interest was separated from the 5-position by a distance of 2.5 Å.

Therefore, to reorient and better position the hydrophobic aryl group into the target pocket, a linker varying from 1–3 carbons was incorporated between the aryl substituent and the 5-position carbon on the A-ring **(Table 5-2)**. Consistent with the molecular modeling results, an evaluation of 5-position substituents of the type $(CH_2)_n Ar$ demonstrated that carbon linkers significantly increased the inhibitory potency versus purified enzyme. Among them, compound 25 had the highest potency (IC_{50} =21 nM), representing a 50-fold increase compared to triclosan. It is the most potent triclosan derivative against purified InhA we have studied to date. Compounds 24 and 26, with similar activities against InhA (24, IC_{50} =51 nM; 26, 50 nM), were also active inhibitors. The crystal structures of InhA bound with 24 and 25 showed that the $(CH_2)_n Ar$ group extended into the pocket and formed hydrophobic interactions with residues Leu218, Ile 215, Phe149, Met199, and Pro 193 **(Figure 5-3)**. The triclosan backbone atoms of compounds 24 and 25 were in nearly identical positions, but their 5-substituents were clearly different. The phenyl group of derivative 24 was positioned in the center of the binding pocket, while the phenyl group of derivative 25 protruded ~2 Å deeper into, and was closer to the end of, the binding pocket due to the longer carbon linker. Thus, compound 25 additionally engaged Trp222 in a hydrophobic interaction. The side chains of Leu 218 and Ile215 rotated 30° and shifted 1.5 Å, respectively, to

accommodate the phenyl ring of compound 25, which also flipped ~70° from the position of the phenyl group of compound 24. All of these conformational changes suggest that there were more hydrophobic interactions between the active site residues and the 5-substituent of compound 25 than that of compound 24, which may explain why compound 25 was twice as potent than compound 24 against InhA.

To further increase the potency of the 5-$(CH_2)_n$Ar series through potential mimicry of the INH–NAD isonicotinoyl moiety, derivatives with 5-CH_2(n-pyridyl) (n =2, 3, and 4) substituents (compounds 17–19) were examined for *in vitro* activity. In the crystal structure of InhA:INH–NAD (Rozwarski et al., 1998), the pyridyl group of the INH–NAD adduct formed a hydrogen-bonding interaction with a buried water molecule. Without a carbon linker, the 5-pyridyl triclosan derivatives 14–16 showed very low activity against InhA (IC_{50} >10 μM), presumably due to a steric clash with Phe 149. In contrast, with a one-carbon linker, pyridyl analogues 17–19 were all potent inhibitors (IC_{50} < 80 nM), and their IC_{50} values were in the order 17 <18 <19. Compared to benzyl derivative 24, the activity of 17 increased approximately twofold while the activity of compound 18 increased only slightly, and compound 19 was slightly less potent. The structure of InhA:NAD$^+$:17 demonstrated that the pyridyl ring of derivative 17 extended into the substrate-binding site to form hydrophobic interactions with the side chains of Tyr 158 and Phe 149 **(Figure 5-4)**. These interactions are similar to those of the phenyl group in InhA:NAD$^+$:24. However, the nitrogen atom on the 2-position of the pyridyl ring also formed a hydrogen-bonding interaction with the side chain carboxylate of Glu 219 through a water molecule. Modeling studies indicated that similar interactions could

203
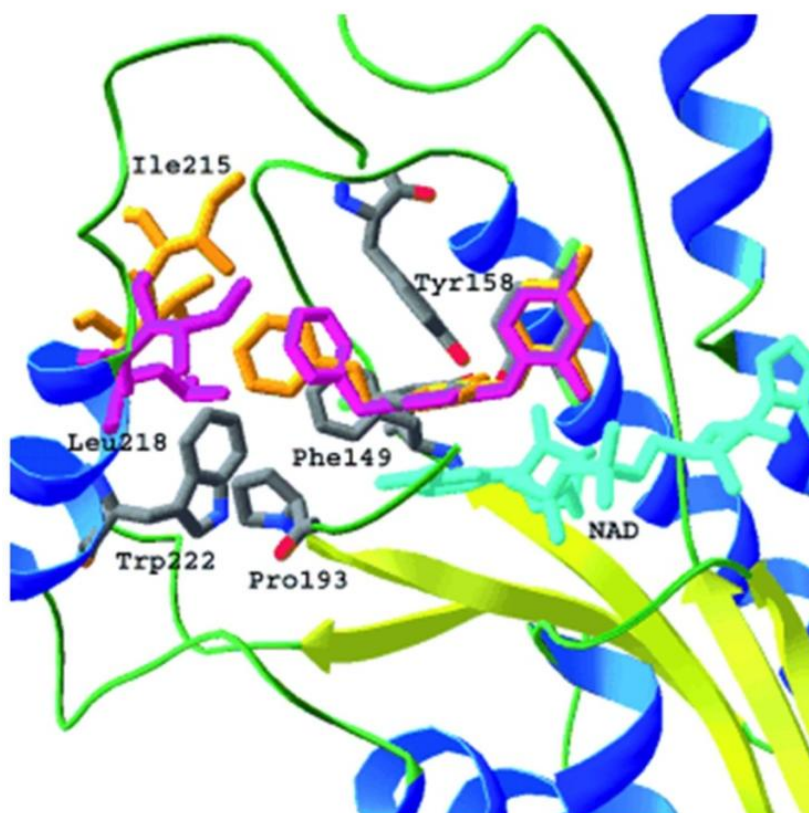
Figure 5-3. The superposition of crystal structures of InhA (ribbons and tubes, key residues in stick format) in complex with triclosan (carbon atoms in gray), compound 24 (purple), and compound 25 (gold) in the presence of NAD$^+$ (cyan). The 5-phenyl groups of compounds 24 and 25 are shown making van der Waals contacts with the side chains of Tyr 158, Pro 193, Leu 218, Ile 215, and Phe 149. The movements of Ile 215 and Leu 218 in response to the triclosan 5-substituent are depicted by showing this residue in stick format with the coloring of the respective ligand. It should be noted that Met 199 is not depicted for the sake of clarity.

also exist for 18, but were not likely for 19, based on the distance between the pyridyl nitrogen atom and the side chain oxygen of Glu 219. The nitrogen atom of 19 potentially pointed to the hydrophobic side chains of residues Met 155 and Leu 218, which may not be favored energetically. The superimposition of InhA:NAD$^+$:17 and

InhA:NAD$^+$:triclosan (Kuo et al., 2003), demonstrated movement of both Glu 219 and Leu 218 to accommodate the 5-pyridylmethyl moiety.

A comparison of the X-ray crystal structures of bound compounds 7, 17, 24, and 25 with that of INH–NAD (PDB code 1ZID) is instructive. At the outset, our goal was to reach from the 5-position of triclosan into the isonicotinoyl-binding pocket. The Phe 149 side chains from the four triclosan analogue structures overlapped well with each other, but not that of 1ZID. In 1ZID, the isonicotinoyl group of INH was stabilized by the rotation of Phe 149. The various triclosan 5-substituents, however, did not occupy the same volume as the isonicotinoyl group, orienting similarly to the C16 chain of the structurally characterized substrate analogue (Rozwarski et al., 1999) and the appended alkyl substituent in the structures reported by Sullivan et al. (Sullivan et al., 2006), and thus did not perturb Phe 149. In the 1ZID structure, the movement of Phe 149 evidently displaced Tyr 158. While this motion was tolerated in 1ZID, in the triclosan structures, the phenol moiety is positioned through an interaction with the hydroxyl of Tyr 158. Loss of the triclosan phenol–Tyr 158 interaction would most likely result in complete loss of inhibitor binding, supported by the observed weak inhibitory activity of des-phenol triclosan analogues against InhA. Thus, while the triclosan analogues can attempt to place substituents proximal to the INH binding pocket, they cannot occupy the exact volume of the isonicotinoyl moiety, as this would ultimately require the movement of Phe 149 and Tyr 158, thus losing a key hydrogen-bonding interaction with Tyr 158.
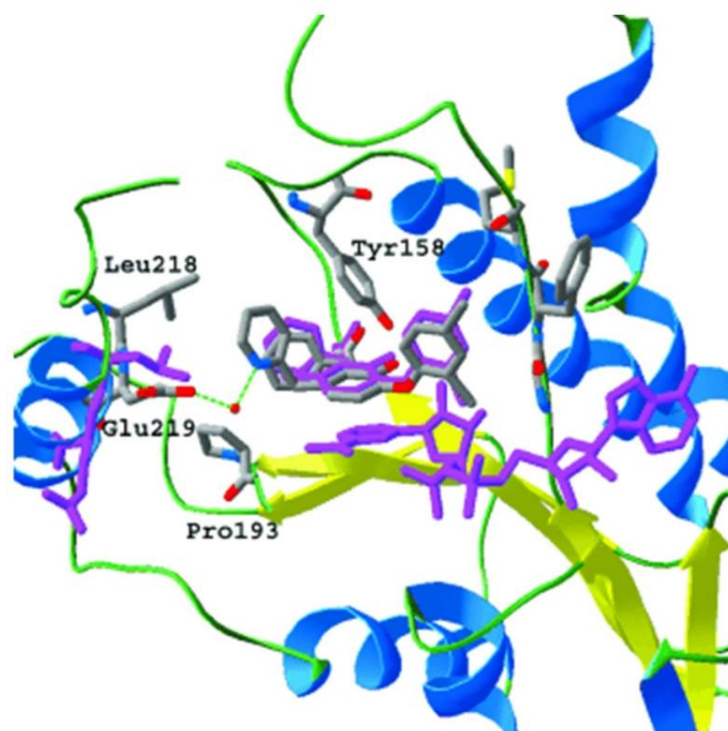
Figure 5-4. The superposition of crystal structures of InhA (ribbons and tubes, key residues in stick format) in complex with triclosan (purple) and compound 17 (carbon atoms in gray) in the presence of NAD⁺ (cyan). The nitrogen atom on the pyridyl ring formed a hydrogen-bonding interaction with the side chain of Glu 219 through a water molecule. The side chain of Glu 219 rotated 90° from its original position (purple) to form this hydrogen-bonding interaction. It should be noted that Phe 149 is not depicted for the sake of clarity.

The activities of three 5-alkyl (7, 10, and 11) and three 5-aryl triclosan derivatives (24, 25, and 26) against six different *M. tuberculosis* strains were assayed (**Table 5-3**). Among these strains, H37Rv is the *M. tuberculosis* wild-type, sensitive to isoniazid. The InhA(S94A) strain, carrying the *inhA(S94A)* allele, has been demonstrated to confer resistance to isoniazid through the weakening of co-factor binding (Vilcheze et al., 2006). The InhA(C-15T) strain is a *M. tuberculosis* spontaneous mutant carrying the

inhA expression region mutation (C-15T) (Vilcheze et al., 2006). It has been shown that the C-15T *inhA* promoter mutation mediates enhanced transcription of *inhA*, resulting in INH and ETH resistance. The Δ*katG* mutant features the complete deletion of the *katG* gene and, thus, INH activation cannot take place. Most significantly, strains 12081 and 5071 are multidrug resistant clinical isolates from Mexico that are INH resistant without known mutations in *katG, inhA, ndh*, and *ahpC*. The 12081 (SCG V) strain was, in our hands, resistant to INH, streptomycin, and rifampicin, while 5071 (SCG 3b) was resistant to only INH and streptomycin.

The selected six compounds with high potency against InhA *in vitro* (IC$_{50}$ ≤ 110 nM) all demonstrated good antitubercular activity against the wild-type H37Rv strain, exhibiting improvements over the whole-cell efficacy of triclosan. Clearly, the 5-chloro group of triclosan is not optimal and the use of larger and more hydrophobic moieties is advantageous both in terms of InhA inhibition and whole-cell potency. The most active compound, 26 (MIC =13 μM), was superior to that of ETH (MIC =18 μM) (Vilcheze et al., 2006), while being greater than tenfold more potent than triclosan. The activities of these compounds against purified InhA and cultured tuberculosis strains generally correlated well. For example, compound 10, which had two-fold higher activity against InhA than compound 11, also exhibited twofold better antitubercular activity. This supports a hypothesis that the straight chain 5-alkyl substituent has an advantage over the analogous branched derivative. In addition, compound 26, which displayed the highest antitubercular activity in the series, also had the second highest potency against InhA. Exceptions may be noted such as with compound 25, which showed the highest

Table 5-3. Minimum Inhibitory Concentration (MIC) values of triclosan derivatives against wild-type and mutant *M. tuberculosis* strains.

Compound	H37Rv [μgmL^{-1}] ([μM])	H37Rv *inhA* (S94 A) [μgmL^{-1}] ([μM])	H37Rv PmabA inhA (C-15T) [μgmL^{-1}] ([μM])	H37Rv Δ*katG* [μgmL^{-1}] ([μM])	Clinical isolate 12081 [μgmL^{-1}] ([μM])	Clinical isolate 5071 [μgmL^{-1}] ([μM])
7	9.4 (27)	37 (110)	37 (110)	19 (55)	37 (110)	4.7 (13)
10	9.4 (30)	37 (120)	37 (120)	19 (60)	37 (120)	4.7 (15)
11	19 (60)	37 (120)	37 (120)	37 (120)	75 (240)	19 (60)
24	9.4 (27)	75 (220)	37 (110)	19 (55)	75 (220)	9.4 (27)
25	19 (52)	75 (210)	75 (210)	37 (110)	75 (210)	19 (52)
26	4.7 (13)	19 (50)	19 (50)	9.4 (26)	37 (100)	4.7 (13)
INH	0.060 (0.44)	0.50 (3.6)	0.80 (5.8)	>200 (>1400)	1.0 (7.3)	8.0 (56)
triclosan	40 (140)					

potency among the $(CH_2)_n$Ar series and yet the worst antitubercular activity. As may be expected, a number of factors other than *in vitro* efficacy against purified InhA, including pharmacokinetic parameters, contribute to the whole-cell inhibition of mycobacterial growth.

It is interesting to compare the H37Rv data for 5-*n*-butyltri-closan (10) and the corresponding des-chloro analogue of Sullivan and co-workers. While compound 10 (IC$_{50}$ =55 nM) was slightly more active than its des-chloro analogue against InhA (IC$_{50}$

=80 nM) (Sullivan et al., 2006), the des-chloro compound (MIC =10.8 μM) was more efficacious against H37Rv compared with compound 10 (MIC =30 μM). It should be noted that the most potent triclosan analogue reported to date is the *n*-octyl analogue reported by Sullivan et al. (InhA, IC_{50} =5.0 nM; H37Rv, MIC =6.5 μM) (Sullivan et al., 2006).

Triclosan has previously been reported to display promising activity against both INH sensitive and resistant *M. tuberculosis* strains (Kuo et al., 2003; Sullivan et al., 2006). We next examined five strains against which the triclosan family of compounds had not been tested. Compared to their efficacy against the wild-type strain, five of the six analogues examined showed 4–8-fold lower activity against the InhA(S94A) and InhA(C-15T) strains. This was slightly lower, yet still comparable in magnitude, to the 8- and 13-fold losses in potency for INH against these two strains, respectively. The InhA(C-15T) strain has a higher expression of InhA, which leads to resistance to compounds that target the protein. The resistance to triclosan by InhA(S94A) strains of either *M. tuberculosis* or other bacteria has not been previously reported to the best of our knowledge. Mutation of the G93V in FabI, a residue conserved in InhA, has previously been found to confer *E. coli* resistance to triclosan (McMurry et al., 1998). The *inhA*(S94A) mutation has been demonstrated to decrease the efficacy of INH by weakening the binding of INH–NAD (Vilcheze et al., 2006). Comparatively, the binding of triclosan and its analogues to InhA depends greatly on the interaction between the diaryl ether scaffold and NAD^+. Quemard and co-workers reported that the S94A mutation weakened the binding of NADH by ~500-fold (Quemard et al., 1995). It is

proposed that the same mutation may decrease the stability of the InhA: NAD$^+$:triclosan derivative complex and, thereby, diminish inhibition. The resistance of the InhA(S94A) and InhA(C-15T) strains to triclosan derivatives is supportive of our proposal that the target of these triclosan relatives is indeed InhA.

The activity of the triclosan analogue subset against the Δ*katG* mutant further substantiates the potential advantage of these InhA inhibitors over INH. While INH was more than 3000-times less active against the Δ*katG* mutant, the triclosan analogues were only twofold less active, as measured by their MIC values. This is consistent with our genetic, enzymatic, and X-ray crystallographic data that demonstrate InhA inhibition by the triclosan family without the need for activation. It, therefore, follows that one would expect a lack of cross-resistance with many INH-resistant *M. tuberculosis* strains, where the *katG* mutations represent the dominant allele conferring resistance. Sullivan and colleagues have also demonstrated their phenol diaryl ethers to not suffer significant losses in potency against clinical isolates with varying degrees of INH resistance (Sullivan et al., 2006).

The two clinical strains shown in Table 5-3 were much more sensitive to the triclosan analogues compared with INH. Strain 12081 afforded 17-fold resistance to INH, which was reduced to a 4–8 factor resistance against the triclosan subset. Against clinical isolate 5071, the triclosan subset maintained or improved (in two cases by twofold) their respective wild-type potencies. These results should be contrasted to the greater than two orders of magnitude resistance conferred against INH. Thus, triclosan analogues may, in general, offer a high degree of activity against INH resistant *M.*

tuberculosis.

Conclusions

The efficacy of these 5-substituted triclosan analogues has been shown to be tunable through a structure-based optimization of the 5-position. Novel analogues afforded a 50-fold increase in InhA inhibition and tenfold in bacteriocidal activity. Most significantly, these triclosan derivatives demonstrated efficacy against INH-resistant laboratory and clinical strains of *M. tuberculosis*. Further studies of triclosan and other small-molecule inhibitors of InhA hold significant promise for the delivery of novel antitubercular agents that are effective against drug-resistant *M. tuberculosis*. Current efforts are underway to examine the physiochemical properties of our most promising analogues. Substitution of the phenol will be investigated to avoid potential liabilities due to rapid Phase II metabolism (Wang et al., 2004), preceding clearance. In addition, other modifications to the A-and/or B-rings of the diaryl ether scaffold can be envisioned that would potentially increase the stability of these derivatives to oxidative metabolism, while not sacrificing efficacy.

Experimental procedures

Triclosan analogue synthesis

The small molecules tested against InhA and the mycobacterial strains were synthesized as described previously in the literature (Freundlich et al., 2007).

Cloning, expression, and purification for **Mycobacterium tuberculosis inhA**

M. tuberculosis inhA was cloned into *E. coli* BL21 (DE3) as described previously (Dessen et al., 1995). The transformed *E. coli* were cultured in Terrific Broth media with 50 μg mL^{-1} carbenicillin at 37 °C until an OD_{600} of 0.8 was observed. *inhA* expression was induced for 20 h at 16 °C through addition of 1 mM isopropyl β-D-thiogalactopyranoside. The resulting cells harvested through centrifugation were re-suspended in 50 mM PIPES (pH 6.8) and 1 mM phenylmethylsulfonyl fluoride and lysed via French press. Exposure to DNAse I was followed by removal of insoluble material through centrifugation. The supernatant was subjected to a HiTrap Blue Sepharose column (AP Biotech), pre-equilibrated with the same buffer, using a fast protein liquid chromatography system and eluted through a NaCl gradient (0–2 M). Elution with 0.9 M NaCl solution afforded fractions containing InhA that were subsequently subjected to an octyl-sepharose column (AP Biotech), pre-equilibrated with 1 M NaCl, and eluted through a NaCl gradient (1–0 M). Pooling of InhA fractions and gel filtration through a Superdex 200 column were carried out to separate monomeric protein from aggregated material. SDS-PAGE and Coomassie Blue staining were consistent with homogeneous InhA, yielding 40 mg L^{-1} from *E. coli* culture.

InhA enzyme assay

Experiments were carried out utilizing a Cary100 Bio spectrophotometer at 25 °C, through monitoring the oxidation of NADH to NAD^+ at 340 nm. Reactions were

initiated via addition of dodecenoyl-CoA (50 μM) substrate to mixtures containing InhA (5 nM), NADH (100 μM), and inhibitor (1–10000 nM). The IC_{50} values were determined from a dose-response plot of enzyme fractional activity versus the concentration of inhibitor.

Antimycobacterial assays

The relevant *M. tuberculosis* strains were obtained from laboratory stocks and grown in Middlebrook 7H9 medium (Difco), supplemented with 10 % (v/v) OADC enrichment (Difco), 0.2% (v/v) glycerol, and 0.05 % (v/v) tyloxapol to an OD_{600} of ~ 1.0. The cultures were diluted 4 logs and 0.1 mL of the resulting dilutions were inoculated into 2 mL of Middlebrook 7H9 media with varying concentrations of inhibitor. The cultures were incubated with shaking at 37 °C for four weeks. The MIC was defined as the concentration of inhibitor that prevented visible mycobacterial growth.

Crystallization of InhA with selected inhibitors

The hanging drop vapor diffusion method was utilized. Typically, InhA was incubated with an inhibitor and NAD^+ in the molar ratio of 1:2:100 for 2 h. Co-crystallization was then attempted in hanging droplets consisting of 2 μL of 10 mg mL^{-1} protein solution and 2 μL of buffer (20% PEG 3350, 6% DMSO, 0.1 M N-(2-acetamido)iminodiacetic acid pH 6.8, 0.08 M NH$_4$OAc) at 16°C in Linbro plates against 0.5 mL of the same buffer. Protein crystals were observed after approximately four days.

213

Data collection and processing

Data were collected at 121 K utilizing cryo-protection solution with reservoir solution and an added 30% ethylene glycol. A crystal of InhA:17 diffracted X-rays to 1.98 Å at beam line 23ID at the Advanced Photo Source (APS), Argonne National Laboratory. Crystals of InhA:24, InhA:25 and InhA:7 diffracted X-rays to 2.30 Å, 2.80 Å and 1.98 Å, respectively, using a Rigaku Raxis detector coupled to an X-ray generator with a copper rotating anode ($Cu_{K\alpha}$, $\lambda=1.54$ Å). Diffraction data was obtained from a single crystal with 0.5° degree oscillation widths for a range of 180°. The data were integrated and reduced using HKL2000 **(Table 5-4)** (Otwinowski and Minor, 1997).

Structure determination and model refinement

Initial phases of the InhA:inhibitor complexes were obtained through molecular replacement utilizing the apo-InhA structure (PDB code: 1ENY) and refined with CCP4 (Collaborative Computational Project, 1994) **(Table 5-4)**. Electron density maps were calculated and additional density was found, consistent with the respective inhibitor. The inhibitor was fit into the additional density and the whole model was rebuilt utilizing XtalView (McRee, 1999). In the final refinement cycles, water molecules were added into peaks above 3-σ of the F_o–F_c electron density maps, such that the water molecules were within hydrogen-bonding distance from the appropriate protein atoms. Final statistics are in Table 5-4.

Table 5-4. Data collection and refinement statistics for reported X-ray structures.[a]

	InhA: **7**	InhA: **17**	InhA: **24**	InhA: **25**
Maximum resolution [Å]	1.97	1.98	2.30	2.80
Space group	*I*4(1)22	*C*2	*C*2	*I*4(1)22
a [Å]	90.0	125.6	125.6	90.0
b [Å]	90.0	92.3	92.4	90.0
c [Å]	183.9	103.0	102.4	183.1
α [°]	90.0	90.0	90.0	90.0
β [°]	90.0	106.4	106.5	90.0
γ [°]	90.0	90.0	90.0	90.0
Unique reflections[b]	27 204 (2964)	71 784 (6260)	43 113 (3799)	9708 (1063)
R_{sym} [%]	6.0 (74.3)	9.9 (70.5)	10.3 (88.3)	12.8 (82.3)
Completeness [%]	99.8 (100)	87.3 (65.5)	99.1 (98.5)	99.8 (100)
Redundancy	10.2 (10.0)	3.7 (2.1)	5.3 (4.9)	8.5 (8.6)
I/σ	42.5 (4.3)	15.7 (1.3)	28.6 (3.3)	22.3 (3.7)
Resolution range [Å]	19.92- 1.97	19.96–1.98	19.85–2.30	19.85–2.8
# reflections	25 755	68885	46 976	9157
# atoms per subunit				
Protein	1994	7826	7739	1994
Cofactor (NAD)	52	176	176	52
Ligand	22	88	88	23
Solvent	168	703	253	11
R_{cryst} [%]	20.7	20.0	20	22.1
R_{free} [%]	25.2	25.3	26	26.8
Average B-factors [Å^2]	35.0	1.98	40.2	54.8
[a] PDB accession code are 3FNG (7), 3FNE (17), 3FNF (24), and 3FNH (25). [b] The value of the highest resolution shell is given in parentheses.				

Additional unpublished results

Activity of JPC2278 against InhA

The 5-*p*-methylphenyl substituted triclosan derivative, named as JPC2278 **(Figure 5-5A)**, which was not tested initially together with the other analogs, was also tested for *in vitro* inhibition against InhA. The compound showed surprisingly good activity against InhA with an IC_{50} of 77.6 ± 2.8 nM **(Figure 5-5B)**. This result was unexpected, because the other close derivatives of JPC2278 with a 5-phenyl (Table 5-1, compound 6), a 5-*o*-methylphenyl (Table 5-2, compound 20), a 5-*m*-methylphenyl (Table 5-2, compound 22), and a 5-*p*-fluorophenyl (Table 5-2, compound 23) group did not display any inhibitory activity ($IC_{50} > 10$ μM) for InhA **(Figure 5-5C)**.

In order to understand the structural basis of the activity difference between these analogs, JPC2278 was manually modeled into the InhA active site. The A-rings of the JPC2278 and compound 25 (the 5-ethylbenzene derivative, Table 5-2) from the InhA:25 complex structure were superimposed to position JPC2278 into the binding site **(Figure 5-6)**. While the initial docking experiments with compounds 6 and 15 suggested that directly attached aryl group at the 5 position will lead to steric clashes with Phe149 (see Results and Discussion for details), it was seen from the JPC2278:InhA model that the methylphenyl group lay in close proximity (<3.0 Å) to the side chains of Phe149 and Pro193 to make favorable hydrophobic and van der Waals interactions. In addition, a π-π stacking aromatic interaction was present between Phe149 and the phenyl ring of the compound.

216
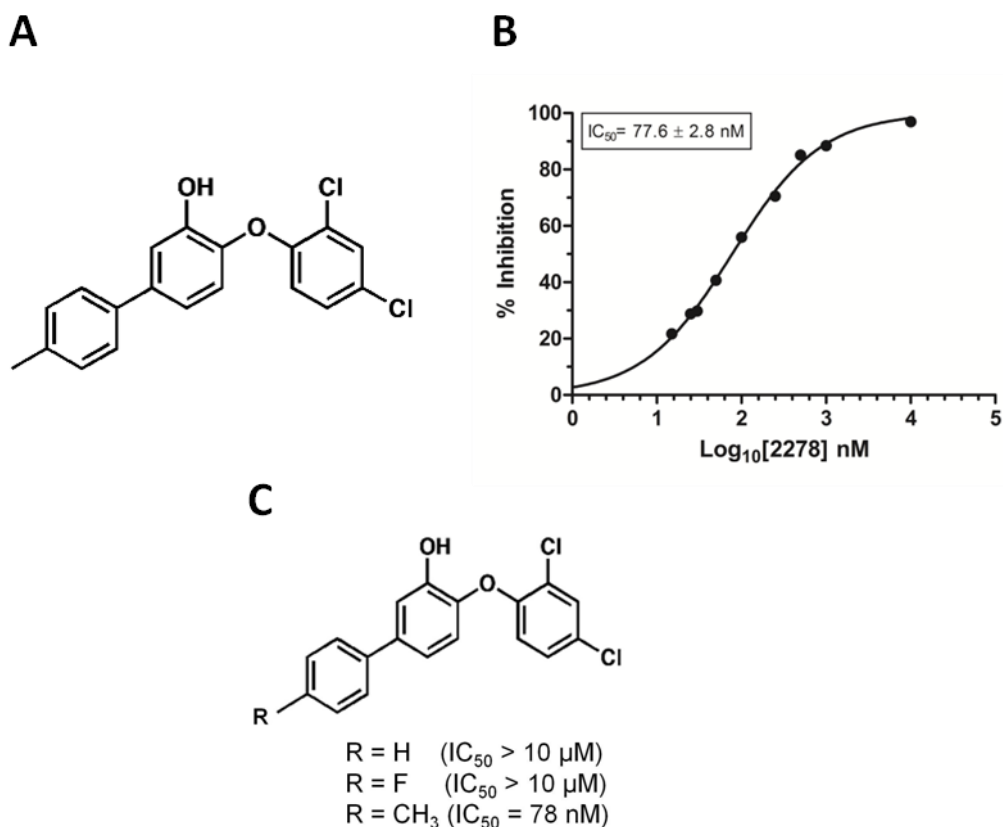
Figure 5-5. Chemical structure and inhibition activity of JPC2278 for InhA. A) Chemical structure of JPC2278. **B)** Inhibition activity of JPC2278 against InhA. **C)** Comparison of the inhibition activity of JPC2278 and its close analogs against InhA.

It is likely that the potent activity of JPC2278 compared with its structural analogs, was due to the *p*-methyl substituent. Based on the model, in the absence of the *p*-methyl substituent, the hydrophobic and van der Waals interactions between the $-CH_3$ group and Phe149 and Pro193 would be lost. This can explain why the 5-*p*-methylphenyl had better potency than the 5-phenyl derivative. A methyl substituent at the -meta or -ortho positions may position too close to Phe149 and cause steric clashes. On the other hand, a single fluorine substitution at the -para position did not restore acti-
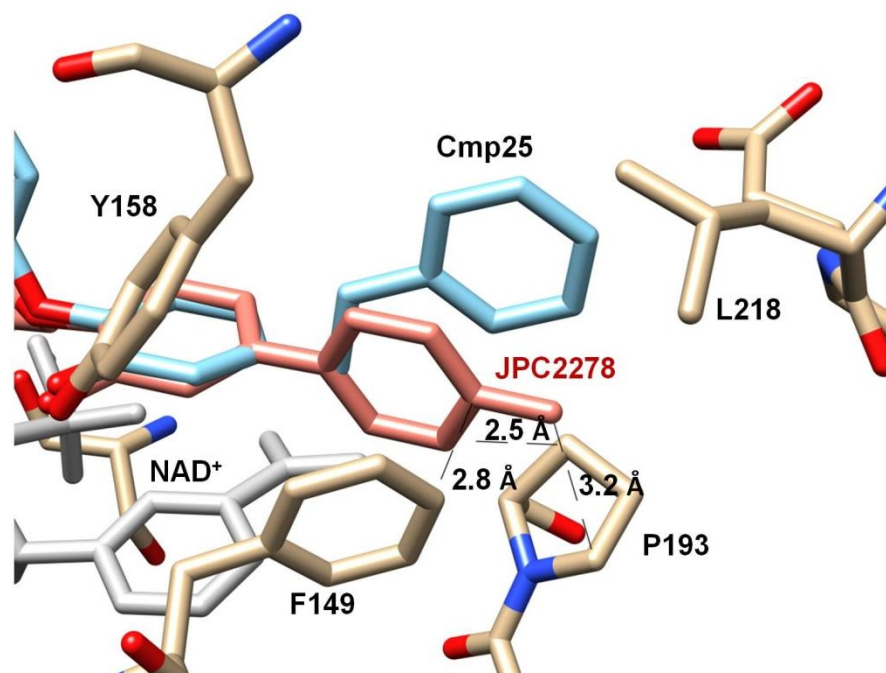
Figure 5-6. Docking of JPC2278 to the InhA active site. JPC2278 (salmon color) is superposed manually with the A-ring of compound 25 (blue color) from the crystal structure of InhA: 25. Based on this model, the methylbenzene substituent lies in close proximity to the side chains of F149 and P193 (< 3.0 Å) and makes favorable hydrophobic and van der Waals interactions.

vity. This could be due to the high negative charge density and low polarizability character of the fluorine atom. The affinity loss might result from the close electrostatically repulsive contacts between the organic fluorine, with its high negative charge density, and the negatively polarized π-clouds of the Phe149 side chain. A similar phenomenon was also observed with neprylisin, a mammalian Zn^{II}-dependent metalloprotease, inhibitors, where any fluorine substituent at a phenyl ring in the binding

pocket caused a decrease in affinity. This was explained by the electrostatically unfavorable close contacts of organic fluorine with the negatively polarized π-surfaces of the surrounding aromatic amino acid side chains (Morgenthalera et al., 2008). In contrast, fluorination of the benzimidazole moiety of the neprylisin inhibitor, which was surrounded by three Arg side chains, led to an increase in the binding affinity. Thus, the authors partitioned the active site of neprylisin to fluorophilic and fluorophobic regions (Morgenthalera et al., 2008). Similarly, the two chlorine substituents on the B ring of triclosan can make favorable interactions with the surrounding residues, whereas the –F of JPC2278 cannot. Organic fluorine prefers dipolar interactions and orientating into the electropositive protein environments (Bissantz et al., 2010; Morgenthalera et al., 2008). This might explain why a hydrophobic –CH_3 group was more potent than the –F substituent for the JPC2278 analogs.

Challenges in obtaining the InhA-ligand complex crystals

Two methods are usually followed to obtain protein-ligand crystals: co-crystallization and soaking. Soaking of the native crystals with the ligand compound solution is usually preferred over co-crystallization, because it's rapid, easy to optimize, and there is no need to search for new crystallization conditions. In contrast, co-crystallization requires the incubation of the ligand with protein prior to crystallization and often results that the crystallization conditions for the native protein do not yield protein-ligand complex crystals. This brings the requirement of searching for new crystallization conditions, which is time and material consuming.

For InhA, the native crystals can be obtained fairly easily in a week using the published crystallization condition of 12% methylpentanediol (MPD), 4% dimethylsulfoxide (DMSO), 100 mM Hepes at pH 7.5, and 50 mM sodium citrate at pH 6.5 (Rozwarski et al., 1998). However, these crystals are very sensitive to soaking with compounds dissolved in DMSO, acetonitrile, ethanol or methanol. The crystals cannot tolerate >2% of any organic solvent that is required to keep the compounds in solution. Even short periods (<30 minutes) of soaking irreversibly damages the crystals causing cracks or dissolving of the crystal, which often results in the loss of diffraction. Furthermore, dissolving the compounds in an aqueous crystallization solution or adding the compound powder to the crystal drop is also not feasible as most of the compounds are hydrophobic and have very poor solubility in aqueous solutions. Even though native InhA crystals can be obtained in some other unpublished crystallization conditions, none of those produce high diffraction quality crystals that can be further used for soaking. Out of the all InhA: inhibitor complex crystal structures described in this dissertation (12 structures in total), only two complex crystals were obtained through soaking. The majority of the complex crystals were obtained by co-crystallization.

