

**STRUCTURAL STUDIES TARGETING ENZYMES INVOLVED IN THE
BIOSYNTHESIS OF COMPLEX CELL SURFACE LIPIDS OF *Mycobacterium
tuberculosis***

A Dissertation

by

ANUP AGGARWAL

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. Sacchetti
Committee Members,	Gregory D. Reinhart
	Paul Straight
	J. Martin Scholtz
Head of Department,	Gregory D. Reinhart

May 2014

Major Subject: Biochemistry

Copyright 2014 Anup Aggarwal

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) is a deadly pathogen that causes the infectious diseases Tuberculosis (TB) in humans. The emergence of drug resistant TB has created an urgent need for the development of new chemotherapeutics that can counter the resistance, and combat the spread of TB by shortening the treatment. One of the major determinants of the virulence of *Mtb* is its thick cell envelope which contains a wide variety of complex lipids and carbohydrates. Thus, the enzymes of cell wall lipid biosynthesis in *Mtb* represent attractive targets for drug discovery. In this dissertation, our studies on the structural and functional characterization of three different proteins from complex lipid synthesis pathways in the *Mtb* are presented.

Firstly, we have crystallized the 225 kDa methyl-branched fatty acid synthesizing mycocerosic acid synthase (MAS) from *Mtb*. We developed methods to improve the diffraction properties of the MAS crystals. A medium resolution dataset allowed initial crystallographic characterization of MAS. The results suggest that these methods could be adapted for crystallographic studies of other type I polyketide synthase from *Mtb*.

Secondly, the crystal structure of *Mtb* acyl carrier protein synthase (AcpS) is presented. This important enzyme activates the acyl carrier protein domains of the FAS I and FAS II systems in *Mtb*. The structure revealed that the *Mtb* AcpS adopts two pH dependent different conformations – open and closed. *In vitro* studies established a structure-function relationship by demonstrating a pH dependent activity profile for AcpS.

Thirdly, we have determined the structural basis for the inhibition of the thioesterase (TE) domain of Pks13 by benzofuran class of inhibitors. Pks13 represents a novel target in the essential mycolic acid biosynthesis pathway in *Mtb*. The crystal structure established that the TE domain belonged to the α/β -hydrolase class of enzymes. We also determined the structural basis for the mechanism of resistance of TE domain towards these inhibitors. We used the structure-guided approach to design new analogs of the benzofuran scaffold, and our best analog exhibited 12-fold increased inhibitory activity against *Mtb*. We also established that the benzimidazole scaffold containing analogs bind to the TE domain in a manner similar to the benzofurans.

DEDICATION

This dissertation is dedicated to my beloved mom and pop, and my dear sister for their unconditional love, constant support and encouragement throughout my life.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Jim Sacchettini for his guidance, generous support and constant encouragement throughout my graduate career. His constant stream of scientific suggestions and contagious enthusiasm for science were invaluable for getting me through the many ups and downs of my Ph.D. research. I would also like to thank the members of my graduate committee, Dr. Gregory D. Reinhart, Dr. J. Martin Scholtz and Dr. Paul Straight, for their helpful suggestions and scientific advice throughout the duration of this research. I also thank Dr. Tom Ioerger for helpful suggestions and advice. Additionally, I would like to thank Dr. Sanghamitra Dey, Dr. A. Arockiasamy and Dr. K. Gokulan for their mentorship and fruitful collaborations.

I would like to extend special thanks to my dearest friend Gulcin. Thank you for all your help and for being the friend when I needed one the most. Many thanks also for all the jokes, emotional support, constant encouragement, and for offering great advice, both scientific and outside of science.

Thanks also go to my friends in the Department of Biochemistry for making me feel welcome and making my time at Texas A&M University a great experience. I would like to thank the members of my lab, past and present, for their support and constructive criticism. I'd like to thank Romano for his support and encouragement, and also for intellectually stimulating discussions on science and spirituality. Thanks for sharing your knowledge with me.

Thank you to Amber, Ashley, Camille and Stephanie for your kindness, true friendship and good food. You've been my family away from home, and I'm very grateful to have you as my friends.

Finally, I wish to convey my deepest gratitude to my family and relatives. Thanks mom and dad for all the encouragement, support and wise counsel over all these years. Your example has helped me to become a better person. I would also like to express my sincere gratitude and thanks to my Guru Ji for shining a spiritual guiding light in my life. I hope I've made you all proud.

NOMENCLATURE

A	Alanine
ACP	Acyl carrier protein
AcpS	Acyl carrier protein synthase
ADP	Adenosine-5'-diphosphate
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
ASU	Asymmetric unit
AT	Acetyl transferase
ATP	Adenosine-5'-triphosphate
BODIPY	Boron-dipyrromethene
CoA	Coenzyme A
Cys	Cysteine
D	Aspartate
DH	Dehydratase
DMSO	Dimethylsulfoxide
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E	Glutamic acid

ER	Enoyl reductase
F	Phenylalanine
G	Glycine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
H	Histidine
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	Histidine
HTS	High throughput screening
I	Isoleucine
Ile	Isoleucine
IPTG	Isopropyl-1-thio -D-galactopyranoside
kbps	Kilo base pairs
kDa	Kilo Dalton
KR	β -ketoacyl reductase
KS	β -ketoacyl synthase
Lys	Lysine
M	Methionine
MAS	Mycocerosic acids synthase
Met	Methionine
Mg	Magnesium

MIC	Minimum inhibitory concentration
MR	Molecular replacement
Mtb	Mycobacterium tuberculosis
MUH	Methylumbelliferyl heptanoate
N	Asparagine
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PDIM	Phthiocerol dimycocerosates
PEG	Polyethylene glycol
PGL	Phenolic glycolipid
Phe	Phenylalanine
PKS	Polyketide synthase
PPTase	4'- phosphopantetheinyl transferase
Q	Glutamine
R	Arginine
RMSD	Root mean square deviation
S	Serine
SAR	Structure-activity relationship
SDS	Sodium dodecyl sulphate
Ser	Serine

SeMet	Selenomethionine
T	Threonine
TB	Tuberculosis
TE	Thioesterase
Thr	Threonine
Tyr	Tyrosine
V	Valine
Val	Valine
VAST	Vector Alignment Search Tool
Y	Tyrosine

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
NOMENCLATURE	vii
TABLE OF CONTENTS	xi
LIST OF FIGURES	xv
LIST OF TABLES	xviii
CHAPTER I INTRODUCTION AND LITERATURE REVIEW	1
Background	1
<i>Mtb</i> cell envelope	6
Mycolic acids of mycobacteria	9
Mycolic acid biosynthesis	12
Polyketide synthases of <i>Mtb</i>	15
Methyl-branched cell surface lipids of <i>Mtb</i>	22
Biosynthesis of phthiocerol dimycocerosate esters (PDIMs) in <i>Mtb</i>	26
Activation of acyl carrier protein (ACP) domains in <i>Mtb</i>	30
CHAPTER II PURIFICATION AND INITIAL CRYSTALLOGRAPHIC CHARACTERIZATION OF MYCOCEROSIC ACID SYNTHASE FROM <i>Mycobacterium tuberculosis</i>	33
Introduction	33
Material and Methods	36
Cloning, expression and purification of <i>Mtb</i> MAS	36
Determination of 4'-PPT modification in recombinantly expressed MAS	39
Determination of oligomeric state of MAS	40
Crystallization of <i>Mtb</i> MAS	40
In situ proteolysis for MAS crystallization	41

	Page
Crystallization of MAS using heterogeneous nucleating agents	41
Optimization of MAS crystals by counter-diffusion and seeding techniques	42
Growing MAS crystals in gel	43
Post-crystallization treatment of MAS crystals	44
Crystallographic screening and data collection	46
Results and Discussion	47
Cloning and purification of <i>Mtb</i> MAS	47
Oligomeric state of <i>Mtb</i> MAS	54
Enhancement of 4'-PPT modification in recombinantly expressed MAS	55
Crystallization of <i>Mtb</i> MAS	57
Summary and Future Directions	71

CHAPTER III <i>Mycobacterium tuberculosis</i> ACYL CARRIER PROTEIN SYNTHASE ADOPTS TWO DIFFERENT pH-DEPENDENT STRUCTURAL CONFORMATIONS	73
Synopsis	73
Introduction	74
Material and Methods	78
Cloning, expression and purification of <i>C. ammoniagenes</i> AcpS and <i>Mtb</i> AcpS	78
Cloning, expression and purification of <i>Mtb</i> ACP (AcpM)	79
AcpS post-translational modification fluorescence assay	80
Crystallization and data collection	81
Structure determination	87
Results and Discussion	89
Sequence and structural comparison of <i>Mtb</i> AcpS and <i>C. ammoniagenes</i> AcpS with other bacterial AcpSs	92
Structural differences between the open and closed structural forms of <i>Mtb</i> AcpS	96
The effect of pH on the structure of the $\alpha 2$ helix	97
Open and closed conformations of the connecting loop between helices $\alpha 3$ and $\alpha 4$	99
Active site of the apo-AcpS structure	102
The binary complex of AcpS and CoA	103
<i>Mtb</i> AcpS can selectively transfer BODIPY-labeled CoA to <i>Mtb</i> apo-ACP	104
Comparison of the <i>B. subtilis</i> AcpS-ACP complex with a proposed model of the <i>Mtb</i> AcpS-AcpM complex	108

	Page
Conclusions	110
Additional Unpublished Results.....	111
Fragment based screening for ligands of <i>Mtb</i> AcpS	114
Additional Unpublished Experimental Methods.....	119
Differential scanning fluorimetry (DSF) of <i>Mtb</i> AcpS.....	119
Future Directions.....	120
Prospects for PPTases as targets for inhibitor discovery	120
 CHAPTER IV TARGETING <i>MTB</i> PKS13 THIOESTERASE DOMAIN FOR ANTITUBERCULAR DRUG DISCOVERY	123
Synopsis	123
Introduction	124
Experimental Methods	127
Cloning and overexpression of <i>Mtb</i> Pks13 TE domain constructs.....	127
Purification of Pks13 TE domain	127
Crystallization and soaking with ligands	128
Data collection and processing.....	129
Determination of TE1451 structures and model refinement.....	130
Enzyme assay	132
IC ₅₀ determination.....	133
Whole cell and cytotoxicity testing.....	134
Results.....	136
Discovery of benzofuran 1 and identification of Pks13 as its target using forward chemical genetics.....	136
Biochemical consequences of TE domain mutations D1607N and D1644G.....	139
Preliminary SAR by inhibition studies with analogs of 1	140
Structural basis of inhibition of TE1451 by 1	142
Active site and binding mode of 1	146
Structural basis for resistance of D1607N and D1644G mutants to 1	153
Structural basis for inhibition of TE1451 by analogs of 1	154
Discussion	159
Additional Unpublished Results.....	162
Extended SAR with the newly synthesized analogs of 1	162
Screening of the whole cell active compounds for identification of novel inhibitors	171
Summary	174
 CHAPTER V CONCLUSIONS	176

	Page
REFERENCES	180

LIST OF FIGURES

FIGURE		Page
1-1	Schematic representation of the <i>Mtb</i> cell envelope.....	8
1-2	Mycolic acids of <i>Mtb</i>	10
1-3	Schematic of mycolic acid biosynthesis pathway in <i>Mtb</i>	14
1-4	Domain organization and reactions of FAS and PKS systems	16
1-5	Chemical structures of multiple methyl-branched fatty acid containing lipids from <i>Mtb</i>	23
1-6	Genetic organization of the PDIM synthesis locus in <i>Mtb</i>	26
1-7	Schematic of PDIM biosynthesis pathway in <i>Mtb</i>	28
1-8	Posttranslational modification of the ACP domain.....	32
2-1	Crystal dehydration methods.....	46
2-2	PCR amplification of MAS	49
2-3	Purification of MAS by Ni ²⁺ affinity chromatography	51
2-4	Gel filtration of MAS	52
2-5	Expression analysis of the ACP domain truncation constructs of MAS	54
2-6	Analysis of 4'-PPT modification in recombinantly expressed MAS.....	56
2-7	Analysis of MAS proteolysis by chymotrypsin	59
2-8	Initial crystallization hit of MAS	61
2-9	Optimization of MAS crystals.....	62
2-10	Initial diffraction from MAS crystals	63
2-11	Improvement in the diffraction properties of MAS crystals	66