For co-crystallization experiments multiple approaches were followed. Since both GSK inhibitors and triclosan derivatives were uncompetitive inhibitors, all of the co-crystallization experiments were performed in the presence of NAD^+. It was observed that at least 10 fold molar excess of NAD^+ was needed for successful crystallization. All of the inhibitor compounds were dissolved in 100% DMSO and 10-100 mM stock solutions were used in experiments. The molar ratio of the compound to InhA was kept between

220

two to five fold; however, a key factor to consider was the final concentration of DMSO, because InhA protein cannot tolerate more than 4% of DMSO in solution. It was also observed that in some cases, instead of mixing the concentrated protein solution (10 mg/ml) with the inhibitor compound, mixing a more diluted (1 mg/ml) protein solution with the compound and then co-concentration helped successful protein-inhibitor complex formation. It also prevented protein precipitation by lowering the local DMSO concentration. Once the protein was incubated with NAD^+ and the inhibitor compound for approximately 20 minutes at room temperature, crystallization plates were set with: i) the published native crystallization condition, ii) the sparse-matrix crystallization screens (~900 crystallization conditions). Among the ten co-crystal structures mentioned in this dissertation, only two of them were obtained in the native crystallization conditions and the other eight required for screening of new crystallization conditions. All of the triclosan derivatives were also co-crystallized in a previously unpublished crystallization condition.

To sum up, obtaining InhA-ligand crystals was not feasible by soaking. Co-crystallization and searching of new crystallization conditions were necessary. Finding the right crystal, which diffracted to high resolution and contained the ligand, required a lot of screening and collection of many data sets.

Summary

Triclosan is composed of two phenyl rings (ring A and B) linked with an oxygen atom. The hydroxyl of the A ring and the chlorines at the 2' and 4' positions of the B

ring are important for the potency of the compound. In the InhA:NAD⁺: triclosan crystal structure, the 5-chlorine of the A ring was positioned at the entrance of the INH-NAD adduct binding pocket (2 Å distance) (Kuo et al., 2003). In this study, the 5-Cl position of the A ring was modified with different substituents of hydrophobic and hydrophilic character in order to occupy the INH-NAD adduct binding pocket **(Figure 5-2)**. The derivatives were first investigated by enzymatic assays and the compounds showing good enzyme inhibition were used for the crystallographic studies.

It was found by enzymatic assays that compounds with hydrophobic substituents at the 5-position such as alkyl groups showed better inhibition than the compounds with hydrophilic substituents like carboxylate groups **(Tables 5-1 and 5-2)**. The potency of the hydrophobic group substituted compounds increased with the chain length of the alkyl substituents up to four carbon chains. A similar trend was also observed by Sullivan et. al. with an optimum chain length of eight carbons for the des-chloro triclosan derivatives. The most potent derivatives were the compounds 17 and 25 with 2-pyridyl-methyl and benzylmethyl substituents **(Table 5-2)**.

The comparison of the InhA:triclosan derivative structures with the InhA:INH-NAD(H) adduct and the InhA:C16 fatty acyl substrate analog structures revealed that the hydrophobic moieties attached to the 5-position of the A ring did not perturb the side chain of Phe149, and thus, they did not occupy the same pocket as the INH-NAD(H) adduct **(Figure 5-7)**. Instead, these substituents were positioned similar to the substrate analog. In the InhA:INH-NAD structure, flipping of the Phe149 side chain was accompanied by the movement of the Tyr158 side chain (Rozwarski et al., 1998)

222

(Figure 5-7B). However, in the InhA:triclosan derivative structures, Tyr158 was a key residue for inhibitor binding due to the direct H-bonding interaction with the hydroxyl of the A ring. Importantly, loss of this interaction resulted in loss of potency. The 5-substituted triclosan analogs were positioned proximal to the isonicotinoyl binding pocket but they were not able to access to this site as this would require the movement of Phe149 and Tyr158 simultaneously, which in turn would cause the loss of the direct H-bonding interaction between the compound and Tyr158. Even though the 5-substituted derivatives could not reach INH-NAD binding pocket, these derivatizations clearly increased the inhibitor potency by 10 to 50 fold compared with the parent compound triclosan. Notably, the 5-substituted derivatives were activation free inhibitors and found to be effective against the isoniazid resistant strains. For instance, derivative #7 displayed an MIC of 19 μg ml^{-1} against the *Mtb* H37Rv Δ*katG* strain compared to INH having an MIC of >200 μg ml^{-1} against the same strain.
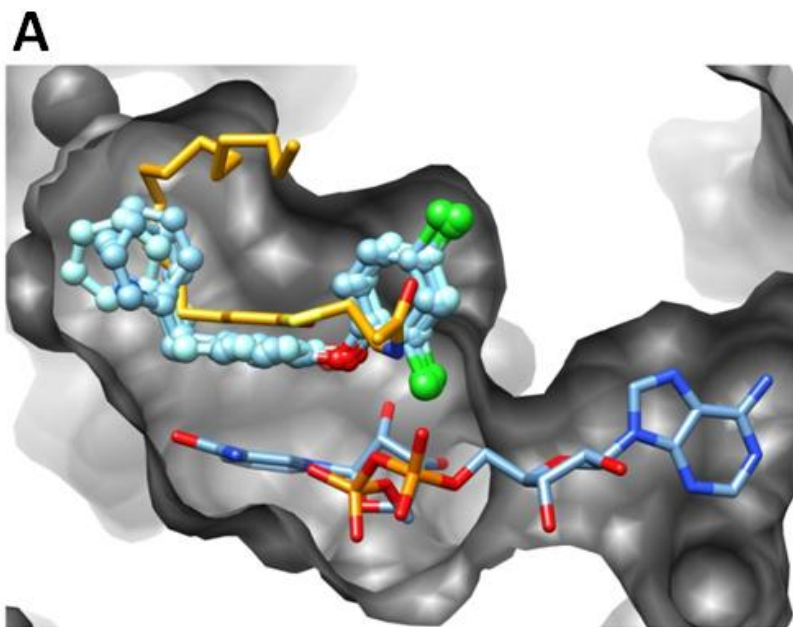
Figure 5-7. Comparison of the InhA:triclosan derivative structures with the InhA: INH-NAD(H) adduct and the InhA:C16 fatty acyl substrate analog structures. A) A cross-section through the surface of the active site of InhA. Four InhA:triclosan analog structures (light blue) were superposed with the InhA:C16 fatty acyl substrate analog structure. NAD$^+$ and the C16-fatty acyl substrate analog are shown in blue and orange sticks, respectively. As seen from the figure, triclosan analogs occupied the substrate binding pocket. **B)** Comparison of the conformation of Tyr158 and Phe149 in the InhA: triclosan analogs structures (light blue) with the InhA:INH-NAD structure (orange) (PDB ID:1ZID). The INH-NAD adduct causes rotation of the Phe149 side chain ~90°, which in turn moves Tyr158 away from NAD(H). In contrast, none of the triclosan analogs perturbed the side chain of Phe149, and the A ring hydroxyl H-bonded with the side chain of Tyr158 (2.3 Å). For both panels atom coloring is: oxygen red; nitrogen blue; sulfur yellow; chlorine green.
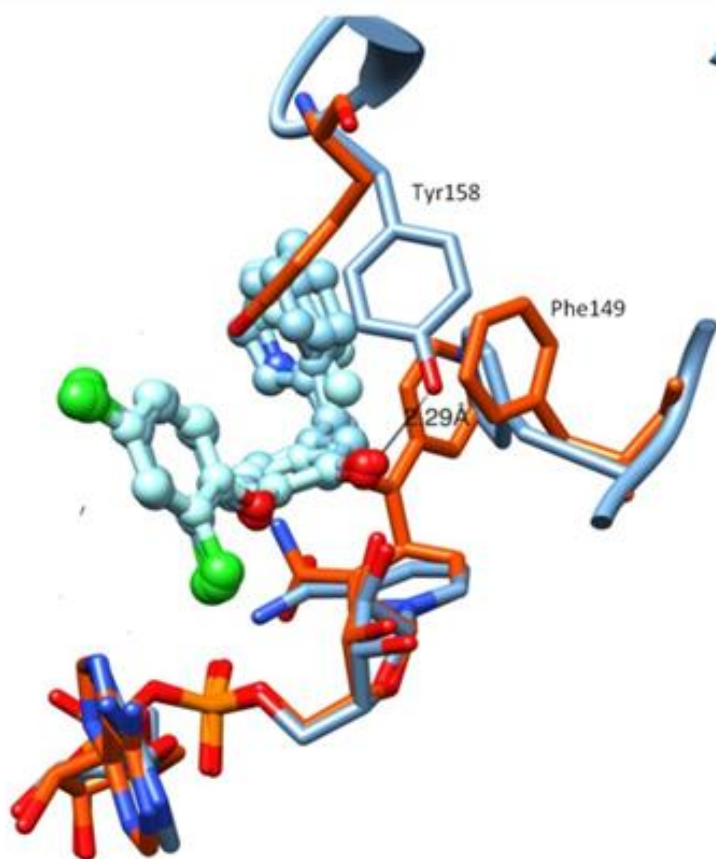
Figure 5-7 Continued.

CHAPTER VI

STRUCTURAL STUDIES ON THE MYCOBACTERIAL TRANSCRIPTION

REGULATOR, CARD*

Overview

CarD from *Mycobacterium tuberculosis* (*Mtb*) is an essential protein shown to be involved in stringent response through downregulation of rRNA and ribosomal protein genes. CarD interacts with the β-subunit of RNAP and this interaction is vital for *Mtb*'s survival during the persistent infection state. We have determined the crystal structure of CarD in complex with the RNAP β-subunit β1 and β2 domains at 2.1 Å resolution. The structure reveals the molecular basis of CarD/RNAP interaction, providing a basis to further our understanding of RNAP regulation by CarD. The structural fold of the CarD N-terminal domain is conserved in RNAP interacting proteins such as TRCF-RID and CdnL, and displays similar interactions to the predicted homology model based on the TRCF/RNAP β1 structure. Interestingly, the structure of the C-terminal domain, which is required for complete CarD function *in vivo*, represents a distinct DNA-binding fold.

Introduction

Tuberculosis (TB) is a global health threat responsible for approximately two million deaths annually (WHO, 2013). The treatment for the primary causative agent of TB, *Mycobacterium tuberculosis* (*Mtb*), is challenging due to the emergence of multidrug (MDR-TB) and extensively drug-resistant (XDR-TB) strains. Because most of the antibiotics currently used for *Mtb* therapy are potent only against replicating bacteria, mycobacteria are able to survive in the host in a nonreplicating, persistent, or chronic state. Identifying new drugs that can target *Mtb* during the persistent stages of infection is very important (Gupta et al., 2009; Raman et al., 2008; Sacchettini et al., 2008).

In *Mtb*, the gene product of *Rv3583c*, annotated CarD, is required for persistence, and has been identified as an essential protein *in vitro* and *in vivo* during normal growth conditions as well as under genotoxic stress and nutrient deprivation (Stallings et al., 2009). Microarray studies have shown that not only is *Rv3583c* upregulated in response to oxidative stress, DNA damage, and starvation, but also that depletion of *Mtb* CarD results in loss of transcriptional regulation of rRNA and ribosomal components, indicating its involvement in stringent response (Stallings et al., 2009). *Mtb* CarD can be used to complement *Escherichia coli* DksA protein, which regulates stringent response alongside the main stringent response element hyperphosphorylated guanine nucleotide ((p)ppGpp), suggesting that the two proteins are functional homologs. DksA directly interacts with the RNA polymerase (RNAP) through the RNAP secondary channel and potentiates the effect of (p)ppGpp (Paul et al., 2004b; Srivatsan and Wang, 2008). *Mtb* produces (p)ppGpp, but does not have a DksA homolog.

227

The first and best studied CarD protein is from *Myxococcus xanthus* and is a transcription regulator involved in carotenogenesis. *M. xanthus* CarD interacts with RNAP, CarG, and the *carQRS* promoter DNA through its N- and C-terminal domains (Nicolas et al., 1996; Penalver-Mellado et al., 2006). *Mtb* CarD shares only a 30% sequence homology with the N terminus of *M. xanthus* CarD, and the C-terminal domain is not similar, suggesting that *Mtb* CarD does not contain the HMG1-like DNA-binding domain (AT hook DNA-binding motif sequence) found in *M. xanthus* CarD **(Figure B-1)**. Bacterial two hybrid assays and immunoprecipitation experiments have shown that *Mtb* CarD associates with the RNAP β-subunit (Stallings et al., 2009; Weiss et al., 2012). All CarD and CarD N-terminal like (CdnL) proteins belong to the transcription-repair coupling factor (TRCF) family of proteins, and share sequence and structural homology with the TRCF RNAP interacting domain (RID). Also, they are thought to interact with RNAP in a homologous manner as TRCF. The previously determined crystal structure of the *Tth* TRCF-RID/*Taq* RNAP β1 complex (Westblade et al., 2010) and the homology models generated for CarD-RNAP interaction based on this structure (Weiss et al., 2012) predicted a similar set of interactions of CarD with RNAP. However, the mechanism CarD uses to regulate RNAP function and transcription, and the role of the *Mtb* CarD C-terminal domain are unknown. The CarD/RNAP interaction is crucial for *Mtb*'s stringent response, viability, and resistance to oxidative stress and loss of the CarD/RNAP interaction sensitizes *Mtb* to the anti-TB drug rifampicin, emphasizing the importance of understanding this protein-protein interaction.

Here, we report the crystal structure of *Mtb* CarD complexed with the *Mtb* RNAP

β-subunit lobe domains at 2.1 Å resolution. The CarD/RNAP β structure reveals that the RNAP CarD binding site is located on the β-subunit arm of the RNAP claws, specifically on the solvent exposed surface of the β1 domain, and is far from the catalytic center of the RNAP. The structure not only provides insight into the molecular basis of RNAP interaction with *Mtb* CarD, but also with other CarD family proteins and CarD homologs. The structural basis for the RNAP regulation through CarD interaction, which is distinct from the DksA regulation mechanism, is presented by comparing the uncomplexed *Mtb* RNAP-β and CarD/RNAP-β complex structures. While the structural fold of the CarD N-terminal domain is conserved among other CarD, CdnL, and TRCF-RID domains, the C-terminal domain structure is not been identified in any other structure in the Protein Data Bank (PDB). We show that *Mtb* CarD is a DNA-binding protein with a distinct DNA-binding domain and that it exhibits a nonsequence-specific DNA-binding mode.

Results and discussion

Structure determination of **Mtb** *Rpobtr and the CarD/RNAP complex*

  E. coli expression plasmids for the full-length *Mtb* RNAP β-subunit (RpoB) and several truncations (based on the secondary structure predictions) were made to test for recombinant protein expression. One truncation containing the β-lobes (consisting of residues 47–433) of RpoB (referred as RpoBtr) yielded soluble protein and when co-expressed with *Mtb* CarD resulted in complex formation, and it was chosen for

subsequent crystallographic studies. The β1 (residues 47–172 and 375–428, corresponding to the *Taq*-β residues 1–130 and 334–395 of the TRCF/β1 structure) and β2 (residues 177–370) domains contained within this truncation are important for RNAP regulation. They form the β-arm of the RNAP claws that cover the DNA/RNA hybrid and dsDNA in the transcription complex. Regulation through these domains often occurs through interaction with various regulatory proteins such as TRCF and sigma70 (Vassylyev et al., 2002; Westblade et al., 2010). RpoB does not have any dispensable regions, and the archaebacterial split site, which maps around residue 570 of RpoB, is not contained within the β-subunit truncation used in this work. Crystals of RpoBtr were determined to be in the $P2_12_12_1$ space group with two molecules in the ASU. The RpoBtr structure was solved by single-wavelength anomalous diffraction using Se-Met derived crystals to a resolution of 2.9 Å. Subsequently, the resolution was improved to 2.5 Å for native (non Se-Met) RpoBtr crystals. The structure was refined to $R_{work} = 21\%$ and $R_{free} = 26\%$, with excellent stereochemistry **(Table 6-1)**.

Table 6-1. Data collection and refinement statistics.

	Se-Met RpoBtr	Native RpoBtr	CarD/RpoBtr complex
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$C222_1$
Cell dimensions			
a, b, c (Å)	52.8, 124.1, 135.5	53.1, 123.9, 135.6	49.1, 128.9, 225.5
α, β, γ (^{0})	90,90,90	90,90,90	90,90,90
Wavelength	0.979	0.979	0.979
Resolution (Å)	2.8 (2.79)	2.5 (2.45)	2.1 (2.11)
Completeness (%)	98.5 (90.3)	99.5 (97)	98.1 (92)
Redundancy	3.1 (3.1)	7.0 (6.3)	4.8 (4.6)
I/Iσ	13.9 (2.4)	15.1 (1.9)	10.1 (2.0)
R_{sym} (%)	5.28 (53.4)	5.45 (65.7)	7.8 (73.14)
Refinement			
Resolution	48.6-2.79	49-2.45	37.6-2.11
No. reflections	22552	33551	40747
R_{work}/ R_{free}	0.24/0.28	0.21/0.26	0.20/0.23
No. atoms			
Protein	5844	6097	4203
Water	19	126	305
B-factors			
Protein	95.2	46.4	31.4
Water	57.1	47.3	45.0
R.m.s deviations			
Bond lengths (Å)	0.010	0.003	0.005
Bond angles ($^\circ$)	1.52	0.68	0.87

Highest resolution shell values are given in parenthesis.

Full-length CarD bound to RpoBtr was crystallized after co-expression and purification of the complex. The crystals belonged to the $C222_1$ space group with a single copy of the heterodimer in the ASU. The structure of the CarD/RpoBtr complex

(hereafter referred to as the CarD/RNAP complex) was determined by molecular replacement, using the uncomplexed RpoBtr β1 and β2 domain structures as two individual search molecules. After locating RpoBtr using molecular replacement and initial refinement, there was clear electron density $|F_o|- |F_c|$ that was unaccounted for **(Figure B-2A)**. This extra electron density belonged to the protein, and the entire atomic model of CarD was manually built into the difference electron density map. The crystal structure of the complex was refined to $R_{work} = 20\%$ and $R_{free} = 23\%$ using diffraction data to 2.1 Å resolution.

Overall structure of Mtb *CarD*

Mtb CarD belongs to the α + β protein class (SCOP) (Murzin et al., 1995). The structure is composed of two distinct domains: an all β-stranded N-terminal domain (residues 1–49) and an all α-helical C-terminal domain (residues 63–160; **Figures 6-1A and 6-2A**). The N- and C-terminal domains are connected by a six-residue twisted α helix (α1) and an eight residue loop. The N-terminal domain has a Tudor-like fold (Selenko et al., 2001) consisting of four antiparallel β strands. Residues Thr26, Ile27, Lys28, and Gly29, which lie on the β-turn connecting the β2 and β3 strands, were the only residues disordered in the N-terminal domain.
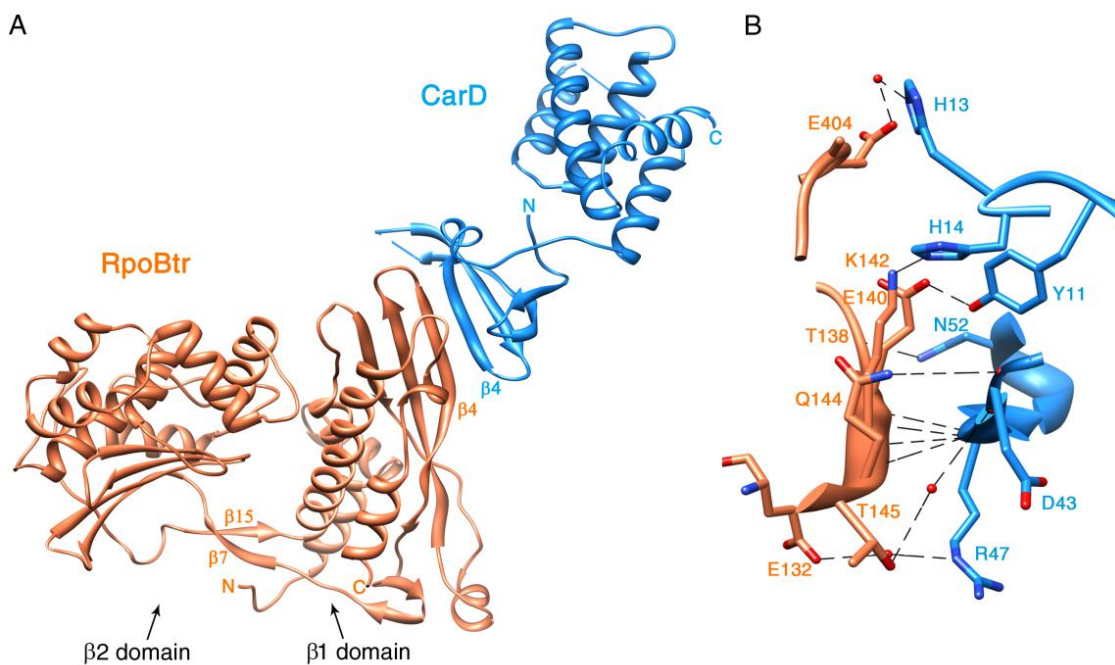
Figure 6-1. Ribbon representation of the *Mtb* CarD/RNAP β1-β2 domain complex structure. RpoBtr is represented by orange ribbons; CarD is represented by blue ribbons. **A)** Overall structure of the complex. β1 and β2 domains of RNAP β-subunit, β4 strands of each protein, and N and C termini of each chain are labeled. The interdomain bridging β sheet of RpoBtr (β7 and β15) is also indicated. **B)** Zoom of the CarD-RNAP interface. Direct and water-mediated H-bonding interactions between side chains and backbone-backbone interactions are shown. Hydrogen bonds and nonbonded contacts between RpoBtr and CarD are formed by the residues located on the β4 strands of both proteins, on the loop connecting α12 and α13 of the RNAP β1 domain, and on the turn between the β1 and β2 strands of CarD. H-bonds are represented by dashed lines and water molecules are shown in red spheres. For the distances, refer to Table 6-2. See also Figure B-2.

The structure of the CarD N-terminal domain is conserved among RNAP interacting proteins such as TRCF-RID and *Thermus thermophilus* (*Tth*) CdnL (CarD N-terminal like protein involved in cell division). Superposition of the *Mtb* CarD N-terminal domain structure with the *Tth* CdnL N-terminal domain (PDB ID: 2LQK)

(Gallego-Garcia et al., 2012) and *Tth* TRCF-RID structures (PDB ID: 3MLQ) (Westblade et al., 2010) gives a root-mean-square deviation (rmsd) of 1.1 Å over the C_α backbone for residues 1–49 **(Figure 6-2A)**.
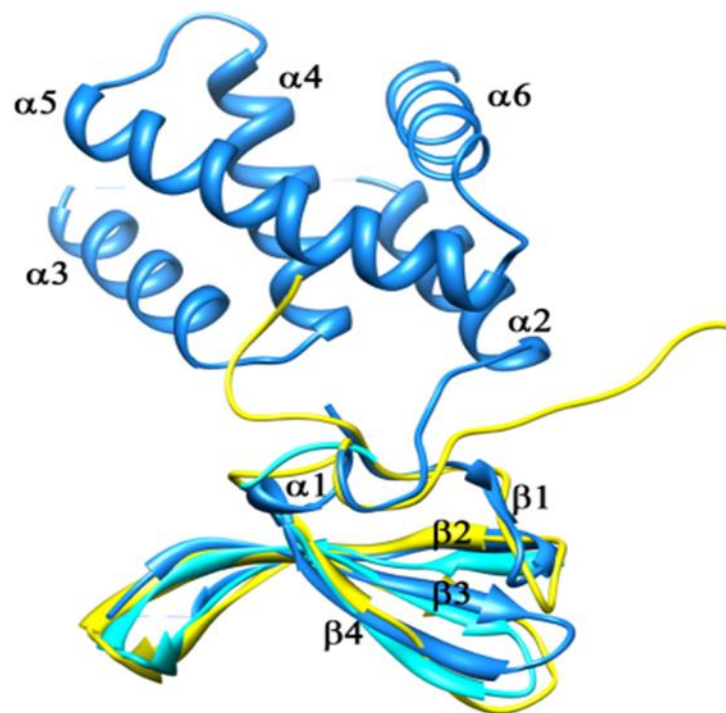
Figure 6-2. Ribbon representation of *Mtb* CarD. A) Superposition of *Mtb* CarD N-terminal domain structure (blue) with the *Tth* CdnL-NMR structure (yellow) and the *Tth* TRCF-RID structure (cyan). The RNAP-interacting β4 strand is labeled. The rmsd over the C_α backbone is 1.1 Å. **B)** Ribbon representation of the CarD C-terminal domain. The helices contain mostly hydrophobic amino acids at their helix-helix interfaces, generating a hydrophobic core. The remainder of the structure is omitted for clarity. **C)** The leucine zipper present between α4 and α5 in the C-terminal domain of CarD is formed by residues Leu120, Leu111, Leu128, and Leu135, and surrounded by hydrophobic residues Val104, Val108, and Ala131. The rest of the structure is omitted for clarity. The leucine zipper lies inside the hydrophobic core of CarD and is not involved in dimerization or DNA interaction. In both panels, hydrophobic residues are shown with yellow sticks. See also Figure B-1.

B
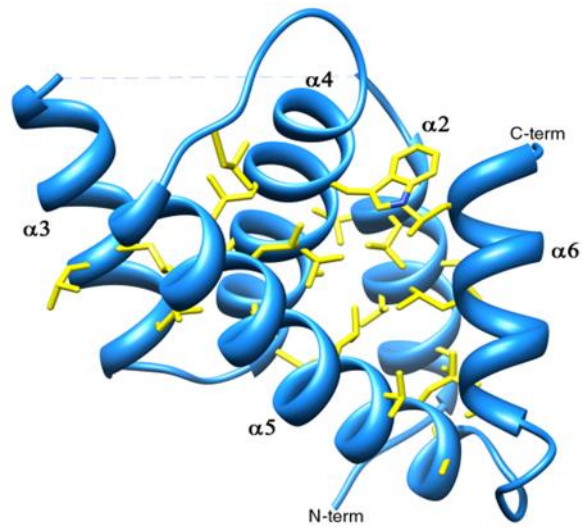
C
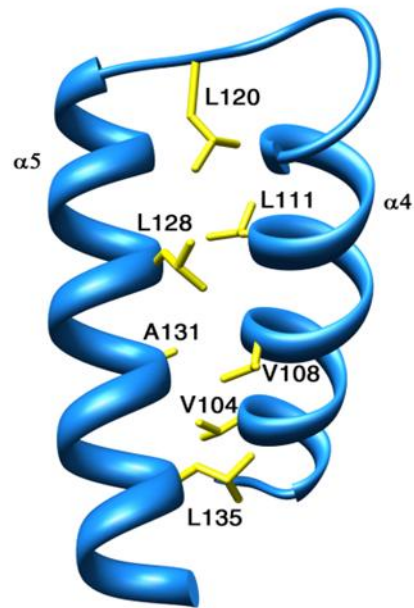
Figure 6-2 Continued.

235

The CarD C-terminal domain is comprised of an α-helical bundle of five α helices that contains an unexpected internal leucine zipper between helices α4 and α5 **(Figures 6-2B and 6-2C)**. Helices α4 and α5 interact only through hydrophobic and van der Waals interactions, afforded by the leucine zipper. Helices α2 and α3 are positioned parallel to each other, whereas α4, α5, and α6 pack orthogonally to each other. Helices α3 and α4 are connected by a γ-turn, while α4-α5 and α5-α6 are connected by β-turns **(Figure 6-2B)**. The loop connecting α2 and α3, spanning residues His78 to Asn83, is completely disordered in the structure. The helices contain mostly hydrophobic amino acids at their helix-helix interfaces, generating a hydrophobic core. There are just two polar interactions inside this compact bundle, the hydrogen bonds between Lys95 (α3)-Glu106 (α4; 2.7 Å) and Arg132 (α5)-Asp155 (α6; 2.9 Å).

The three-helix bundle of α3, α4, and α5 is involved in DNA binding (see DNA-binding studies on *Mtb* CarD section for details). The structure of this three-helix bundle is unlike other DNA-binding proteins in the PDB, including all of the other HTH motifs and leucine zipper domains. DNA-binding proteins with classical HTH motifs usually insert their second (recognition) α helix in the major groove for base-specific DNA interaction. Leucine zippers containing DNA-binding proteins usually dimerize through the hydrophobic leucine zipper region, while also interacting with the major groove of the DNA; however, this is not the case for CarD. The DNA-interacting region of *Mtb* CarD is mapped to the N termini of α3 and α5, the C terminus of α4, and the β-turn connecting α4 and α5. The leucine zipper motif of CarD appears to stabilize the conformation of α4 and α5 inside the hydrophobic core and is not involved directly in

dimerization or DNA interaction (**Figure 6-2C**).

Neither the AT-hook DNA-binding motif of *M. xanthus* CarD nor any other recognizable DNA-binding motif is present in the *Mtb* CarD protein sequence. Structural similarity searches of the CarD C-terminal domain structure, using the PDBeFold, VAST, and DALI servers against the PDB and Structural Classification of Proteins database, did not identify any significant structural homologs. The structural alignment scores were well below the threshold of significance (VAST score < 5.5, VAST -log(p) value < 4.0, and Q-score < 0.49) for each alignment program (Gibrat et al., 1996; Holm and Rosenstrom, 2010; Krissinel and Henrick, 2004). It has been observed that the VAST hits do not share any common functional or structural features with *Mtb* CarD, besides being α-helical proteins.

Overall structure of Mtb *RNAP β1-β2 domains*

The RpoBtr structure reported here consists of the β1 and β2 domains of the *Mtb* RNAP β-subunit (corresponding to the protrusion and lobe domains, respectively, of eukaryotic RNAP II) that form the RNAP claws, together, with the β' subunit. RNAP interacts with the transcription bubble nontemplate strand, especially with the G+2 base, through the β2 domain residues, and these interactions are critical for sequence-specific promoter recognition of RNAP along with the transcription bubble formation and stability (Zhang et al., 2012a).

The RpoBtr β1 domain aligns well with the β1 domain from the TRCF/β1 structure (rmsd 1.04 Å). However, superposition of the RpoBtr β1 domain with the *E.*

coli, Tth, and *Taq* RNAP β-lobe structures gives an rmsd of approximately 10 Å over the C_α atoms of the β2 domain **(Figures B-3A and B-3B)**. β-lobes are known to have conformational flexibility (Tagami et al., 2010), and the relative conformation adopted by the β1 and β2 domains of RpoBtr has not been observed in any other bacterial core or holo RNAP structure. The two molecules in the ASU of the uncomplexed RpoBtr structure, RpoBtr_A and RpoBtr_B, are also in different conformations. When the β1 domains of RpoBtr_A and RpoBtr_B are aligned, the rmsd of the C_αs of the β2 domains is 5.2 Å (over 191 atom pairs). The conformational difference observed in the β1-β2 domain-domain orientation can be explained by rotation around the hinge axis centered on the two stranded antiparallel β sheet connecting the two domains **(Figure B-3C)**.

The RNAP β-subunits are structurally highly conserved among different kingdoms, even though sequence conservation is low (Lane and Darst, 2010; Severinov et al., 1996). As expected, the secondary and the tertiary structure of the *Mtb* RNAP β1-β2 domains are almost identical to the *E. coli* and *Tth* RNAP β1-β2 domains **(Figures B-3A and B-3B)**. The β1 domain (residues 47–172 and 375–428) consists of four antiparallel β strands flanked by five α helices on one side and one α-helix and a β-hairpin from the other side **(Figure 6-3A)**. The β2 domain (residues 177–370) is composed of four antiparallel β strands flanked by seven α helices. The two domains are connected by a two-stranded antiparallel β sheet (the β7 strand and β15 strand), positioned like a bridge **(Figure 6-3A)**. In contrast to other bacterial RNAPs, RpoBtr has an additional 12-residue β-hairpin connecting α11 and the bridge strand β15 on the β2 domain.
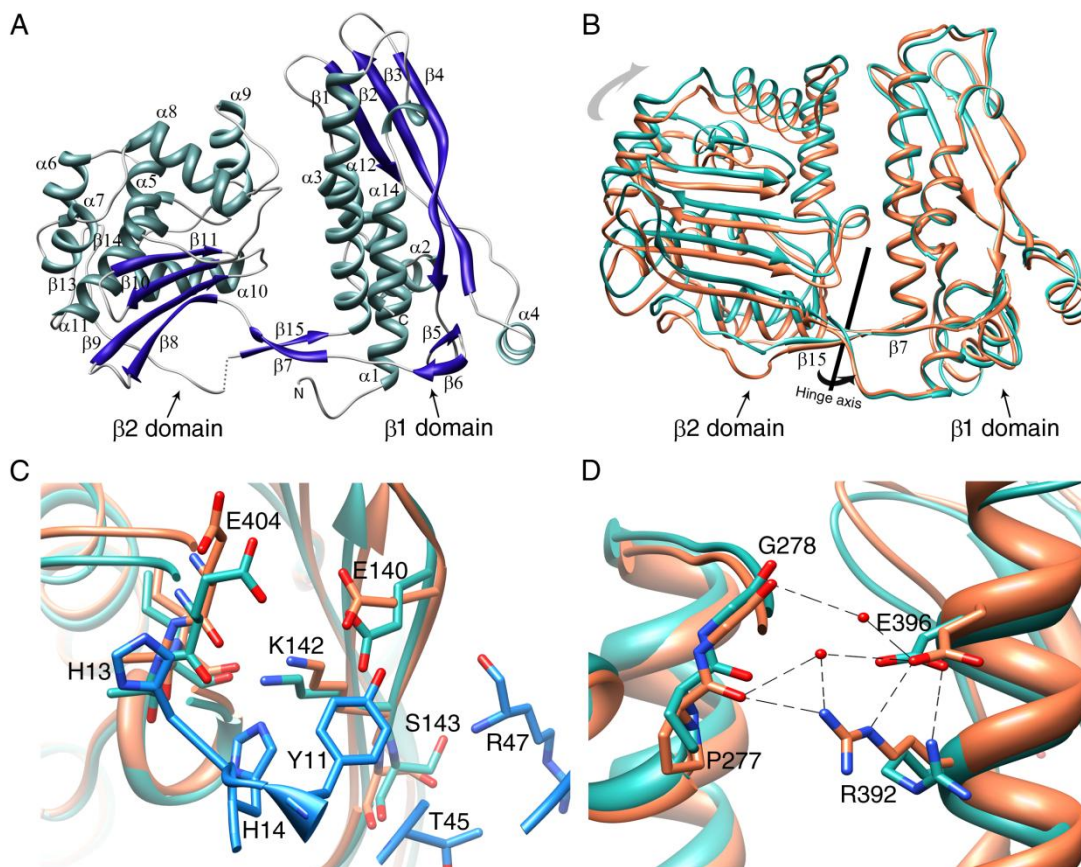
Figure 6-3. Conformational differences in RpoBtr in the uncomplexed and complexed forms. A) Ribbon representation of an uncomplexed RpoBtr molecule. The secondary structure assignments were done with PDBsum server (Laskowski, 2009). **B)** The conformational differences observed in the β1-β2 domain-domain orientation between uncomplexed RpoBtr (green) and RpoBtr complexed with CarD (orange). Superposition of the β1 domains yields an rmsd of 2.8 Å over the C_αs of the β2 domains. The hinge axis centered on the two-stranded antiparallel β sheet (β7 and β15) bridging the two domains is also shown. **C)** Local conformational changes of the RNAP β1 domain residues at the CarD/RNAP interface. β1-E404, S143, and E140 change conformation to interact with CarD-H13, R47, T45, and Y11. **D)** Local conformational changes of the RNAP β1 domain residues at the β1-β2 domain interface. E396 and R392 of the β1 domain make additional water-mediated interactions with the β2 domain residues P277 and G278 in the CarD/RNAP complex. The CarD residues are shown in blue. Coloring of RpoBtr is the same as in (B). Molecular visualization and analysis, including rmsd calculations, were performed with the UCSF Chimera package (Pettersen et al., 2004). See also Figure B-3.