FIGURE	Page
2-12 Post-crystallization treatment of MAS crystals.....	67
3-1 Overview of the <i>C. ammoniagenes</i> and <i>Mtb</i> AcpS structures with ribbon representation of the homotrimer.....	91
3-2 Sequence alignment of AcpS	93
3-3 Backbone superposition of the AcpS enzymes	95
3-4 Overview of the α 2-helix movement and the open and closed conformations of the α 3– α 4 connecting loop.....	98
3-5 ADP binds in the dimer interface between two subunits	105
3-6 AcpS post-translational modification activity assay	106
3-7 Comparison of AcpS-AcpM complex	109
3-8 Schematic representation of DSF experiment.....	112
3-9 Typical fluorescence intensity plots from DSF.....	113
3-10 A representative plot of the compound titration assay.....	115
4-1 Chemical structure of benzofuran 1	126
4-2 Comparison of IC ₅₀ values of 1 against wt TE1451 and D1607N, D1644G mutants.....	139
4-3 Cytotoxicity profile of 1 and its analogs	142
4-4 Overall structure of Pks13 TE domain.....	144
4-5 Overall fold and structural elements of the TE domain	145
4-6 Identification of the catalytic residues of TE1451	146
4-7 Surface groove on the lid domain in TE1451 structure	148
4-8 Polypropylene glycol binding in the TE1451 active site	149

FIGURE	Page
4-9 Cartoon representation of the binding interactions of 1 with the TE lid domain.....	152
4-10 Structural basis for resistance of the TE1451:D1607N mutant to 1	155
4-11 Fitting of the inhibitors in the TE1451-analog complex structures	156
4-12 Binding mode of analogs of 1 to TE domain	158
4-13 Binding mode of methylamide containing analog MKP-128	167
4-14 Binding mode of benzimidazole analog AA-8.....	172

LIST OF TABLES

TABLE		Page
1-1	Current recommendations for TB treatment	3
1-2	Taxonomic classification of <i>Mtb</i>	4
1-3	Polyketide synthases of <i>Mtb</i> and their predicted domain organizations	18
2-1	Crystal dehydration protocol	45
2-2	Data-collection and processing statistics for MAS crystals	70
3-1	List of ACPS structures from several pathogenic bacteria and their sequence similarity to <i>Mtb</i> AcpS	77
3-2	Anomalous data-collection and phasing statistics for <i>C. ammoniagenes</i> apo-AcpS	84
3-3	Data-collection, refinement and geometry statistics	85
3-4	Fragment hits from DSF screening of <i>Mtb</i> AcpS	115
4-1	Data collection and refinement statistics for TE1451	131
4-2	Data collection and refinement statistics for analogs of 1	133
4-3	Kinetic parameters of wild-type TE1451 and D1607N, D1644G mutants	137
4-4	Preliminary SAR of 1 and its analogs	138
4-5	Interactions of compound 1 with the residues of TE domain	151
4-6	Pks13 TE domain SAR of the synthesized benzofuran analogs of 1	164
4-7	Pks13 TE domain SAR of the benzimidazole analogs of 1	171
4-8	Hits from the whole cell actives screening of TE1451	173

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

BACKGROUND

Tuberculosis (TB) is a deadly infectious disease that remains a serious threat to the public health worldwide. Despite the availability of active antibiotics, TB continues to cause significant morbidity and mortality globally. According to the latest World Health Organization (WHO) report, in 2012 there were about 9 million new cases and about 1.3 million deaths due to TB. In addition, it has been estimated that nearly two billion people worldwide are latently infected with the bacillus *Mycobacterium tuberculosis* (*Mtb*), the causative agent of TB. In this infected population, about 5-10% of the immunocompetent individuals are expected to develop active disease; although the population retains a lifetime risk of reactivation of the disease. Furthermore, the latently infected population represents a significant reservoir of this obligate human pathogen. Therefore, the WHO recommends prompt detection and treatment of the infectious cases through implementation of the DOTS (Directly Observed Treatment, Short-Course) strategy for TB control.

The current therapy for TB involves a minimum of six months of intensive treatment with four first-line antitubercular drugs isoniazid (INH), rifampicin (RIF), ethambutol (EMB) and pyrazinamide (Table 1-1) (Zumla et al., 2013b). With proper

compliance and access to good health care, this treatment regimen can achieve a successful cure rate of >95%. However, patient non-compliance due to the long duration of the treatment coupled with resource poor settings in places with the highest burden of TB has resulted in the emergence of multidrug-resistant (MDR-) and extensively drug-resistant (XDR-) TB. MDR-TB is resistant to INH and RIF, as a result, its treatment is much longer (20 months), expensive and uses toxic drugs. XDR-TB is even more difficult to treat as the *Mtb* bacilli are resistant to the first-line drugs INH and RIF, and to the second-line fluoroquinolones and one of the three injectable TB drugs (Zumla et al., 2013a). The emergence of resistant strains of *Mtb* and the co-infection of HIV patients with TB can seriously undermine global efforts to control TB pandemic. Thus, there is an urgent need for the discovery and development of novel anti-TB agents that can help shorten the treatment duration and are effective against drug-resistant TB.

Mtb is one of the most successful and persistent human pathogen in the world. It was identified by Robert Koch in 1882 as the principal etiological agent of TB in humans. *Mtb* belongs to class Actinobacteria and is a member of the *Mycobacterium tuberculosis* complex which includes *M. bovis*, *M. bovis* BCG, *M. africanum*, *M. canetti*, *M. caprae*, *M. microti*, and *M. pinnipedii* (Table 1-2) (Shenai and Rodrigues, 2011). *Mtb* is an aerobic, rod-shaped bacterium that has no capsule or flagellum. It is classified as a non-sporulating, weak Gram-positive bacterium with a G+C-rich genome. The *Mtb* cells have a thick waxy cell wall that retains Carbol fuchsin dye even in the presence of acidic alcohol, giving the bacterium its acid fast property. It is a slow-growing intracellular and intravacuolar pathogen with a doubling time of ~24 h.

Table 1-1. Current recommendations for TB treatment.

Type of Infection	Recommended Regimen	Comments
Active disease		
Newly diagnosed cases that are not multidrug-resistant	Isoniazid, rifampin, ethambutol, and pyrazinamide for 2 mo (intensive phase), followed by isoniazid and rifampin for 4 mo (continuation phase)	Pyridoxine supplementation recommended to prevent isoniazid-induced neuropathy
Multidrug-resistant disease	Four second-line antituberculosis drugs (as well as pyrazinamide), including a fluoroquinolone, a parenteral agent, ethionamide or prothionamide, and either cycloserine or para-aminosalicylic acid if cycloserine cannot be used	Initial treatment based on local disease patterns and pending drug-susceptibility results; later-generation fluoroquinolones (e.g., moxifloxacin or levofloxacin) preferred
Latent infection		
	Isoniazid at a dose of 300 mg daily for at least 6 mo and preferably for 9 mo	Recommended for 9 mo or more in HIV-infected persons; daily administration for 6 mo also an option but with lower efficacy; extension to 36 mo further reduces risk among HIV-positive patients in regions in which tuberculosis is endemic
	Isoniazid at a dose of 900 mg plus rifapentine at a dose of 900 mg weekly for 3 mo (directly observed therapy)	Studied with directly observed therapy in predominantly HIV-uninfected persons; higher completion rates and equal efficacy, as compared with isoniazid for 9 mo
	Rifampin at a dose of 600 mg daily for 4 mo	Shown to be effective in persons with silicosis
	Isoniazid at a dose of 300 mg plus rifampin at a dose of 600 mg daily for 3 mo	Effective alternative for HIV-infected persons
	Isoniazid at a dose of 900 mg plus rifampin at a dose of 600 mg twice weekly for 3 mo	Another effective alternative for HIV-infected persons

Table 1-2. Taxonomic classification of *Mtb*.

Kingdom	Bacteria
Phylum	Actinobacteria
Order	Actinomycetales
Suborder	Corynebacterineae
Family	Mycobacteriaceae
Genus	<i>Mycobacterium</i>
Clinically relevant species	<i>M. tuberculosis</i> complex. <i>M. leprae</i> , <i>M. avium</i> , <i>M. kansasii</i> , <i>M. marinum</i> , <i>M. ulcerans</i> , <i>M. fortuitum</i> complex, <i>M. abscessus</i> . <i>M. intracellulare</i> , etc.

Transmission of TB occurs from person to person via aerosolized droplets containing *Mtb* that are generated when an individual with an active disease coughs or sneezes. When these droplets are inhaled, the tubercular bacilli enter the terminal air spaces of the lungs to establish the primary site of infection. After the bacilli have lodged into the lungs, they are phagocytosed by the resident alveolar macrophages, which recruit other immune cells through cytokine and chemokine signaling. These cells surround infected macrophages to form containment structures called the granuloma, or tubercle, which is the pathological hallmark of TB (Russell, 2007). Once the macrophage engulfs *Mtb*, there can be four potential outcomes: (i) the initial immune response can be completely effective in eradicating the bacteria, and in such a case the patient will not develop active TB; (ii) the bacteria can immediately establish primary

TB infection and begin to multiply and grow in the lungs of the infected person; (iii) the bacilli may enter a non- or slowly-replicating dormant state in the granuloma and do not cause the disease, a condition referred to as the latent infection; (iv) there could be reactivation of the latent bacteria causing a post-primary infection (Schluger and Rom, 1998). However, the eventual outcome of the TB infection and its clinical manifestation is strongly dependent on the host immune response and the virulence of *Mtb*, both of which are influenced by the many complex lipid effectors in the mycobacterial cell envelope (Astarie-Dequeker et al., 2010). As the cell envelope is critical for the bacterial physiology, targeting the enzymes involved in its biosynthesis has been an effective strategy for promoting rapid killing of the bacterial cells, as exemplified by the mechanism of action of β -lactam antibiotics. Similarly, the first-line TB drugs INH and EMB target enzymes essential for the biosynthesis of the *Mtb* cell wall core (Banerjee et al., 1994; Belanger et al., 1996). However, increasing development of resistance to these drugs has led to intensive efforts over the last few decades for elucidating the biosynthesis pathways of several pathogenesis related cell surface lipids of *Mtb* (Neyrolles and Guilhot, 2011). This understanding of the biogenesis and the structural organization of the *Mtb* cell envelope is critical for the research focused on targeting cell wall biogenesis for antitubercular drug discovery.

***Mtb* cell envelope**

The cell envelope of *Mtb* is the primary interface between the bacterium and its surrounding environment including the host. Therefore, to thrive under the harsh intracellular environment of the human macrophage, the bacterium produces a uniquely thick cell wall containing a wide range of complex lipids that also impart an intrinsic resistance to many antimycobacterial agents (Brennan and Nikaido, 1995). Due to its importance for the virulence and the survival of the bacterium, the *Mtb* cell wall has been studied intensively to determine its unique architecture and chemical composition (Brennan, 2003). Application of advanced NMR and mass spectrometric techniques has resulted in the identification and complete chemical characterization of many novel complex metabolites in the *Mtb* envelope (Chopra and Gokhale, 2009). The genetic and biochemical investigations have also furthered our understanding of the synthesis and the distribution of complex lipids like PDIMs, mycolic acids and sulfolipids in the mycobacterial envelope (Chopra and Gokhale, 2009; Gokhale et al., 2007; Portevin et al., 2004). These studies have revealed many of the previously unidentified enzymes and transporters involved in the mycolic acid and PDIM biosynthetic pathways that can provide highly prioritized targets for the drug discovery against TB (Favrot and Ronning, 2012). In addition, sequencing of the *Mtb* genome has revealed many new genes specifically involved in the cell envelope biogenesis, thereby identifying more potential targets for the design and development of new anti-tubercular drugs (Cole, 1999). These analytical, biochemical and genetic studies have established that the

mycobacterial cell envelope exists as a highly complex network of sugars, lipids and proteins.