239

Overall structure of the CarD/RNAP-β1-β2 complex

The CarD binding site of RNAP is located at the solvent exposed surface of the β1 domain, which is approximately 70 Å away from the RNAP active site Mg^{+2} (based on the *Tth* EC; PDB ID: 2O5I) **(Figure 6-4A)**. Despite the long distance between the binding site and the active site, this domain serves as an interaction module for various regulatory proteins, including sigma factors at different stages of transcription, and is important for RNAP DNA binding and open complex stability (Trinh et al., 2006; Vassylyev et al., 2002).

At the CarD-RNAP interface, the primary four-stranded β sheet of the CarD N-terminal domain forms an extended eight-stranded β sheet with the β1 domain of the RNAP β-subunit (see **Figure B-2** for a stereo image). Specifically, the β4 strand of CarD comprising residues Leu44 to Pro49 forms an antiparallel β sheet with the β4 strand residues Thr138 to Gln144 of RpoBtr (1:1 heterodimer; **Figures 6-1A and B-2B**). This results in a mixed β sheet in the $1_\uparrow$ $2_\downarrow$ $3_\uparrow$ $4_\downarrow$ $4'_\uparrow$ $3'_\downarrow$ $2'_\uparrow$ $1'_\downarrow$ topology. Surprisingly, association of RNAP with CarD results in only a 500 $Å^2$ (otherwise solvent exposed) buried surface area, which is below average (1,500–2,000 $Å^2$) for heteromeric protein-protein complexes (Kleanthous, 2000). While the buried surface is relatively small, it is rich in intermolecular hydrogen bonds. There are eight hydrogen bonds and 69 nonbonded contacts between RpoBtr and CarD formed by the residues located on the β4 strands of both proteins, on the loop connecting α12 and α13 of the RNAP β1 domain, and on the turn between the β1 and β2 strands of CarD. Specifically, β1-Ile141 interacts with CarD-Arg47 (2.8 Å), and β1-Ser143 interacts with CarD-Thr45 (2.7 Å) through

240

four backbone-backbone hydrogen bonds (**Figure 6-1B**). Interestingly, the side chain-specific hydrogen bonding interactions are present only between β1-Lys142:CarD-His14 (2.9 Å), β1-Glu140:CarD-Tyr11 (2.4 Å), β1-Thr138:CarD-Asn52 (2.8 Å), and β1-Gln144:CarD-Gly42 (3.0 Å; **Figure 6-1B**).
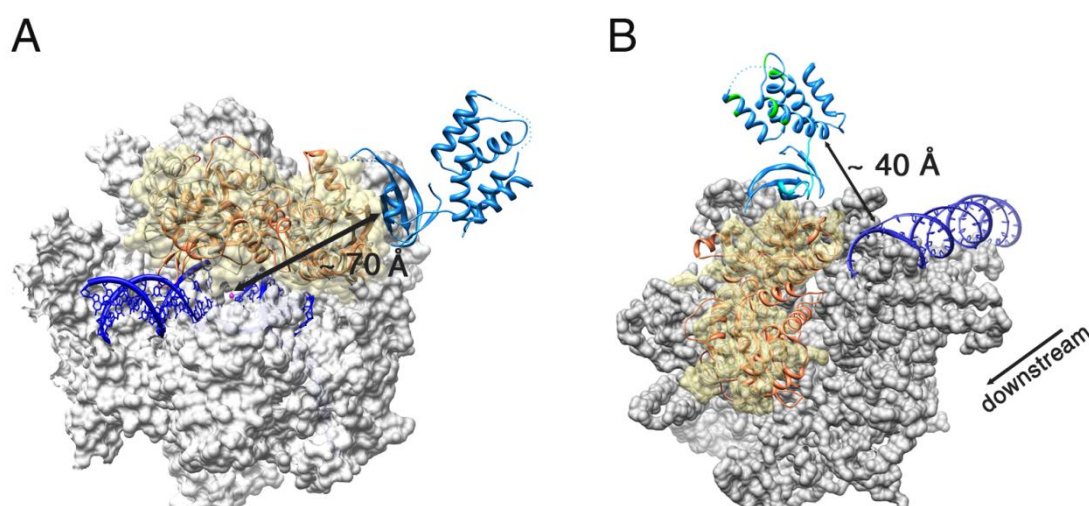
Figure 6-4. Superposition of the *Mtb* CarD/RNAP complex structure with the bacterial elongation and initiation complex structures. The *Mtb* CarD/RNAP complex is colored as previously (RpoBtr is orange and CarD is blue). **A)** Superposition of the *Mtb* CarD/RNAP complex structure with the *Tth* RNAP EC structure (PDB ID: 2O5I). *Tth* RNAP is gray, the DNA duplex and DNA-RNA hybrid are dark blue, and the active site Mg^{+2} is shown as a magenta sphere. The molecular surface of the α and β' subunits is shown in gray, where the β1-β2 domains of the *Tth* RNAP are represented as ribbons under a transparent yellow surface. CarD binds to the solvent exposed surface of the β1 domain, ~70 Å away from the catalytic center. **B)** Superposition of the *Mtb* CarD/RNAP complex structure with the *Taq* RNAP IC structure (PDB ID: 1L9Z). The molecular surface of the *Taq* RNAP is gray except for the β1-β2 domains, which are represented as a transparent yellow surface. The DNA duplex is colored dark blue. The flexible α1 and eight-residue loop of CarD are cyan. CarD's DNA interacting patches as determined by EMSA are green. The direction of transcription is also indicated. The CarD C-terminal DNA interaction domain lies in proximity to the downstream end of the dsDNA in the initiation complex (~ 40 Å).

The intermolecular interface is also stabilized by electrostatic, hydrophobic, and van der Waals interactions **(Figure B-2)**. In fact, electrostatic forces contribute significantly to the CarD/RNAP interaction because altering the local charge distribution at the interface was reported to abolish CarD/RNAP interaction completely (Weiss et al., 2012). A more detailed analysis of the intermolecular contacts is provided in **Table 6-2**.

In contrast to the structural model generated by homology modeling (based on *Tth* TRCF-RID/β1 structure), mutagenesis and two-hybrid assays (Weiss et al., 2012), which suggested that β1-Glu132 interacts with both Arg25 and Arg47 directly through hydrogen bonding and that these residues are critical for intermolecular interaction, we observed from the *Mtb* CarD/RNAP structure that β1-Glu132 is not in direct contact with CarD-Arg25 and CarD-Arg47 (5.0 Å and 6.1 Å, respectively). Arg25 interacts with β1-Ile141 only through van der Waals interactions and does not appear crucial for CarD/RNAP interaction. Similarly, Glu132 and Arg47 interact only through a water molecule in the CarD/RNAP crystal structure, and Arg47 is engaged in other hydrogen bonding and van der Waals interactions with β1-Ile141, β1-Glu140, and β1-Gly139 **(Table 6-2)**. Therefore, loss of the CarD/RNAP interaction, as suggested by two-hybrid assays, upon E132R, R25E, and R47E mutations, should be due to these factors rather than the disruption of the direct interaction between Glu132-Arg25 and Glu132-Arg47.

Table 6-2. Details of the intermolecular interactions between *Mtb* β1 domain residues and *Mtb* CarD. See also Figure B-4.

RNAP β1	CarD	Distance (Å)	Interaction type
Thr138	Asn52	2.8	H-bond and van der Waals
	Pro49	3.5	van der Waals
	Val56	3.9	van der Waals
Gly139	Pro49	3.5	van der Waals
	Arg47	3.5	van der Waals
Glu140	Arg47	3.4	van der Waals
	Val48	3.8	van der Waals
	Tyr11	2.4	H-bond and van der Waals
	Val56	3.6	van der Waals
Ile141	Arg47	2.8	H-bond and van der Waals
	Thr45	3.4	van der Waals
	Val46	3.2	van der Waals, hydrophobic
	Arg25	3.9	van der Waals
Lys142	Thr45	3.3	van der Waals
	His14	2.9	H-bond and van der Waals
Ser143	Thr45	2.8	H-bond and van der Waals
	Leu44	3.6	van der Waals
Gln144	Leu44	3.8	van der Waals
	Asp43	3.8	van der Waals
	Gly42	3.0	H-bond and van der Waals
Glu404	His14	3.8	van der Waals
	His13	3.4	van der Waals
Ala405	His13	3.7	van der Waals
	His14	3.7	van der Waals
Thr407	His14	3.3	van der Waals

Comparison of *Mtb* CarD/RNAP and *Tth* TRCF-RID/β1 complex structures reveals that CarD and TRCF-RID display a similar set of interactions with RNAP, even though there is no sequence conservation between the CarD β4- ([43]DLTVRVP[49]) and TRCF β4- ([358]EGKLYLP[364]) strands that interact with the RNAP β1 domain (except for the last proline residues; **Figure B-4A**). CarD residues Tyr11, His13, and His14, located on the turn between β1 and β2 strands, also interact with the RNAP-β1 domain, which was not observed in the TRCF-RID/β1 structure. We have tested the contribution of Y11 and H14 to the CarD/RNAP interaction by generating CarD-Y11A-H14A mutant, and comparing the thermal stability of CarD-Y11A-H14A/RpoBtr and CarD-wt/RpoBtr complexes by ThermoFluor (DSF) experiments. The thermal denaturation profiles suggested that the CarD-Y11A-H14A/RpoBtr complex is less stable than the CarD-wt/RpoBtr complex ($Tm_{C11A-14A} = 37.9 \pm 0.1$ °C versus $Tm_{Cwt} = 39.2 \pm 0.1$ °C; **Figure B-4B**), which was also supported by the size exclusion chromatography (data not shown) consistent with our structure that these residues are involved in CarD/RNAP interaction. On the other hand, the salt bridge interaction observed in the TRCF-RID/β1 structure between residues RNAP Glu110 and TRCF Tyr362 and Arg341 (Westblade et al., 2010) is not present in CarD/RNAP structure.

DNA binding studies on **Mtb** *CarD*

Mtb CarD is classified as a CdnL protein due to the lack of a DNA-binding motif in its protein sequence. It has been proposed that CdnL proteins do not interact with DNA directly (Garcia-Moreno et al., 2010). We have tested whether *Mtb* CarD can

interact with DNA by electrophoretic mobility shift assay (EMSA).

Because CarD is required for stringent response in mycobacteria, we tested CarD binding to ribosomal protein and rRNA operons. The 200–300 base pairs (bp) upstream of the *rpsH*, 16S, 23S, and 5S rRNA genes were amplified for gel shift assays. Our results showed a clear shift of electrophoretic mobility between the free DNA and the CarD-bound DNA for these probes (**Figure 6-5A**). The gel shift assays done with various random DNA probes, as well as DNase footprinting experiments (data not shown), suggested that *Mtb* CarD does not show a sequence preference for DNA binding, indicating a nonspecific DNA-binding mode.

To elucidate the *Mtb* CarD-DNA interaction further, four different N- and C-terminally truncated CarD variants ($CarD_{61-162}$, $CarD_{83-162}$, $CarD_{1-53}$, $CarD_{1-74}$) were cloned and expressed to test each domain's DNA-binding activity. $CarD_{61-162}$ and $CarD_{83-162}$ contain the α-helical C-terminal domain and exhibited a gel shift, although with different mobilities, which could be due to the charge, size, and shape differences of the two constructs. EMSA results verified that the C-terminal domain is the DNA interaction domain (**Figure 6-5B**). The N-terminal domain is not involved in DNA interaction and is required only for RNAP interaction as observed from the CarD/RNAP structure.
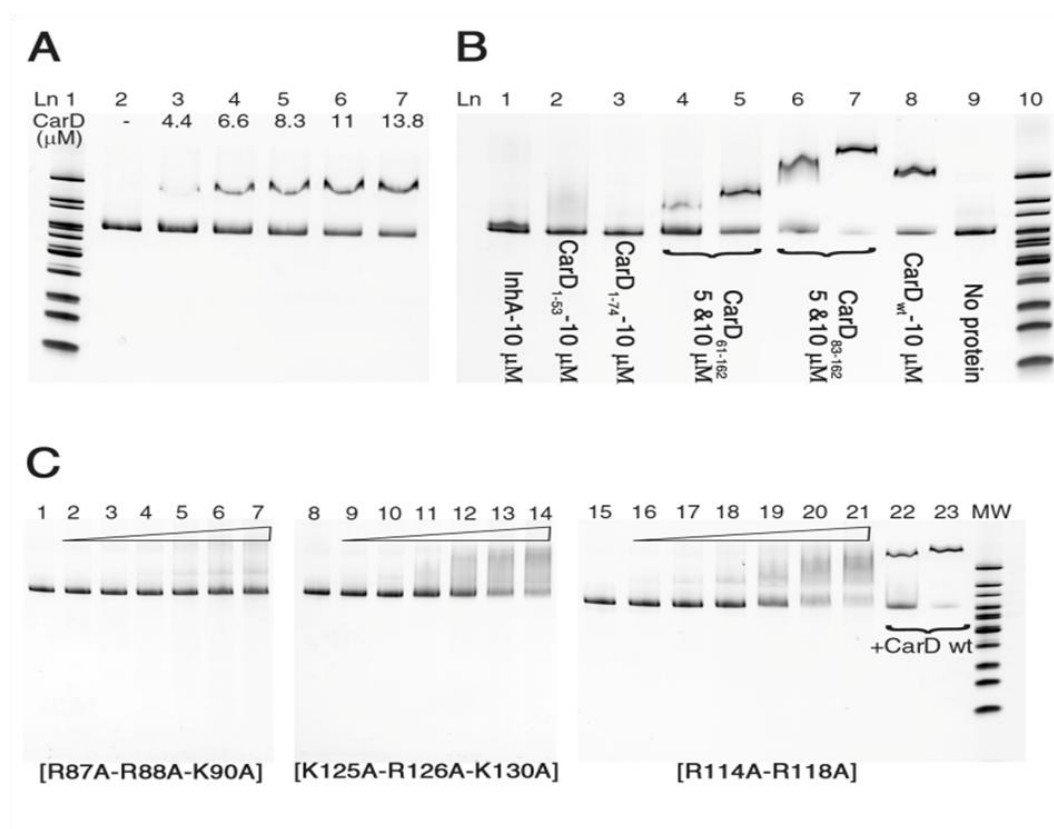
Figure 6-5. DNA binding activity of *Mtb* CarD determined by EMSA. A) CarD interaction with upstream DNA of 16S rRNA gene (313 bp). Lane 1: MW (molecular weight) marker. Lane 2: DNA probe, no protein. Lanes 3–7: 4.4–13.8 µM CarD. **B)** Interaction of CarD domains and a well-known non-DNA-binding protein with 16S rRNA upstream DNA probe. Lane 1: InhA (10 µM). Lane 2: $CarD_{1-53}$ (10 µM). Lane 3: $CarD_{1-74}$ (10 µM). Lanes 4 and 5: $CarD_{61-162}$ (5 and 10 µM). Lanes 6 and 7: $CarD_{83-162}$ (5 and 10 µM). Lane 8: CarD full length (10 µM). Lane 9: 16S rRNA upstream DNA probe, no protein. Lane 10: MW marker. **C)** EMSA experiments with CarD mutant proteins. Mutation of Arg and Lys residues to Ala significantly reduced the DNA-binding activity of *Mtb* CarD. R87A-R88A-K90A showed the greatest effect. Lanes 1, 8, 15: 16S rRNA upstream DNA probe, no protein. Lanes 2–7, 9–14, and 16–21: 0–44 µM mutant CarD protein (as labeled on the gel). Lanes 22 and 23: 11 and 22 µM native CarD. **(D and E)** Mutations are mapped on the ribbon representation and electrostatic potential surface of CarD. R87-R88-K90 are red, K125-R126-K130 are dark orange, and R114-R118 are light orange. Gel imaging was done using the Bio-Rad Chemidoc XRS+ molecular imager, by excitation at 255 nm and emission at 520 nm. Electrostatic potential surface calculations were done with PyMol (The PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger) using APBS as the macromolecular electrostatics calculation program (Baker et al., 2001).
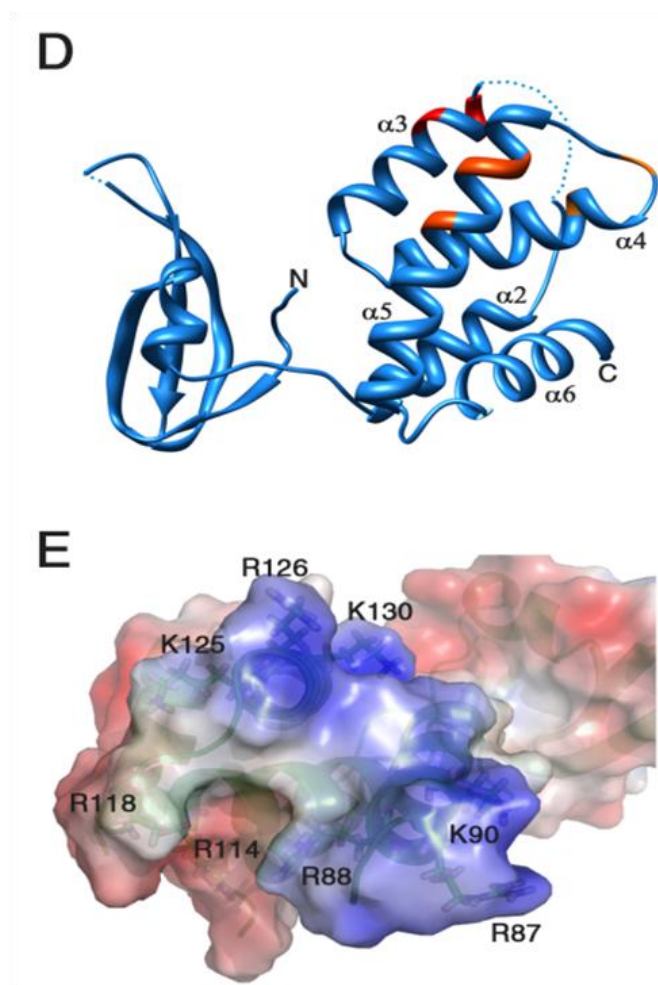
Figure 6-5 Continued.

Electrostatic potential surface calculations on the *Mtb* CarD structure revealed a single positively charged patch in the C-terminal domain formed by helices α3, α4, and α5 **(Figures 6-5D and 6-5E)**. The basic residues contributing to this positively charged surface, i.e., Arg87-Arg88-Lys90 on α3, Arg114-Arg118 on α4, and Lys125-Arg126-Lys130 on α5, were mutated to alanine and subjected to EMSA. As anticipated, mutation of all the aforementioned Arg and Lys residues to Ala significantly reduced the DNA-

binding activity of CarD, with the R87A-R88A-K90A mutation located on the solvent exposed surface of α3, showing the greatest effect **(Figure 6-5C)**, suggesting that the CarD-DNA interaction is mainly electrostatically driven, as expected.

Sequence independent DNA-binding modes are commonly seen in bacterial nucleoid associated proteins, which are involved in chromosome compaction and structuring, DNA replication, repair, and transcription (Basu et al., 2009). The *M. xanthus* CarD protein, which has affinity for AT-rich DNA sequences, and the *M. xanthus* CdnL protein, which does not have a DNA-binding sequence motif, were both localized to the nucleoid, but this localization was proposed to occur through protein-protein interactions with RNAP (Elias-Arnanz et al., 2010; Garcia-Moreno et al., 2010). It is plausible that *Mtb* CarD also localizes to the nucleoid in the same manner as the *M. xanthus* CdnL and *M. xanthus* CarD proteins, but considering the sequence-independent DNA-binding activity, we suggest that this localization might be provided by the DNA-binding ability of CarD rather than by associating and tailing with RNAP. The C-terminal domain, and thereby the DNA-binding activity of *Mtb* CarD, is crucial for mycobacterial viability because CarD depletion cannot be complemented with the RID domain alone (Weiss et al., 2012).

Conformational changes in RNAP upon CarD binding

The conformation of the β1 and β2 domains observed in the CarD/RNAP complex differs from the conformations observed in the uncomplexed RpoBtr structure. Superposition of the uncomplexed RpoBtr_A and RpoBtr_B β1 domains with the

CarD/RpoBtr β1 domain structure gives an rmsd of 6.8 Å and 2.8 Å, respectively, over the C_αs of the β2 domains **(Figure 6-3B)**. In this context, the conformation of RpoBtr in complex with CarD is closer to the conformation of the uncomplexed RpoBtr_B molecule. It was reported in the TRCF/β1 structure that the RNAP β4 strand undergoes a ''register shift'' with respect to the β3 strand in the complex structure (Westblade et al., 2010). In contrast, CarD does not cause a register shift or conformational rearrangement in the β4 strand upon RNAP binding.

The CarD/RNAP β1 domain interaction causes local conformational changes primarily in the nearby RNAP side chains that are propagated through the water-mediated network of interactions and transferred to the β1-β2 domain interface and β2 domain residues. In particular, in the CarD/RNAP complex, β1-Glu404, Ser143, and Glu140 change conformation to interact with CarD-His13, Arg47, Thr45, and Tyr11, respectively **(Figure 6-3C)**. The side chain of β1-Lys142 also moves 1.4 Å and loses direct H-bonding interaction with β1-Ile406, instead forming hydrogen bonds with CarD-His14 and β1-Glu140. Consecutively, the β1 domain residues Glu396 and Arg392, located at the β1-β2 domain interface of RpoBtr, adopt different conformations in CarD/RNAP complex and make additional water mediated interactions with the β2 domain residues Pro277 and Gly278 **(Figure 6-3D)**. This can explain the particular conformation adopted by the two domains in the complex structure.

RNAP regulation by CarD

The CarD/RNAP structure indicates that CarD and the functional homolog DksA

regulate RNAP through different mechanisms. DksA is proposed to bind to the RNAP secondary channel, very close to the active site, to coordinate to a (p)ppGpp-bound active site Mg^{2+} ion through its coiled-coil Asp residues, and stabilize the (p)ppGpp-RNAP complex (Perederina et al., 2004). In contrast, CarD interacts with the β1 domain of the β-subunit, approximately 70 Å away from the active site, through its Tudor-fold N-terminal domain. It is interesting that even though CarD and DksA do not share sequence and structural homology, CarD can complement DksA function in a Δ*dksA E. coli* strain (Stallings et al., 2009). Furthermore, DksA is not a DNA-binding protein, whereas we have shown that CarD can interact with DNA. Whether CarD functions synergistically with (p)ppGpp the same way as DksA needs to be determined experimentally.

It is not known whether CarD regulates RNAP function during transcription initiation or elongation, or has any effect on the rate of transcription. Based on our structural data, we propose that CarD might be involved in RNAP regulation in three different ways. The first is by inducing conformational changes in the β-lobes and affecting the open complex stability and the downstream non-specific DNA-binding activity of RNAP. This can explain how *Mtb* CarD can complement DksA, which destabilizes the open complex together with (p)ppGpp during stringent response in *E. coli* (Paul et al., 2004a; Stallings et al., 2009). Bacterial RNAP β1 and β2 domains (equivalently eukaryotic RNAP II protrusion and lobe domains) are involved in various processes during transcription such as downstream DNA binding and selection of the transcription initiation site, formation and stabilization of the open complex, maintaining

250

the proper transcription bubble via downstream DNA gripping, keeping the template and nontemplate strand-separated DNA in place during transcription initiation, and covering the DNA/RNA hybrid inside the RNAP active-site channel (**Figures 6-4A and 6-4B)** (Lane and Darst, 2010; Murakami et al., 2002a; Nechaev et al., 2000; Trautinger and Lloyd, 2002; Trinh et al., 2006). Therefore, conformational changes in this region may likely alter critical interactions of RNAP with DNA and DNA/RNA hybrid. To test if CarD interaction with the β1 domain would affect RNAP's DNA-binding affinity, we compared the nonspecific DNA-binding activity of RpoBtr in both the uncomplexed form and in complex with CarD by EMSA. Our results suggest that the CarD/RpoBtr complex has a higher affinity than the uncomplexed RpoBtr for the same DNA probe **(Figure 6-6)**. We propose that CarD might affect the DNA-binding affinity of the β-lobes, and the affinity change of RpoBtr for DNA may result primarily from the conformational changes of the β-lobes induced by CarD interaction.

Overlay of the *Taq* RNAP initiation complex and CarD/RNAP complex structures show that the CarD C-terminal DNA interaction domain lies in proximity (<40 Å) to the downstream end of dsDNA in the initiation complex **(Figure 6-4B)**, suggesting that CarD may interact with the promoter DNA together with RNAP during initiation. The CarD C-terminal domain is connected to the CarD N-terminal domain by a twisted α1 helix and a short loop and may adopt a different relative conformation in solution than the one observed in the crystal structure. Therefore, a second possibility is that this domain can either function as an anchor on DNA to hold CarD in place and strengthen the CarD/ RNAP interaction or it can have a direct functional role in

transcription regulation such as promoter selection and binding. The role of the C-terminal domain on CarD function is currently under investigation.
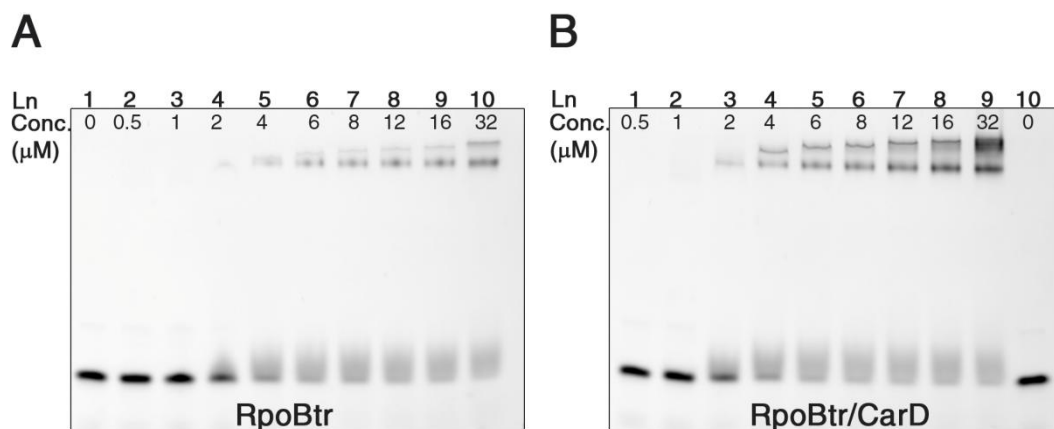
Figure 6-6. Comparison of the DNA binding activity of RpoBtr by EMSA. A) Uncomplexed form. Lane 1: dsDNA probe, no protein. Lanes 2–10: 0.5–32 μM RpoBtr. **B)** In complex with CarD. The CarD/RpoBtr complex has higher affinity than the uncomplexed RpoBtr for the same DNA probe. Lanes 1–9: 0.5–32 μM RpoBtr/CarD complex. Lane 10: dsDNA probe, no protein.

Another possible mechanism is allosteric regulation by inducing conformational changes around the RNAP active site despite the distance of the CarD binding site from the RNAP catalytic center. In fact, mutations at the interface that weaken the CarD/RNAP interaction were reported to make *Mtb* more susceptible to rifampicin (Rif), which binds to the RNAP active site and inhibits transcription elongation. This suggests that CarD/RNAP interaction is able to induce conformational changes not only in the

β1-β2 domains but also in the Rif binding pocket, causing Rif to bind more weakly to RNAP. The clinically isolated Rif-resistant *Mtb* strains carrying mutations on distant β residues, such as Val170, which do not interact directly with Rif, affect the conformation of the Rif binding pocket, and alter the affinity of RNAP for the drug (Campbell et al., 2001). Similarly, CarD interaction with the β-lobes may result in a reduced affinity of RNAP for Rif. A more complete understanding of the effect of CarD on RNAP both structurally and functionally must await the solution of the full-length RNAP/CarD structure.

The CarD/RNAP structure presented here reveals the molecular basis of this protein-protein interaction and provides insights into RNAP regulation by CarD. EMSA experiments revealed an unexpected DNA-binding activity for *Mtb* CarD, which is required for complete *in vivo* function and mycobacterial viability, and is provided by a distinct domain not associated with RNAP interaction. Determination of the CarD/DNA complex and RNAP/CarD/DNA ternary complex crystal structures is needed to further characterize the transcriptional regulation by CarD.

Experimental procedures

Generation of expression constructs and cloning

Rv3583c, encoding the *Mtb* CarD protein, and DNA encoding the *Mtb* RNAP β-subunit (*Rv0667*) residues 47–433 (labeled RpoBtr), were amplified from *Mtb* H37Rv genomic DNA by PCR. The genes were inserted into pET15b and pET30b (Novagen)

expression vectors using the NdeI-BamHI and NdeI-HindIII restriction sites. The pET15b construct contained an N-terminal 6X-His tag and labeled RpoBtr-NHis, while the pET30b construct had a stop codon at the end of the gene sequence, generating an untagged protein (CarD-notag). DNA encoding the full length and truncated CarD proteins ($CarD_{1-74}$, $CarD_{1-53}$, $CarD_{61-162}$ and $CarD_{83-162}$) were amplified from *Mtb* H37Rv genomic DNA by PCR with the NdeI-HindIII restriction sites and inserted into the pET28b vector (Novagen). CarD_R87A-R88A-K90A, CarD_R114A-R118A, and CarD_K125A-R126A-K130A plasmids were generated using a site directed mutagenesis kit (Stratagene). The sequence of each construct was verified by DNA sequencing. The primers used in this study are provided in **Table B-1**.

Expression and purification

Expression plasmids for the uncomplexed RpoBtr, native, and mutant CarD proteins were transformed to *E. coli* BL21(DE3) cells, and recombinant protein expression was induced with 1 mM IPTG. For co-expression of *Mtb* RNAP β1-β2 domains and *Mtb* CarD, the plasmids RpoBtr-NHis and CarD-notag were cotransformed into *E. coli* Rosetta2(DE3)pLysS cells, and expression was induced with 0.75 mM IPTG. Proteins were extracted with French press and purified by metal affinity and size exclusion chromatography. The RpoBtr-NHis:CarD complex eluted as a single peak from the size exclusion column. Co-elution of RpoBtr and CarD from the IMAC and size-exclusion columns was verified by SDS-PAGE. Finally, the purified proteins were

concentrated to 10 mg/ml and stored at -80°C for further use. Details of the expression and purification process are provided in Appendix B.

Crystallization

RpoBtr-NHis-SeMet and native protein crystals were obtained using hanging-drop vapor diffusion method by incubating 2 µl of purified protein solution with 2 µl of crystallization solution (0.1 M MgCl$_2$, 0.1 M HEPES [pH 7.5], 10% [w/v] PEG4000, and 0.2 M potassium citrate tribasic monohydrate, and 20% [w/v] PEG3350, respectively) at 16°C. The RpoBtr-NHis:CarD complex was crystallized by mixing 2 µl of protein solution with 2 µl of mother liquor (2% [v/v] tacsimate [pH 5.0], 0.1 M sodium citrate tribasic dihydrate [pH 5.6], and 14% [w/v] PEG3350) by hanging-drop vapor diffusion. Crystals were cryo-protected with 20% (v/v) ethylene glycol and flash-frozen prior to data collection. Data were collected at the Advanced Light Source (ALS; Lawrence Berkeley National Laboratory) and at the Advanced Photon Source (APS beamlines 23ID and 19ID; Argonne National Laboratory) at 0.979 Å.

Data collection and structure determination

The structure of the *Mtb* RNAP β-subunit β1-β2 domains was solved by single-wavelength anomalous diffraction (SAD) using RpoBtr-NHis-SeMet crystals. Crystals belonged to the P2$_1$2$_1$2$_1$ space group and diffracted to 2.9 Å. Subsequently, resolution was improved to 2.5 Å by diffraction data obtained from native (non-SeMet) RpoBtr-NHis crystals. Refinement and iterative manual model building was performed with

Phenix (Adams et al., 2010) and COOT (Emsley et al., 2010), and the final model had R_{work} and R_{free} values of 0.21 and 0.26, respectively.

The RpoBtr-NHis:CarD complex crystals belonged to the $C222_1$ space group and the diffraction data to 2.1 Å resolution was processed with Denzo/Scalepack. The structure was solved by MR using the RpoBtr β1 and β2 domains as two different search ensembles (Phaser, CCP4) (McCoy et al., 2007; Winn et al., 2011). After locating one copy of RpoBtr in the ASU, CarD was built into the additional $|F_o|-|F_c|$ density manually. The final model included one RpoBtr:CarD complex in the ASU, and the structure was refined with Phenix Refine to a R_{work} = 0.20 and an R_{free} = 0.23. Data collection and processing statistics are provided in **Table 6-1**. Details of the data collection and structure determination are provided in Appendix B.

EMSAs

For EMSA, DNA 200–300 bp upstream of the *rpsH*, 16S, 23S, and 5S rRNA promoters were amplified from H37Rv genomic DNA with PCR (for primers, see Table **S1**) and purified by gel-extraction. At room temperature, 40 ng of double-stranded DNA was incubated with different amounts of protein (0–8 μg) for 30 min in 25 mM Tris (pH 7.5) and 50 mM NaCl. As a negative control, a known non-DNA-binding protein, enoyl-ACP reductase InhA from *Mtb*, was used to confirm that binding of *Mtb* CarD to DNA is protein-specific. The mixture was loaded on a precast 10% nondenaturing polyacryl-amide gel and the gel was run at a constant voltage (120 V) with prechilled 0.5X TBE (89 mM Tris base, 89 mM boric acid, 1 mM EDTA [pH 8.0]) buffer at 4°C. After the run

was completed, the gel was stained with 1X Syber green (Invitrogen) DNA stain solution for 30 min in the dark and imaged (Jing et al., 2003).

ThermoFluor measurements

Differential scanning fluorimetry experiments were carried out with 1 μM of CarD/RpoBtr or CarD-Y11A-H14A/RpoBtr complex in 200 mM Tris (pH 7.5) 100 mM NaCl buffer, in the presence of 5X Sypro orange dye (Molecular Probes), in a 20 μl reaction volume. The temperature of the samples was changed from 25°C to 95°C at a heating rate of 0.5°C/min, and the fluorescence was monitored with an Mx3005P qPCR instrument (Agilent). The melting point (Tm) was calculated as the lowest point of the first derivative plot (DeSantis et al., 2012).

Additional unpublished results

Degradation of CarD protein in solution

It was observed that CarD protein (with N-terminal His-tag or without any tag) underwent proteolytic cleavage over time in solution **(Figure 6-7)**. Addition of EDTA, protease inhibitor cocktail or PMSF did not help in preventing the truncation of the protein. To eliminate any possible protease contamination coming along the purification, the protein was passed through high-salt washes and several columns including gel filtration and benzamidine FF columns. Even though 99% pure protein was obtained, the proteolytic cleavage still persisted.
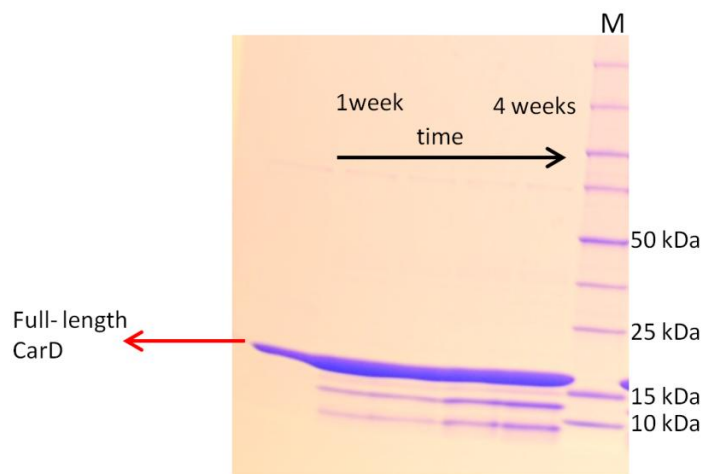
Figure 6-7. Degradation of the CarD protein in solution over time. Purified CarD (full length- 18 kDa) was incubated in 25 mM Tris 7.5, 150 mM NaCl buffer at 18 °C for four weeks. At the end of each week, a sample was aliquoted for SDS-PAGE. Consistent with the mass-spectrometry analysis, the truncation products at around 8 and 14 kDa were observed. As judged from the SDS-PAGE, CarD is >98% pure and further purifications did not resolve the truncation issue. Addition of protease inhibitors or EDTA also did not prevent the degradation.

The degradation of the protein was analyzed by mass spectrometry. The ESI-FT mass analysis resulted in seven major peaks indicating that CarD was truncated at three different aminoacid positions; i.e., Leu37, Phe71, and Leu74 **(Figure 6-8)**, which suggested a chymotrypsin like cleavage. The single, double and triple alanine mutants of the protein at these residues (L37A, F71A-L74A, L37A-F71A-L74A) were generated to increase the protein stability; however, none of the mutant proteins prevented the truncation of the protein (data not shown). In parallel with these observations, a similar truncation behavior for the *Mtb* CarD protein was reported in a recent publication. In those experiments, only a C-terminally His-tagged CarD construct was observed to be stable in solution and yielded full length protein crystals. The crystal structure revealed that the C-terminal His-tag induced dimerization through the N-terminal domain of the

protein and prevented the proteolysis (Kaur et al., 2013).
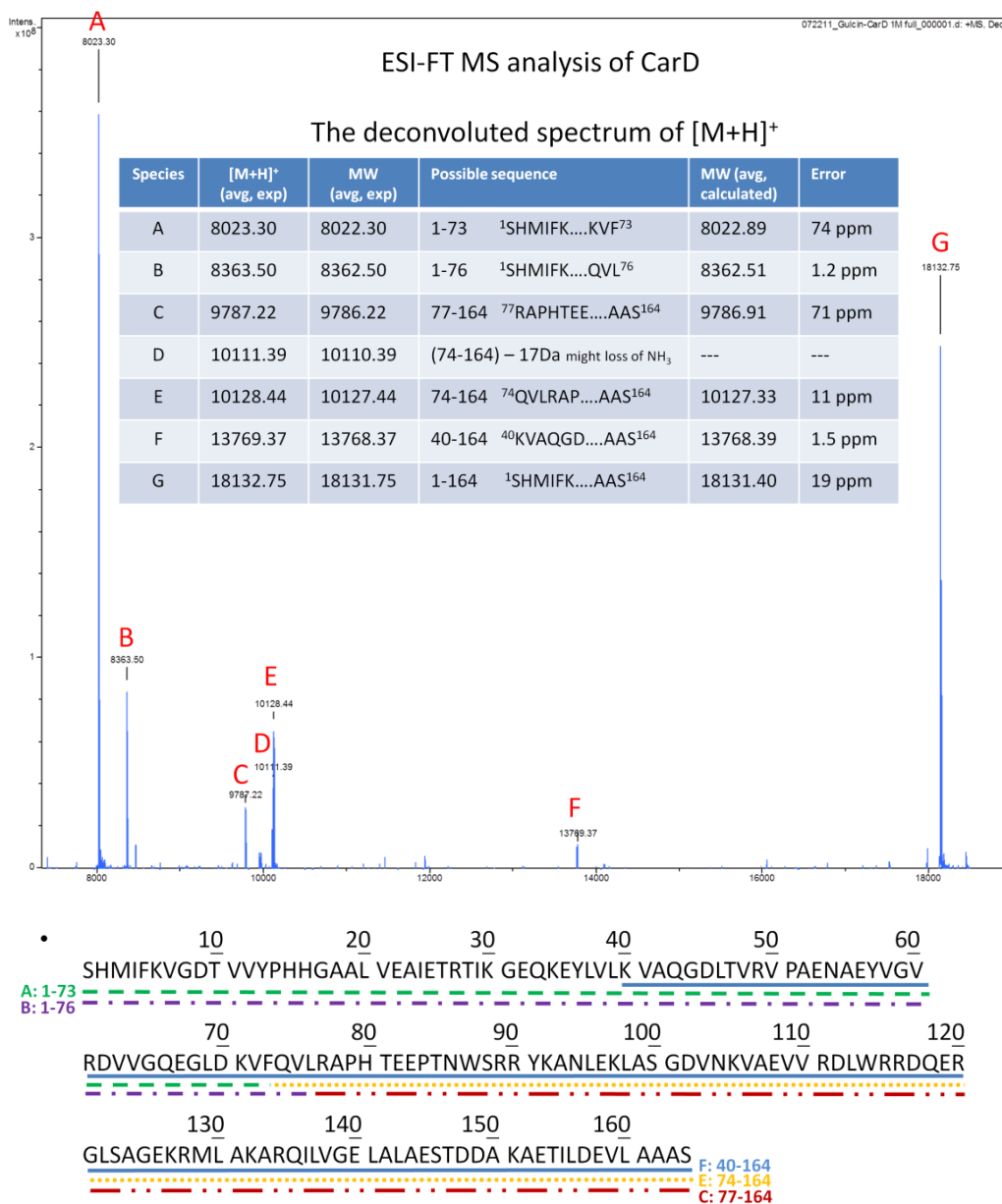
Figure 6-8. ESI-FT mass spectrometry analysis of the CarD protein. As concluded from the spectrum, seven major peptides were produced as a result of truncation. Below, the peptide masses were mapped to the protein sequence, showing the corresponding region for each peak (A-F). CarD got truncated at three different aminoacid positions, i.e., Leu37, Phe71 and Leu74.

Structure determination of uncomplexed CarD

Despite several attempts, the crystals for the full length protein could not be obtained due to the proteolytic cleavage of CarD over time in solution. None of the single, double or triple mutant proteins (L37A, F71A-L74A, and L37A-F71A-L74A) yielded full length protein crystals. Crystallization trials set with the full length protein resulted in formation of the truncated CarD crystals, which covered only the C-terminal of the protein (residues 82 -141). Mass spectrometry analysis (MSWIFT) also confirmed that these crystals were the truncation products with molecular weights of 10 and 14 kDa **(Figure 6-9C)**. Interestingly, complex formation between the *Mtb* CarD and RNAP proteins prevented the proteolysis of CarD, and hence, the full length CarD bound to the β-subunit of RNAP was crystallized after co-expression and purification of the complex (as described in the results and discussion section of this chapter). Compared with the uncomplexed CarD protein, CarD complexed with RpoBtr was stable in solution, and could be stored without any degradation.

The truncated CarD crystals were obtained in the presence of nucleating agents like seaweed, hydroxyapatite and bioglass. Wild-type and Se-met derivatized fine hair-like crystals were obtained in 0.2 M trisodium citrate monohydrate and 20% (w/v) Peg3350 in 2 to 3 weeks **(Figure 6-9B)**. The crystals belonged to the $P6_1$ space group with unit cell dimensions of 67.6, 67.6, 33.7 Å, containing one molecule in the asymmetric unit. Diffraction data was collected at Argonne National Laboratory Advanced Photon Source (APS) at 0.97 Å to a resolution of 2.0 Å. The structure was solved by single wavelength anomalous dispersion (SAD) method, using the diffraction

data obtained from the Se-Met crystals. Model building and structure refinement was done with Coot and Phenix Refine, respectively, yielding a final model with R_{work}=22% and R_{free}=27% (**Table 6-3**).
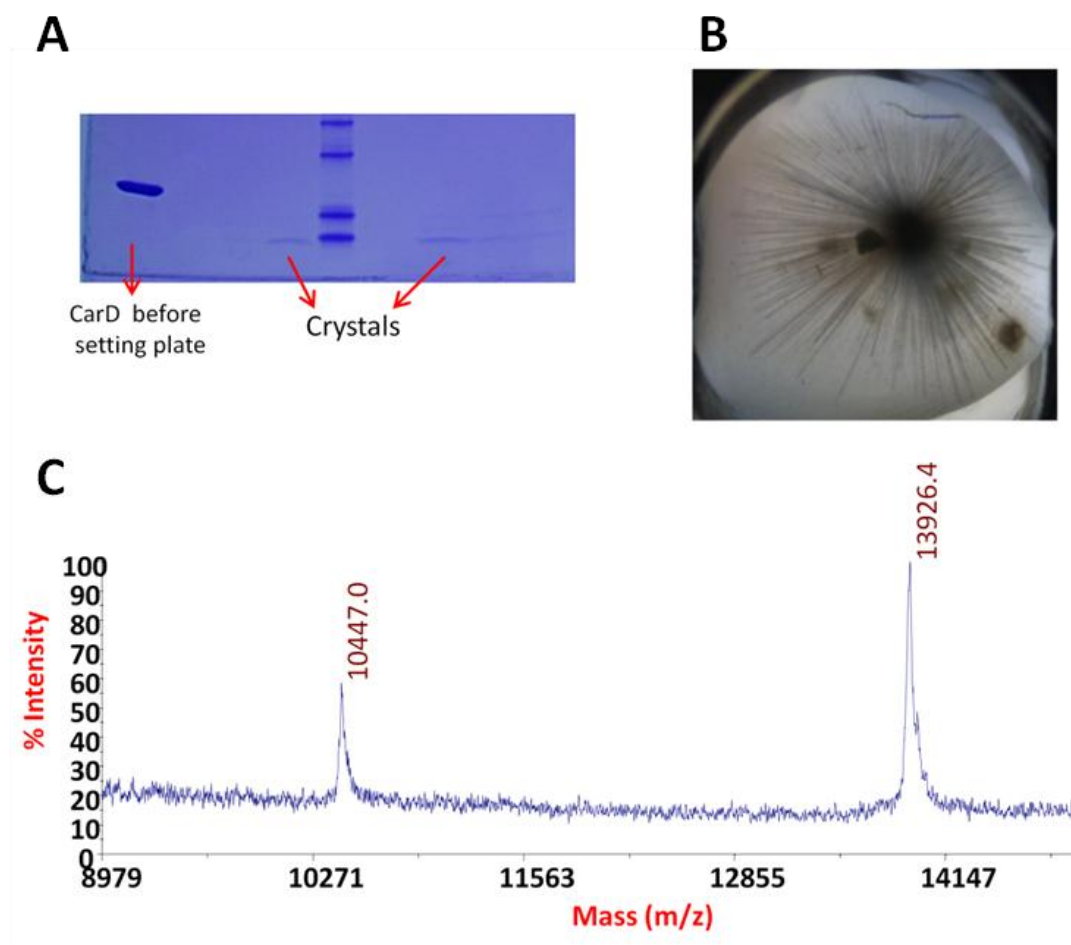
Figure 6-9. Truncated CarD crystals. A) SDS-PAGE of the CarD protein before setting the crystal plate (lane 1) and the truncated CarD crystals obtained after two weeks (lanes 4-9). **B)** Fine hair-like crystals of the truncated CarD protein. Crystals were obtained in the presence of seaweed, which can be seen next to the crystals as the brownish spots. **C)** MSWIFT analysis of the crystals verified that they were truncation products of the CarD protein with molecular weights of 10 and 14 kDa.

Table 6-3. Data collection, phasing and refinement statistics for truncated CarD structure.

	Se-Met CarD C-term. Truncation
Data collection	
Space group	$P6_1$
Cell dimensions	
a, b, c (Å)	67.6, 67.6, 33.7
α, β, γ (^{0})	90, 90, 120
Wavelength	0.979
Resolution (Å)	1.95 (1.95)
Completeness (%)	99.6 (99.9)
Redundancy	6.6 (6.6)
I/Iσ	13.0 (2.6)
R_{sym}	0.39 (0.04)
Refinement	
Resolution	58.5-1.95
No. reflections	6266
R_{work}/ R_{free}	0.22/0.27
No. atoms	
Protein	491
B factors	36.7
RMSD deviations	
Bond distances (Å)	0.022
Bond angles (^{0})	1.96

Highest resolution shell values are given in parenthesis.

No significant conformational differences were observed between the protease truncated CarD C-terminal and the full-length CarD C-terminal domain structures. The structure of the truncated CarD C-terminal domain (residues 82-141) was comprised of three α-helices with two helix-turn-helix (HTH) motifs **(Figure 6-10)**. This three-helix bundle covered α3, α4, and α5 (numbered according to the full length structure), and comprised the leucine zipper motif between the α4 and α5. Also, it represented the DNA interaction domain of the protein as described in the results and discussion section of this chapter.

Characterization of the CarD-DNA interaction

The CarD-DNA interaction was examined extensively by using several DNA probes with different sequence, length, shape and secondary structures. It was observed that CarD had similar binding affinity to the dsDNAs generated from the 200-300 bp upstream of the *rpsH*, 16S, 23S and 5S rRNA genes. Next, to narrow down the specific DNA region needed for the CarD interaction, the 16S, 23S and 5S rRNA upstream dsDNAs were digested at a single restriction site with the restriction enzymes, BsrI or NheI, generating pieces of DNA varying in length from 76 bp to 237 bp **(Figure 6-11)**. These restrictedly digested fragments were used in EMSA experiments as the DNA probes. *Mtb* CarD interacted with each restricted fragment, showing a slightly higher affinity for the longest (237 bp) fragment. In fact, it was observed that longer the length of the dsDNA, the higher was the affinity of CarD.

263
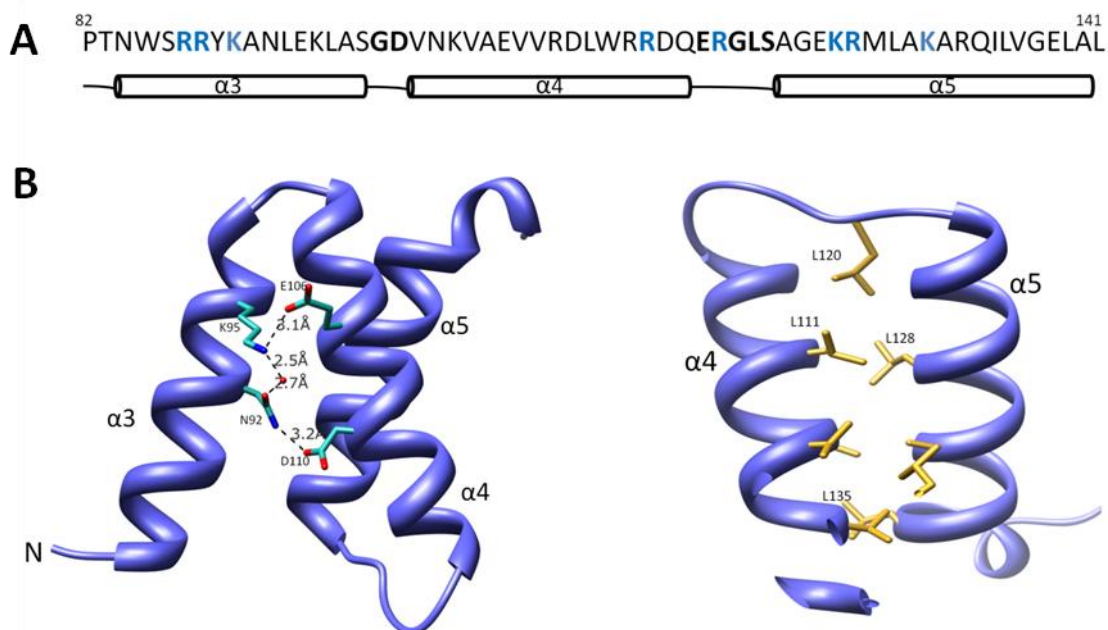
Figure 6-10. Structure of the truncated CarD protein. A) The amino acid sequence of the *Mtb* CarD C-terminus that was crystallized is shown. Positively charged residues that were tested for DNA interaction are colored blue. The secondary structure elements are presented below the sequence. **B)** Cartoon representation of the structure of the three helix bundle of CarD C-terminal domain. Only two polar interactions were present between K95-E106 and N92-D110. On the right, the leucine zipper formed by the side chains of the residues Leu120, Leu111, L128, and L135 on the α4 and α5 is shown. α3 is omitted from the structure for clarity.
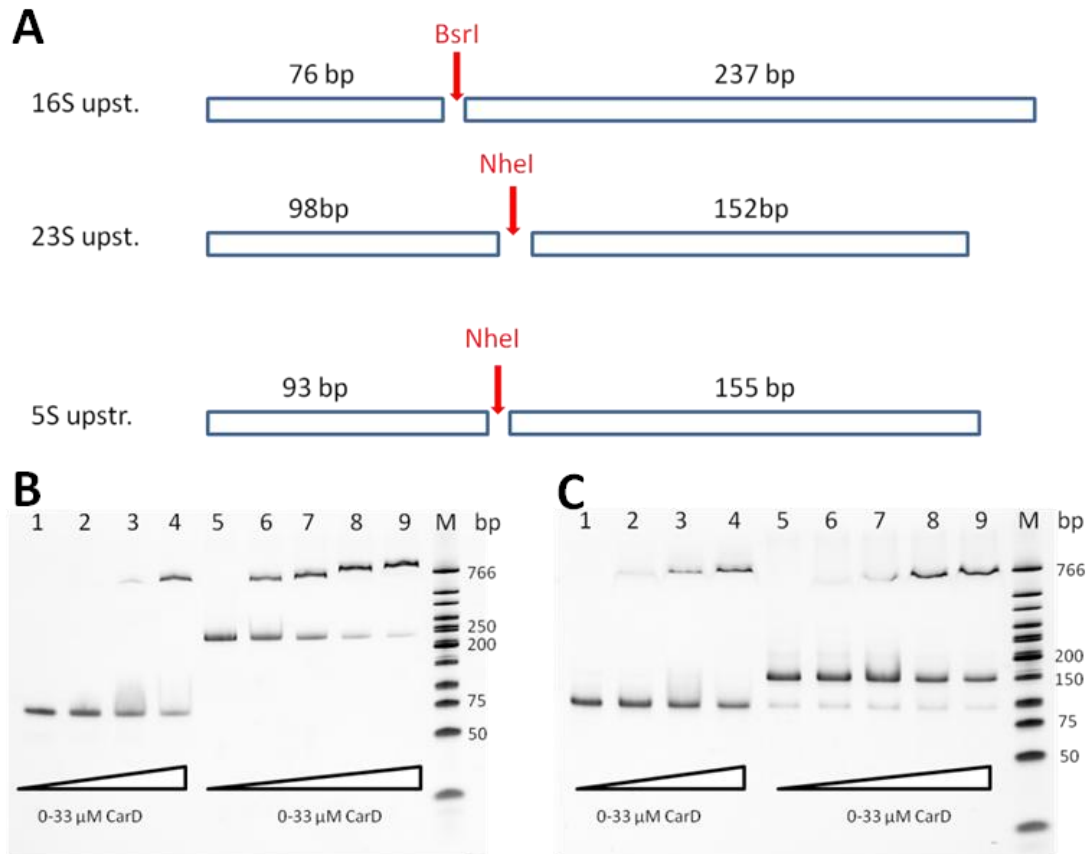
Figure 6-11. EMSA experiments with the restrictedly digested 16S, 23S and 5S rRNA upstream DNA probes. The 16S, 23S and 5S rRNA upstream dsDNAs were digested at a single site with the restriction enzymes, BsrI or NheI, generating pieces of DNA varying in length from 76 bp to 237 bp. CarD binding to each fragment was tested. **A)** A diagram showing the restriction digestion map of the 16S, 23S and 5S rRNA upstream dsDNA sequences. **B)** EMSA with the BsrI digested 16S rRNA upstream DNA probe. Lanes 1-4: 76 bp fragment + 0-33 μM CarD. Lanes 5-9: 237 bp fragment + 0-33 μM CarD. **C)** EMSA with the NheI digested 23S rRNA upstream DNA probe. Lanes 1-4: 98 bp fragment + 0-33 μM CarD. Lanes 5-9: 152 bp fragment + 0-33 μM CarD. M: Molecular marker.

Intriguingly, there was no conserved sequence motif in the DNA sequences tested for the EMSA experiments leading us to suspect the sequence specificity of the *Mtb* CarD-DNA interaction. The DNA binding specificity of *Mtb* CarD was further tested by using randomly generated linear dsDNA sequences of similar length that were

not related to the stringent response. EMSA experiments demonstrated that *Mtb* CarD was also able bind to these non-specific DNA sequences, supporting the idea of sequence independent DNA-binding **(Figure 6-12A)**. CarD was even able to bind to the dsDNA sequences of the molecular markers **(Figure 6-12A)**, confirming the non-sequence specific binding mode. Moreover, the CarD-DNA interaction was strongly dependent on salt concentration, indicating mainly electrostatic interactions rather than the base specific interactions are present **(Figure 6-12B)**. In addition, the CarD-DNA interaction was not affected by the reaction buffer pH (pH range 5.0-9.0) (data not shown), supporting the involvement of the arginine or lysine residues in DNA interaction. At the tested pHs, Arg and Lys residues would be positively charged due to their high pKa values and can contribute to electrostatic interactions with the negatively charged DNA backbone.

Next, it was investigated if the DNA binding activity of CarD was dependent on the secondary and tertiary structure of the DNA probe. Cruciform DNA and bulged DNA probes were used in EMSA experiments to test if CarD could interact with DNA in different conformations. *Mtb* CarD did not show any binding to these probes. Furthermore, CarD did not show any binding to the single stranded DNAs (data not shown). Consequently, based on the EMSA results it was concluded that the DNA binding mode of *Mtb* CarD was not sequence dependent and the protein was able to bind to any linear duplex DNA sequence longer than approximately 80 bp with similar affinity.
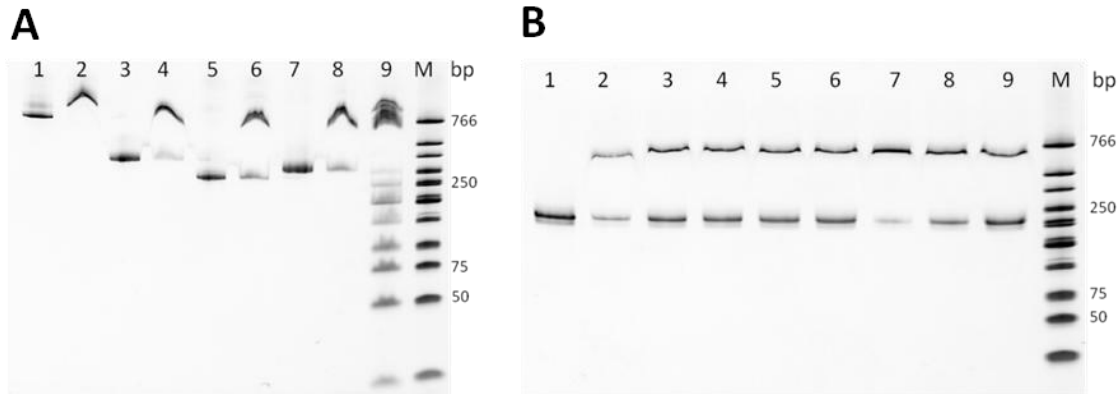
Figure 6-12. Non-specific DNA binding, and salt and buffer effect on the DNA binding. A) EMSA experiments performed with the DNA probes in different sizes that are not related to the stringent response. Lanes 1-2: *Mtb* MurI gene sequence (816 bp) + 0-11 μM CarD. Lanes 3-4: *Mtb* WhiB6 gene sequence (351 bp) + 0-11 μM CarD. Lanes 5-6: *Mtb* Rv1993 gene sequence (273 bp) + 0-11 μM CarD. Lanes 7-8: 16S rRNA upstream dsDNA probe (313 bp) + 0-11 μM CarD. Lane 9: molecular marker + 11 μM CarD. Lane 10: molecular marker. **B)** pH and salt dependence of the CarD-DNA interaction. Lane 1: 23S upstream DNA probe. Lanes 2-6: 23S upstream DNA probe + 11 μM CarD in acetate pH 5.0, Tris pH 6.8, Tris pH 7.5, Tris pH 8.5 and Tris pH 9.0, respectively. Lanes 7-9: 23S upstream DNA probe + 11 μM CarD in 0, 100 and 500 mM NaCl, respectively, in the reaction buffer.

Complex formation between CarD and RNAP

Before we started the complex formation trials of *Mtb* CarD and RNAP, there was a single report in the literature suggesting that CarD interacts with the N-terminus of the RNAP β-subunit (Stallings et al., 2009). A subsequent report on the *Tth* TRCF-RID/RNAP β1 structure had predicted that the *Mtb* CarD/RNAP interaction region would also be located on the β1 domain of the β-subunit, between the conserved residues 123 to 153 of RNAP (Westblade et al., 2010). Even though *Mtb* RNAP displays ~50% identity and ~70% similarity to the *E.coli* and *Taq* RNAPs, whose structures have been solved (Murakami et al., 2002a; Opalka et al., 2010), there is no reported structure of the

full length *Mtb* RNAP or its β subunit in the PDB till date. It is well accepted that it is difficult to crystallize large proteins or protein-protein complexes due to the presence of high number of flexible domains and loops in these macromolecules (Dale et al., 2003). Therefore, to facilitate crystallization and increase the probability of getting high resolution diffracting crystals of the complex, we focused on the RpoB domains that were needed for the complex formation with *Mtb* CarD. Different constructs were designed for the β-subunit covering the predicted interaction region based on the secondary structure predictions and homology models generated using the published *Taq* and *Tth* RNAP structures. The homology models suggested that the first 40 residues of *Mtb* RpoB would be disordered without any secondary structure, and this disordered region could significantly lower the chances of obtaining high-resolution diffracting crystals. Therefore, the residue 40 was selected as the N-terminal starting point for the β-subunit constructs. Four different constructs of the β-subunit spanning the residues 40-403, 47-174, 47-433, and 1-1172 (full length β-subunit) were cloned and tested for soluble recombinant protein expression.

Mtb CarD and the β-subunit variants were co-expressed in *E. coli* to investigate the complex formation between the two proteins. In order to facilitate the purification of the complex, a co-expression system was selected that allowed us to put an N-terminal His$_6$-tag on one of the protein constructs, while keeping the other protein tagless. The CarD-no-tag plasmid was co-transformed with the RpoB-N-His plasmids in *E. coli* Rosetta2(DE3)pLysS cells. Separately, RpoB-no-tag plasmids were co-transformed with the CarD-N-His plasmid. This allowed us to examine *in vivo* complex formation

between the proteins by analysis of the co-elution profiles from the IMAC purification on SDS-PAGE gels. Among these co-expression systems, CarD-no-tag co-expression with the RpoB-tr-N-His (47-433) construct showed that both proteins co-eluted from the IMAC with gradient elution. The band corresponding to CarD was verified by mass spectrometry and SDS-PAGE gel sequencing. To confirm that the co-elution of the two proteins represented a stable complex, size-exclusion chromatography was performed. Both proteins co-eluted from the gel-filtration column giving a single peak for the complex **(Figure 6-13)**. Even though the full length RpoB was able to associate with CarD *in vivo* as judged by the IMAC results, this association was found to be weak since both of the proteins eluted under separate peaks from the gel-filtration column (data not shown). The RpoB-47-174 construct did not give any soluble protein expression, while the RpoB-40-403 construct dissociated from CarD on the gel-filtration column similar to the full length RpoB construct. As the primary objective of this study was to determine the crystal structure of the complex, the stable complex formed between the RpoB-tr-N-His (47-433) and CarD-no-tag constructs (RpoBtr:CarD complex) was selected for crystallographic studies.

Conversely, there was no complex formation upon co-expression of the CarD-NHis and RpoBtr-notag constructs, which suggested that the N-terminal His_6-tag on CarD interfered with the protein-protein interaction.
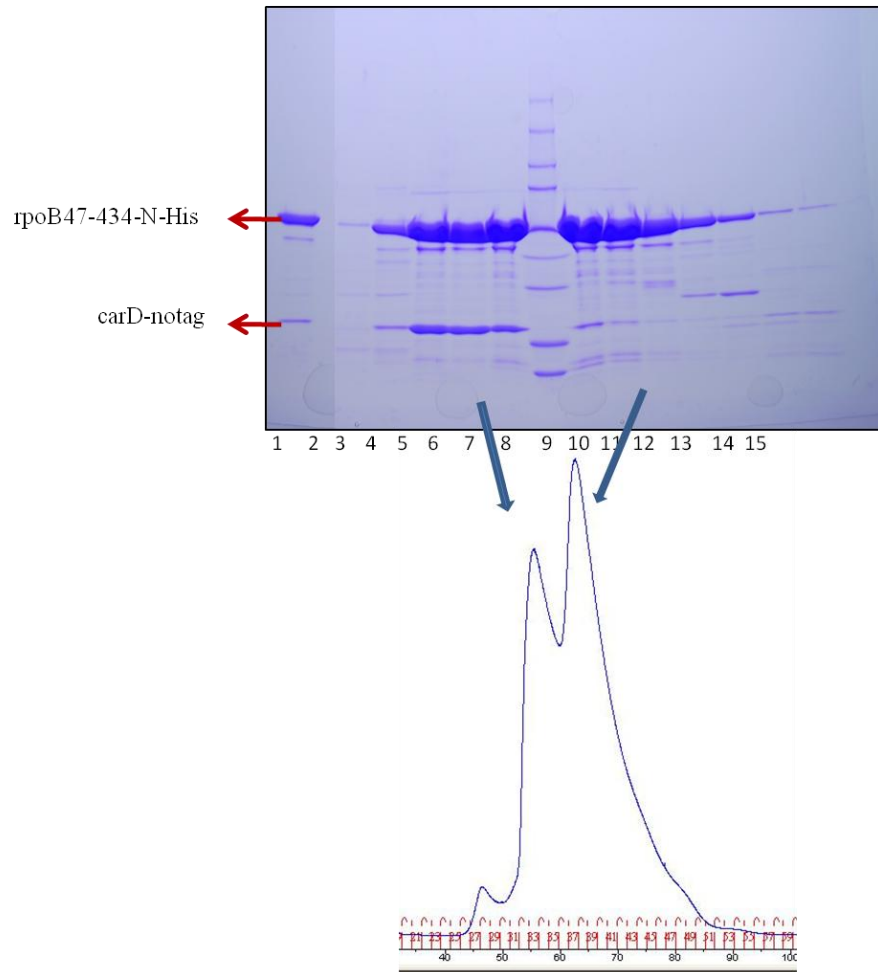
Figure 6-13. Co-purification of the CarD-RNAP complex. After co-elution from the IMAC, fractions were combined, concentrated and loaded to the gel-filtration column (lane 1). Both proteins co-eluted in a single peak (lanes 5-8). Uncomplexed RpoBtr eluted later than the CarD/RpoBtr complex as expected (lanes 11-15). The elution chromatogram is shown below the SDS-PAGE gel.

270

CHAPTER VII

CONCLUSIONS

Tuberculosis is caused by the bacterium *Mycobacterium tuberculosis*, and continues to be a global health threat. It is estimated that one-third of the world's population is infected by latent *Mtb*. Emergence of the multi-drug resistant *Mtb* strains along with the ability of the bacterium to persist in host for years, despite the use of antibiotics, exacerbates the problem and threatens the efforts to keep TB under control. Current TB therapy continues for at least six months and requires administration of at least four drugs, which were discovered decades ago. Furthermore, co-infection with HIV raises the mortality rates to close to 100% in the case of extensively-drug resistant TB. Therefore, it is incredibly urgent to discover new anti-tubercular agents that are effective against drug-resistant and persistent bacteria.

Chapter I of this dissertation introduced the facts about TB disease, provided a brief explanation of TB drugs, current TB therapy, drug-resistance, and persistence, as it aimed to further our understanding of these concepts. The important components of the mycolic acid biosynthesis pathway, which represent valuable drug targets, were described in detail. Also, the transcriptional response and regulation of *Mtb* during adaptation to persistent state was examined in order to understand the factors required for the establishment of this state. Lastly, an overview of transcription and regulation of transcription by RNAP interacting proteins of *Mtb* was given.

Chapter II described the mechanism of drug action of isoniazid and ethionamide against *Mtb* and *M. leprae*. By developing a cell-based activation method, the active forms of ethionamide and prothionamide were isolated and characterized by mass-spectrometry and X-ray crystallography. It was found that isoniazid and ethionamide targeted and inhibited InhA in a similar manner as both formed adducts with the cofactor NAD^+. The crystal structures of InhA complexed with the adducts revealed the binding mode of these prodrugs. Using the same cell-based activation method, it was shown that InhA and not Dhfr was the primary cellular target of isoniazid. Absence of mutations in the *dfrA* gene in isoniazid resistant clinical isolates and the fact that overexpression of Dhfr did not cause any resistance to isoniazid confirmed that InhA was the clinically relevant target of isoniazid, bringing an end to the controversies about the target of this anti-TB drug.