The cell envelope of *Mtb* is a lipid rich multilayered structure composed of a plasma membrane surrounded by a covalently linked cell wall skeleton and a polysaccharide-rich capsule (Fig. 1-1). The mycobacterial cell wall core is composed of three covalently linked macromolecules – peptidoglycan, arabinogalactan and mycolic acids, termed the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex (Brennan, 2003) (Fig. 1-1). The peptidoglycan (PG) layer is a polymer of the repeating units of β -1,4-linked N-acetylglucosamine and N-acetyl/glycolylmuramic acid which are cross-linked *via* L-alanyl-D-isoglutaminyl-*meso*-diaminopimelyl-D-alanine tetrapeptides attached to the muramic acid. The arabinogalactan (AG) is a branched chain heteropolysaccharide composed of covalently linked D-furanosyl galactan and arabinan polymers. The AG is linked to the PG through the reducing residue of the galactan chain, whereas the α -D-arabinofuranosyl residues at the nonreducing terminus of the arabinan chains are esterified to the mycolic acids, which are very long chain α -branched and β -hydroxylated fatty acids that are unique to mycobacteria (McNeil et al., 1990, 1991). In addition, the *Mtb* mAGP layer is surrounded on the exterior by a mobile layer consisting of free glycolipids, carbohydrates, and proteins like porins and the antigen 85 complex. The lipids of this outer layer include free mycolic acids, phenolic glycolipids (PGL), phthiocerol dimycocerosates (PDIM), diacyl trehaloses (DAT), sulfolipids and trehalose esters of mycolates called trehalose monomycolate (TMM) and dimycolate (TDM), also referred

to as the cord factor. These intercalate into the cell wall through their interactions with the mycolyl moiety of the mAGP complex (Brennan, 2003) (Fig. 1-1).

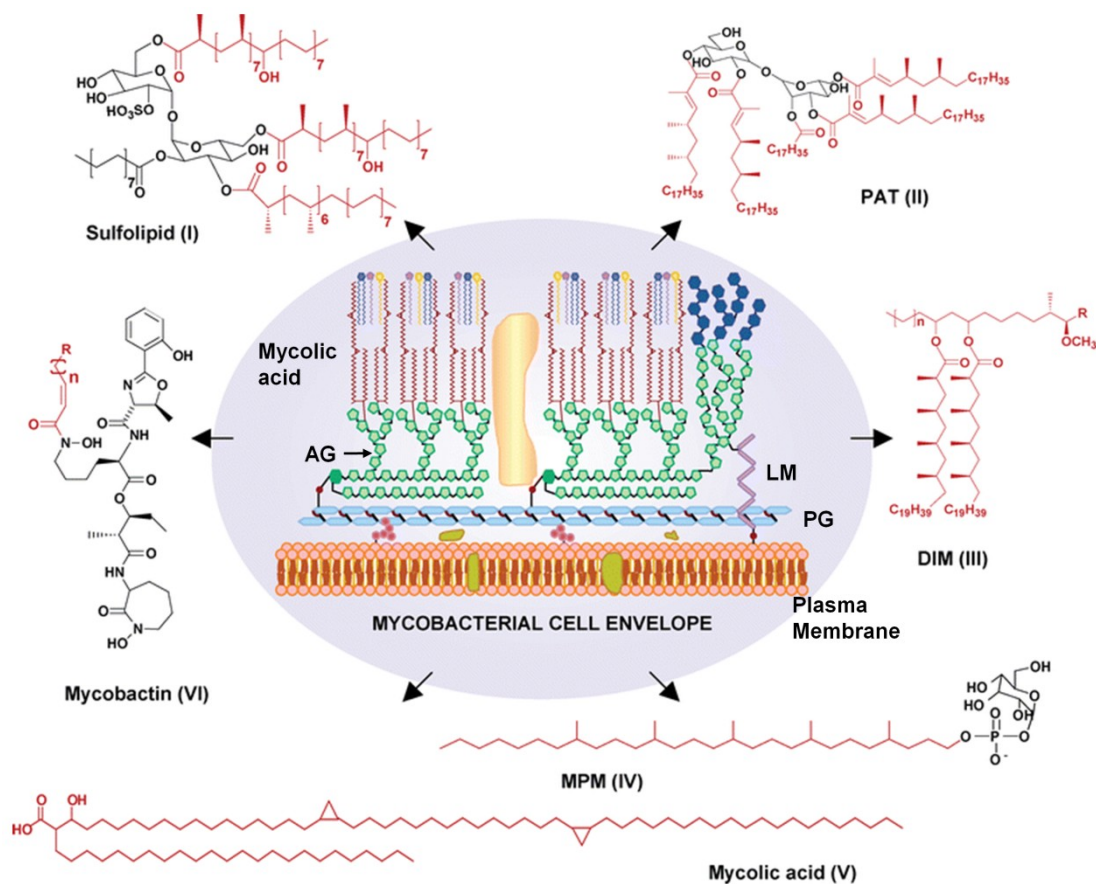


Figure 1-1. Schematic representation of the *Mtb* cell envelope. Architecture of the mycobacterial cell wall shown along with the chemical structures of the major lipid components identified in the *Mtb* envelope. AG, arabinogalactan; PG, peptidoglycan; LM, lipomannan; PAT, polyacyltrehalose; DIM, phthiocerol dimycocerosates; MPM, mannosyl- β -1-phosphomycoketides. Figure reproduced from Gokhale et al., (2007), Nat. Prod. Rep. 24, 267-277, with the permission of © [2007] The Royal Society of Chemistry. <http://dx.doi.org/10.1039/b616817p>

Mycolic acids of mycobacteria

Mycolic acids are major lipids of the mycobacterial cell envelope which are critical for maintaining its cell wall integrity. Though they are primarily found esterified to the terminal units of arabinogalactan as part of the mAGP complex, they are also present in the unbound forms in the cell envelope as trehalose esters TMM and TDM (Takayama et al., 2005). Mycolic acids are long chain fatty acids comprised of a saturated α -branch of C_{22-26} atoms condensed to a longer C_{40-60} meromycolate chain. The meromycolate chain contains two *cis*- or *trans*-double bonds, and it can be further functionalized with different groups, such as keto, ester, methoxy and cyclopropane rings, which define the different classes of mycolic acids (Barry et al., 1998). In *Mtb*, the mycolic acids are comprised of a C_{22-24} α -branch condensed to a very hydrophobic C_{54-63} meromycolate chain (Fig. 1-2a). Based on the chemical modifications, they are divided into three major classes - α -, methoxy- and keto-mycolic acids (Takayama et al., 2005) (Fig. 1-2b). In *Mtb* cell envelope, highly apolar α -mycolates are the most abundant (>70%), followed by the oxygenated methoxy- and keto-mycolic acids (10-15%). The structural features and the content of mycolic acids have been determined to be critical for the virulence of *Mtb*. In a study with mutant *Mtb* H37Rv and *M. bovis* BCG strains defective for the ketomycolate synthesis, it was shown that the mutants were severely defective for growth within macrophages (Yuan et al., 1998). Analysis of a mutant of *Mtb* that caused deletion in the proximal cyclopropane ring in the α -mycolates indicated that despite the normal initial replication, the mutant was significantly attenuated and failed to persist in the infected mice (Glickman et al., 2000). In another study, a mutant

of *Mtb* defective in the biosynthesis of oxygenated keto- and methoxymycolic acids was found to have a significantly altered envelope permeability, and displayed profound growth attenuation in the infected mice (Dubnau et al., 2000).

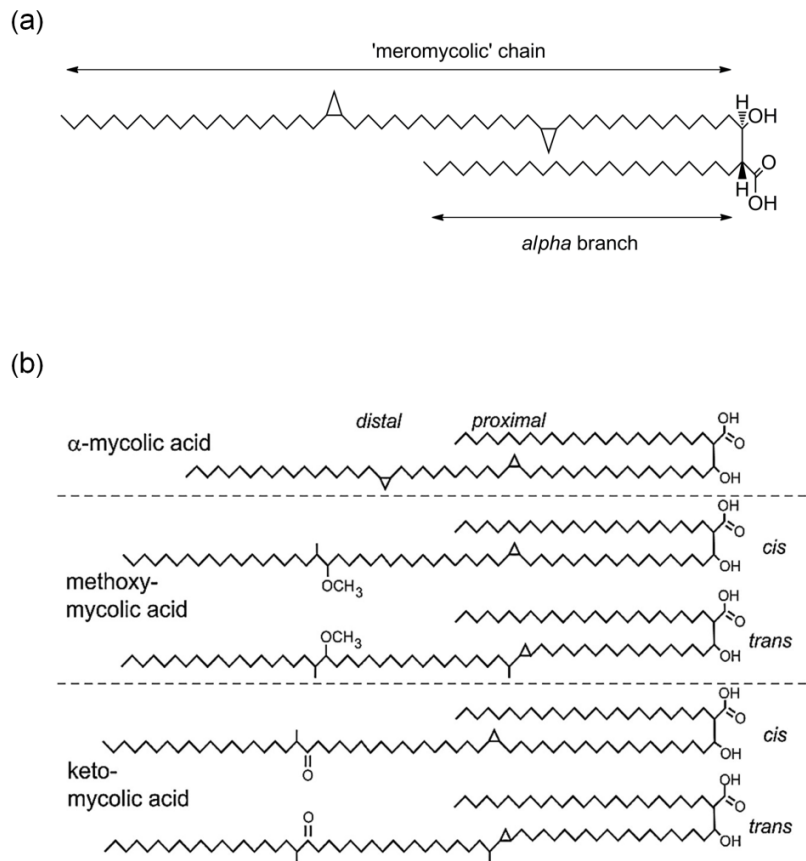


Figure 1-2. Mycolic acids of *Mtb*. (a) The chemical features of the dicyclopropanated α -mycolic acid of *Mtb*. (b) Chemical structures of the different classes of mycolic acids found in *Mtb*. Figure (b) is adapted with permission from Takayama, et al., (2005), Clin. Microbiol. Rev. 18, 81-101, © [2005] American Society for Microbiology.

In addition to their role in the virulence, mycolic acids also play an important role as the structural components of the cell wall and envelope. They provide a highly hydrophobic barrier around the bacterium, and studies with the mycobacterial mutants defective in mycolate synthesis have established a strong link between the cell wall mycolic acid content and resistance to antibiotics. In a study on the effect of mycolate composition on the cell wall function, it was found that a deficiency of the keto-mycolic acids altered envelope fluidity and made the mutant cells hypersensitive to ampicillin and rifampicin (Yuan et al., 1998). A *M. smegmatis* mutant defective in mycolic acid biosynthesis displayed enhanced permeability to hydrophobic compounds, and became hypersensitive to hydrophobic drugs like chloramphenicol and novobiocin (Liu and Nikaido, 1999). Disruption of the mycolyltransferase *fbpA* in *M. smegmatis* caused a 45% decrease in the TDM (cord factor) production in the mutant which resulted in a smooth colony phenotype in the mutant. The *fbpA* mutants also displayed increased sensitivity to both front-line antitubercular drugs as well as other broad-spectrum antibiotics (Nguyen et al., 2005). These studies have strongly established the importance of mycolic acids for the viability and pathogenesis of *Mtb*. This has raised significant interest in understanding the enzymes of the mycolic acid biosynthesis pathway. Indeed, inhibition of mycolic acid biosynthesis has been the mainstay of TB therapy since the introduction of INH as a first-line antitubercular drug in 1952 (Takayama et al., 1972). Therefore, this metabolic pathway represents an important reservoir of novel targets for the development of new anti-*Mtb* drugs especially in the context of the emergence of drug resistant TB.

Mycolic acid biosynthesis

In *Mtb*, mycolic acid biosynthesis occurs by the concerted action of more than 20 different proteins that are components of different multi-enzyme complexes (Takayama et al., 2005). Even though some details of the pathway remain to be elucidated, it has been well established that the pathway uses both FAS I and II systems, as well as a PKS for the mycolic acid biosynthesis (Barry et al., 1998; Portevin et al., 2004). The mycobacteria are atypical in that they possess both type I and type II FAS systems (Barry et al., 1998). The FAS I of *Mtb* perform the *de novo* synthesis of fatty acids containing 16 and 24 carbon atoms as fatty acyl-CoAs in a bimodal distribution. In contrast, the *Mtb* FAS II system is unable to perform *de novo* biosynthesis of fatty acids; instead it preferentially extends palmitoyl-CoA derived from FAS I to synthesize C₁₈₋₃₀ long fatty acids (Barry et al., 1998).