The biochemical and structural investigations on regulation of the InhA enzyme was addressed in Chapter III. Since InhA is an essential protein and is a well validated drug target, deciphering its post-translational regulation is quite important to understand the regulation of mycolic acid production in mycobacteria. It was found that InhA function was regulated negatively by phosphorylation. *In vitro* phosphorylation assays indicated that InhA was phosphorylated by multiple *Mtb* serine-threonine kinases at a single threonine residue (T266) close to the C-terminal of the protein. Enzymatic assays and fluorescence quenching experiments done with the phosphomimetic aspartate and glutamate mutants suggested that the decreased enzymatic activity was mainly due to the decreased affinity of the enzyme for NADH. The crystal structures of the wild-type and

phosphomimetic InhA proteins revealed that the mutation of T266 to aspartate or glutamate caused local structural changes by perturbing the H-bonding network and introducing water molecules. It was noteworthy that even though T266 was 15 Å away from the active site and not involved in catalysis, mutation of this residue caused significant activity loss. In addition, transfer of the T266D or T266E InhA alleles to mycobacteria was lethal confirming the required InhA activity for cell viability.

Chapters IV and V covered the drug discovery studies for InhA. A target based high-throughput screening (HTS) was performed in collaboration with GlaxoSmithKline (GSK), and a structure based drug design study on triclosan scaffold was performed in collaboration with the Jacobus Pharmaceutical Company that led to the discovery of highly potent InhA inhibitors. The HTS at GSK resulted in the identification of multiple hits belonging to different chemical series. Among these, the thiadiazole series were the most potent and featured the best pharmacokinetic properties; therefore, these series of compounds were further studied. The optimized GSK lead compound was a potent inhibitor of InhA with an IC_{50} of 7 nM and was effective against drug-sensitive and drug-resistant clinical strains. It also showed better pharmacokinetic properties than the mother compound. In addition, the broad spectrum antibacterial compound triclosan was used as a template for structure-based drug design. Triclosan derivatives modified at the 5-position were tested *in vitro* and *in vivo* for anti-tubercular activity. It was seen that hydrophobic alkyl and aryl substituents increased the potency of the derivatives toward InhA, with the most potent compound having an IC_{50} of 21 nM. Triclosan derivatives were also active against the drug-resistant *Mtb* strains. The binding mode of the GSK

inhibitors and triclosan analogs were determined by X-ray crystallography. It was seen that both sets of inhibitors bound to the hydrophobic substrate binding pocket of the enzyme interacting with NAD^+ and the protein backbone. In regard to the GSK inhibitors, InhA residue Met98 was involved in direct H-bonding with the inhibitor compound, whereas for triclosan derivatives Tyr158 made an H-bond with the inhibitor compounds. It is noteworthy that both GSK inhibitors and triclosan derivatives are activation-free inhibitors unlike isoniazid, which *Mtb* develops a majority of the drug resistance against due to the mutations in the activator gene *katG*.

Chapter VI discussed the structural and functional studies on CarD, which is an *Mtb* transcription regulator that is essential for bacterial viability and the establishment of persistence. CarD is involved in stringent response through downregulation of rRNA and ribosomal protein genes, and it interacts with the β-subunit of RNAP. The crystal structure of the CarD/RNAP complex revealed the molecular basis of this important protein-protein interaction and suggested possible mechanisms CarD can employ on RNAP for transcription regulation. Furthermore, it was found that CarD could bind to dsDNA through its C-terminal α-helical domain.

REFERENCES

Abubakar, I., Zignol, M., Falzon, D., Raviglione, M., Ditiu, L., Masham, S., Adetifa, I., Ford, N., Cox, H., Lawn, S.D.*, et al.* (2013). Drug-resistant tuberculosis: time for visionary political leadership. Lancet Infect Dis *13*, 529-539.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Ainsa, J.A., Martin, C., Cabeza, M., De la Cruz, F., and Mendiola, M.V. (1996). Construction of a family of Mycobacterium/Escherichia coli shuttle vectors derived from pAL5000 and pACYC184: their use for cloning an antibiotic-resistance gene from Mycobacterium fortuitum. Gene *176*, 23-26.

Alahari, A., Trivelli, X., Guerardel, Y., Dover, L.G., Besra, G.S., Sacchettini, J.C., Reynolds, R.C., Coxon, G.D., and Kremer, L. (2007). Thiacetazone, an antitubercular drug that inhibits cyclopropanation of cell wall mycolic acids in mycobacteria. PLoS One *2*, e1343.

am Ende, C.W., Knudson, S.E., Liu, N., Childs, J., Sullivan, T.J., Boyne, M., Xu, H., Gegina, Y., Knudson, D.L., Johnson, F.*, et al.* (2008). Synthesis and in vitro antimycobacterial activity of B-ring modified diaryl ether InhA inhibitors. Bioorg Med Chem Lett *18*, 3029-3033.

Anand, R.S., Somasundaram, S., Doble, M., and Paramasivan, C.N. (2011). Docking studies on novel analogues of 8 methoxy fluoroquinolones against GyrA mutants of Mycobacterium tuberculosis. BMC Struct Biol *11*, 47.

Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H.W., Neefs, J.M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E.*, et al.* (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. Science *307*, 223-227.

Argyrou, A., Jin, L., Siconilfi-Baez, L., Angeletti, R.H., and Blanchard, J.S. (2006a). Proteome-wide profiling of isoniazid targets in Mycobacterium tuberculosis. Biochemistry *45*, 13947-13953.

Argyrou, A., Vetting, M.W., Aladegbami, B., and Blanchard, J.S. (2006b). Mycobacterium tuberculosis dihydrofolate reductase is a target for isoniazid. Nat Struct Mol Biol *13*, 408-413.

Av-Gay, Y., and Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol *8*, 238-244.

275

Avarbock, D., Salem, J., Li, L.S., Wang, Z.M., and Rubin, H. (1999). Cloning and characterization of a bifunctional RelA/SpoT homologue from Mycobacterium tuberculosis. Gene *233*, 261-269.

Bacon, J., and Marsh, P.D. (2007). Transcriptional responses of Mycobacterium tuberculosis exposed to adverse conditions in vitro. Curr Mol Med *7*, 277-286.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., Collins, D., de Lisle, G., and Jacobs, W.R., Jr. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science *263*, 227-230.

Banerjee, A., Sugantino, M., Sacchettini, J.C., and Jacobs, W.R., Jr. (1998). The mabA gene from the inhA operon of Mycobacterium tuberculosis encodes a 3-ketoacyl reductase that fails to confer isoniazid resistance. Microbiology *144 (Pt 10)*, 2697-2704.

Bardarov, S., Bardarov Jr, S., Jr., Pavelka Jr, M.S., Jr., Sambandamurthy, V., Larsen, M., Tufariello, J., Chan, J., Hatfull, G., and Jacobs Jr, W.R., Jr. (2002). Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. Microbiology *148*, 3007-3017.

Barthe, P., Mukamolova, G.V., Roumestand, C., and Cohen-Gonsaud, M. (2010). The structure of PknB extracellular PASTA domain from mycobacterium tuberculosis suggests a ligand-dependent kinase activation. Structure *18*, 606-615.

Barthe, P., Roumestand, C., Canova, M.J., Kremer, L., Hurard, C., Molle, V., and Cohen-Gonsaud, M. (2009). Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism. Structure *17*, 568-578.

Bartos, M., Falkinham, J.O., and Pavlik, I. (2004). Mycobacterial catalases, peroxidases, and superoxide dismutases and their effects on virulence and isoniazid-susceptibility in mycobacteria - a review. Veterinarni Medicina *49*, 161-170.

Basu, D., Khare, G., Singh, S., Tyagi, A., Khosla, S., and Mande, S.C. (2009). A novel nucleoid-associated protein of Mycobacterium tuberculosis is a sequence homolog of GroEL. Nucleic Acids Res *37*, 4944-4954.

Baulard, A.R., Betts, J.C., Engohang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., Locht, C., and Besra, G.S. (2000). Activation of the pro-drug ethionamide is regulated in mycobacteria. J Biol Chem *275*, 28326-28331.

276

Belardinelli, J.M., and Morbidoni, H.R. (2012). Mutations in the essential FAS II beta-hydroxyacyl ACP dehydratase complex confer resistance to thiacetazone in Mycobacterium tuberculosis and Mycobacterium kansasii. Mol Microbiol *86*, 568-579.

Belardinelli, J.M., and Morbidoni, H.R. (2013). Recycling and refurbishing old antitubercular drugs: the encouraging case of inhibitors of mycolic acid biosynthesis. Expert Rev Anti Infect Ther *11*, 429-440.

Bergval, I.L., Schuitema, A.R., Klatser, P.R., and Anthony, R.M. (2009). Resistant mutants of Mycobacterium tuberculosis selected in vitro do not reflect the in vivo mechanism of isoniazid resistance. J Antimicrob Chemother *64*, 515-523.

Bermudez, L.E., Reynolds, R., Kolonoski, P., Aralar, P., Inderlied, C.B., and Young, L.S. (2003). Thiosemicarbazole (thiacetazone-like) compound with activity against Mycobacterium avium in mice. Antimicrob Agents Chemother *47*, 2685-2687.

Bernstein, J., Lott, W.A., Steinberg, B.A., and Yale, H.L. (1952). Chemotherapy of experimental tuberculosis. V. Isonicotinic acid hydrazide (nydrazid) and related compounds. Am Rev Tuberc *65*, 357-364.

Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., and Duncan, K. (2002). Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol *43*, 717-731.

Bhatt, A., Fujiwara, N., Bhatt, K., Gurcha, S.S., Kremer, L., Chen, B., Chan, J., Porcelli, S.A., Kobayashi, K., Besra, G.S.*, et al.* (2007a). Deletion of kasB in Mycobacterium tuberculosis causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. Proc Natl Acad Sci U S A *104*, 5157-5162.

Bhatt, A., Kremer, L., Dai, A.Z., Sacchettini, J.C., and Jacobs, W.R., Jr. (2005). Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. J Bacteriol *187*, 7596-7606.

Bhatt, A., Molle, V., Besra, G.S., Jacobs, W.R., Jr., and Kremer, L. (2007b). The Mycobacterium tuberculosis FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. Mol Microbiol *64*, 1442-1454.

Bissantz, C., Kuhn, B., and Stahl, M. (2010). A Medicinal Chemist's Guide to Molecular Interactions. J Med Chem *53*, 5061-5084.

Bloch, K. (1975). Fatty acid synthases from Mycobacterium phlei. Methods Enzymol *35*, 84-90.

Bloch, K. (1977). Control mechanisms for fatty acid synthesis in Mycobacterium smegmatis. Adv Enzymol Relat Areas Mol Biol *45*, 1-84.

Bold, T.D., and Ernst, J.D. (2009). Who benefits from granulomas, mycobacteria or host? Cell *136*, 17-19.

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

Boyne, M.E., Sullivan, T.J., amEnde, C.W., Lu, H., Gruppo, V., Heaslip, D., Amin, A.G., Chatterjee, D., Lenaerts, A., Tonge, P.J.*, et al.* (2007). Targeting fatty acid biosynthesis for the development of novel chemotherapeutics against Mycobacterium tuberculosis: evaluation of A-ring-modified diphenyl ethers as high-affinity InhA inhibitors. Antimicrob Agents Chemother *51*, 3562-3567.

Brennan, P.J., and Nikaido, H. (1995). The envelope of mycobacteria. Annu Rev Biochem *64*, 29-63.

Brossier, F., Veziris, N., Truffot-Pernot, C., Jarlier, V., and Sougakoff, W. (2011). Molecular investigation of resistance to the antituberculous drug ethionamide in multidrug-resistant clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother *55*, 355-360.

Brown, A.K., Bhatt, A., Singh, A., Saparia, E., Evans, A.F., and Besra, G.S. (2007). Identification of the dehydratase component of the mycobacterial mycolic acid-synthesizing fatty acid synthase-II complex. Microbiology *153*, 4166-4173.

Brown, A.K., Sridharan, S., Kremer, L., Lindenberg, S., Dover, L.G., Sacchettini, J.C., and Besra, G.S. (2005). Probing the mechanism of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III mtFabH: factors influencing catalysis and substrate specificity. J Biol Chem *280*, 32539-32547.

Brown, M.S., Akopiants, K., Resceck, D.M., McArthur, H.A., McCormick, E., and Reynolds, K.A. (2003). Biosynthetic origins of the natural product, thiolactomycin: a unique and selective inhibitor of type II dissociated fatty acid synthases. J Am Chem Soc *125*, 10166-10167.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Calgin, M.K., Sahin, F., Turegun, B., Gerceker, D., Atasever, M., Koksal, D., Karasartova, D., and Kiyan, M. (2013). Expression analysis of efflux pump genes among drug-susceptible and multidrug-resistant Mycobacterium tuberculosis clinical isolates and reference strains. Diagn Microbiol Infect Dis *76*, 291-297.

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

Campbell, P.J., Morlock, G.P., Sikes, R.D., Dalton, T.L., Metchock, B., Starks, A.M., Hooks, D.P., Cowan, L.S., Plikaytis, B.B., and Posey, J.E. (2011). Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis. Antimicrob Agents Chemother *55*, 2032-2041.

Canova, M.J., Kremer, L., and Molle, V. (2008). pETPhos: a customized expression vector designed for further characterization of Ser/Thr/Tyr protein kinases and their substrates. Plasmid *60*, 149-153.

Canova, M.J., Kremer, L., and Molle, V. (2009). The Mycobacterium tuberculosis GroEL1 chaperone is a substrate of Ser/Thr protein kinases. J Bacteriol *191*, 2876-2883.

CDC (2006). Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. MMWR Morb Mortal Wkly Rep *55*, 301-305.

Chan, J., Silver, R.F., Kampmann, B., and Wallis, R.S. (2005). Intracellular Models of Mycobacterium tuberculosis Infection. In Tuberculosis and the tubercle bacillus, S.T. Cole, K.D. Eisenach, D.N. McMurray, and J. William R. Jacobs, eds. (Washington, D.C.: ASM Press).

Chan, J.N., Nislow, C., and Emili, A. (2010). Recent advances and method development for drug target identification. Trends Pharmacol Sci *31*, 82-88.

Chen, T., He, L., Deng, W., and Xie, J. (2013). The Mycobacterium DosR regulon structure and diversity revealed by comparative genomic analysis. J Cell Biochem *114*, 1-6.

Chim, N., Habel, J.E., Johnston, J.M., Krieger, I., Miallau, L., Sankaranarayanan, R., Morse, R.P., Bruning, J., Swanson, S., Kim, H.*, et al.* (2011). The TB Structural Genomics Consortium: a decade of progress. Tuberculosis (Edinb) *91*, 155-172.

Choi, K.H., Kremer, L., Besra, G.S., and Rock, C.O. (2000). Identification and substrate specificity of beta -ketoacyl (acyl carrier protein) synthase III (mtFabH) from Mycobacterium tuberculosis. J Biol Chem *275*, 28201-28207.

Chorine, V. (1945). Action de l'amide nicotinique sur les bacilles du genre Mycobacterium. Comptes Rendus de l'Academie des Sciences *220*, 150-151.

Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual) (The Center for Macromolecular Design, Texas A&M University, College Station, TX).

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, *et al.* (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature *393*, 537-544.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Crofton, J., Chaulet, P., Maher, D., Grosset, J., Harris, W., Norman, H., Iseman, M., and Watt, B. (1997). Guidelines for the Management of Multidrug-Resistant Tuberculosis (Geneva, Switzerland World Health Organization).

Dahl, J.L., Kraus, C.N., Boshoff, H.I., Doan, B., Foley, K., Avarbock, D., Kaplan, G., Mizrahi, V., Rubin, H., and Barry, C.E., 3rd (2003). The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice. Proc Natl Acad Sci U S A *100*, 10026-10031.

Dale, G.E., Oefner, C., and D'Arcy, A. (2003). The protein as a variable in protein crystallization. J Struct Biol *142*, 88-97.

Dao, D.N., Sweeney, K., Hsu, T., Gurcha, S.S., Nascimento, I.P., Roshevsky, D., Besra, G.S., Chan, J., Porcelli, S.A., and Jacobs, W.R. (2008). Mycolic acid modification by the mmaA4 gene of M. tuberculosis modulates IL-12 production. PLoS Pathog *4*, e1000081.

Darnell, J.E., and Lodish, H.F. (2000). Molecular cell biology, 4th edn (New York: W.H. Freeman).

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

Darst, S.A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. Proc Natl Acad Sci U S A *99*, 4296-4301.

Davidson, P.T., and Le, H.Q. (1992). Drug treatment of tuberculosis--1992. Drugs *43*, 651-673.

Davis, J.M., and Ramakrishnan, L. (2009). The role of the granuloma in expansion and dissemination of early tuberculous infection. Cell *136*, 37-49.

DeBarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G., and Barry, C.E., 3rd (2000). Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *97*, 9677-9682.

DeSantis, K., Reed, A., Rahhal, R., and Reinking, J. (2012). Use of differential scanning fluorimetry as a high-throughput assay to identify nuclear receptor ligands. Nucl Recept Signal *10*, e002.

Dessen, A., Quemard, A., Blanchard, J.S., Jacobs, W.R., Jr., and Sacchettini, J.C. (1995). Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. Science *267*, 1638-1641.

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

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

Dias, M.V., Vasconcelos, I.B., Prado, A.M., Fadel, V., Basso, L.A., de Azevedo, W.F., Jr., and Santos, D.S. (2007). Crystallographic studies on the binding of isonicotinyl-NAD adduct to wild-type and isoniazid resistant 2-trans-enoyl-ACP (CoA) reductase from Mycobacterium tuberculosis. J Struct Biol *159*, 369-380.

Dover, L.G., Alahari, A., Gratraud, P., Gomes, J.M., Bhowruth, V., Reynolds, R.C., Besra, G.S., and Kremer, L. (2007). EthA, a common activator of thiocarbamide-containing drugs acting on different mycobacterial targets. Antimicrob Agents Chemother *51*, 1055-1063.

Dover, L.G., Alderwick, L., Bhowruth, V., Brown, A.K., Kremer, L., and Besra, G.S. (2008). Antibiotics and new inhibitors of the cell wall. In Mycobacterial Cell Envelope, M. Daffé, and J.M. Reyrat, eds. (Washington, DC: ASM Press), pp. 107-131.

Dubnau, E., Chan, J., Raynaud, C., Mohan, V.P., Laneelle, M.A., Yu, K., Quemard, A., Smith, I., and Daffe, M. (2000). Oxygenated mycolic acids are necessary for virulence of Mycobacterium tuberculosis in mice. Mol Microbiol *36*, 630-637.

Elias-Arnanz, M., Padmanabhan, S., and Murillo, F.J. (2010). The regulatory action of the myxobacterial CarD/CarG complex: a bacterial enhanceosome? FEMS Microbiol Rev *34*, 764-778.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G.S., Locht, C., and Baulard, A.R. (2004). EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. Mol Microbiol *51*, 175-188.

Fajardo, T.T., Guinto, R.S., Cellona, R.V., Abalos, R.M., Dela Cruz, E.C., and Gelber, R.H. (2006). A clinical trial of ethionamide and prothionamide for treatment of lepromatous leprosy. Am J Trop Med Hyg *74*, 457-461.

Fang, J.L., Stingley, R.L., Beland, F.A., Harrouk, W., Lumpkins, D.L., and Howard, P. (2010). Occurrence, efficacy, metabolism, and toxicity of triclosan. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev *28*, 147-171.

Fattorini, L., Iona, E., Ricci, M.L., Thoresen, O.F., Orru, G., Oggioni, M.R., Tortoli, E., Piersimoni, C., Chiaradonna, P., Tronci, M.*, et al.* (1999). Activity of 16 antimicrobial agents against drug-resistant strains of Mycobacterium tuberculosis. Microb Drug Resist *5*, 265-270.

Fenton, M.J., Vermeulen, M.W., Kim, S., Burdick, M., Strieter, R.M., and Kornfeld, H. (1997). Induction of gamma interferon production in human alveolar macrophages by Mycobacterium tuberculosis. Infect Immun *65*, 5149-5156.

Fiuza, M., Canova, M.J., Zanella-Cleon, I., Becchi, M., Cozzone, A.J., Mateos, L.M., Kremer, L., Gil, J.A., and Molle, V. (2008). From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of Corynebacterium glutamicum toward the role of PknA and PknB in cell division. J Biol Chem *283*, 18099-18112.

Flipo, M., Willand, N., Lecat-Guillet, N., Hounsou, C., Desroses, M., Leroux, F., Lens, Z., Villeret, V., Wohlkonig, A., Wintjens, R.*, et al.* (2012). Discovery of novel N-phenylphenoxyacetamide derivatives as EthR inhibitors and ethionamide boosters by combining high-throughput screening and synthesis. J Med Chem *55*, 6391-6402.

Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T.R., Sacchettini, J.C., Lipsitch, M.*, et al.* (2011). Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet *43*, 482-486.

Fraaije, M.W., Kamerbeek, N.M., Heidekamp, A.J., Fortin, R., and Janssen, D.B. (2004). The prodrug activator EtaA from Mycobacterium tuberculosis is a Baeyer-Villiger monooxygenase. J Biol Chem *279*, 3354-3360.

Freundlich, J.S., Wang, F., Tsai, H.C., Kuo, M., Shieh, H.M., Anderson, J.W., Nkrumah, L.J., Valderramos, J.C., Yu, M., Kumar, T.R.*, et al.* (2007). X-ray structural analysis of Plasmodium falciparum enoyl acyl carrier protein reductase as a pathway toward the optimization of triclosan antimalarial efficacy. J Biol Chem *282*, 25436-25444.

Freundlich, J.S., Wang, F., Vilcheze, C., Gulten, G., Langley, R., Schiehser, G.A., Jacobus, D.P., Jacobs, W.R., Jr., and Sacchettini, J.C. (2009). Triclosan derivatives: towards potent inhibitors of drug-sensitive and drug-resistant Mycobacterium tuberculosis. ChemMedChem *4*, 241-248.

Gallego-Garcia, A., Mirassou, Y., Elias-Arnanz, M., Padmanabhan, S., and Jimenez, M.A. (2012). NMR structure note: N-terminal domain of Thermus thermophilus CdnL. J Biomol NMR *53*, 355-363.

Gangjee, A., Kurup, S., and Namjoshi, O. (2007). Dihydrofolate reductase as a target for chemotherapy in parasites. Curr Pharm Des *13*, 609-639.

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

Geszvain, K., and Landick, R. (2005). The Structure of bacterial RNA polymerase. In The Bacterial Chromosome, N.P. Higgins, ed. (Washington, D.C.: ASM Press), pp. 283-296.

Gibrat, J.F., Madej, T., and Bryant, S.H. (1996). Surprising similarities in structure comparison. Curr Opin Struct Biol *6*, 377-385.

Gill, S.K., and Garcia, G.A. (2011). Rifamycin inhibition of WT and Rif-resistant Mycobacterium tuberculosis and Escherichia coli RNA polymerases in vitro. Tuberculosis (Edinb) *91*, 361-369.

Glickman, M.S., Cox, J.S., and Jacobs, W.R., Jr. (2000). A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. Mol Cell *5*, 717-727.

Godfrey, H.P., Bugrysheva, J.V., and Cabello, F.C. (2002). The role of the stringent response in the pathogenesis of bacterial infections. Trends Microbiol *10*, 349-351.

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

Gomez, J.E., and McKinney, J.D. (2004). M. tuberculosis persistence, latency, and drug tolerance. Tuberculosis (Edinb) *84*, 29-44.

Green, K.D., and Garneau-Tsodikova, S. (2013). Resistance in tuberculosis: what do we know and where can we go? Front Microbiol *4*, 208.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Gulten, G., and Sacchettini, J.C. (2013). 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 *21*, 1859-1869.

Guo, H., Seet, Q., Denkin, S., Parsons, L., and Zhang, Y. (2006). Molecular characterization of isoniazid-resistant clinical isolates of Mycobacterium tuberculosis from the USA. J Med Microbiol *55*, 1527-1531.

Gupta, R.K., Thakur, T.S., Desiraju, G.R., and Tyagi, J.S. (2009). Structure-based design of DevR inhibitor active against nonreplicating Mycobacterium tuberculosis. J Med Chem *52*, 6324-6334.

Gupta, S., and Chatterji, D. (2005). Stress responses in mycobacteria. IUBMB Life *57*, 149-159.

Hale, K.J., Hummersone, M.G., Manaviazar, S., and Frigerio, M. (2002). The chemistry and biology of the bryostatin antitumour macrolides. Nat Prod Rep *19*, 413-453.

Hartkoorn, R.C., Sala, C., Neres, J., Pojer, F., Magnet, S., Mukherjee, R., Uplekar, S., Boy-Rottger, S., Altmann, K.H., and Cole, S.T. (2012). Towards a new tuberculosis drug: pyridomycin - nature's isoniazid. EMBO Mol Med *4*, 1032-1042.

Hazbon, M.H., Brimacombe, M., Bobadilla del Valle, M., Cavatore, M., Guerrero, M.I., Varma-Basil, M., Billman-Jacobe, H., Lavender, C., Fyfe, J., Garcia-Garcia, L.*, et al.* (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother *50*, 2640-2649.

He, X., Alian, A., and Ortiz de Montellano, P.R. (2007). Inhibition of the Mycobacterium tuberculosis enoyl acyl carrier protein reductase InhA by arylamides. Bioorg Med Chem *15*, 6649-6658.

He, X., Alian, A., Stroud, R., and Ortiz de Montellano, P.R. (2006). Pyrrolidine carboxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from Mycobacterium tuberculosis. J Med Chem *49*, 6308-6323.

Heath, R.J., Su, N., Murphy, C.K., and Rock, C.O. (2000). The enoyl-[acyl-carrier-protein] reductases FabI and FabL from Bacillus subtilis. J Biol Chem *275*, 40128-40133.

Heath, R.J., Yu, Y.T., Shapiro, M.A., Olson, E., and Rock, C.O. (1998). Broad spectrum antimicrobial biocides target the FabI component of fatty acid synthesis. J Biol Chem *273*, 30316-30320.

Ho, Y.M., Sun, Y.J., Wong, S.Y., and Lee, A.S. (2009). Contribution of dfrA and inhA mutations to the detection of isoniazid-resistant Mycobacterium tuberculosis isolates. Antimicrob Agents Chemother *53*, 4010-4012.

Hoang, T.T., and Schweizer, H.P. (1999). Characterization of Pseudomonas aeruginosa enoyl-acyl carrier protein reductase (FabI): a target for the antimicrobial triclosan and its role in acylated homoserine lactone synthesis. J Bacteriol *181*, 5489-5497.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, W545-549.

Honer zu Bentrup, K., and Russell, D.G. (2001). Mycobacterial persistence: adaptation to a changing environment. Trends Microbiol *9*, 597-605.

Hopewell, P.C., and Jasmer, R.M. (2005). Overview of clinical Tuberculosis. In Tuberculosis and the Tubercle Bacillus, S.T. Cole, K.D. Eisenach, D.N. McMurray, and J. William R. Jacobs, eds. (Washington , DC: ASM Press).

Hu, Y., and Coates, A.R. (2011). Mycobacterium tuberculosis acg gene is required for growth and virulence in vivo. PLoS One *6*, e20958.

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

Ioerger, T.R., and Sacchettini, J.C. (2009). Structural genomics approach to drug discovery for Mycobacterium tuberculosis. Curr Opin Microbiol *12*, 318-325.

Jarlier, V., and Nikaido, H. (1994). Mycobacterial cell wall: structure and role in natural resistance to antibiotics. FEMS Microbiol Lett *123*, 11-18.

Jing, D., Agnew, J., Patton, W.F., Hendrickson, J., and Beechem, J.M. (2003). A sensitive two-color electrophoretic mobility shift assay for detecting both nucleic acids and protein in gels. Proteomics *3*, 1172-1180.

Katoch, K., Natarajan, M., Bhatia, A.S., and Yadav, V.S. (1992). Treatment of paucibacillary leprosy with a regimen containing rifampicin, dapsone and prothionamide. Indian J Lepr *64*, 303-312.

Kaufmann, S.H. (2001). How can immunology contribute to the control of tuberculosis? Nat Rev Immunol *1*, 20-30.

Kaur, G., Dutta, D., and Thakur, K.G. (2013). Crystal structure of Mycobacterium tuberculosis CarD, an essential RNA polymerase binding protein, reveals a quasidomain-swapped dimeric structural architecture. Proteins.

Khan, S., Nagarajan, S.N., Parikh, A., Samantaray, S., Singh, A., Kumar, D., Roy, R.P., Bhatt, A., and Nandicoori, V.K. (2010). Phosphorylation of enoyl-acyl carrier protein reductase InhA impacts mycobacterial growth and survival. J Biol Chem *285*, 37860-37871.

Kim, P., Zhang, Y.M., Shenoy, G., Nguyen, Q.A., Boshoff, H.I., Manjunatha, U.H., Goodwin, M.B., Lonsdale, J., Price, A.C., Miller, D.J*., et al.* (2006). Structure-activity relationships at the 5-position of thiolactomycin: an intact (5R)-isoprene unit is required for activity against the condensing enzymes from Mycobacterium tuberculosis and Escherichia coli. J Med Chem *49*, 159-171.

Kleanthous, C. (2000). Protein-protein recognition (Oxford, NY: Oxford University Press).

Koch, R. (1982). Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882. Rev Infect Dis *4*, 1270-1274.

Koff, W.C., Burton, D.R., Johnson, P.R., Walker, B.D., King, C.R., Nabel, G.J., Ahmed, R., Bhan, M.K., and Plotkin, S.A. (2013). Accelerating next-generation vaccine development for global disease prevention. Science *340*, 1232910.

Kohanski, M.A., Dwyer, D.J., and Collins, J.J. (2010). How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol *8*, 423-435.

Kompis, I.M., Islam, K., and Then, R.L. (2005). DNA and RNA synthesis: antifolates. Chem Rev *105*, 593-620.

Kondreddi, R.R., Jiricek, J., Rao, S.P., Lakshminarayana, S.B., Camacho, L.R., Rao, R., Herve, M., Bifani, P., Ma, N.L., Kuhen, K*., et al.* (2013). Design, Synthesis, and Biological Evaluation of Indole-2-carboxamides: A Promising Class of Antituberculosis Agents. J Med Chem *56*, 8849-8859.

Kremer, L., Baulard, A.R., and Besra, G.S. (2000a). Genetics of mycolic acid biosynthesis. In Molecular Genetics of Mycobacteria, W.R. Jacobs, Jr., and G.F. Hatfull, eds. (Washington, DC: ASM Press), pp. 173-190.

Kremer, L., Douglas, J.D., Baulard, A.R., Morehouse, C., Guy, M.R., Alland, D., Dover, L.G., Lakey, J.H., Jacobs, W.R., Jr., Brennan, P.J*., et al.* (2000b). Thiolactomycin and related analogues as novel anti-mycobacterial agents targeting KasA and KasB condensing enzymes in Mycobacterium tuberculosis. J Biol Chem *275*, 16857-16864.

Kremer, L., Dover, L.G., Carrere, S., Nampoothiri, K.M., Lesjean, S., Brown, A.K., Brennan, P.J., Minnikin, D.E., Locht, C., and Besra, G.S. (2002). Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from Mycobacterium tuberculosis. Biochem J *364*, 423-430.

Kremer, L., Dover, L.G., Morbidoni, H.R., Vilcheze, C., Maughan, W.N., Baulard, A., Tu, S.C., Honore, N., Deretic, V., Sacchettini, J.C*., et al.* (2003). Inhibition of InhA activity, but not KasA activity, induces formation of a KasA-containing complex in mycobacteria. J Biol Chem *278*, 20547-20554.

Krieger, I.V., Freundlich, J.S., Gawandi, V.B., Roberts, J.P., Sun, Q., Owen, J.L., Fraile, M.T., Huss, S.I., Lavandera, J.L., Ioerger, T.R*., et al.* (2012). Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. Chem Biol *19*, 1556-1567.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.

Kumar, A., Farhana, A., Guidry, L., Saini, V., Hondalus, M., and Steyn, A.J. (2011). Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert Rev Mol Med *13*, e39.

Kuo, M.R., Morbidoni, H.R., Alland, D., Sneddon, S.F., Gourlie, B.B., Staveski, M.M., Leonard, M., Gregory, J.S., Janjigian, A.D., Yee, C*., et al.* (2003). Targeting tuberculosis and malaria through inhibition of Enoyl reductase: compound activity and structural data. J Biol Chem *278*, 20851-20859.

Lacave, C., Laneelle, M.A., Daffe, M., Montrozier, H., and Laneelle, G. (1989). Mycolic acid metabolic filiation and location in Mycobacterium aurum and Mycobacterium phlei. Eur J Biochem *181*, 459-466.

Lacave, C., Laneelle, M.A., Daffe, M., Montrozier, H., Rols, M.P., and Asselineau, C. (1987). Structural and metabolic study of the mycolic acids of Mycobacterium fortuitum. Eur J Biochem *163*, 369-378.

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

Larsen, M.H., Biermann, K., Tandberg, S., Hsu, T., and Jacobs, W.R., Jr. (2007). Genetic Manipulation of Mycobacterium tuberculosis. Curr Protoc Microbiol *Chapter 10*, Unit 10A 12.

Larsen, M.H., Vilcheze, C., Kremer, L., Besra, G.S., Parsons, L., Salfinger, M., Heifets, L., Hazbon, M.H., Alland, D., Sacchettini, J.C*., et al.* (2002). Overexpression of inhA,

but not kasA, confers resistance to isoniazid and ethionamide in Mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. Mol Microbiol *46*, 453-466.

Laskowski, R.A. (2009). PDBsum new things. Nucleic Acids Res *37*, D355-359.

Leger, J., Kempf, M., Lee, G., and Brandt, R. (1997). Conversion of serine to aspartate imitates phosphorylation-induced changes in the structure and function of microtubule-associated protein tau. J Biol Chem *272*, 8441-8446.

Lei, B., Wei, C.J., and Tu, S.C. (2000). Action mechanism of antitubercular isoniazid. Activation by Mycobacterium tuberculosis KatG, isolation, and characterization of inha inhibitor. J Biol Chem *275*, 2520-2526.

Levy, C.W., Roujeinikova, A., Sedelnikova, S., Baker, P.J., Stuitje, A.R., Slabas, A.R., Rice, D.W., and Rafferty, J.B. (1999). Molecular basis of triclosan activity. Nature *398*, 383-384.

Li, B., and Lin, S.X. (1996). Fluorescence-energy transfer in human estradiol 17 beta-dehydrogenase-NADPH complex and studies on the coenzyme binding. Eur J Biochem *235*, 180-186.

Lu, H., and Tonge, P.J. (2008). Inhibitors of FabI, an enzyme drug target in the bacterial fatty acid biosynthesis pathway. Acc Chem Res *41*, 11-20.

Lu, X.Y., You, Q.D., and Chen, Y.D. (2010). Recent progress in the identification and development of InhA direct inhibitors of Mycobacterium tuberculosis. Mini Rev Med Chem *10*, 181-192.