The biosynthesis of mycolic acids in *Mtb* can be divided into three major steps (Takayama et al., 2005) (Fig. 1-3a). First, the FAS I synthesizes alpha-alkyl and meromycolate precursors as the C₁₆-C₁₈ and C₂₄-C₂₆ fatty acids, respectively. Next, the FAS II system extends the C₁₆-C₁₈ fatty acids to generate very long (C₄₈-C₆₄) meromycolic chain which is further modified by different enzymes that introduce *cis* or *trans* unsaturation, methylation, *cis* or *trans* cyclopropanation and oxygenated functional groups. The antituberculosis drug INH targets the FAS II in the mycolic acid synthesis pathway by inhibiting the enoyl reductase InhA *via* formation of a covalent adduct with NAD⁺ (Dessen et al., 1995). The final condensation step is performed by the modular type I PKS, Pks13 of *Mtb*. A fatty acyl-adenylate ligase, FadD32, activates the meroacyl

precursor to meroacyl-AMP and transfers it to the N-terminal ACP domain of Pks13. The C₂₄-C₂₆ fatty acids from FAS-I are activated through carboxylation to generate 2-carboxyl-C₂₆-S-CoA which are transferred by the AT domain of Pks13 to its C-terminal ACP domain. The KS domain of Pks13 then performs the final assembly step through a decarboxylative Claisen-type condensation of the meroacyl and the alpha chain to form α -alkyl β -ketoacids, which are then reduced by the ketoacyl reductase CmrA to form the mature mycolic acids (Takayama et al., 2005). *Mtb* Pks13 is a large modular PKS enzyme which has domain architecture of NACP-KS-AT-ACP-TE (Portevin et al., 2004) (Fig. 1-3b). This protein is essential for the viability of mycobacteria; therefore, it represents an attractive target for the development of novel inhibitors targeting the essential mycolic acid biosynthesis pathway, especially against MDR-TB. Indeed, a recent report indicated that Pks13 is susceptible to chemical inhibition and that its inhibition could kill *Mtb* (Wilson et al., 2013). We have identified a novel benzofuran inhibitor of the essential Pks13 enzyme that targets the functionality of its TE domain. In addition, we have performed structural studies on the TE domain in complex with the inhibitor to elucidate the molecular basis of inhibition. These studies are described in more detail in the Chapter IV of this dissertation.

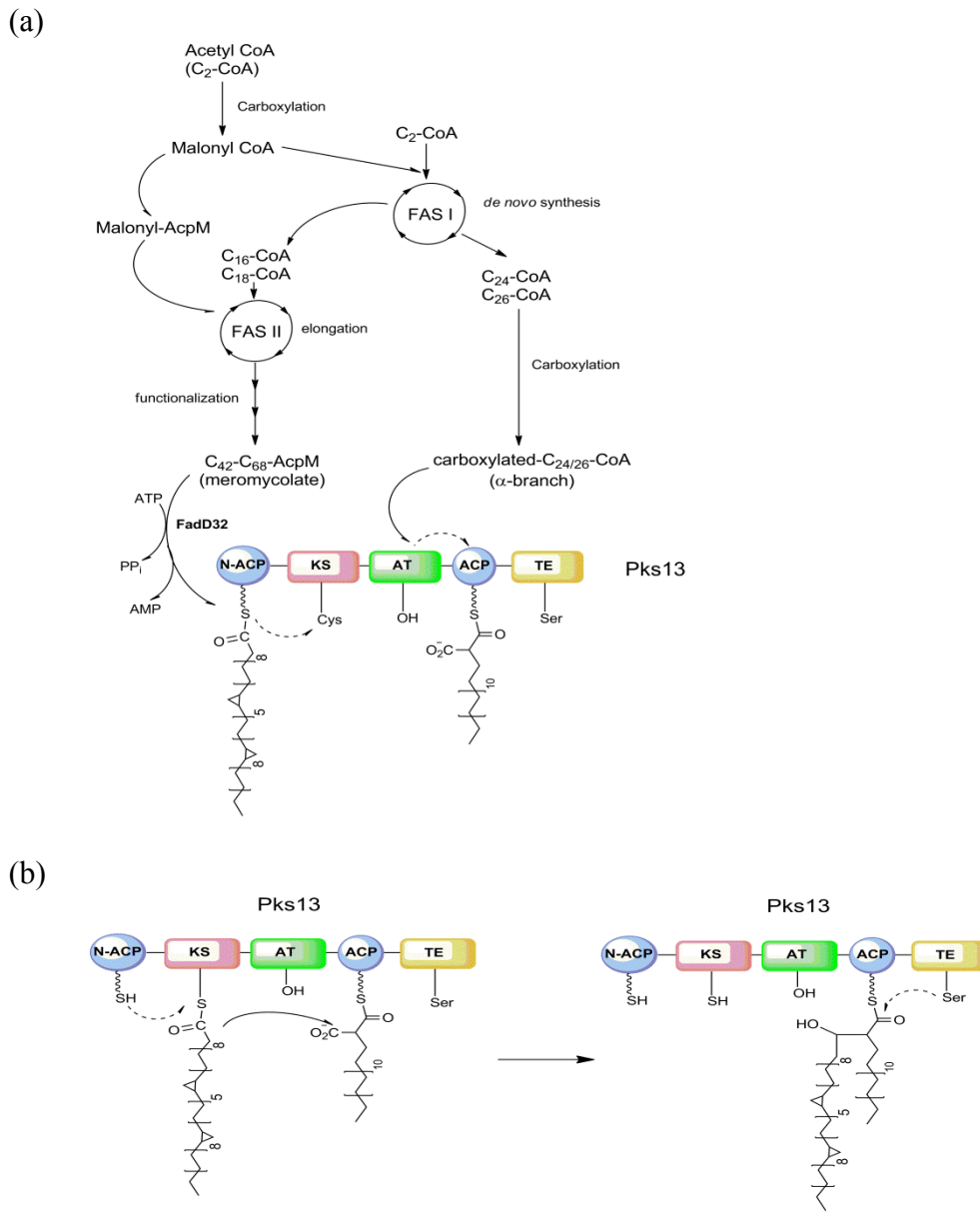


Figure 1-3. Schematic of mycolic acid biosynthesis pathway in *Mtb*. (a) Mycolic acid biosynthesis in *Mtb* starts with the *de novo* synthesis of fatty acid precursors by FAS I. The C_{16-18} fatty acids are elongated by FAS II system and further functionalized to produce the very long meromycolic precursors. The C_{24-26} fatty acids synthesized by FAS I are carboxylated to form acyl-CoAs that provide the α -branch. The activated meromycolic chain is subsequently condensed with the C_{24-26} -CoA by the condensing enzyme Pks13. (b) Pks13 is a type I modular PKS that performs a single round of Claisen-like condensation of the meromycolic chain with the α -branch and the product is then released by the TE domain for reduction to yield the characteristic mycolic acids. AT, acyl transferase; KS, β -ketoacyl synthase; N-ACP, N-terminal acyl carrier protein; ACP, C-terminal acyl carrier protein; TE, thioesterase.

Polyketide synthases of *Mtb*

The polyketide synthases (PKSs) are a class of multifunctional enzymes with a structural organization similar to fatty acid synthase (FAS) systems (Khosla et al., 1999; Smith and Tsai, 2007). The PKS can be synthesized as a multifunctional enzyme in which all the constituent catalytic domains are present on a single polypeptide (type I PKS) similar to FAS I, or they can occur as enzyme clusters of discrete monofunctional subunits (type II PKS) similar to FAS II (Fig. 1-4a). In addition, there are type III PKS systems belonging to the plant chalcone synthase family of enzymes that are present as freestanding homodimeric condensing enzymes (Shen, 2003). In the PKS and FAS systems, the domains are defined based on their catalytic function and they are generally dedicated to a single transformation or transfer function in the multifunctional assembly line. The catalytic domains in type I PKS and FAS I systems are defined based on the chemical transformation they perform and include acyl transferase (AT), β -ketoacyl synthase (KS), dehydratase (DH), β -ketoreductase (KR), enoyl reductase (ER), acyl carrier protein (ACP) and thioesterase (TE) activities. The FAS I is comprised of a single copy of all the seven domains, whereas the minimal type I PKS module can consist of a core set of KS, AT and ACP domains essential for elongating a growing polyketide by one $-\text{CH}_2\text{-CO}-$ unit (Smith and Tsai, 2007) (Fig. 1-4b). Since the reducing domains (KR, DH and ER) are optional, the PKS I systems can alter the extent of the β -carbon processing to retain it in β -keto, β -hydroxy, enoyl or fully saturated form to introduce chirality and generate different functionalities in the polyketide product (Fig. 1-4c). In addition, the products of PKS can be further elaborated by glycosylation, methylation,

by incorporation of phosphate or peptidyl moieties, or *via* cyclization etc. (Gokhale et al., 2007; Walsh, 2004).

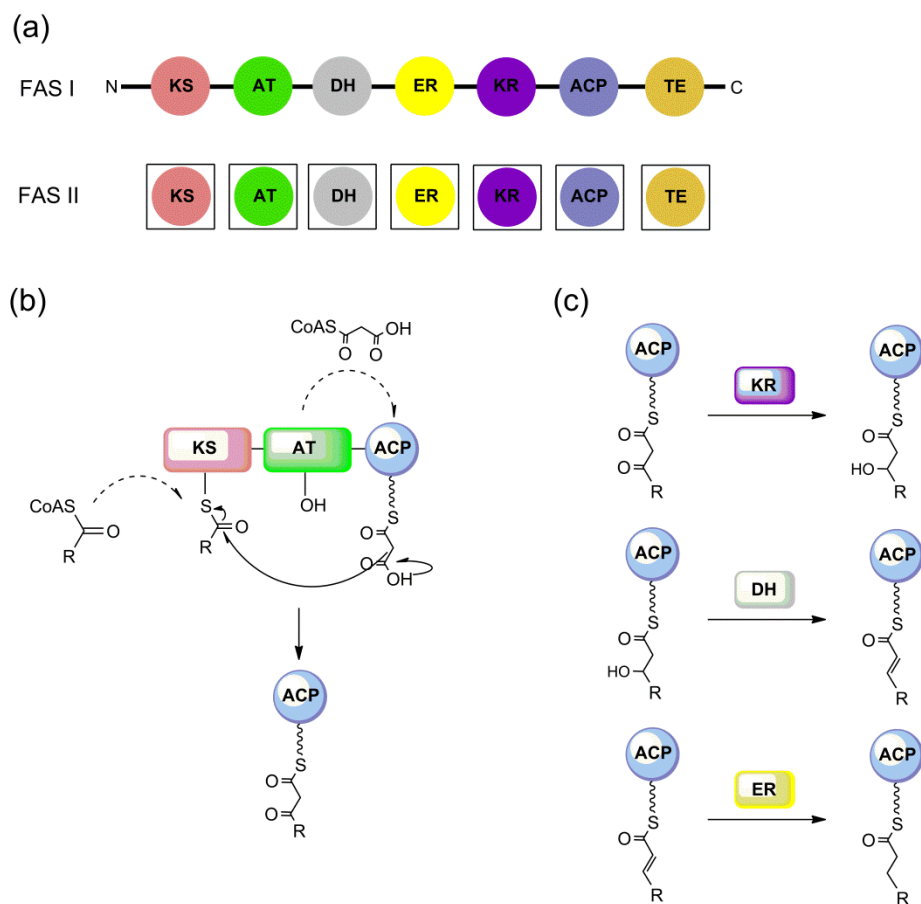


Figure 1-4. Domain organization and reactions of FAS and PKS systems. (a) Domain organization of FAS I and FAS II systems. (b) A minimal PKS module that can extend the growing polyketide chain by the addition of one extender unit/building block. (c) Reactions catalyzed by the ancillary domains in FAS and PKS systems to process the β -keto center. Different reductions states (hydroxyl, enol, fully reduced alkyl) can result from the action of one or more of these domains introducing chirality and functionalities in the product. AT, acyl transferase; KS, β -ketoacyl synthase; DH, dehydratase; KR, β -ketoreductase; ER, enoyl reductase; ACP, acyl carrier protein; TE, thioesterase.

The earliest studies suggesting that the *Mtb* utilized PKS systems to synthesize complex lipids were done by Kolattukudy and coworkers (Kolattukudy et al., 1997). Prior to their work, decades of intensive analysis of the waxy components of the *Mtb* cell wall had led to elucidation of the structures and the chemical nature of many complex lipids, however, their biosynthetic pathways remained unknown (Demartreau-Ginsburg et al., 1959; Goren et al., 1976; Marks and Polgar, 1955). The first indication that the PKS are involved in the biosynthesis of cell wall lipids in *Mtb* came from the studies on the biosynthesis of the methyl-branched fatty acyl chains of phthiocerol wax esters (PDIM) (Azad et al., 1997; Kolattukudy et al., 1997). A multifunctional synthase, MAS was identified as an iterative type I PKS containing all the six domains required for the biosynthesis of multiple methyl-branched mycocerosic acids in *M. bovis* BCG (Mathur and Kolattukudy, 1992). Further, probes for the KS and AT domains of the *mas* and *fas* genes identified a cluster of five genes (*pps1-5*) encoding for type I PKSs in the *M. bovis* BCG genomic cosmid library. The homologues for these were also identified in *Mtb* (*ppsA-E*) and they were shown to be modular type I PKSs involved in the biosynthesis of the phthiocerol branch of the PDIM (Kolattukudy et al., 1997). Subsequently, the sequencing of the *Mtb* genome revealed a number genes encoding for the enzymes of lipid metabolism, consistent with the presence of a diverse range of complex lipids in *Mtb* (Cole et al., 1998). Analysis of the *Mtb* genome also identified the presence of a *fas* I gene, *fas* II system genes, two non-ribosomal peptide synthase clusters (*nrps*) and 25 *pks* genes encoded in its genome (Cole, 1999; Cole et al., 1998). Bioinformatic analysis

of the pks genes suggested that the *Mtb* contains examples of all three types of PKS systems (Table 1-3).

Table 1-3. Polyketide synthases of *Mtb* and their predicted domain organizations.

Gene name	Accession number	Domain ^a organization	Reference
ppsA	Rv2931	ACP-KS-AT-KR-ACP	(Camacho et al., 1999), (Trivedi et al., 2005)
ppsB	Rv2932	KS-AT-KR-ACP	(Azad et al., 1997), (Trivedi et al., 2005)
ppsC	Rv2933	KS-AT-DH-ER-KR-ACP	(Azad et al., 1997)
ppsD	Rv2934	KS-AT-DH-KR-ACP	
ppsE	Rv2935	KS-AT-ACP-C	(Trivedi et al., 2005)
mas	Rv2940c	KS-AT-DH-ER-KR-ACP	(Mathur and Kolattukudy, 1992)
pks1	Rv2946c	AT-DH-ER-KR-ACP	(Constant et al., 2002)
pks15	Rv2947c	KS	(Constant et al., 2002)
pks2	Rv3825c	KS-AT-DH-ER-KR-ACP	(Sirakova et al., 2001)
pks3	Rv1180	KS	(Rousseau et al., 2003a)
pks4	Rv1181	AT-DH-ER-KR-ACP	(Rousseau et al., 2003a)
pks5	Rv1527c	KS-AT-DH-ER-KR-ACP	(Rousseau et al., 2003b)
pks6	Rv0405	ACP-KS-AT- ACP-TE	(Camacho et al., 1999)
pks7	Rv1661	KS-AT-DH-ER-KR-ACP	(Rousseau et al., 2003b)
pks8	Rv1662	KS-AT-DH-ER	(Dubey et al., 2003)
pks9	Rv1664	KS-AT-ACP	

Table 1-3. Continued.

Gene name	Accession number	Domain ^a organization	Reference
pks10	Rv1660	KS (type III)	(Saxena et al., 2003; Sirakova et al., 2003a)
pks11	Rv1665	KS (type III)	(Saxena et al., 2003)
pks12	Rv2048c	KS-AT-DH-ER-KR-ACP-KS-AT-DH-ER-KR-ACP	(Chopra et al., 2008; Sirakova et al., 2003b)
pks13	Rv3800c	ACP-KS-AT-ACP-TE	(Portevin et al., 2004)
pks17	Rv1663	ER-KR-ACP	(Dubey et al., 2003)
pks18	Rv1372	KS (type III)	(Saxena et al., 2003)
MbtC	Rv2382c	KS	(Chopra and Gokhale, 2009)
MbtD	Rv2381c	AT-KR-ACP	(Chopra and Gokhale, 2009)
pks14	Rv1342c	wrongly annotated	(Onwueme et al., 2005)
pks16	Rv1013	wrongly annotated (FadD)	(Onwueme et al., 2005)

^a Different domains are annotated based on the PKSDB prediction server.

Since the cell wall lipids of *Mtb* provide a thick hydrophobic shell around the bacterium imparting it resistance against hostile environments and also actively modulate host immune response, intensive studies have been done to decipher their biosynthesis as well as to identify novel drug targets. In the *Mtb* genome, more genes with potential functionalities in lipid metabolism were discovered than the number of

known metabolites, raising the possibilities for uncovering novel mechanisms or metabolites (Cole, 1999). Indeed, studies over the last decade have identified other enzymes and functional domains in *Mtb* that either mediate cross-talk between multi-enzyme systems involved in the biosynthesis of PKS synthase derived lipids (Mohanty et al., 2011; Trivedi et al., 2004), or functionalize core structures with specific chemical groups (Boissier et al., 2006; Glickman et al., 2000; Keating et al., 2002). For example, Trivedi et al. used sequence analysis to identify homologues of *E. coli* FadD proteins, which convert fatty acids to their corresponding fatty acyl-CoAs for β -oxidation, in the *Mtb* genome in close proximity to *pks* or *nrps* gene clusters. A total of 34 such genes were identified in the *Mtb* genome that clustered in two groups with distinct biochemical activities, fatty acyl-CoA ligases (FACLs) and fatty acyl-adenylate ligases (FAALs) (Arora et al., 2009; Trivedi et al., 2004). Further, it was shown that the FAALs have a novel acyl loading mechanism whereby they can transfer long-chain fatty acids to their cognate PKS for further extension, in an ATP-dependent manner (Trivedi et al., 2004). Similarly, comparative sequence analysis of the *Mtb* genome was used to identify *pks13* gene as the one encoding for the elusive central condensing enzyme of the mycolic acid biosynthesis pathway (Portevin et al., 2004). More than 40 years of biochemical and genetic investigations by several groups had failed to identify this central condensase. Moreover, analysis of the nucleotide sequences flanking the *pks13* gene also helped to identify two genes predicted to encode for an acyl-CoA synthase (*fadD32*) and a subunit of an acyl-CoA carboxylase (*accD4*) involved in the activation of the substrates of the Pks13 enzyme (Gavalda et al., 2009; Portevin et al., 2004).

In recent years, studies on the cell envelope lipids of *Mtb* have clearly established that their production involves novel biosynthetic mechanisms in which PKS and FAS multienzyme systems work in concert to synthesize a remarkable diversity of complex lipids (Chopra and Gokhale, 2009; Takayama et al., 2005). However, despite their importance for the virulence and pathogenesis of *Mtb*, there has been no structural data available on the type I PKSs of *Mtb*. Till date, there has been only a single study on the inhibitor discovery targeting PKS systems in *Mtb* which reported identification of an inhibitor that disrupted the interaction between FadD32 and the N-terminal ACP domain of Pks13 (Wilson et al., 2013). However, in the absence of the structural data, the inhibitors could not be rationally modified to improve their potency. Moreover, the proposed mechanism of inhibition derived from the modeling studies was speculative. For a proper understanding of these megasynthases, it is essential to obtain atomic level structural details. Even though a porcine FAS I crystal structure has become available to provide a general framework for the type I PKSs (Maier et al., 2006), it was found to be lacking in fine atomic level details as determined from the comparison with a high resolution structure of the KS-AT didomeric domain of 6-deoxyerythronolide B synthase (Smith and Tsai, 2007). Moreover, the available structural data does not provide a model for the stacking of a series of PKS modules that lack one or more of the ancillary domains. Therefore, it is critical to determine structures of these complex multienzymes to understand their mechanism, and to obtain insights into their intra- and inter-domain recognition mechanisms, which are currently lacking for most of the PKSs in *Mtb*. The structural data can help elucidate sequence and structural motifs required for the

substrate recognition and provide insights into the protein-protein recognition mechanisms (Tsai and Ames, 2009). In addition, the structural details of the active sites and protein-protein interaction interfaces can help guide structure-based inhibitor discovery against these megasynthases which are essential for the virulence and viability of the *Mtb* in its host (Cox et al., 1999). In this dissertation I describe our efforts towards crystallographic studies on the two type I PKS synthases of *Mtb*, mycocerosic acid synthase (MAS) from the PDIM biosynthesis pathway, and Pks13 of the mycolic acid biosynthesis pathway. These studies are discussed in Chapters II and IV, respectively.

Methyl-branched cell surface lipids of *Mtb*

Mtb contains many PKS derived methylated complex lipids and lipoglycans in its cellular envelope (Fig. 1-5). The PKSs involved in their biosynthesis introduce methyl branches on the esterified fatty acyl chains by using methylmalonyl-CoA, instead of malonyl-CoA, as the chain extending substrate (Jackson et al., 2007). These surface exposed lipids are non-covalently associated with the cell wall and form a waxy low permeability barrier that imparts resistance against therapeutics (Brennan and Nikaido, 1995). In addition, these diffusible complex lipids are critical for the virulence and the pathogenesis of *Mtb*, since they also play important roles in mediating specific host-pathogen interactions (Forrellad et al., 2013; Rousseau et al., 2004). The phthiocerol diesters (PDIM) of *Mtb* consist of long-chain β -diols (C₃₃-C₄₁) esterified to multimethyl-branched fatty acids (C₂₇-C₃₄) called mycocerosic acids. Phenolic glycolipids (PGL) are

composed of long-chain phenolic β -diols (C_{33} - C_{41}) with a variable sugar moiety consisting of one to four O-methylated deoxysugars, and esterified by mycocerosic acids or phthioceranic acids (Onwueme et al., 2005). Both, PDIM and PGL, are located in the outermost layer of the *Mtb* capsule partly exposed to the surface of the cell, and partly contributing to the structure of the outer layer of the envelope (Ortalo-Magne et al., 1996).

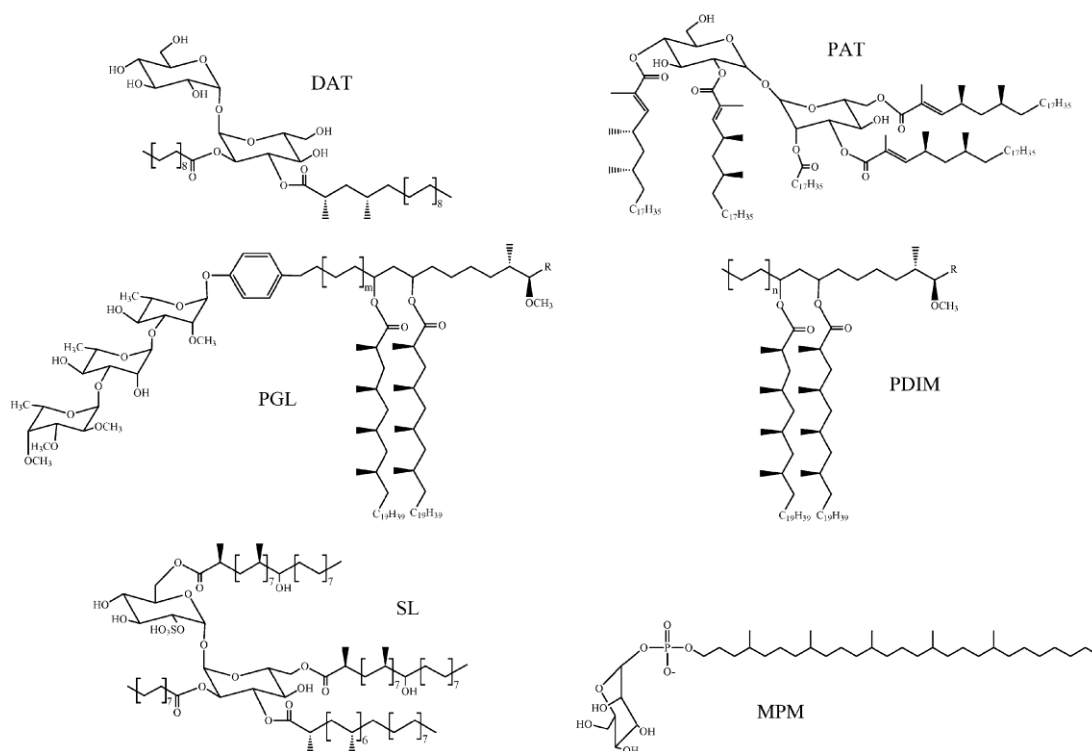


Figure 1-5. Chemical structures of multiple methyl-branched fatty acid containing lipids from *Mtb*. DAT, diacyl trehalose; PAT, polyacylthralose; PGL, phenolic glycolipids; PDIM, phthiocerol dimycocerosates; SL, sulfolipids; MPM, mannosyl- β -1-phosphomycoketides. Reprinted from Jackson et al., *Tuberculosis* 87, 78-86, with permission from © [2007] Elsevier.