Luckner, S.R., Machutta, C.A., Tonge, P.J., and Kisker, C. (2009). Crystal structures of Mycobacterium tuberculosis KasA show mode of action within cell wall biosynthesis and its inhibition by thiolactomycin. Structure *17*, 1004-1013.

Lyadova, I. (2012). Inflammation and Immunopathogenesis of Tuberculosis Progression, Understanding Tuberculosis - Analyzing the Origin of Mycobacterium Tuberculosis Pathogenicity. In Understanding Tuberculosis - Analyzing the Origin of Mycobacterium Tuberculosis Pathogenicity, P.-J. Cardona, ed. (Rijeka, Croatia: InTech), pp. 560.

Magnet, S., Hartkoorn, R.C., Szekely, R., Pato, J., Triccas, J.A., Schneider, P., Szantai-Kis, C., Orfi, L., Chambon, M., Banfi, D.*, et al.* (2010). Leads for antitubercular compounds from kinase inhibitor library screens. Tuberculosis (Edinb) *90*, 354-360.

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

Manjunatha, U.H., Boshoff, H., Dowd, C.S., Zhang, L., Albert, T.J., Norton, J.E., Daniels, L., Dick, T., Pang, S.S., and Barry, C.E., 3rd (2006). Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *103*, 431-436.

Marrakchi, H., Ducasse, S., Labesse, G., Montrozier, H., Margeat, E., Emorine, L., Charpentier, X., Daffe, M., and Quemard, A. (2002). MabA (FabG1), a Mycobacterium tuberculosis protein involved in the long-chain fatty acid elongation system FAS-II. Microbiology *148*, 951-960.

Martin, A., Morcillo, N., Lemus, D., Montoro, E., Telles, M.A., Simboli, N., Pontino, M., Porras, T., Leon, C., Velasco, M.*, et al.* (2005). Multicenter study of MTT and resazurin assays for testing susceptibility to first-line anti-tuberculosis drugs. Int J Tuberc Lung Dis *9*, 901-906.

Mathew, R., and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol *14*, 450-455.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McKinney, J.D., Honer zu Bentrup, K., Munoz-Elias, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., Jacobs, W.R., Jr., and Russell, D.G. (2000). Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature *406*, 735-738.

McMurry, L.M., Oethinger, M., and Levy, S.B. (1998). Triclosan targets lipid synthesis. Nature *394*, 531-532.

McNeil, M., Daffe, M., and Brennan, P.J. (1991). Location of the mycolyl ester substituents in the cell walls of mycobacteria. J Biol Chem *266*, 13217-13223.

McRee, D.E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. J Struct Biol *125*, 156-165.

McShane, H., Jacobs, W.R., Fine, P.E., Reed, S.G., McMurray, D.N., Behr, M., Williams, A., and Orme, I.M. (2012). BCG: myths, realities, and the need for alternative vaccine strategies. Tuberculosis (Edinb) *92*, 283-288.

Mdluli, K., Slayden, R.A., Zhu, Y., Ramaswamy, S., Pan, X., Mead, D., Crane, D.D., Musser, J.M., and Barry, C.E., 3rd (1998). Inhibition of a Mycobacterium tuberculosis beta-ketoacyl ACP synthase by isoniazid. Science *280*, 1607-1610.

Miesel, L., Weisbrod, T.R., Marcinkeviciene, J.A., Bittman, R., and Jacobs, W.R., Jr. (1998). NADH dehydrogenase defects confer isoniazid resistance and conditional lethality in Mycobacterium smegmatis. J Bacteriol *180*, 2459-2467.

Milano, A., Forti, F., Sala, C., Riccardi, G., and Ghisotti, D. (2001). Transcriptional regulation of furA and katG upon oxidative stress in Mycobacterium smegmatis. J Bacteriol *183*, 6801-6806.

Minnikin, D.E., Kremer, L., Dover, L.G., and Besra, G.S. (2002). The methyl-branched fortifications of Mycobacterium tuberculosis. Chem Biol *9*, 545-553.

Mitchison, D., and Davies, G. (2012). The chemotherapy of tuberculosis: past, present and future. Int J Tuberc Lung Dis *16*, 724-732.

Mitra, P.P. (2012). Drug discovery in tuberculosis: a molecular approach. Indian J Tuberc *59*, 194-206.

Moadebi, S., Harder, C.K., Fitzgerald, M.J., Elwood, K.R., and Marra, F. (2007). Fluoroquinolones for the treatment of pulmonary tuberculosis. Drugs *67*, 2077-2099.

Molle, V., Brown, A.K., Besra, G.S., Cozzone, A.J., and Kremer, L. (2006). The condensing activities of the Mycobacterium tuberculosis type II fatty acid synthase are differentially regulated by phosphorylation. J Biol Chem *281*, 30094-30103.

Molle, V., Gulten, G., Vilcheze, C., Veyron-Churlet, R., Zanella-Cleon, I., Sacchettini, J.C., Jacobs, W.R., Jr., and Kremer, L. (2010). Phosphorylation of InhA inhibits mycolic acid biosynthesis and growth of Mycobacterium tuberculosis. Mol Microbiol *78*, 1591-1605.

Molle, V., and Kremer, L. (2010). Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way. Mol Microbiol *75*, 1064-1077.

Molle, V., Kremer, L., Girard-Blanc, C., Besra, G.S., Cozzone, A.J., and Prost, J.F. (2003). An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from Mycobacterium tuberculosis. Biochemistry *42*, 15300-15309.

Monack, D.M., Mueller, A., and Falkow, S. (2004). Persistent bacterial infections: the interface of the pathogen and the host immune system. Nat Rev Microbiol *2*, 747-765.

Morgenthalera, M., Aebib, J.D., Grüningerb, F., Monab, D., Wagnerb, B., Kansyb, M., and Diederich, F. (2008). A fluorine scan of non-peptidic inhibitors of neprilysin: Fluorophobic and fluorophilic regions in an enzyme active site. Journal of Fluorine Chemistry *129*, 852-865.

Morlock, G.P., Metchock, B., Sikes, D., Crawford, J.T., and Cooksey, R.C. (2003). ethA, inhA, and katG loci of ethionamide-resistant clinical Mycobacterium tuberculosis isolates. Antimicrob Agents Chemother *47*, 3799-3805.

Muller, B., Streicher, E.M., Hoek, K.G., Tait, M., Trollip, A., Bosman, M.E., Coetzee, G.J., Chabula-Nxiweni, E.M., Hoosain, E., Gey van Pittius, N.C.*, et al.* (2011). inhA promoter mutations: a gateway to extensively drug-resistant tuberculosis in South Africa? Int J Tuberc Lung Dis *15*, 344-351.

Murakami, K.S. (2013). X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. J Biol Chem *288*, 9126-9134.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science *296*, 1285-1290.

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

Murray, K.D., and Bremer, H. (1996). Control of spoT-dependent ppGpp synthesis and degradation in Escherichia coli. J Mol Biol *259*, 41-57.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol *247*, 536-540.

Nechaev, S., Chlenov, M., and Severinov, K. (2000). Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. J Biol Chem *275*, 25516-25522.

Nicolas, F.J., Cayuela, M.L., Martinez-Argudo, I.M., Ruiz-Vazquez, R.M., and Murillo, F.J. (1996). High mobility group I(Y)-like DNA-binding domains on a bacterial transcription factor. Proc Natl Acad Sci U S A *93*, 6881-6885.

North, E.J., Jackson, M., and Lee, R.E. (2013). New Approaches to Target the Mycolic Acid Biosynthesis Pathway for the Development of Tuberculosis Therapeutics. Curr Pharm Des.

Ojha, A.K., Baughn, A.D., Sambandan, D., Hsu, T., Trivelli, X., Guerardel, Y., Alahari, A., Kremer, L., Jacobs, W.R., Jr., and Hatfull, G.F. (2008). Growth of Mycobacterium

tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol Microbiol *69*, 164-174.

Oliveira, J.S., Vasconcelos, I.B., Moreira, I.S., Santos, D.S., and Basso, L.A. (2007). Enoyl reductases as targets for the development of anti-tubercular and anti-malarial agents. Curr Drug Targets *8*, 399-411.

Opalka, N., Brown, J., Lane, W.J., Twist, K.A., Landick, R., Asturias, F.J., and Darst, S.A. (2010). Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. PLoS Biol *8*, e1000483.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307-326.

Parikh, S., Moynihan, D.P., Xiao, G., and Tonge, P.J. (1999). Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-ACP reductase from Mycobacterium tuberculosis. Biochemistry *38*, 13623-13634.

Parikh, S.L., Xiao, G., and Tonge, P.J. (2000). Inhibition of InhA, the enoyl reductase from Mycobacterium tuberculosis, by triclosan and isoniazid. Biochemistry *39*, 7645-7650.

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

Paul, B.J., Ross, W., Gaal, T., and Gourse, R.L. (2004b). rRNA transcription in Escherichia coli. Annu Rev Genet *38*, 749-770.

Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov *6*, 29-40.

Payne, D.J., Warren, P.V., Holmes, D.J., Ji, Y., and Lonsdale, J.T. (2001). Bacterial fatty-acid biosynthesis: a genomics-driven target for antibacterial drug discovery. Drug Discov Today *6*, 537-544.

Penalver-Mellado, M., Garcia-Heras, F., Padmanabhan, S., Garcia-Moreno, D., Murillo, F.J., and Elias-Arnanz, M. (2006). Recruitment of a novel zinc-bound transcriptional factor by a bacterial HMGA-type protein is required for regulating multiple processes in Myxococcus xanthus. Mol Microbiol *61*, 910-926.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I., and Vassylyev, D.G. (2004). Regulation through the secondary

channel--structural framework for ppGpp-DksA synergism during transcription. Cell *118*, 297-309.

Perozzo, R., Kuo, M., Sidhu, A., Valiyaveettil, J.T., Bittman, R., Jacobs, W.R., Jr., Fidock, D.A., and Sacchettini, J.C. (2002). Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. J Biol Chem *277*, 13106-13114.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Prisic, S., Dankwa, S., Schwartz, D., Chou, M.F., Locasale, J.W., Kang, C.M., Bemis, G., Church, G.M., Steen, H., and Husson, R.N. (2010). Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc Natl Acad Sci U S A *107*, 7521-7526.

Quemard, A., Laneelle, G., and Lacave, C. (1992). Mycolic acid synthesis: a target for ethionamide in mycobacteria? Antimicrob Agents Chemother *36*, 1316-1321.

Quemard, A., Sacchettini, J.C., Dessen, A., Vilcheze, C., Bittman, R., Jacobs, W.R., Jr., and Blanchard, J.S. (1995). Enzymatic characterization of the target for isoniazid in Mycobacterium tuberculosis. Biochemistry *34*, 8235-8241.

Raman, K., Yeturu, K., and Chandra, N. (2008). targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. BMC Syst Biol *2*, 109.

Raman, S., Hazra, R., Dascher, C.C., and Husson, R.N. (2004). Transcription regulation by the Mycobacterium tuberculosis alternative sigma factor SigD and its role in virulence. J Bacteriol *186*, 6605-6616.

Ramaswamy, S.V., Reich, R., Dou, S.J., Jasperse, L., Pan, X., Wanger, A., Quitugua, T., and Graviss, E.A. (2003). Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother *47*, 1241-1250.

Rao, S.P., Lakshminarayana, S.B., Kondreddi, R.R., Herve, M., Camacho, L.R., Bifani, P., Kalapala, S.K., Jiricek, J., Ma, N.L., Tan, B.H.*, et al.* (2013). Indolcarboxamide is a preclinical candidate for treating multidrug-resistant tuberculosis. Sci Transl Med *5*, 214ra168.

Rao, V., Gao, F., Chen, B., Jacobs, W.R., Jr., and Glickman, M.S. (2006). Trans-cyclopropanation of mycolic acids on trehalose dimycolate suppresses Mycobacterium tuberculosis -induced inflammation and virulence. J Clin Invest *116*, 1660-1667.

Rawat, R., Whitty, A., and Tonge, P.J. (2003). The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: adduct affinity and drug resistance. Proc Natl Acad Sci U S A *100*, 13881-13886.

Reddy, P.V., Puri, R.V., Khera, A., and Tyagi, A.K. (2012). Iron storage proteins are essential for the survival and pathogenesis of Mycobacterium tuberculosis in THP-1 macrophages and the guinea pig model of infection. J Bacteriol *194*, 567-575.

Remuinan, M.J., Perez-Herran, E., Rullas, J., Alemparte, C., Martinez-Hoyos, M., Dow, D.J., Afari, J., Mehta, N., Esquivias, J., Jimenez, E.*, et al.* (2013). Tetrahydropyrazolo[1,5-a]pyrimidine-3-carboxamide and N-benzyl-6',7'-dihydrospiro[piperidine-4,4'-thieno[3,2-c]pyran] analogues with bactericidal efficacy against Mycobacterium tuberculosis targeting MmpL3. PLoS One *8*, e60933.

Rodrigue, S., Provvedi, R., Jacques, P.E., Gaudreau, L., and Manganelli, R. (2006). The sigma factors of Mycobacterium tuberculosis. FEMS Microbiol Rev *30*, 926-941.

Rodriguez-Herrera, A.E., and Jordan-Salivia, G. (1999). Macrophage-Activating and Tissue-Damaging Immune Responses to M. tuberculosis. In Biometrics Unit Technical Reports (Ithaca, NY: Cornell University Library), pp. 1-12.

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., and Smith, I. (2002). ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun *70*, 3371-3381.

Roujeinikova, A., Levy, C.W., Rowsell, S., Sedelnikova, S., Baker, P.J., Minshull, C.A., Mistry, A., Colls, J.G., Camble, R., Stuitje, A.R.*, et al.* (1999). Crystallographic analysis of triclosan bound to enoyl reductase. J Mol Biol *294*, 527-535.

Rouse, D.A., Li, Z., Bai, G.H., and Morris, S.L. (1995). Characterization of the katG and inhA genes of isoniazid-resistant clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother *39*, 2472-2477.

Rozwarski, D.A., Grant, G.A., Barton, D.H., Jacobs, W.R., Jr., and Sacchettini, J.C. (1998). Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. Science *279*, 98-102.

Rozwarski, D.A., Vilcheze, C., Sugantino, M., Bittman, R., and Sacchettini, J.C. (1999). Crystal structure of the Mycobacterium tuberculosis enoyl-ACP reductase, InhA, in complex with NAD+ and a C16 fatty acyl substrate. J Biol Chem *274*, 15582-15589.

Rullas, J., Garcia, J.I., Beltran, M., Cardona, P.J., Caceres, N., Garcia-Bustos, J.F., and Angulo-Barturen, I. (2010). Fast standardized therapeutic-efficacy assay for drug discovery against tuberculosis. Antimicrob Agents Chemother *54*, 2262-2264.

Russell, D.G. (2005). Mycobacterium tuberculosis: the Indigestible Microbe. In Tuberculosis and Tubercle Bacillus, S.T. Cole, K.D. Eisenach, D.N. McMurray, and J. William R. Jacobs, eds. (Washington, D.C.: ASM Press), pp. 427-437.

Russell, D.G. (2011). Mycobacterium tuberculosis and the intimate discourse of a chronic infection. Immunol Rev *240*, 252-268.

Sacchettini, J.C., Rubin, E.J., and Freundlich, J.S. (2008). Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. Nat Rev Microbiol *6*, 41-52.

Sacco, E., Covarrubias, A.S., O'Hare, H.M., Carroll, P., Eynard, N., Jones, T.A., Parish, T., Daffe, M., Backbro, K., and Quemard, A. (2007). The missing piece of the type II fatty acid synthase system from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *104*, 14628-14633.

Sakamoto, K. (2012). The pathology of Mycobacterium tuberculosis infection. Vet Pathol *49*, 423-439.

Sandgren, A., Strong, M., Muthukrishnan, P., Weiner, B.K., Church, G.M., and Murray, M.B. (2009). Tuberculosis drug resistance mutation database. PLoS Med *6*, e2.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol *48*, 77-84.

Sassetti, C.M., and Rubin, E.J. (2003). Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A *100*, 12989-12994.

Scarsdale, J.N., Kazanina, G., He, X., Reynolds, K.A., and Wright, H.T. (2001). Crystal structure of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III. J Biol Chem *276*, 20516-20522.

Schaeffer, M.L., Agnihotri, G., Volker, C., Kallender, H., Brennan, P.J., and Lonsdale, J.T. (2001). Purification and biochemical characterization of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthases KasA and KasB. J Biol Chem *276*, 47029-47037.

Schluger, N.W., and Rom, W.N. (1998). The host immune response to tuberculosis. Am J Respir Crit Care Med *157*, 679-691.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C.*, et al.* (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. J Exp Med *198*, 693-704.

Schroeder, E.K., de Souza, N., Santos, D.S., Blanchard, J.S., and Basso, L.A. (2002). Drugs that inhibit mycolic acid biosynthesis in Mycobacterium tuberculosis. Curr Pharm Biotechnol *3*, 197-225.

Scorpio, A., and Zhang, Y. (1996). Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat Med *2*, 662-667.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol *8*, 27-31.

Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S.A., and Goldfarb, A. (1996). Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. J Biol Chem *271*, 27969-27974.

Shah, I.M., Laaberki, M.H., Popham, D.L., and Dworkin, J. (2008). A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. Cell *135*, 486-496.

Sharma, V., Sharma, S., Hoener zu Bentrup, K., McKinney, J.D., Russell, D.G., Jacobs, W.R., Jr., and Sacchettini, J.C. (2000). Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis. Nat Struct Biol *7*, 663-668.

Silva Miranda, M., Breiman, A., Allain, S., Deknuydt, F., and Altare, F. (2012). The tuberculous granuloma: an unsuccessful host defence mechanism providing a safety shelter for the bacteria? Clin Dev Immunol *2012*, 139127.

Singh, R., Manjunatha, U., Boshoff, H.I., Ha, Y.H., Niyomrattanakit, P., Ledwidge, R., Dowd, C.S., Lee, I.Y., Kim, P., Zhang, L.*, et al.* (2008). PA-824 kills nonreplicating Mycobacterium tuberculosis by intracellular NO release. Science *322*, 1392-1395.

Singh, V., Mani, I., Chaudhary, D.K., and Somvanshi, P. (2011). The beta-ketoacyl-ACP synthase from Mycobacterium tuberculosis as potential drug targets. Curr Med Chem *18*, 1318-1324.

Slayden, R.A., and Barry, C.E., 3rd (2002). The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in Mycobacterium tuberculosis. Tuberculosis (Edinb) *82*, 149-160.

Slayden, R.A., Lee, R.E., Armour, J.W., Cooper, A.M., Orme, I.M., Brennan, P.J., and Besra, G.S. (1996). Antimycobacterial action of thiolactomycin: an inhibitor of fatty acid and mycolic acid synthesis. Antimicrob Agents Chemother *40*, 2813-2819.

Smith, A.J., and Savery, N.J. (2005). RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. Nucleic Acids Res *33*, 755-764.

Smith, I., Bishai, W.R., and Nagaraja, V. (2005). Control of Mycobacterial Transcription. In Tuberculosis and the Tubercle Bacillus, S.T. Cole, K.D. Eisenach, D.N. McMurray, and J. William R. Jacobs, eds. (Washington, D.C.: ASM Press), pp. 219-235.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Mol Microbiol *4*, 1911-1919.

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

Srivatsan, A., and Wang, J.D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. Curr Opin Microbiol *11*, 100-105.

Stallings, C.L., Stephanou, N.C., Chu, L., Hochschild, A., Nickels, B.E., and Glickman, M.S. (2009). CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. Cell *138*, 146-159.

Steyn, A.J., Collins, D.M., Hondalus, M.K., Jacobs, W.R., Jr., Kawakami, R.P., and Bloom, B.R. (2002). Mycobacterium tuberculosis WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. Proc Natl Acad Sci U S A *99*, 3147-3152.

Stock, J.B., Ninfa, A.J., and Stock, A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol Rev *53*, 450-490.

Stokes, R.W., and Waddell, S.J. (2009). Adjusting to a new home: Mycobacterium tuberculosis gene expression in response to an intracellular lifestyle. Future Microbiol *4*, 1317-1335.

Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F.*, et al.* (1991). New use of BCG for recombinant vaccines. Nature *351*, 456-460.

Suarez, J., Ranguelova, K., Schelvis, J.P., and Magliozzo, R.S. (2009). Antibiotic resistance in Mycobacterium tuberculosis: peroxidase intermediate bypass causes poor isoniazid activation by the S315G mutant of M. tuberculosis catalase-peroxidase (KatG). J Biol Chem *284*, 16146-16155.

Sullivan, T.J., Truglio, J.J., Boyne, M.E., Novichenok, P., Zhang, X., Stratton, C.F., Li, H.J., Kaur, T., Amin, A., Johnson, F.*, et al.* (2006). High affinity InhA inhibitors with activity against drug-resistant strains of Mycobacterium tuberculosis. ACS Chem Biol *1*, 43-53.

Surolia, N., and Surolia, A. (2001). Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum. Nat Med *7*, 167-173.

Symeonidis, A., and Marangos, M. (2012). Iron and Microbial Growth In Insight and Control of Infectious Disease in Global Scenario, R. Priti, ed. (Rijeka, Croatia: InTech), pp. 289-330.

Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R.A., Ghosh, S., Parsy, M.L., Simpson, P.J., O'Dwyer, K., Matthews, S.J., and Paget, M.S. (2013). The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. Nucleic Acids Res *41*, 5679-5691.

Tagami, S., Sekine, S., Kumarevel, T., Hino, N., Murayama, Y., Kamegamori, S., Yamamoto, M., Sakamoto, K., and Yokoyama, S. (2010). Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. Nature *468*, 978-982.

Takayama, K., Schnoes, H.K., Armstrong, E.L., and Boyle, R.W. (1975). Site of inhibitory action of isoniazid in the synthesis of mycolic acids in Mycobacterium tuberculosis. J Lipid Res *16*, 308-317.

Takayama, K., Wang, C., and Besra, G.S. (2005). Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. Clin Microbiol Rev *18*, 81-101.

Takayama, K., Wang, L., and David, H.L. (1972). Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis. Antimicrob Agents Chemother *2*, 29-35.

Tan, M.P., Sequeira, P., Lin, W.W., Phong, W.Y., Cliff, P., Ng, S.H., Lee, B.H., Camacho, L., Schnappinger, D., Ehrt, S.*, et al.* (2010). Nitrate respiration protects hypoxic Mycobacterium tuberculosis against acid- and reactive nitrogen species stresses. PLoS One *5*, e13356.

Telenti, A. (1998). Genetics and pulmonary medicine. 5. Genetics of drug resistant tuberculosis. Thorax *53*, 793-797.

Timmins, G.S., and Deretic, V. (2006). Mechanisms of action of isoniazid. Mol Microbiol *62*, 1220-1227.

Timmins, G.S., Master, S., Rusnak, F., and Deretic, V. (2004). Nitric oxide generated from isoniazid activation by KatG: source of nitric oxide and activity against Mycobacterium tuberculosis. Antimicrob Agents Chemother *48*, 3006-3009.

Tokuyama, H., Yamashita, T., Reding, M.T., Kaburagi, Y., and Fukuyama, T. (1999). Radical Cyclization of 2-Alkenylthioanilides: A Novel Synthesis of 2,3-Disubstituted Indoles. J Am Chem Soc *121*, 3791–3792.

Tomioka, H., Tatano, Y., Yasumoto, K., and Shimizu, T. (2008). Recent advances in antituberculous drug development and novel drug targets. Expert Rev Respir Med *2*, 455-471.

Tonge, P.J., Kisker, C., and Slayden, R.A. (2007). Development of modern InhA inhibitors to combat drug resistant strains of Mycobacterium tuberculosis. Curr Top Med Chem *7*, 489-498.

Trautinger, B.W., and Lloyd, R.G. (2002). Modulation of DNA repair by mutations flanking the DNA channel through RNA polymerase. EMBO J *21*, 6944-6953.

Trinh, V., Langelier, M.F., Archambault, J., and Coulombe, B. (2006). Structural perspective on mutations affecting the function of multisubunit RNA polymerases. Microbiol Mol Biol Rev *70*, 12-36.

Trivedi, A., Singh, N., Bhat, S.A., Gupta, P., and Kumar, A. (2012). Redox biology of tuberculosis pathogenesis. Adv Microb Physiol *60*, 263-324.

Trucksis, M. (2000). Mycobacterium infections. In Persistent Bacterial Infections, J.P. Nataro, M.J. Blaser, and S. Cunningham-Rundles, eds. (Washington, D.C.: ASM Press), pp. 327-337.

Ung, K.S., and Av-Gay, Y. (2006). Mycothiol-dependent mycobacterial response to oxidative stress. FEBS Lett *580*, 2712-2716.

Vannelli, T.A., Dykman, A., and Ortiz de Montellano, P.R. (2002). The antituberculosis drug ethionamide is activated by a flavoprotein monooxygenase. J Biol Chem *277*, 12824-12829.

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

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. Nature *448*, 157-162.

Vazquez, M.J., Ashman, S., Ramon, F., Calvo, D., Bardera, A., Martin, J.J., Rudiger, M., Tew, D., and Dominguez, J.M. (2006). Utilization of substrate-induced quenching for screening targets promoting NADH and NADPH consumption. J Biomol Screen *11*, 75-81.

Venkataswamy, M.M., Goldberg, M.F., Baena, A., Chan, J., Jacobs, W.R., Jr., and Porcelli, S.A. (2012). In vitro culture medium influences the vaccine efficacy of Mycobacterium bovis BCG. Vaccine *30*, 1038-1049.

Veyron-Churlet, R., Molle, V., Taylor, R.C., Brown, A.K., Besra, G.S., Zanella-Cleon, I., Futterer, K., and Kremer, L. (2009). The Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single threonine residue. J Biol Chem *284*, 6414-6424.

Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Molle, V., and Kremer, L. (2010). Phosphorylation of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein reductase MabA regulates mycolic acid biosynthesis. J Biol Chem *285*, 12714-12725.

Vilcheze, C., Av-Gay, Y., Attarian, R., Liu, Z., Hazbon, M.H., Colangeli, R., Chen, B., Liu, W., Alland, D., Sacchettini, J.C.*, et al.* (2008). Mycothiol biosynthesis is essential for ethionamide susceptibility in Mycobacterium tuberculosis. Mol Microbiol *69*, 1316-1329.

Vilcheze, C., and Jacobs, W.R., Jr. (2007). The mechanism of isoniazid killing: clarity through the scope of genetics. Annu Rev Microbiol *61*, 35-50.

Vilcheze, C., Morbidoni, H.R., Weisbrod, T.R., Iwamoto, H., Kuo, M., Sacchettini, J.C., and Jacobs, W.R., Jr. (2000). Inactivation of the inhA-encoded fatty acid synthase II (FASII) enoyl-acyl carrier protein reductase induces accumulation of the FASI end products and cell lysis of Mycobacterium smegmatis. J Bacteriol *182*, 4059-4067.

Vilcheze, C., Wang, F., Arai, M., Hazbon, M.H., Colangeli, R., Kremer, L., Weisbrod, T.R., Alland, D., Sacchettini, J.C., and Jacobs, W.R., Jr. (2006). Transfer of a point mutation in Mycobacterium tuberculosis inhA resolves the target of isoniazid. Nat Med *12*, 1027-1029.

Vilcheze, C., Weisbrod, T.R., Chen, B., Kremer, L., Hazbon, M.H., Wang, F., Alland, D., Sacchettini, J.C., and Jacobs, W.R., Jr. (2005). Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. Antimicrob Agents Chemother *49*, 708-720.

Voskuil, M.I., Bartek, I.L., Visconti, K., and Schoolnik, G.K. (2011). The response of mycobacterium tuberculosis to reactive oxygen and nitrogen species. Front Microbiol *2*, 105.

Voskuil, M.I., Visconti, K.C., and Schoolnik, G.K. (2004). Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis (Edinb) *84*, 218-227.

Wall, M.D., Oshin, M., Chung, G.A., Parkhouse, T., Gore, A., Herreros, E., Cox, B., Duncan, K., Evans, B., Everett, M.*, et al.* (2007). Evaluation of N-(phenylmethyl)-4-[5-(phenylmethyl)-4,5,6,7-tetrahydro-1H-imidazo[4,5-c]pyridin- 4-yl]benzamide inhibitors of Mycobacterium tuberculosis growth. Bioorg Med Chem Lett *17*, 2740-2744.

Wang, F., Langley, R., Gulten, G., Dover, L.G., Besra, G.S., Jacobs, W.R., Jr., and Sacchettini, J.C. (2007). Mechanism of thioamide drug action against tuberculosis and leprosy. J Exp Med *204*, 73-78.

Wang, L.Q., Falany, C.N., and James, M.O. (2004). Triclosan as a substrate and inhibitor of 3'-phosphoadenosine 5'-phosphosulfate-sulfotransferase and UDP-glucuronosyl transferase in human liver fractions. Drug Metab Dispos *32*, 1162-1169.

Wang, X., Wang, H., and Xie, J. (2011). Genes and regulatory networks involved in persistence of Mycobacterium tuberculosis. Sci China Life Sci *54*, 300-310.

Wehenkel, A., Bellinzoni, M., Grana, M., Duran, R., Villarino, A., Fernandez, P., Andre-Leroux, G., England, P., Takiff, H., Cervenansky, C.*, et al.* (2008). Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. Biochim Biophys Acta *1784*, 193-202.

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

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

WHO (2013). Tuberculosis (TB).www.who.int/tb/en

Willand, N., Dirie, B., Carette, X., Bifani, P., Singhal, A., Desroses, M., Leroux, F., Willery, E., Mathys, V., Deprez-Poulain, R.*, et al.* (2009). Synthetic EthR inhibitors boost antituberculous activity of ethionamide. Nat Med *15*, 537-544.

Wilming, M., and Johnsson, K. (1999). Spontaneous Formation of the Bioactive Form of the Tuberculosis Drug Isoniazid. Angew Chem Int Ed Engl *38*, 2588-2590.

Wilson, T.M., and Collins, D.M. (1996). ahpC, a gene involved in isoniazid resistance of the Mycobacterium tuberculosis complex. Mol Microbiol *19*, 1025-1034.

Winder, F.G., Collins, P., and Rooney, S.A. (1970). Effects of isoniazid on mycolic acid synthesis in Mycobacterium tuberculosis and on its cell envelope. Biochem J *117*, 27P.

Winder, F.G., Collins, P.B., and Whelan, D. (1971). Effects of ethionamide and isoxyl on mycolic acid synthesis in Mycobacterium tuberculosis BCG. J Gen Microbiol *66*, 379-380.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A*., et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Wirth, T., Hildebrand, F., Allix-Beguec, C., Wolbeling, F., Kubica, T., Kremer, K., van Soolingen, D., Rusch-Gerdes, S., Locht, C., Brisse, S*., et al.* (2008). Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog *4*, e1000160.

Yajko, D.M., Nassos, P.S., and Hadley, W.K. (1987). Therapeutic implications of inhibition versus killing of Mycobacterium avium complex by antimicrobial agents. Antimicrob Agents Chemother *31*, 117-120.

Zaman, K. (2010). Tuberculosis: a global health problem. J Health Popul Nutr *28*, 111-113.

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

Zhang, J.H., Chung, T.D., and Oldenburg, K.R. (1999b). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen *4*, 67-73.

Zhang, M., Yue, J., Yang, Y.P., Zhang, H.M., Lei, J.Q., Jin, R.L., Zhang, X.L., and Wang, H.H. (2005a). Detection of mutations associated with isoniazid resistance in Mycobacterium tuberculosis isolates from China. J Clin Microbiol *43*, 5477-5482.

Zhang, Y. (2005). The magic bullets and tuberculosis drug targets. Annu Rev Pharmacol Toxicol *45*, 529-564.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012a). Structural basis of transcription initiation. Science *338*, 1076-1080.

Zhang, Y., Garbe, T., and Young, D. (1993). Transformation with katG restores isoniazid-sensitivity in Mycobacterium tuberculosis isolates resistant to a range of drug concentrations. Mol Microbiol *8*, 521-524.

Zhang, Y., Vilcheze, C., and William R. Jacobs, J. (2005b). Mechanism of Drug Resistance in Mycobacteirum tuberculosis. In Tuberculosis and Tubercle Bacillus, S.T. Cole, K.D. Eisenach, D.N. McMurray, and J. William R. Jacobs, eds. (Washington, D.C. : ASM Press).

Zhang, Y., Yew, W.W., and Barer, M.R. (2012b). Targeting persisters for tuberculosis control. Antimicrob Agents Chemother *56*, 2223-2230.

Zhao, X., Yu, H., Yu, S., Wang, F., Sacchettini, J.C., and Magliozzo, R.S. (2006). Hydrogen peroxide-mediated isoniazid activation catalyzed by Mycobacterium tuberculosis catalase-peroxidase (KatG) and its S315T mutant. Biochemistry *45*, 4131-4140.

Zimhony, O., Vilcheze, C., and Jacobs, W.R., Jr. (2004). Characterization of Mycobacterium smegmatis expressing the Mycobacterium tuberculosis fatty acid synthase I (fas1) gene. J Bacteriol *186*, 4051-4055.

Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., and Nerlich, A.G. (2003). Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol *41*, 359-367.

Zumla, A., Nahid, P., and Cole, S.T. (2013a). Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov *12*, 388-404.

Zumla, A., Raviglione, M., Hafner, R., and von Reyn, C.F. (2013b). Tuberculosis. N Engl J Med *368*, 745-755.