PDIM and PGL are primarily found in the pathogenic species of mycobacteria and are implicated in promoting *Mtb* virulence in the early stage of infection (Daffe and Laneelle, 1988; Rousseau et al., 2004). PDIM were shown to insert into the human macrophage membrane thereby altering its lipid organization and facilitating the phagocytic uptake of *Mtb* into the host macrophage (Astarie-Dequeker et al., 2009). PGL were also found to enhance *Mtb* virulence through modulation of host immune response. A knock-out mutant of the *pks15/1* gene involved in the PGL synthesis in a highly virulent *Mtb* clinical isolate of the *W*-Beijing family, resulted in a mutant strain attenuated in its ability to kill mice following aerosol infection (Reed et al., 2004). Synthesis of PGL was also found to inhibit proinflammatory cytokine production in the infected mouse macrophages and the effect was attributed to the saccharide moiety of the PGL (Reed et al., 2004). Sulfolipids (SL) are found exclusively in the human pathogenic *Mtb*. SLs are the second most abundant free lipids in the *Mtb* cell envelope which are found associated with the mycolic acid layer in the envelope (Chopra and Gokhale, 2009). Gene disruption and complementation studies have identified Pks2 as the synthase involved in the synthesis of methyl-branched phthioceronic acids (PA) or hydroxyl phthioceronic acids (HPA) of SL-1 (Sirakova et al., 2001). The SL-1 family was determined to be sulfated trehalose esters of a straight chain (C₁₆ or C₁₈) fatty acid at the 2-position and different combinations of the three chains of PA or HPA at the 3-/4-, 6- and 6'-positions (Goren et al., 1976; Jackson et al., 2007) (Fig. 1-5). Although SL have been suggested to mediate host-pathogen interactions due to their presence on the cell surface, evidence of their direct involvement in *Mtb* virulence has not yet been

established (Forrellad et al., 2013). PATs (polyacyltrehaloses) and DATs (diacyl trehalose) are trehalose ester lipids of waxy nature that are localized on the outer surface of the cell wall. PATs are penta-acylated trehaloses consisting of one medium-chain (C₁₆ or C₁₉) fatty acid and different combinations of trimethyl C₂₇-mycolipenic and C₂₇-mycolipanic acids at the 2, 2', 3', 4 and 6' positions (Jackson et al., 2007) (Fig. 1-5). A PAT mutant strain with defects in capsule attachment was shown to cause clumping of the *Mtb* cells (Dubey et al., 2002), and in another mutant strain PAT deficiency was correlated with the efficiency of binding and entry into the host cells (Rousseau et al., 2003a). However, their contribution to the bacterial virulence remains to be determined. The MPMs (mannosyl- β -1-phosphomycoketides) are a group of antigenic phospholipids that are synthesized by the bimodular synthase Pks12 in *Mtb* (Chopra et al., 2008). The MPMs have been found only in the slow growing pathogenic mycobacteria. They have a mannose-phosphate head group identical to that found in the mammalian mannosyl- β -1-phosphodolichols, however, they differ in their alkyl chain which is proposed to contribute to their antigenicity (Matsunaga et al., 2004). The mycoketides have been proposed to function as virulence factors, or as bioactive compounds mediating biological signaling due to their extremely low levels in the *Mtb* cells. Though MPMs were shown to interact with human CD1c proteins, their precise function in the mycobacterial infection remains to be determined (Matsunaga and Sugita, 2012). The methyl-branched complex lipids of the mycobacterial cell envelope are important virulence effectors many of which are found only in pathogenic mycobacteria, such as PDIMs, SLs and MPMs of *Mtb*. Therefore, the protein components involved in the

metabolism and transport of these complex cell wall lipids could potentially offer important new targets for the antitubercular drug design.

Biosynthesis of phthiocerol dimycocerosate esters (PDIMs) in *Mtb*

The genes encoding the various enzymes of the PDIM biosynthesis pathway constitute a large operon comprised of ~15 open reading frames, and they are found clustered on a 50 kbp fragment of the *Mtb* genome (Cole et al., 1998; Trivedi et al., 2005) (Fig. 1-6). The five type I modular PKSs (PpsA-E) and the iterative type I PKS (MAS) encoded by this cluster exemplify the biochemical logic of the modular PKS based lipid biosynthesis.

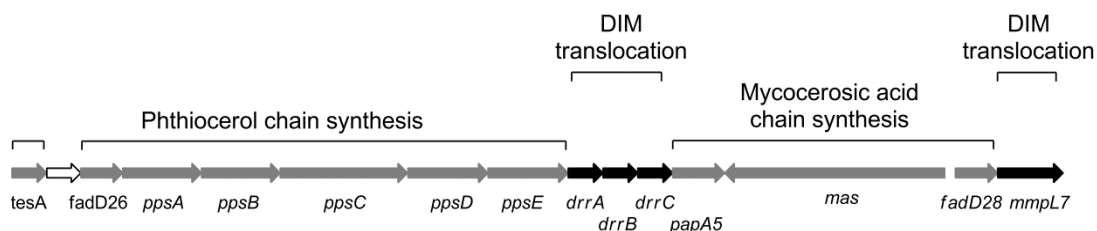


Figure 1-6. Genetic organization of the PDIM synthesis locus in *Mtb*. The genes involved in the production of PDIM constitute a large operon (~50 kbp) and include a type I iterative PKS (*mas*) and five type I modular PKSs (*ppsA-E*). The genes are depicted as arrows with the genes encoding the proteins involved in PDIM synthesis in grey, and the genes encoding proteins involved in the translocation in black.

The biosynthesis of PDIMs can be divided into four key steps: (i) priming of PpsA with specific long-chain fatty acid; (ii) extension of the primer by PpsA-E, leading to the synthesis of phthiocerol; (iii) synthesis of mycocerosic acids by MAS; and (iv) esterification of mycocerosic acids to the diol of phthiocerol and final assembly (Gokhale et al., 2007) (Fig. 1-7). The first step in phthiocerol biosynthesis involves the transfer of a fatty acyl starter unit to the PCP domain of PpsA by FadD26. The starter unit is then loaded onto the active site Cys of the KS domain of PpsA, which then performs a decarboxylative condensation with a malonyl group attached to the phosphopantetheine arm of the ACP domain. This extends the fatty acyl chain by two carbon units, and the KR domain of PpsA then reduces the β -keto group to a hydroxyl group. After performing one extension cycle, PpsA covalently transfers the chain from its ACP domain to the active site of the KS domain of PpsB through an acyl transfer reaction (Gokhale et al., 2007; Trivedi et al., 2005). The PpsB protein also utilizes malonyl-CoA to extend the chain by two carbon units, and generates another hydroxyl group on the growing chain. Similarly, PpsC and PpsD also extend the acyl chain; however, they completely reduce the β -keto groups to saturated carbons due to the presence of ancillary dehydratase and enoyl reductase domains. In contrast to PpsC, the ER domain of PpsD is a *trans*-ER encoded by the Rv2953 gene in *Mtb* (Simeone et al., 2007) (Fig. 1-7). The final step in the phthiocerol biosynthesis is catalyzed by PpsE, which can utilize both malonyl- as well as methylmalonyl-CoA as extender units. This catalytic flexibility of PpsE has been proposed to produce structural diversity in the PDIMs (Trivedi et al., 2005).

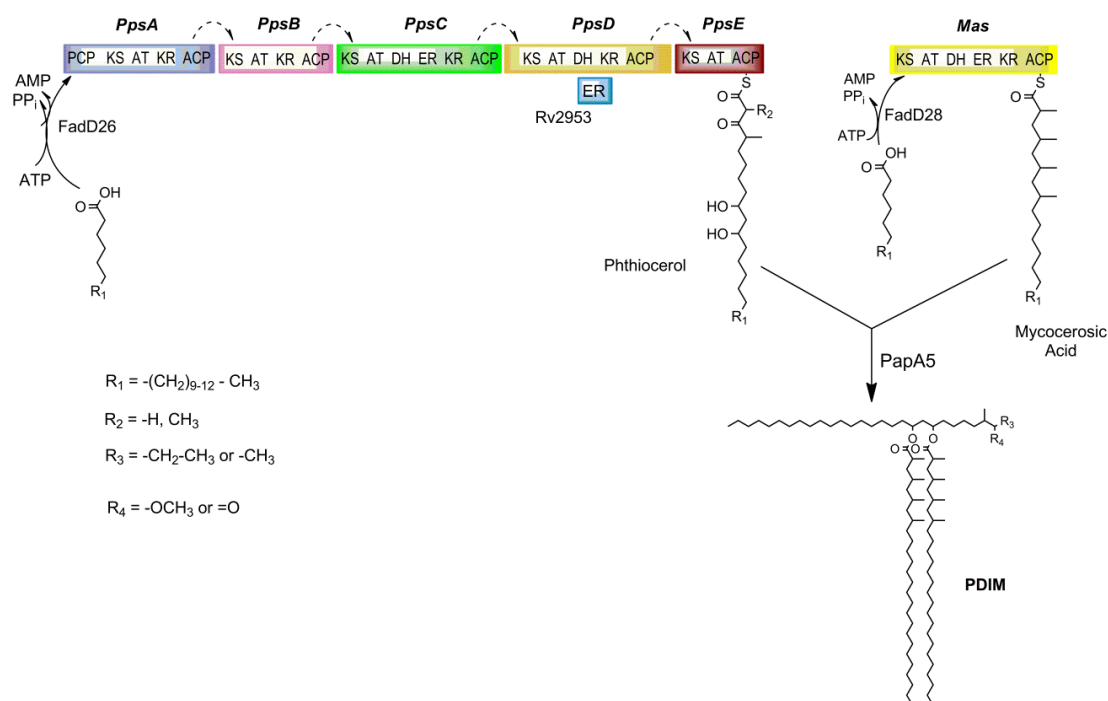


Figure 1-7. Schematic of PDIM biosynthesis pathway in *Mtb*. FadD26 provides activated fatty acyl precursors to the modular PpsA-E PKS machinery for the synthesis of the phthiocerol branch of PDIM. The ER domain (Rv2953) is not a part of the PpsD polypeptide, but acts in *trans* with it. The type I iterative PKS, MAS synthesizes methyl-branched mycocerosic acids using the activated acyl precursors provided by FadD28. PapA5 mediates the final condensation of phthiocerol with mycocerosic acids to form PDIM. AT, acyl transferase; KS, β -ketoacyl synthase; DH, dehydratase; KR, β -ketoreductase; ER, enoyl reductase; ACP, acyl carrier protein.

The methyl-branched mycocerosic acids of the PDIMs are synthesized by MAS by using variable length fatty acyl-CoA starters with methylmalonyl-CoA as extender units (described in more detail in Chapter II) (Chopra and Gokhale, 2009; Mathur and Kolattukudy, 1992). The final assembly of PDIM is performed by the polyketide-associated protein A5 (PapA5) which interacts with MAS and transfers the mycocerosic acids onto the diol component of phthiocerol (Onwueme et al., 2004; Trivedi et al.,

2005). The fully synthesized PDIMs are further processed before being transported to the cell wall periphery by the transmembrane protein MmpL7 and other accessory proteins (Jain and Cox, 2005; Onwueme et al., 2005).

Even though the combination of biochemical and genetic studies over the last couple of decades have provided a detailed picture of the biosynthesis of PDIMs in *Mtb* (Fig. 1-7), it still remains to be elucidated how these complex enzymes work at the molecular level (Guilhot et al., 2008). For example, it is not known what structural motifs of the linkers are involved in conferring specificity for the hand-to-hand transfer of the acyl chain from one Pps protein to another. Moreover, there is no information available on the supramolecular organization of the PKS complexes, and how they interact with other partner proteins of the PDIM biosynthetic pathway. Also, the structural basis for substrate specificity, and the molecular mechanisms for generating subtle structural variations in the PDIM core are yet to be deciphered. Since the PDIMs of *Mtb* are critical for its virulence and pathogenesis, it has been suggested that inhibition of the enzymes of this pathway could synergize with, or potentiate other anti-tubercular drugs to improve the treatment of TB (Favrot and Ronning, 2012). The PKSs are excellent targets for drug discovery since they are not found in the human host, and they provide multiple active sites for chemical inhibition on a single polypeptide. For inhibitor development, three dimensional structural studies are critical for elucidating the structure-function relationships in these enzymes. Such an approach was successfully applied in determining the structural basis for the acyltransferase activity of PapA5 that transfers mycocerosic acids from MAS to phthiocerol (Buglino et al., 2004). Another

structural study identified LppX as the protein involved in PDIM transport and provided a structural basis for its activity (Sulzenbacher et al., 2006). However, despite the identification of the first PKS of this pathway (MAS) more than 25 years ago (Rainwater and Kolattukudy, 1985), so far no structural studies have been reported for any of the PKSs of the PDIM biosynthesis pathway. This can be largely attributed to the challenges associated with the crystallographic studies of large macromolecular complexes, as exemplified by the fact that except for FAS I, structure of no other full length type I PKS from any organism has been solved (Maier et al., 2010; Mueller et al., 2007). In this dissertation, I report our efforts on crystallization and initial X-ray crystallographic characterization of the crystals of the *Mtb* MAS protein. We identified conditions for successful recombinant expression of this large protein in *E. coli*, and developed protocols for the improvement of the diffraction qualities of the protein crystals. These studies are described in detail in Chapter II.

Activation of acyl carrier protein (ACP) domains in *Mtb*

The ACP domains function as flexible swinging arms that deliver the covalently tethered biosynthetic intermediates for sequential processing to different active sites of PKS and FAS enzymes, either in *cis* (type I systems) or in *trans* (type II systems) (Smith and Tsai, 2007). In all PKS and FAS systems, a posttranslational modification of the constituent ACP domains by the attachment of a 4'-phosphopantetheine (4'-ppt) prosthetic group to the hydroxyl of a conserved serine residue, is required for the enzymes to become functionally active (Meier and Burkart, 2009) (Fig. 1-8). This

posttranslational modification is mediated by 4'-phosphopantetheinyl transferases (PPTases) that convert inactive *apo*-ACP to the active *holo*-form by catalyzing transfer of the 4'-ppt arm from coenzyme A (CoA) to the ACP domain (Lambalot et al., 1996). During the biosynthesis cycles of PKS and FAS systems, the growing acyl chains and the biosynthetic intermediates remain covalently attached to the ACP domain by a thioester linkage with the terminal thiol of the 4'-ppt arm. After the synthesis is complete, the finished polyketide or the fatty acid chain is released from the ACP by a thioesterase or an acyltransferase domain (Smith and Tsai, 2007).

The genome of *Mtb* contains two PPTase encoding genes, *acpS* and *pptT* that encode the corresponding PPTases AcpS and PptT in *Mtb* (Cole et al., 1998). The AcpS (*holo*-acyl carrier protein synthase) enzyme was determined to be essential for the activation of ACP domains in both the FAS I and FAS II systems in *Mtb*, whereas the PPTase PptT was proposed to activate the carrier domains of PKS I and non-ribosomal peptide synthases in *Mtb*. Thus, both the enzymes play crucial roles in the biosynthesis of complex lipids that are essential for the survival and virulence of *Mtb*. (Chalut et al., 2006). Since *Mtb* AcpS activates the acyl carrier protein AcpM of the FAS II system which is essential for the synthesis of mycolic acids, it represents a potential target for drug development. Therefore, to determine the structure-function relationship for the inhibitor development, we have solved the three dimensional structure of *Mtb* AcpS. The structure revealed that the *Mtb* AcpS can adopt two different pH-dependent conformations. In addition, we screened the enzyme against a fragment library to

identify lead fragments for further structure-guided development. These studies are described in detail in the Chapter III of this dissertation.

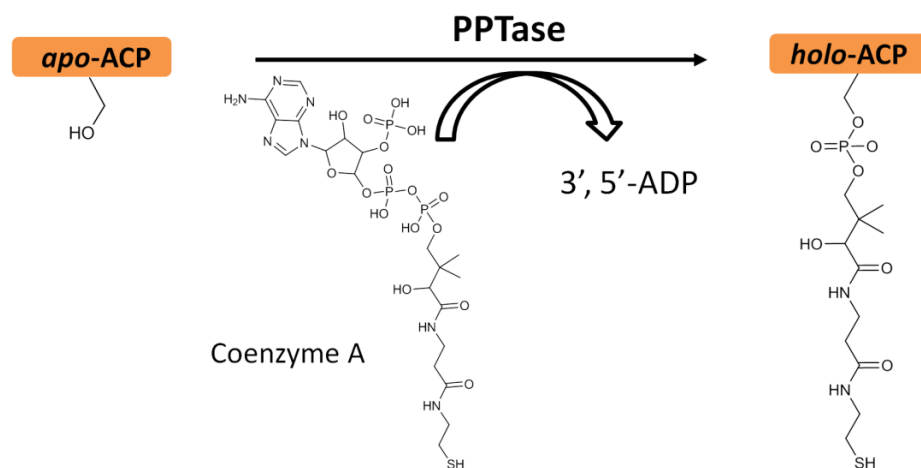


Figure 1-8. Posttranslational modification of the ACP domain. The PPTase catalyzes the transfer of the 4'-phosphopantetheine moiety from CoA onto a conserved serine residue of the *apo*-ACP, thereby converting it to the activated *holo*-ACP form.

CHAPTER II

PURIFICATION AND INITIAL CRYSTALLOGRAPHIC CHARACTERIZATION OF MYCOCEROSIC ACID SYNTHASE FROM *Mycobacterium tuberculosis*

INTRODUCTION

Mycocerosic acids synthase (MAS, Rv2940c), is a type I PKS that synthesizes multimethyl-branched fatty acids (mycocerosic acids) in mycobacteria (Rainwater and Kolattukudy, 1985). The mycocerosic acids are an important component of the complex lipids PDIM and PGL that are found primarily in the cell walls of the pathogenic mycobacteria (Onwueme et al., 2005). MAS performs iterative condensation of medium- to long-chain fatty acyl-CoA primers with methylmalonyl-CoA as the extender units to synthesize fully reduced and saturated mycocerosic acids in a manner analogous to the vertebrate FAS I system (Mathur and Kolattukudy, 1992; Trivedi et al., 2005). In addition, it was also determined that the MAS protein required NADPH as a cofactor for the KR and ER activities (Rainwater and Kolattukudy, 1985).

The cloning and sequencing of the *mas* gene identified a modular domain organization for the multifunctional MAS protein with a domain order: β -ketoacyl synthase (KS), acyl transferase (AT), dehydratase (DH), enoyl reductase (ER), β -ketoreductase (KR) and malonyl/palmitoyl transferase-acyl carrier protein (ACP)

domains, similar to the FAS I enzyme (Mathur and Kolattukudy, 1992). Thus, MAS possesses all the ancillary domains (KR, DH and ER) that are necessary for the complete reduction of the resulting β -ketoacyl moiety, before another round of chain extension occurs. Sequence analysis revealed that, unlike FAS-1, the MAS polypeptide did not contain a thioesterase (TE) domain that could be involved in the release of the synthesized mycosates. This was further confirmed by the experimental observation that *in vitro* the synthesized mycosates were not released from the protein (Mathur and Kolattukudy, 1992; Trivedi et al., 2005). It was postulated that another specific enzyme may be involved in transferring the mycosates from MAS onto the hydroxyl groups of phthiocerol and phenolphthiocerol chains (Kolattukudy et al., 1997). Indeed, a subsequent study identified the enzyme catalyzing this acyl transfer as PapA5, encoded by the gene adjacent to the *mas* gene in the PDIM operon (Onwueme et al., 2004) (Fig. 1-6). It was shown that *in vitro* PapA5 was able to transfer the MAS bound mycosatic acid analogs to the 2-dodecanol (a phthiocerol surrogate), suggesting a hand-to-hand transfer mechanism for the mycosates (Trivedi et al., 2005). However, the structural basis for the PapA5 interaction with the MAS protein, and the molecular mechanism of chain termination in MAS are presently not known.

In the studies performed with the purified MAS, it was observed that *in vitro* MAS was able to utilize many different primers ranging from C₆-C₂₀ acyl-CoAs and generated di-, tri- or tetra-methyl branched mycosatic acids (Onwueme et al., 2005; Trivedi et al., 2005). A similar structural diversity in the mycosatic acids was also observed *in vivo* for the *Mtb* cultured under different growth conditions (Jain et al.,

2007). These studies established that MAS could utilize primer units of varying lengths, and modulate the number of iterative cycles to generate chemical diversity depending on the growth conditions. But the molecular basis for such substrate promiscuity has not been determined.

Initial studies with MAS suggested that its preferred substrates were acyl-CoAs of C₁₆-C₂₀ fatty acids (Rainwater and Kolattukudy, 1985). The FadD28 protein, encoded by a gene adjacent to *mas*, was proposed to be the activating enzyme for converting the fatty acids to their corresponding acyl-CoA forms, since many genetic studies had shown it to be essential for the biosynthesis of both PGL and PDIM in *Mtb* (Camacho et al., 2001; Cox et al., 1999; Fitzmaurice and Kolattukudy, 1998). It was later demonstrated in an *in vitro* study that FadD28 did activate fatty acids but as fatty acyl-AMP, and not as acyl-CoA (Trivedi et al., 2004). These findings and the close chromosomal proximity of FadD28 and MAS strongly suggest that FadD28 is the activating enzyme for the substrates of MAS. However, a direct evidence for the interaction between MAS and FadD32 is not yet available, and the molecular mechanism for the acyl transfer also remains to be elucidated.

Thus, on the basis of its importance in PDIM biosynthesis and the critical role of this complex lipid group in *Mtb* virulence, the multifunctional MAS protein presents an important target for structure-function studies. Determination of MAS structure will enable us to understand the structural and mechanistic aspects underlying its function, and the structural information could also aid in the subsequent inhibitor discovery studies.

MATERIAL AND METHODS

Cloning, expression and purification of *Mtb* MAS

The 6336 base pair *mas* gene was PCR amplified from the genomic DNA of *Mtb* strain H37Rv. Due to the large size of the gene, a combination of FailSafe PCR reaction mix D (Epicenter) and Phusion high fidelity DNA polymerase (NEB) was used to ensure fidelity of the amplified gene and eliminate non-specific products in the PCR reaction. The PCR primers for full length *mas* gene (forward primer, 5'-CATTCGCTAGCATGGAATCACGTGTCACTCCCGTTGC 3'; reverse primer, 5'-CATTCGGTACCCTATGATGCCGCCGGTGCCGC 3') were used to create restriction sites for *NheI* and *KpnI* (underlined), respectively. The amplified DNA fragment was cloned into the corresponding restriction sites in a chimeric pET28b_pET30b expression vector (pET28_30b) to create N-terminally His₆-tagged full length MAS (2111 residues) to simplify purification.