A
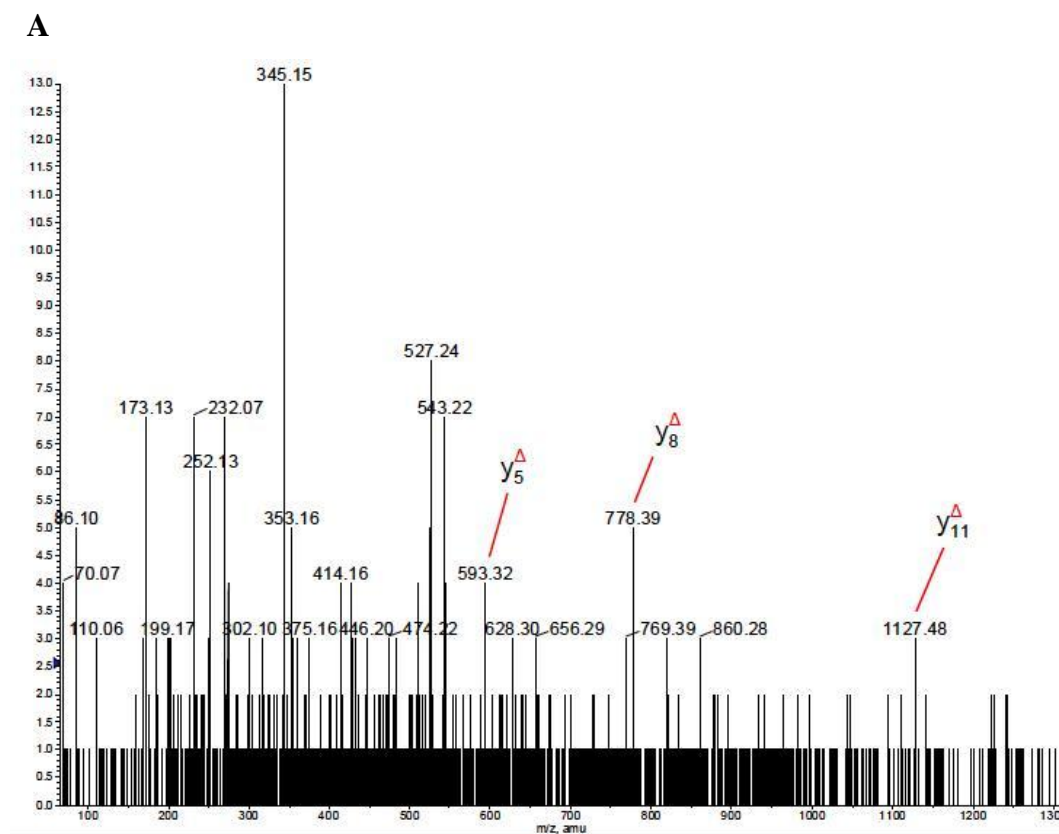
Figure A-1. Mass spectrometric analysis of phosphorylated InhA. A) Mass spectrometric analysis of *in vitro* PknB-phosphorylated InhA**.** MS/MS spectrum of the triply charged ion [M +3H]$^{3+}$ at m/z 1051.51of peptide [241-269] (monoisotopic mass: 3151.52 Da). Unambiguous location of the phosphate group on Thr266 was shown by observation of the "y" C-terminal daughter ion series. Starting from the C-terminal residue, all "y" ions lose phosphoric acid (-98 Da) after the Thr266 phosphorylated residue. **B)** Mass spectrometric analysis of InhA purified *in vivo*. MS/MS spectrum of the triply charged ion [M+3H]$^{3+}$ at m/z 1051.51 of peptide [241-269] (monoisotopic mass: 3151.52Da). The phosphate group on T266 was unambiguously located by observing the "y" C-terminal daughter ion series. Starting from the C-terminal residue, all "y" ions lose phosphoric acid (-98 Da) after the T266 phosphorylated residue.
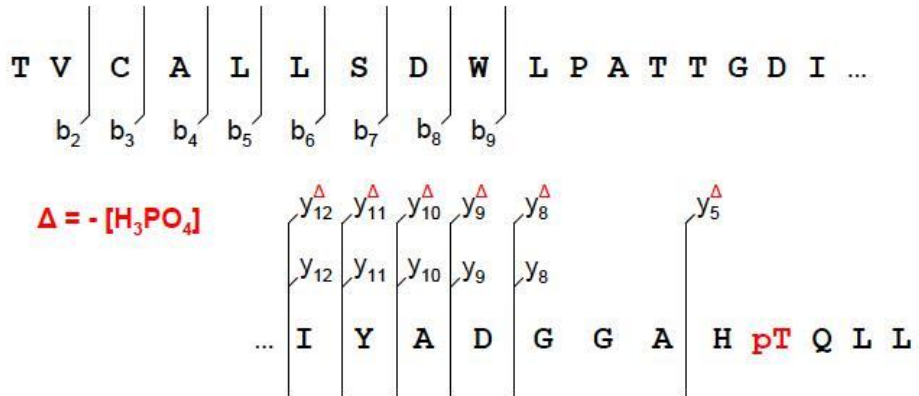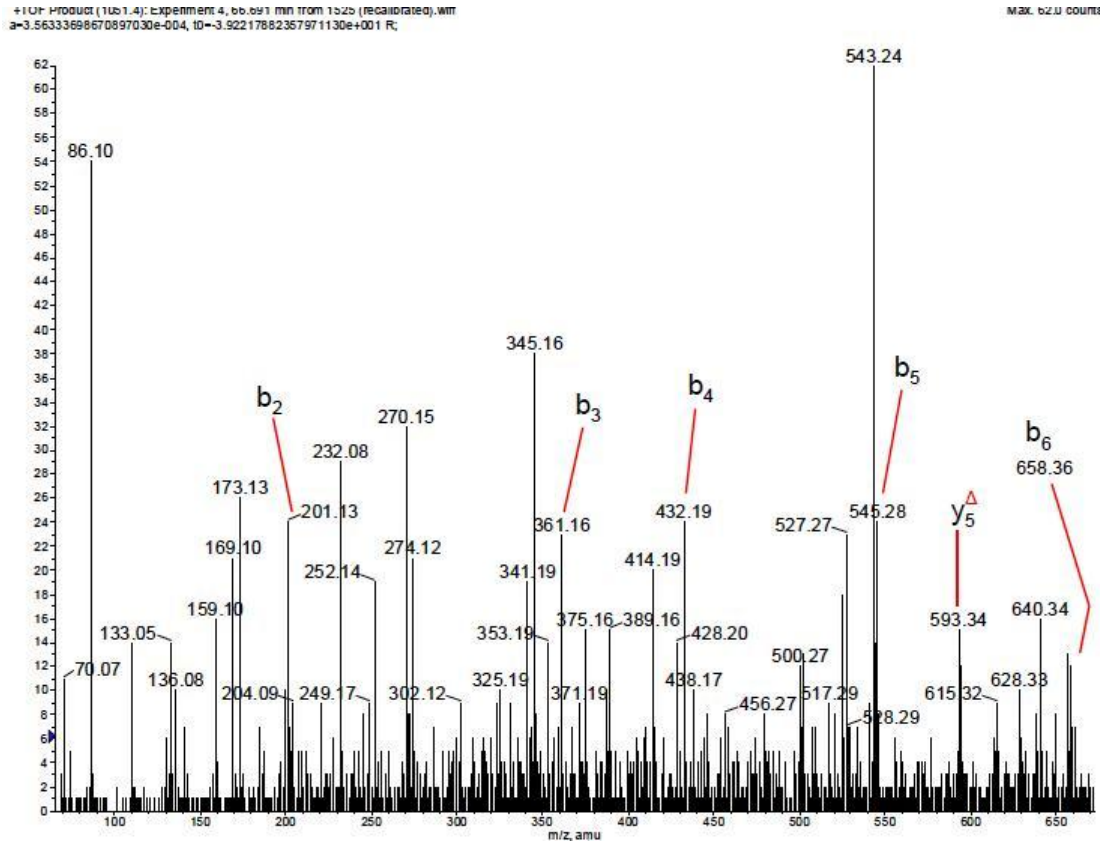
Figure A-1 Continued.
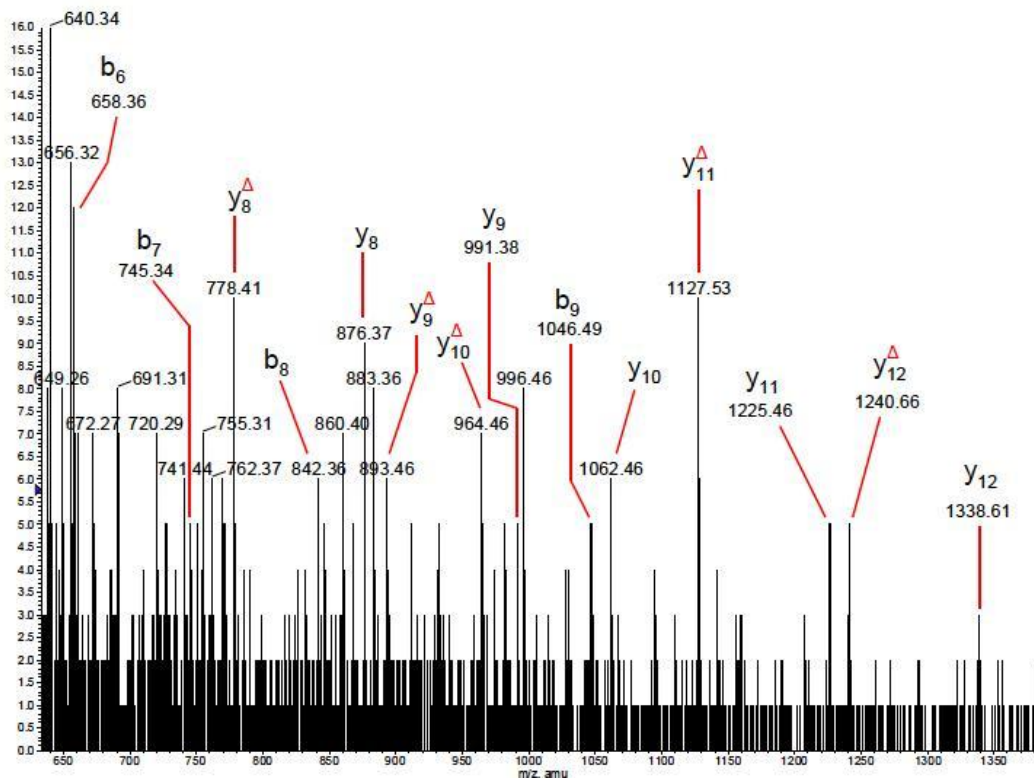
Figure A-1 Continued.
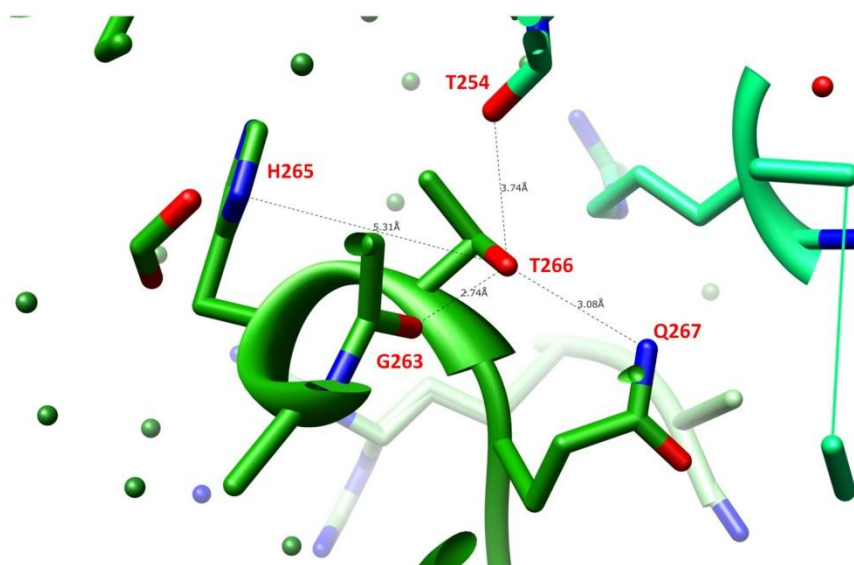
Figure A-2. Structures of binary complexes of InhA_WT:NADH, InhA_T266D: NADH and InhA_T266E:NADH. Crystals were obtained using the hanging drop vapor-diffusion method by mixing 2 μl of 5-10 mg/ml protein solution containing NADH (1:2 ratio of enzyme to NADH) with an equal volume of reservoir solution (100 mM Hepes 7.2, 50 mM sodium citrate 6.5, 8-12% methylpentanediol and 4% DMSO). Crystals were cryoprotected by 30% MPD mixed with reservoir solution prior to flash-cooling at 100K in a cold nitrogen stream. Data sets were collected at 0.97947 Å using synchrotron radiation on GM/CA-CAT 23ID beamline at Advanced Photon Source, Argonne National Laboratory. Data were processed and scaled by DENZO/Scalepack (Otwinowski and Minor (1997) *Methods in Enzymology*, 307-326) and initial phases were obtained by molecular replacement using previously published wild-type_InhA:NAD+ structure (PDB code: 2AQ8) as a starting model. Model building and refinement of the structures were done by Coot (Emsley and Cowtan (2004) *Acta Crystallogr D Biol Crystallogr*. 60, 2126-32) and Phenix (Adams *et al.* (2002) *Acta Cryst. D*. 58, 1948-1954), respectively. In the final refinement cycles, water molecules were added to the peaks above 3 σ of the Fo-Fc electron density map such that the water molecules were within H-bonding distance to the surrounding protein/water molecules. Data collection and processing statistics are provided in Table A-3. Comparative three-dimensional structures and H-bonding interaction of InhA-WT, InhA_T266D and InhA_T266 are represented in the up, middle and low down panels, respectively. A new water molecule (W9), not present in InhA_WT, is introduced between residues D266 and N267 residues. As observed in T266D structure, a new water molecule (W5) is found to be present introduced between D266 and N267 residues in the T266E structure. Atomic coordinates and structure factors of wild-type InhA:NADH, T266D:NADH and T266E:NADH were deposited in the PDB with the accession codes 3OEW, 3OF2 and 3OEY, respectively.
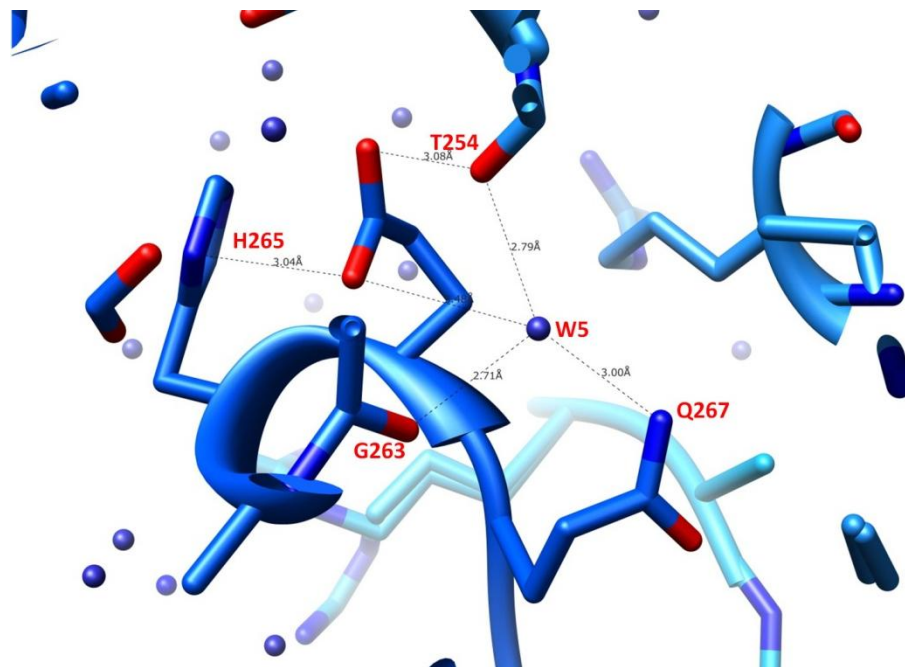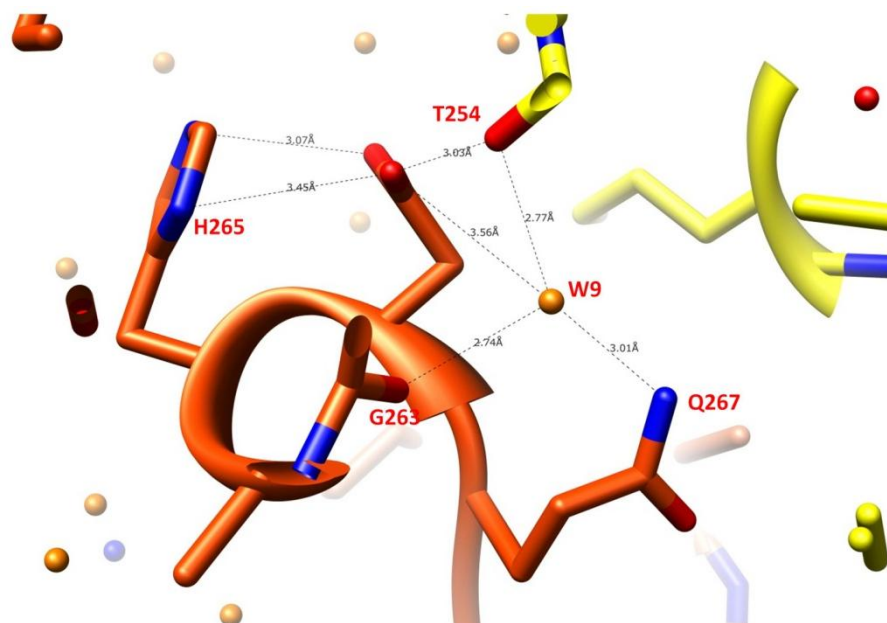
Figure A-2 Continued.
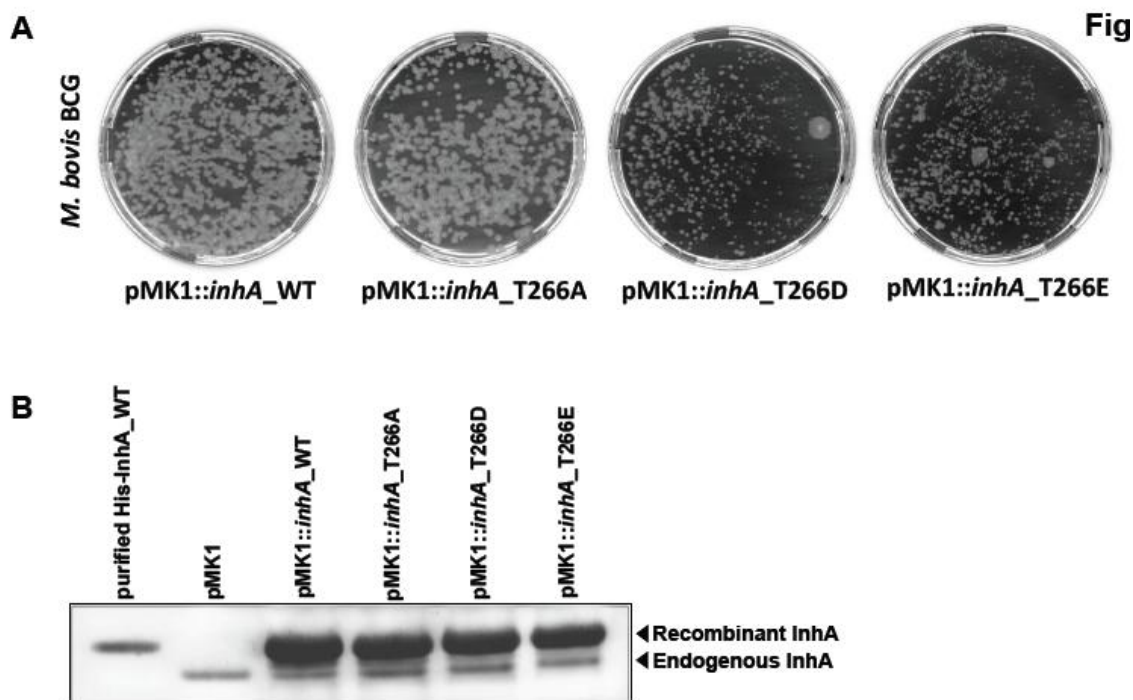
Figure A-3. Overexpression of the various InhA variants and effect on growth in *M. bovis* BCG. A) Electrocompetent *M. bovis* BCG were transformed with the empty pMK1 construct, the pMK1::*inhA_WT*, pMK1::*inhA_T266A,* pMK1::*inhA_T266D* or pMK1::*inhA_T266E* to allow constitutive expression of the various *inhA* alleles under the control of the strong *hsp60* promoter. Transformed mycobacteria were plated and incubated at 37°C for 3 weeks. **B)** InhA expression levels in the InhA-overproducing *M. bovis* BCG strains. Western blot analysis of *M. bovis* BCG cells overexpressing the phosphoablative *inhA*_T266A and phosphomimetic *inhA*_T266D and *inhA*_T266E alleles were harvested, resuspended in PBS and disrupted. Equal amounts of crude lysates (20 μg) were then loaded onto a 4-12% acrylamide gel, subjected to electrophoresis and transferred onto a membrane for immunoblot analysis using rabbit anti-InhA antibodies. Endogenous and recombinant InhA proteins are indicated by arrowheads.
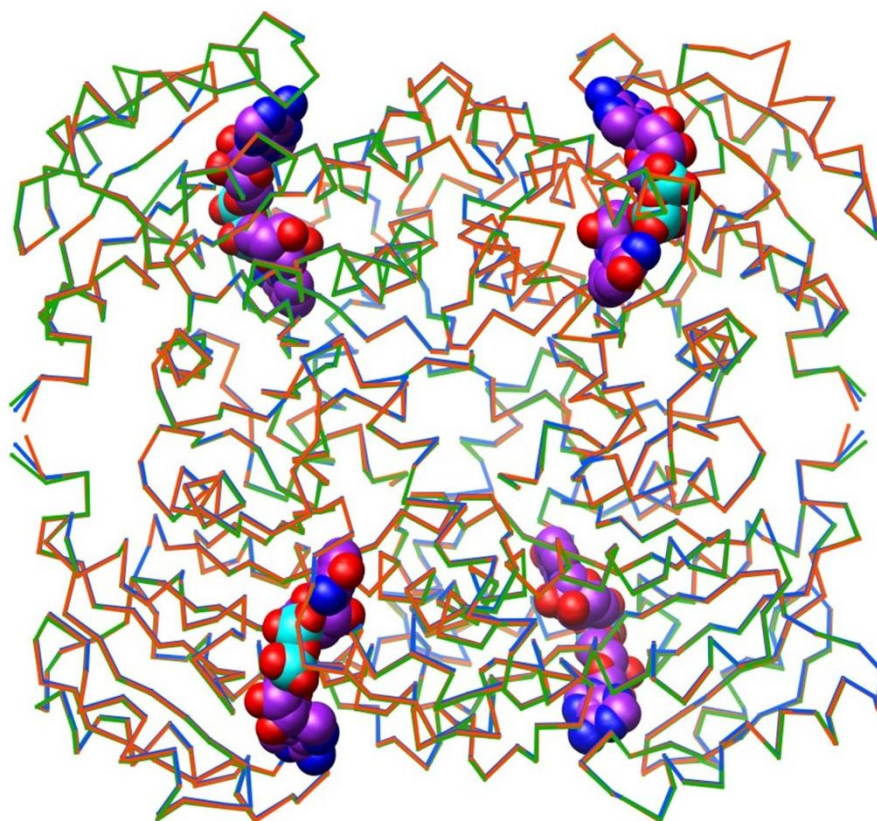
Figure A-4. Superimposition of InhA_WT (green), InhA_T266D (orange) and InhA_ T266E (blue) tetrameric structures. The Rmsd is 0.3 Å. NADH is represented as spheres (carbon purple, oxygen red, nitrogen blue and phosphate cyan).
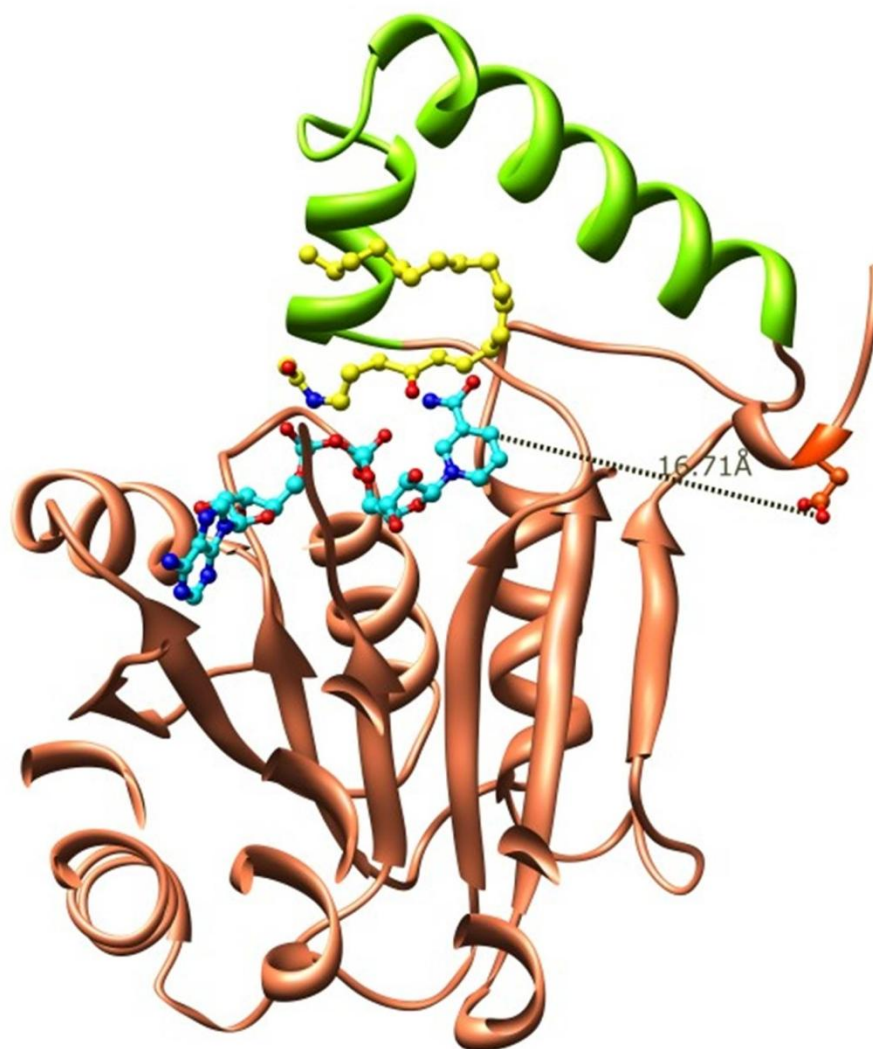
Figure A-5. The crystal structure of InhA_T266D mutant. The figure shows the proximity of the mutation site to the active site. NADH is represented with cyan sticks, C16 fatty acid substrate analog (PDB:1BVR) with yellow sticks. Substrate binding loop is colored green.

Table A-1. Bacterial strains and plasmids used in this study.

Strains or Plasmids	Genotype or Description	Source or Reference
E. coli TOP10	F⁻ *mcr*A (*mrr-hsd*RMS-*mcr*BC) φ80*lacZ M15 lacX74 deoR recA1 araD139* (*ara-leu*)7697 *galU galK rpsL endA1 nupG*; used for general cloning	Invitrogen
E. coli BL21(DE3)Star	F2 *ompT hsdSB*(rB2 mB2) *gal dcm* (DE3); used to express recombinant proteins in *E. coli*	Stratagene
M. bovis BCG 1173P2 *M. smegmatis* mc²155	Vaccine strain *ept-1*	WHO, Stockholm (Snapper *et al.*, 1990)
M. smegmatis mc²2359	*ept-1 attB*::pYUB412::*ndh inhA40* INH-R, Ts mutant of mc²2354	(Vilcheze *et al.*, 2000)
pETPhos	pET15b (Novagen) derivative including the replacement of the thrombin site coding sequence with a tobacco etch virus (TEV) protease site and Ser to Gly mutagenesis in the Nterm His-tag	(Canova *et al.*, 2008)
pETPhos_*inhA_WT*	pETPhos derivative used to express His-tagged fusion of WT InhA in *E.coli*	This work
pETPhos_*inhA_T266A*	pETPhos derivative used to express His-tagged fusion of InhA_T266A in *E.coli*	This work
pETPhos_*inhA_T266D*	pETPhos derivative used to express His-tagged fusion of InhA_T266D in *E.coli*	This work
pETPhos_*inhA_T266E*	pETPhos derivative used to express His-tagged fusion of InhA_T266E in *E.coli*	This work
pMK1	*E. coli*/mycobacterial shuttle vector, allows expression of N-term His-tagged proteins, derived from pMV261 (Stover *et al.*, 1991) and containing the pET28a polylinker	(Veyron-Churlet *et al.*, 2010)
pMK1_*inhA_WT*	pMK1 derivative used to express His-tagged fusion of WT InhA in mycobacteria	This work
pMK1_*inhA_T266A*	pMK1 derivative used to express His-tagged fusion of InhA_T266A in mycobacteria	This work
pMK1_*inhA_T266D*	pMK1 derivative used to express His-tagged fusion of InhA_T266D in mycobacteria	This work

Table A-1 Continued.

Strains or Plasmids	Genotype or Description	Source or Reference
pMK1_*inhA_T266E*	pMK1 derivative used to express His-tagged fusion of InhA_T266E in mycobacteria	This work
pSD26	Acetamide-inducible expression vector	(Daugelat *et al.*, 2003)
pSD26_*inhA_WT*	pSD26 derivative used to induce His-tagged fusion of WT InhA in mycobacteria	This work
pSD26_*inhA_T266A*	pSD26 derivative used to induce His-tagged fusion of InhA_T266A in mycobacteria	This work
pSD26_*inhA_T266D*	pSD26 derivative used to induce His-tagged fusion of InhA_T266D in mycobacteria	This work
pSD26_*inhA_T266E*	pSD26 derivative used to induce His-tagged fusion of InhA_T266E in mycobacteria	This work
p004	pJSC347 (Sambandamurthy *et al.*, 2002) derivative containing lambda phage *cos* sites, a *hyg* resistance marker and a *sacB* cassette.	Hsu and Jacobs, manuscript in preparation

Table A-2. Primers used in this study.

Primers	5' to 3' Sequence[ab]
pMK1 InhA dir	tcat<u>catatg</u>acaggactgctggacgg (NdeI)
pMK1 InhA rev	tac<u>ggaattc</u>tagagcaattgggtgtgcgc (EcoRI)
pSD26 InhA dir	tata<u>ggatcc</u>atgacaggactgctggacggc (BamHI)
pSD26 InhA WT rev	gagcaattgggtgtgcgcgc
pSD26 InhA T266A rev	gagcaattgggcgtgcgcgc
pSD26 InhA T266D rev	gagcaattggtcgtgcgcgc
pSD26 InhA T266E rev	gagcaattgctcgtgcgcgc
InhA T266A dir	cgccgacggcggcgcgcacgcccaattgctctagaattcg
InhA T266A rev	cgaattctagagcaattgggcgtgcgcgccgccgtcggcg
InhA T266D dir	cgccgacggcggcgcgcacgaccaattgctctagaattcg
InhA T266D rev	cgaattctagagcaattggtcgtgcgcgccgccgtcggcg
InhA T266E dir	cgccgacggcggcgcgcacgagcaattgctctagaattcg
InhA T266E rev	cgaattctagagcaattgctcgtgcgcgccgccgtcggcg
inhA_forward	agagaag<u>catatg</u>atgacaggactgctggacgg caaacg ttctggttagcg (ndei)
inhA_T266A reverse	ccc<u>aagctt</u>ttactagagcaattgggcgtgcgcgccgcc gtcggcgtagatg (hindiii)
inhA_T266D reverse	ccc<u>aagctt</u>ttactagagcaattggtcgtgcgcgccgcc gtcggcgtagatg (hindiii)
inhA_T266E reverse	ccc<u>aagctt</u>ttactagagcaattgctcgtgcgcgccgcc gtcggcgtagatg (hindiii)

[a]Restriction sites are underlined and specified into brackets.
[b]Mutagenized bases are shown in bold.

Table A-3. Data collection and processing statistics for InhA_WT:NADH, InhA_T266D:NADH and InhA_T266E:NADH crystal structures.

	InhA_WT	InhA_T266D	InhA_T266E
Unit cell dimensions a,c (Å)	98.551, 139.714	98.366, 139.887	98.154, 139.439
Space group	$P6_222$	$P6_222$	$P6_222$
Number of unique reflections	23240	39230	27393
Data collection resolution range	40.88-2.1	40.9-1.75	40.78-2.0
Refinement resolution range (Å)	40.88- 2.1	40.9-2.0	40.78-2.0
Completeness	96.7 (98.3)	99.5 (97.4)	99.6 (99.9)
I/σ	19.45(5.28)	14.3 (7.4)	36.97 (9.39)
R*sym*	3.5 (18.85)	4.36 (10.05)	1.79 (9.56)
R*factor*/ R*free*	0.1667/ 0.2086	0.1756/ 0.2033	0.1819/0.2143
Redundancy	11.1 (11.3)	9.3 (9.5)	17.5 (16.7)
Total protein residues	268	268	268
Heteroatoms	2	2	2
	(MPD)	(MPD)	(MPD)
	(NADH)	(NADH)	(NADH)
Water molecules	186	207	191
Average B-factors (A^2)			
Protein	42.57	34.26	35.80
NADH	47.32	35.88	37.22
Water	51.52	43.51	42.94
total	49.95, 45.69	40.96	41.43

Table A-3 Continued.

	InhA_WT	InhA_T266D	InhA_T266E
Ramachandran analysis			
MolProbity (Davis *et al.*, 2007)			
Favorable regions	96.99	96.24	96.24
Allowed region	100	100	100
Disallowed	0	0	0
r.m.s.d			
Bond lengths (Å)	0.008	0.007	0.007
Bond angles (^{0})	1.211	1.094	1.19

$R_{sym} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$ where I_{hi} is the intensity of observation *I* of reflection *h*. $R_{factor} \sum = _h ||F_o| - |F_c|| \sum |F/_o|$ where F_o and F_c are observed and calculated structure factors, respectively. R_{free} is calculated on 5% of the data omitted at random.

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER IV

Table B-1. Primers that are used in this work.

Construct	Forward 5'→ 3'	Reverse 5'→ 3'
CarD-notag	GGGAAT TCCATATGATTT TCAAGGTCGGAGACACCGTTGTC	GATCCCAAGCTTTCAAGACGCGGCGGC TAAAACCTCGTCAAG
CarD-Nterm-1-53 (CarD$_{1-53}$)	GGGAATTCCATATGATTTTCAAGG TCGGAGACACCGTTGTC	CCCAAGCTTTTAGGCGTTTTCAGCGGG AACTCGTA
CarD-Nterm-1-74 (CarD$_{1-74}$)	GGGAATTCCATATGATTTTCAAGG TCGGAGACACCGTTGTC	CCCAAGCTTTTACAACACCTGGAAAAC CTTGTCCAGG
CarD-Cterm-61-162 (CarD$_{61-162}$)	AGAGAAGCATATGATGGTCGTCG GGCAGGAAGGCCTGG	GATCCCAAGCTTTCAAGACGCGGCGGC TAAAACCTCGTCAAG
CarD-Cterm-83-162 (CarD$_{83-162}$)	AGAGAAGCATATGATGACGAACT GGTCACGTCGTTACAAGGC	GATCCCAAGCTTTCAAGACGCGGCGGC TAAAACCTCGTCAAG
CarD_R87A-R88A-K90A	ACCGAGGAGCCGACGAACTGGTC AGCTGCTTACGCGGCGAACCTCG AG	CTCGAGGTTCGCCGCGTAAGCAGCTGA CCAGTTCGTCGGCTCCTCGGT
CarD_R114A-R118A	GCGATTTGTGGCGTGCCGACCAG GAGGCTGGCTTGTC	GACAAGCCAGCCTCCTGGTCGGCACG CCACAAATCGC
CarD_K125A-R126A-K130A	CTTGTCGGCCGGTGAGGCGGCCA TGCTGGCCGCGGCCCGGCAGATT	GGCCGCGGCCAGCATGGCCGCCTCACC GGCCGACAAG
CarD_Y11A-H14A	CGGAGACACCGTTGTCGCTCCAC ACGCCGGTGCTGCGTTAGTC	GACTAACGCAGCACCGGCGTGTGGAGC GACAACGGTGTCTCCG
RpoBtr-NHis	AGAGAAGCATATGATGCTCCTTG ACGTCCAGACCGATTCG	CGCGGATCCTTACATGAATTGGCTCAG CTGGCT
16S-upstream	AGACTGGCAGGGTCGCCCCGAAG CGG	CGCCAGCGTTCGTCCTGAGCCAGGATC

317

β1   β2   β3   β4   α1   α2   α3   α4

Mtb	1	--------MIFKVGDTVVYPHHGAALVEAIETRTIKGEQSYLVLKVAQGDLTVRVPAENADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.leprae	1	----MNLMIFKVGDTVVYPHHGAALVEAIETRTIKGEQSYLVLKVAQGDLTVRVPAENADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.bovis	1	--------MIFKVGDTVVYPHHGAALVEAIETRTIKGEQSYLVLKVAQGDLTVRVPAENADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.parascrofulaceum	1	--------MIFKVGDTVVYPHHGAALVEAIETRTIKGEQSYLVLKVAQGDLTVRVPADNADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.smegmatis	1	--------MIFKVGDTVVYPHHGAALIEAIETRTIKGEQSYLVLKVAQGDLTVRVPADNADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.gilvum	1	--------MIFKVGDTVVYPHHGAALIEAIETRTIKGEQSYLVLKVAQGDLTVRVPADNADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.fortuitum	1	--------MIFKVGDTVVYPHHGAALIEAIETRTIKGEQSYLVLKVSQGDLTVRVPADNADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.thermoresistibile	1	--------MIFKVGDTVVYPHHGAALIEAIETRTIKGEQSYLVLKVAQGDLTVRVPAENADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.hassiacum	1	--------MIFKVGDTVVYPHHGAALIEAIETRTIKGEQSYLVLKVAQGDLTVRVPAENADYVGVRDVVGQEGLDKVFQVLRAPHTEEPTNWSRRYKANLEKLASGDVN
Actinosynnema	1	--------MVFKVGDTVVYPHHGAALIEAIETRTLKGERKYLVLKVAQGDLTVRVPADNADIVGVRDVVGQEGLDKVFEVLRAPHTEEPTNWSRRYKANLEKLASGDVN
Saccharothrix	1	--------MVFKVGDTVVYPHHGAALIEAIETRLLKGERKYLVLKVAQGDLTVRVPADNADIVGVRDVVGQEGLDKVFEVLRAPHTEEPTNWSRRYKANLEKLASGDVN
Saccharopolyspora	1	--------MVFKVGETVVYPHHGAALIEAIETRVIKGERKQLVLKVAQGDLTVRVPADNABYVGVRDVVGQEGLDKVFDVLRAPHTEEPTNWSRRYKANLEKLASGDVN
Streptomyces	1	--------MVFKVGDTVVYPHHGAALIEAIETRTIKGERKYLVLKVAQGDLTVRVPADNADVVGVRDVVGQEGLDKVFEVLRAPHTEEPTNWSRRYKANLEKLASGDVN
Pseudonocardia	1	--------MVFKVGDTVVYPHHGAALIEAIETRTIKGERKYLVLKVAQGDLTVRVPADNADVVGVRDVVGQEGLDKVFEVLRAPHTEEPTNWSRRYKANLEKLASGDVN
M.xanthus_DK1622_CdnL	1	-----MQTSEKIGDKAVYGQVGEVMGLHTEVASQRQSFYVLRILENGMRIMIHINKVGSVGRGIISERCVKQVYSILEKDISVDSLTWNRRAREYMEIKTLSVF
M.xanthus_DK1622_CarD	1	MPEGSASLQLAVGDRVVYPNQGVCRVSKLDVKEVAGQKLTFVTMRREEDGAVVMVPEGKVLAIGVNASABQVEAIETLRSD-SDKADLDWKQRARTNLDRMTQGIM
Clustal Consensus		: .*: .*** :*. . :: : *:. : :.: .: .:*:*.: .*. :: .: .*:

α4   α5   α6

Mtb	103	KVAEVVRDLWRRDQDRGLSAGEKRMLAKARQILVGELALABSTDDAKAETLDEVLAAAS--
M.leprae	106	KVSEVVRDLWRRDQDRGLSAGEKRMLAKARQILVGELALABSTDDAKAETILDEVLAAAS---
M.bovis	103	KVAEVVRDLWRRDQDRGLSAGEKRMLAKARQILVGELALABSTDDAKAETILDEVLAAAS---
M.parascrofulaceum	103	KVAEVVRDLWRRDQDRGLSAGEKRMLAKARQILVGELALABSTDDAKAETILDEVLAAAS---
M.smegmatis	103	KVAEVVRDLWRRDQERGLSAGEKRMLAKARQILVGELALABNTDDAKAETILDEVLAAAS---
M.gilvum	103	KVAEVVRDLWRRDQERGLSAGEKRMLAKARQILVGELALABNTDDAKASILDEVLAAAS--
M.fortuitum	103	KVAEVVRDLWRRDQERGLSAGEKRMLAKARQILVGELALABNTDDAKAETILDEVLAAAS---
M.thermoresistibile	103	KVAEVVRDLWRRDQERGLSAGEKRMLAKARQILVGELALABNTDDAKAETILDEVLAAAS---
M.hassiacum	103	KVAEVVRDLWRRDQERGLSAGEKRMLAKARQILVGELALABNTDDAKAESLLDEVLAAAS---
Actinosynnema	103	KVAEVVRDLWRREKDRGLSAGEKRMLAKARQILVSELALABGTDEDKAEVLLDEVLAAAS---
Saccharothrix	103	KVAEVVRDLWRREKDRGLSAGEKRMLAKARQILVSELALABGTDEDKAEVLLDEVLAAAS---
Saccharopolyspora	103	KVAEVVRDLWRREKDRGLSAGEKRMLAKARQILVSELALABGTDEGKAETLDEVLAAAV---
Streptomyces	103	KVAEVVRDLWRREKDRGLSAGEKRMLAKARQILVSELALABGTDEGKAETLDEVLETATV---
Pseudonocardia	103	KVAEVVRDLWRREKDRGLSAGEKRMLAKARQILVSELALABGTDERAEVLLDEVLAAATA--
M.xanthus_DK1622_CdnL	106	ETAEVVRDLYLLKGDKDLSFLSRKMDTARSTLKPLLAKDCSEDEIVSLLKIFNLA---
M.xanthus_DK1622_CarD	110	GLAEVVRLQVLSELDLTPTKRRELYDNARHILVFLGAEVNAEDAIRTYTFPPGRERPKRTAAEFQREEGDLDLDSDLLGLDSDLDLPSDEEPSEPSEEESSEE
Clustal Consensus		::**::.* . : *. *:.: .** :*: *:: * . : . : . .:: .

|	
Mtb	162 --
M.leprae	165 --
M.bovis	162 --
M.parascrofulaceum	162 --
M.smegmatis	162 --
M.gilvum	162 --
M.fortuitum	162 --
M.thermoresistibile	162 --
M.hassiacum	162 --
Actinosynnema	162 --
Saccharothrix	162 --
Saccharopolyspora	162 --
Streptomyces	163 --
Pseudonocardia	164 --
M.xanthus_DK1622_CdnL	164 --
M.xanthus_DK1622_CarD	220 GESEEAASSEESAPKKRGRPPKAKTEAPEGAEPPAPKKRGRPPKPKPEATAEGAEPPAPKKRGRPPKPKPPEVEGAAPAAPKKRGRPPKVKPPEGES
Clustal Consensus	

318

Figure B-1. Protein sequence alignment of *Mtb* CarD. *Mtb* CarD sequence is aligned with the closest homologs (determined by the BLAST search) and *M. xanthus* CarD and CdnL proteins. Identical residues are colored dark gray and boxed, similar residues are shaded with light gray. *M. xanthus* CarD has additional 135 residues at the C-terminus compared to *Mtb* CarD. The AT-hook DNA binding motif (four 'RGRP' motifs) of *M. xanthus* CarD are shaded in yellow boxes. The secondary structure of *Mtb* CarD is shown above its sequence. The missing residues in the final model are indicted by dashed lines. Alignment is generated by ClustalW. Figure B-1 is related to Figure 6-2.
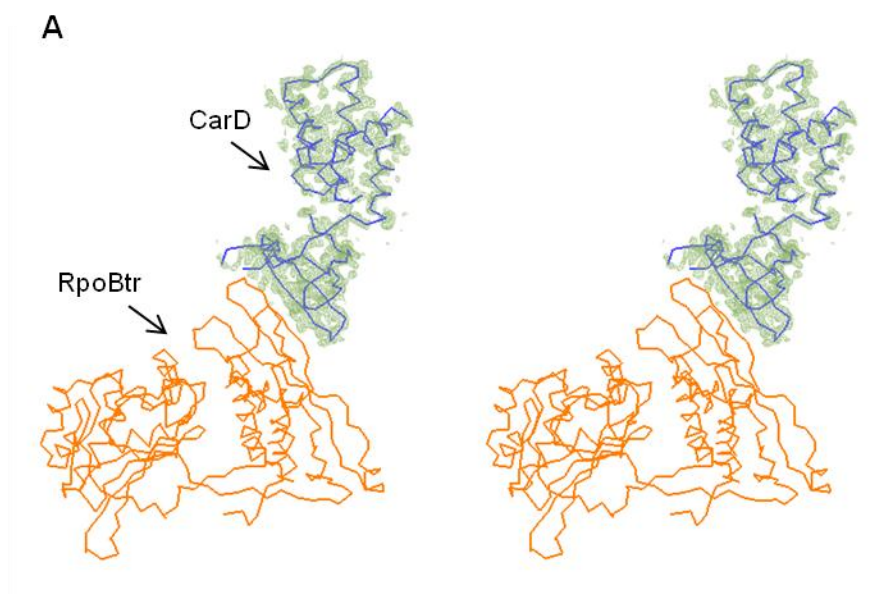
Figure B-2. Stereo images of the CarD/RNAP complex and electrostatic potential surface representation of the CarD/RNAP interface. A) The |Fo|-|Fc| density (shown at 1.8 sigma in green) for CarD at the protein-protein interface was unambiguous after locating RpoBtr in the complex structure by MR. Cα backbone of the final model of the complex is included for clarity. **B)** The 2|Fo|-|Fc| map at 2.0 sigma of the CarD/RNAP interface after refinement. RpoBtr residues are shown in orange; CarD residues are shown in gray sticks. Images were prepared by Pymol. **C)** Wall-eyed stereo image of the CarD/RNAP complex in cartoon representation. **D)** CarD displays a positively charged surface (left), and RNAP has a negatively charged surface (right) at the interface. Electrostatic potential molecular surface calculations were done and images were prepared with PyMol using APBS as the macromolecular electrostatics calculation program. In all panels, RpoBtr is shown in orange, CarD is shown in blue ribbons. Figure B-2 is related to Figure 6-1.
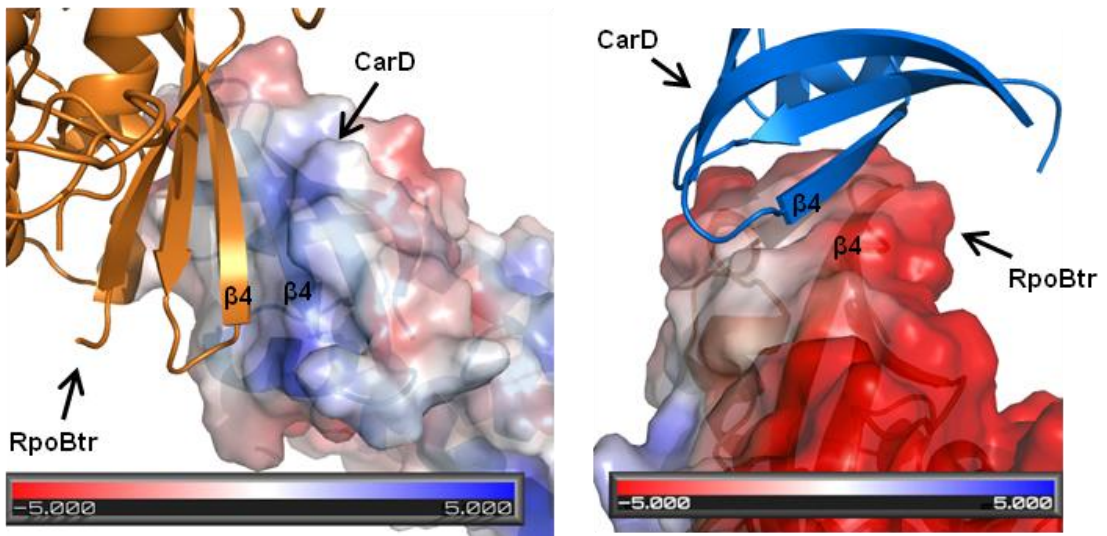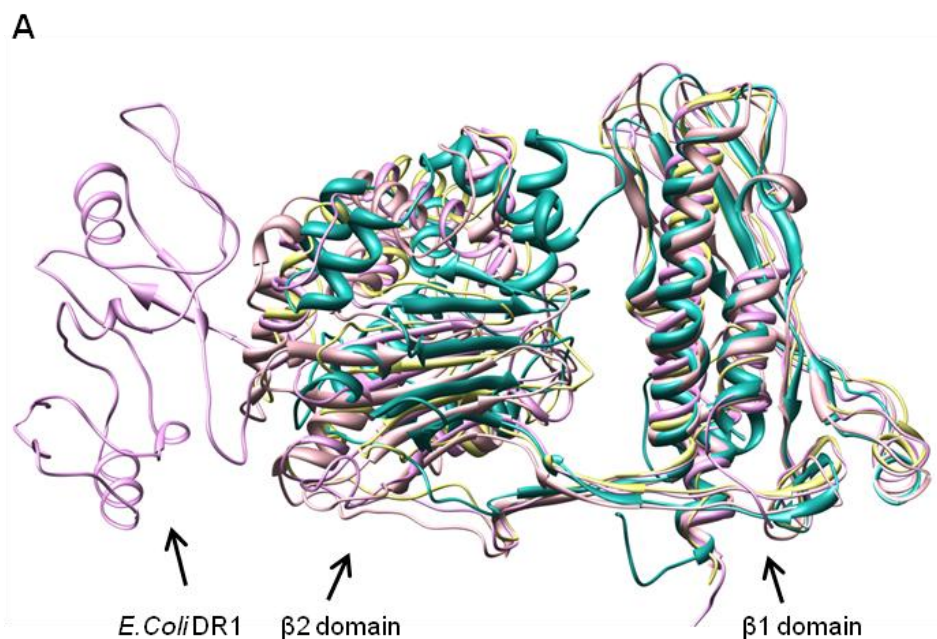
Figure B-2 Continued.

321
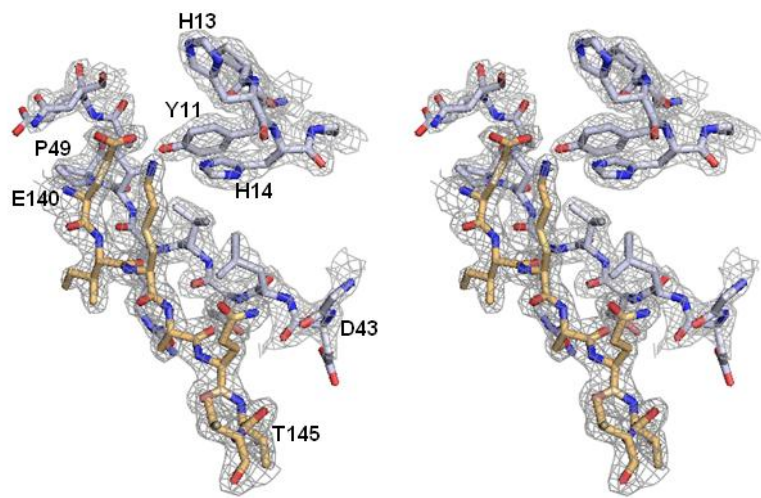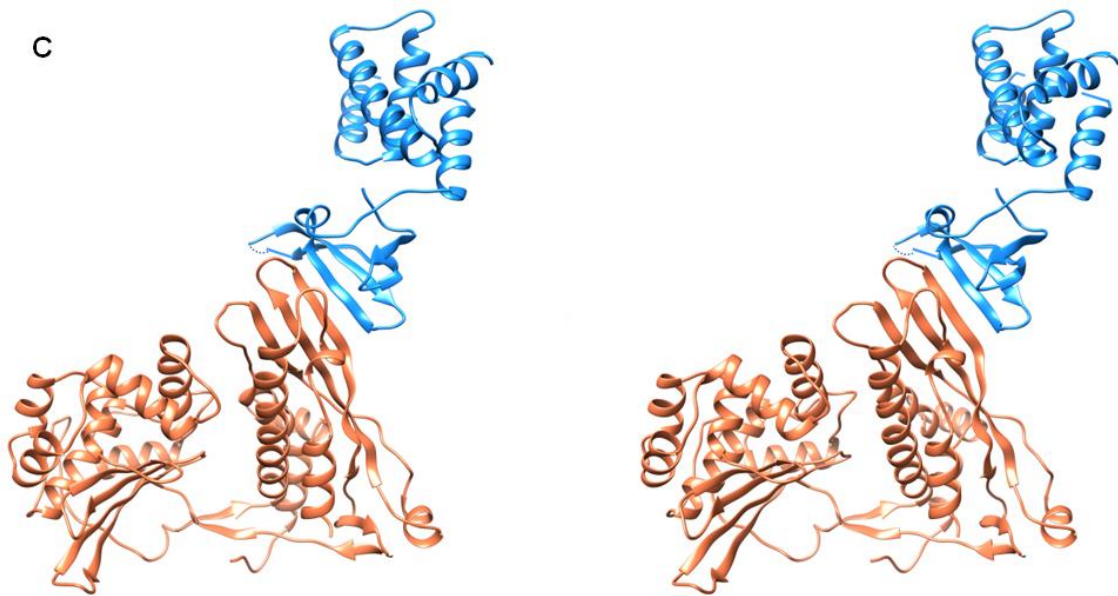
D

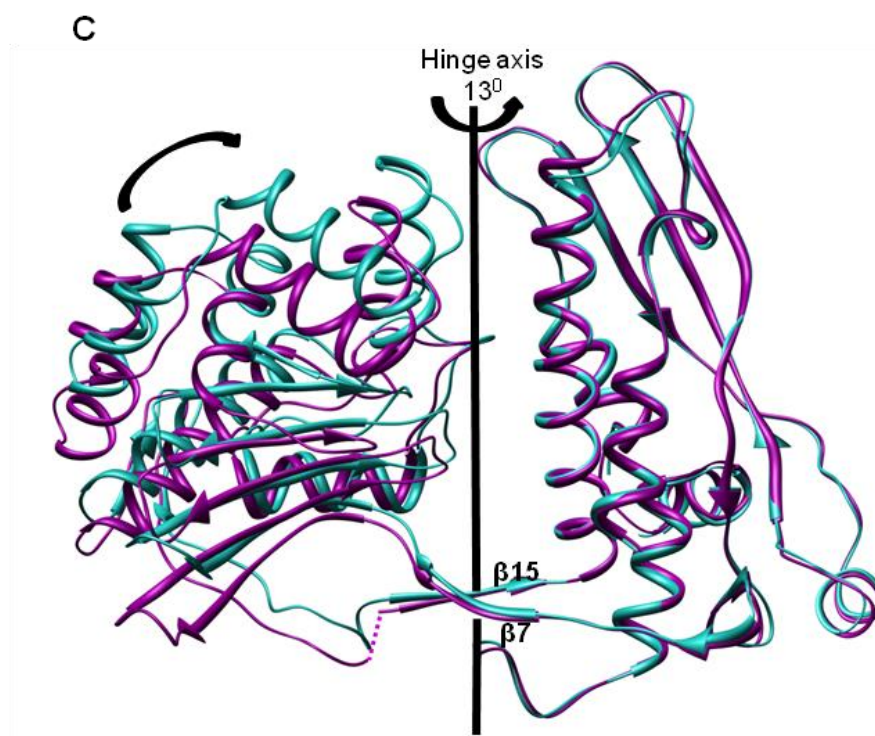**Figure B-2** Continued.

Figure B-3. Comparison of the uncomplexed *Mtb* β1 and β2 domain structure (green) with the *Tth* (dark pink, PDB ID: 2O5I), *Taq* (yellow, PDB ID: 1I6V) and *E. coli* (light pink, PDB ID: 4IGC) RNAP β1 and β2 domain structures. Rest of the *Tth, Taq* and *E. coli* RNAP structures is omitted for clarity. Superposition of the β1 **(A)** and β2 **(B)** domains results ~10 Å RMSD over the Cα atoms of the other domain, which indicates that the relative conformation adopted by the two domains in RpoBtr is different than other bacterial core or holo RNAP structures. **C)** Superposition of the uncomplexed RpoBtr β1 domains of the two molecules in the ASU (RpoBtr_A (magenta) and RpoBtr_B (green)) gives an RMSD of 5.2 Å (over 191 atom pairs) on the Cαs of the β2 domains. The conformational difference observed in the β1-β2 domain-domain orientation can be explained by a 13° rotation around the hinge axis centered on the two stranded anti-parallel β-sheet connecting the two domains. Figure B-3 is related to Figure 6-3.
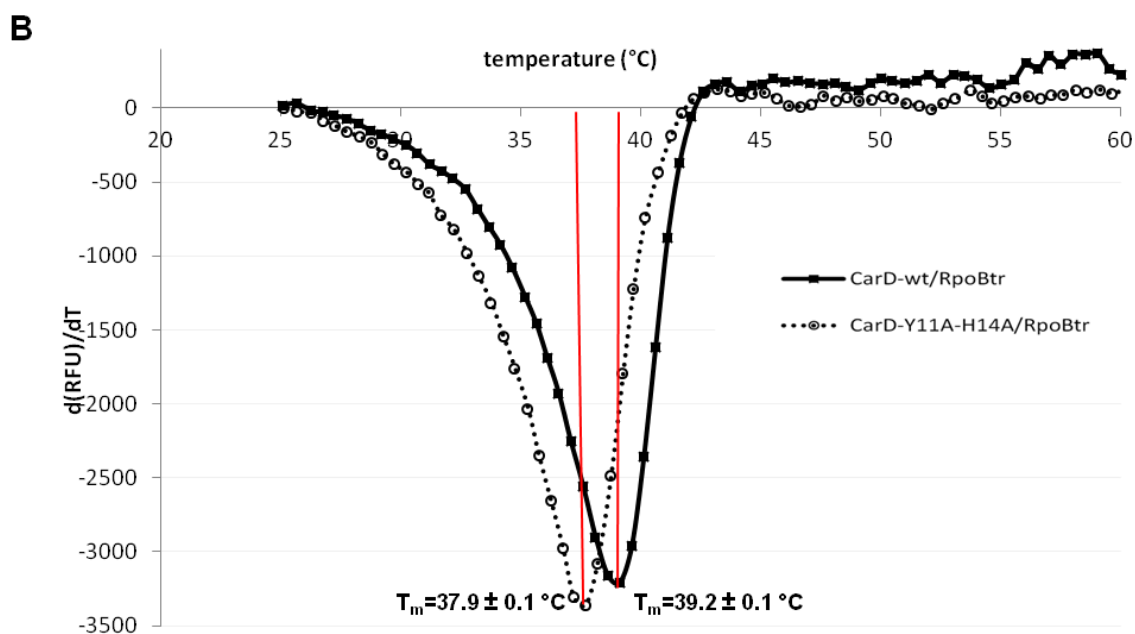
Figure B-3 Continued.
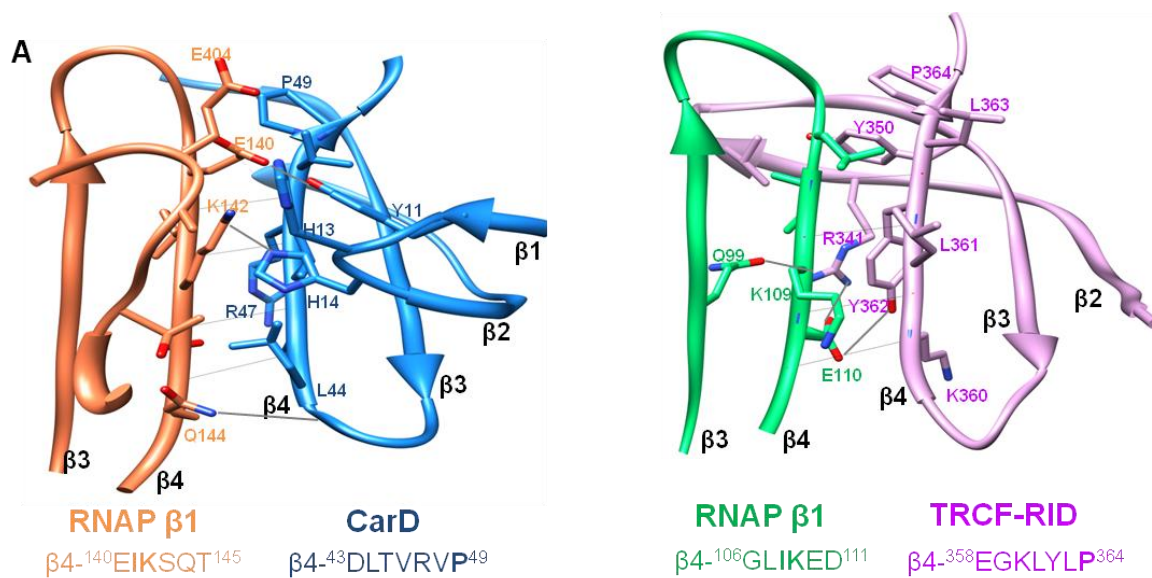
Figure B-4. Comparison of CarD/RNAP interaction with TRCF-RID/RNAP interaction and the DSF experiments on CarD/RpoBtr and CarD-Y11A-H14A/RpoBtr complexes. A) The CarD/RNAP interface (this study, left panel) and the TRCF-RID/RNAP interface (PDB ID: 3MLQ, right panel) are displayed. The sequence of the β4-strands is given below each chain with conserved residues in bold. Important residues for intermolecular interaction are labeled, and H-bonds are shown with gray lines. **B)** Differential scanning fluorimetry (DSF) experiments on CarD/RpoBtr (filled squares, solid line) and CarD-Y11A-H14A/RpoBtr (open circles, dashed line) complexes. Equal amount of each complex (1 μM) in 200 mM Tris pH 7.5, 100 mM NaCl buffer was mixed with 5X Sypro orange dye. The temperature of the samples was changed from 25 to 95 °C at a heating rate of 0.5 °C/min, and the fluorescence was monitored. The first derivative of fluorescence d(RFU)/dT is plotted against temperature and the melting point (Tm, indicated with red lines) was calculated as the lowest point of the first derivative plot. Experiments were repeated five times and standard errors are calculated. Figure B-4 is related to Table 6-2.

B

Figure B-4 Continued.

Supplemental experimental procedures

Expression and purification

For recombinant expression of uncomplexed RpoBtr, native and mutant CarD proteins, the corresponding plasmids were transformed into *E. coli* BL21(DE3) cells. The cells were grown at 37 ºC in the presence of antibiotic to an OD600 of 0.6, and protein expression was induced with 1 mM IPTG at 18 ºC overnight. The cells were harvested, resuspended in buffer A (25 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole, 2 mM β-ME, 2 mM MgCl2, 40 μg/ml DNAse, and protease inhibitor

cocktail) and lysed by French press. After centrifuging the lysate, the supernatant was loaded onto a His Trap FF column pre-equilibrated with buffer A. After washing the column with at least 5 column volumes of buffer A, column bound proteins were eluted with a gradient of buffer B (15-100%) (25 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole, 2 mM β-ME). Fractions containing RpoBtr-NHis or CarD were pooled, dialyzed against buffer C (25 mM Tris 7.5, 100 mM NaCl, 2 mM DTT), concentrated, and loaded onto a Superdex 200 gel filtration column pre-equilibrated with buffer C. Final pure protein was concentrated to 15 mg/ml and stored at -80 ℃.

Selenomethionine (Se-Met) incorporated RpoBtr-NHis was expressed by using a methionine auxotrophic *E. coli* strain B834(DE3) in minimal media supplemented with Se-Met (Doublié, S. (1997). *Methods Enzymol*.276, 523–530). Purification of the Se-Met labeled protein was performed the same way as the native protein, except that the DTT concentration in buffer C was increased to 5 mM.

For co-expression of *Mtb* RNAP β1-β2 domains and *Mtb* CarD, the plasmids RpoBtr-NHis and CarD-notag were cotransformed into *E. coli* Rosetta2(DE3)pLysS cells. The cells were grown at 37 ℃ in the presence of kanamycin and carbenecillin to an OD600 of 0.6 and were induced with 0.75 mM IPTG at 18 ℃ overnight. The cells were harvested by centrifugation at 3000 x g and resuspended in buffer D (25 mM Tris pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 5 mM imidazole, 2 mM β-ME, 2 mM MgCl2, 40 µg/ml DNAse, and protease inhibitor cocktail), lysed by French press, centrifuged at 16 K rpm for 45 min, and the supernatant was loaded onto a His Trap FF column pre-equilibrated with buffer D. The column was washed with at least 5 column volumes of

buffer D and a stepwise gradient (5-25%) of buffer E (25 mM Tris pH 7.5, 200 mM NaCl, 300 mM imidazole, 2 mM β-ME). Column bound proteins were eluted with a gradient of buffer E (25-100%). Fractions containing the RpoBtr-NHis:CarD complex were pooled, concentrated, and loaded onto a Superdex 200 gel filtration column pre-equilibrated with buffer C (25 mM Tris 7.5, 100 mM NaCl, 2 mM DTT). The RpoBtr-NHis:CarD complex eluted as a single peak from the size exclusion column. Co-elution of RpoBtr and CarD from the IMAC and size-exclusion columns was verified by SDS-PAGE. Protein concentrations were estimated using the extinction coefficients of 17000 $M^{-1}cm^{-1}$ for CarD, 32490 $M^{-1}cm^{-1}$ for RpoBtr, and 49000 $M^{-1}cm^{-1}$ for RpoBtr/CarD complex at 280 nm using NanoDrop. Finally, the purified complex was concentrated to 10 mg/ml and stored at -80 ℃ for further use.

Data collection and structure determination

Molecular replacement (MR) trials to determine the structure of *Mtb* RpoBtr using homologous *E. coli, Tth*, and *Taq* RNAP structures did not yield a solution. The structure of the *Mtb* RNAP β-subunit β1-β2 domains was solved by single-wavelength anomalous diffraction (SAD) using RpoBtr-NHis-SeMet crystals. Crystals belonged to the $P2_12_12_1$ space group and diffracted to 2.9 Å. Data was collected at the selenium peak wavelength (0.979 Å), processed with Denzo/Scalepack (Minor et al. (2006) *Acta Crystallogr D Biol Crystallogr.* 62, 859-66) and Phenix AutoSol, and Autobuilt modules were used to determine the initial phases and obtain the first atomic model (Adams et al. (2010) *Acta Crystallogr D Biol Crystallogr.* 66, 213-21). Subsequently, resolution of the

RpoBtr structure was improved to 2.5 Å by diffraction data obtained from native (non-SeMet) RpoBtr-NHis crystals, which also belonged to the $P2_12_12_1$ space group with two molecules in the ASU. The structure of native RpoBtr-NHis was obtained by MR using the RpoBtr-NHis-SeMet protein structure as the search model. The structure comprised residues 47-426 of the β-subunit and included six additional residues at the N-terminus between the His-tag and the first methionine residue. Refinement was performed with Phenix Refine. NCS restraints were applied to the atomic positions and individual B-factors subsequent to simulated annealing and rigid-body refinements. Iterative manual model building was done with COOT (Emsley et al. (2010) *Acta Crystallogr D Biol Crystallogr.* 66, 486-501). Water molecules were added to |Fo-Fc| peaks >3σ and the model was further refined using Phenix Refine. The final model had Rwork and Rfree values of 0.21 and 0.26, respectively.

The RpoBtr-NHis:CarD complex crystals belonged to the $C222_1$ space group and the diffraction data to 2.1 Å resolution was processed with Denzo/Scalepack. The structure was solved by MR using the RpoBtr β1 and β2 domains as two different search ensembles (Phaser, CCP4) (McCoy et al. (2007) *J Appl Crystallogr.* 40, 658-674; Winn et al. (2011) *Acta Crystallogr D Biol Crystallogr.* 67, 235-42). After locating one copy of RpoBtr in the ASU, the additional |Fo-Fc| density for the CarD N-terminal domain β-strands and the C-terminal domain helices were visible with enough clarity to allow us to build CarD manually. After a few cycles of rigid-body and simulated-annealing refinements, the |Fo-Fc| density for the side chains gradually improved. The model was further refined with Phenix Refine and water molecules were added to |Fo-Fc| peaks

>3σ. In the final stages of refinement, TLS refinement (four TLS groups: β1 domain, β2 domain, CarD N-terminal domain, and CarD C-terminal domain) was included. Iterative cycles of model building and refinement resulted in a final model which included one RpoBtr:CarD complex in the ASU with an Rwork= 0.20 and an Rfree= 0.23. Data collection and processing statistics are provided in Table 6-1.

APPENDIX C

AUTHOR CONTRIBUTION FOR CHAPTERS II-VI

For the publications reprinted in Chapter II, Gulten G. worked on cloning of constructs, InhA-EthA and Dhfr-KatG co-expressions, protein purification, InhA enzymatic assays, crystallization and data analysis.

For the publication reprinted in Chapter III, the first two authors contributed equally to the work. Jacobs and Kremer labs performed genetics, kinase assays, and mass spectrometry analysis. Gulten G. did cloning, expression, purification, InhA enzymatic assays and crystallographic studies with the wild-type and T266D, T266E, and T266A InhA mutant proteins, and also contributed in manuscript preparation.

Chapter IV was done in collaboration with GlaxoSmithKline (GSK). Gulten G. performed the crystallographic studies. GSK performed the high-throughput screening, hit identification and optimization, enzymatic assays, *in vivo* and pharmacokinetic studies. The compounds used for crystallographic studies were supplied by GSK. This manuscript is going to be submitted for publication.

For the publication reprinted in Chapter V, JSF and GAS synthesized the triclosan analogs, Jacobs lab did genetic and *in vivo* studies, RL did enzymatic assays, FW did data collection, processing, enzymatic assays and manuscript preparation, Gulten G. did protein purification, crystallization of the InhA-inhibitor complexes, data collection, processing and manuscript review.

For the publication reprinted in Chapter VI, Gulten G. performed all the experiments, data collection and analysis, and wrote the manuscript.