The C-terminal ACP domain truncations of MAS were cloned in the pET28_30b vector in the same way as was done for the full length MAS from the PCR fragments amplified using the following primers:

For all five truncations, the forward primer was,

5'-CATTCGCTAGCATGGAATCACGTGTCACTCCCGTTGCGGTGATC 3'

Reverse primer for C-terminus truncation 1 of MAS (1-2011, T₁),

5'-CATTCGGTCCCTCAGCCCCTTGACCGCTGCCAGTGGATGCGAA 3'

Reverse primer for C-terminus truncation 2 of MAS (1-1974, T₂),

5'-CATTCGGTACCTCAGTCGTGGCGCACGAGCGTCTCGAAGGCATA 3'

Reverse primer for C-terminus truncation 3 of MAS (1-1906, T₃),

5'-CATTCGGTACCTCAGCCGAGCAATGCCGCTCCCGAGGAGAACAA 3'

Reverse primer for C-terminus truncation 4 of MAS (1-1842, T₄),

5'-CATTCGGTACCTCACCCGGTAGCGGTCGCCGCACTCACCAGCCG 3'

Reverse primer for C-terminus truncation 5 of MAS (1-1780, T₅),

5'-CATTCGGTACCTCAGGGCTGGGAACGTGCGGTCAGCACGATCCG 3'

In the MAS protein expressed in the *Mtb* cells, its ACP domain is posttranslationally modified with the covalent attachment of a 4'-phosphopantetheine (4'-PPT) group that functions to tether the fatty acyl substrate during the processing cycles. The MAS protein over-expressed in *E. coli* does not get this modification because of the non-specificity of the *E. coli* phosphopantetheinyl transferase (PPTase) for MAS. So in order to get the post-translational modification on MAS expressed in *E. coli*, a PPTase from *Mtb* (PptT, Rv2794c) (Chalut et al., 2006) was PCR amplified from the *Mtb* genomic DNA (forward primer, 5'-GGAATTCCATATGATGACGGTAGGCACGCTGGTGGCGTC 3'; reverse primer, 5'-CCGCTCGAGTCATAGCACGTACGCGGTCAGCACCAAGTC 3') and cloned into the second multiple cloning site (MCS2) of the pCDFDuet-1 vector (Novagen, for plasmid compatibility and antibiotic selectivity) using *Nde*I and *Xho*I restriction enzymes, generating a tag-less expression construct. The integrity of all the cloned genes was confirmed by DNA sequencing.

For the protein production, the expression plasmid for the full length MAS protein (pET28_30b-MAS) was co-transformed with the pCDFDuet-1-Rv2794c plasmids into *E. coli* expression strain Rosetta 2(DE3)pLysS (Novagen) to compensate for rare codons in *mas* gene. The ACP domain truncations were also transformed (without Rv2794c) in *E. coli* Rosetta 2(DE3)pLysS cells for the protein expression. The transformed cells were cultured in the LB-Miller media containing 50 µg/ml kanamycin, 34 µg/ml chloramphenicol and 50 µg/ml streptomycin for the full length MAS, and with 50 µg/ml kanamycin, 34 µg/ml chloramphenicol for the truncations, until OD₆₀₀ reached 0.6. The culture was then induced with 0.5 mM IPTG and grown at 18 °C for 14-16 hrs. Cells were harvested by centrifugation at 4,000g and the cell pellet was re-suspended in the lysis buffer (50 mM HEPES, pH7.5, 200 mM NaCl, 40 µg/ml DNase, 2 mM MgCl₂ and EDTA-free protease inhibitor cocktail (Calbiochem)). The cells were lysed using a French press at 11,000 psi and the lysate was centrifuged with Sorvall ss-34 rotor at 15,000 rpm for 60 min to remove cell debris. The supernatant was filtered and loaded onto a HisTrap FF Ni column (GE Healthcare) that was pre-equilibrated with Buffer A (20 mM HEPES, pH7.5, 250 mM NaCl, 10% (v/v) glycerol, 2 mM β-mercaptoethanol). The protein was eluted with 40% Buffer B (20 mM HEPES, pH7.5, 250 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, 2 mM β-mercaptoethanol) using a step gradient purification protocol. Fractions containing the protein were pooled, concentrated and loaded on a Superdex 200 gel-filtration column (GE healthcare) pre-equilibrated with 20 mM HEPES (pH7.5) buffer containing 100 mM NaCl, 10% (v/v) glycerol and 10 mM DTT, for further purification. The eluted protein was >95% pure as determined by

Coomassie Blue stained SDS-PAGE gel. The fractions containing pure protein were pooled and concentrated using a 100K molecular weight cutoff Amicon concentrator (Millipore) to ~35-40 mg/mL, and used immediately for crystallization trials or stored at 193 K for later use.

For the production of selenomethionine (Se-Met) incorporated full length MAS protein, the metabolic inhibition method was used (Doublie, 2007). The protocol for the purification of Se-Met-incorporated MAS was same as that used for the native full length MAS, however, the final concentration of DTT in the gel filtration and the storage buffers was 20 mM.

Determination of 4'-PPT modification in recombinantly expressed MAS

The extent of 4'-ppt modification of the full length MAS protein expressed in *E. coli* was assessed by an *in vitro* 4'-ppt transferase assay using *B. subtilis* PPTase Sfp and fluorescent BODIPY-labeled CoA (La Clair et al., 2004). In this assay, 50-100 μ M of purified MAS expressed without a PPTase, as well as MAS co-expressed with either *Mtb* AcpS or PptT in *E. coli*, was incubated with 2 μ M Sfp for 1 h at room temperature in the dark in a reaction mixture containing 20 mM HEPES pH 7.5 buffer with 20-40 μ l BODIPY-S-CoA (120-240 μ M final concentration) and 10 mM MgCl₂. As a negative control, an ACP-domain truncated MAS protein without a 4'-ppt attachment site was also included in the assay. After incubation, cold 20% TCA was added to the mixture to stop the reaction. The samples were incubated for 15 min on ice and then spun at 10,000g for 10 min. The pellets were washed twice with 300 μ l ice-cold acetone. The

pellets were dried under vacuum, dissolved in a small amount of buffer and analyzed by SDS–PAGE. The resulting bands on the gel were analyzed for BODIPY-CoA labeling of the protein under a UV transilluminator.

Determination of oligomeric state of MAS

The full length purified MAS (30 μ L of 10 mg/mL) was injected on a SuperoseTM 6 10/300 GL (GE Healthcare) gel filtration column which was pre-equilibrated with 20 mM HEPES (pH7.5) buffer containing 100 mM NaCl, 10% (v/v) glycerol and 10 mM DTT. The protein was eluted at the rate of 0.2 mL/min at 4 °C in 2 mL fractions. Gel filtration standards from BioRad (bovine thyroglobulin (670 kDa), bovine γ -globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa) and vitamin B₁₂ (1.35 kDa)) were run on the column using the same buffer system to obtain the calibration curve.

Crystallization of *Mtb* MAS

Initial crystallization trials for the purified MAS were carried out at 18 °C by sitting drop vapor diffusion method using a nanoliter robotic screening setup. The screening was done using commercial sparse-matrix and grid crystallization screens, as well as custom made factorial and random screens, giving >1,200 conditions per sample. The crystallization trials were setup with a 1:1 ratio of MAS protein (10 mg/ml) and crystallization condition in 96 well Intelli-Plates (Hampton Research) using a Mosquito nanoliter robot (TTP Labtech). The crystallization screening was further expanded by

setting up trials at two different temperatures (4 °C and 18 °C) using varying concentrations of MAS (12-35 mg/ml) in different drop ratios with the screening conditions, and by addition of substrates methylmalonyl CoA, NADPH and fatty acyl-CoAs (C₁₂-C₂₀) to the crystallization drops. In addition, the crystallization trials were also performed under oil (paraffin and silicon oils, Hampton Research) using the microbatch method. The two truncations of MAS (T₁ and T₂) that expressed as soluble proteins were also screened for crystallization in the same way as was done for the full length protein.

In situ proteolysis for MAS crystallization

To facilitate crystallization of MAS, in situ proteolysis was performed by either pre-incubating the protein with trace amounts of a protease (trypsin or chymotrypsin or thrombin) (Sigma) before setting crystallization trials, or by direct addition of the protease to the trial drops. The protease was added in three different ratios to the protein (1:2,000, 1:10,000 and 1: 100,000, *w/w*) and the crystallization trials were setup at 4 °C and 18 °C. The protease working stocks were prepared at 1-2 mg/mL according to the manufacturer's instructions, and added immediately prior to setting up of crystallization trials.

Crystallization of MAS using heterogeneous nucleating agents

Seaweed and hydroxyapatite were selected for assessment as heterogeneous nucleating agents in the crystallization trials of MAS. Both seaweed and hydroxyapatite

were ground to a fine powder using a pestle and mortar, and dried at 50 °C for 2-4 h to remove any moisture. For crystallization screening, each nucleating agent (seaweed, hydroxyapatite and a 1:1 (w/w) mixture of the two) was added separately to the protein (at 0.5 µg nucleating agent per µl of the protein) and the drops were set up at 4 °C and 18 °C in 96 well Intelli-plates.

Optimization of MAS crystals by counter-diffusion and seeding techniques

For counter-diffusion technique, Granada crystallization box (GCB) (Hampton Research) was used to grow single crystals of MAS inside capillaries (Garcia-Ruiz et al., 2002). In the GCB, a reservoir was filled with melted 1.5% (w/v) agarose in 0.1 M HEPES, pH 7.5 buffer and allowed to cool to room temperature. A glass capillary (0.5 mm internal diameter, 100 mm length (Hampton research)) was filled with MAS solution (20-30 mg/mL) to a height of 5-6 cm and inserted into the agarose layer to a depth of 2-3 mm. The top open end of the capillary was sealed with superglue and crystallization buffer (0.1 M HEPES Na, pH 7.5, 10-8% (v/v) 2-propanol, 20-16% (w/v) PEG 4,000) was layered over the agarose (volume equal to the agarose volume). The GCB was then covered and placed at 18 °C.

Seeding technique was used to obtain single, large crystals of MAS (microseeding) and to screen for new crystallization conditions for MAS (microseed matrix screening) (Bergfors, 2003; D'Arcy et al., 2007). Seed stock was prepared by transferring MAS crystals to a Seed Bead tube (Hampton research) containing 50 µl of the well solution and vortexing it 3-5 times at 30 sec per cycle. The temperature of the

seed solution was maintained at 4 °C by cooling the tube on ice between the vortexing cycles. Microseeding was carried out by streaking of the seeds through the crystallization drops by a cat whisker. The crystallization drops contained 1 µl of MAS solution (15-20 mg/mL) and 1 µl of crystallization buffer (0.1 M HEPES Na, pH 7.5, 8% (v/v) 2-propanol, 15-20% (w/v) PEG 3,350 and 5% (v/v) glycerol) and were equilibrated against 50-100 µl crystallization buffer in the reservoir in a sitting drop vapor diffusion set up. The plates were incubated at 4 °C and 18 °C, and large single crystals of MAS were obtained in 7-10 days. Microseed matrix screening was done at 18 °C in a 96 well format using the Mosquito nanoliter robot to set up 1 µl crystallization drops containing 0.5 µl of MAS solution (20-25 mg/mL) mixed with 0.4 µl of the screening condition and 0.1 µl of the seed stock.

Growing MAS crystals in gel

To obtain single, well-developed crystals with perfect habit, MAS was crystallized in a matrix of low melting agarose (LMA, Hampton Research) (Lorber et al., 2009). To prepare the crystallization drops, equal volumes of the protein solution (20-25 mg/mL) and the crystallization condition (0.1 M HEPES Na, pH 7.5, 8-10% (v/v) 2-propanol, 15% (w/v) PEG 3,350 and 5% (v/v) glycerol) were mixed, and an appropriate volume of a 2% (w/v) stock of LMA was added to give a final concentration of 0.1-0.4% (w/v) agarose in the drop. The crystallization was carried out at 18 °C.

Post-crystallization treatment of MAS crystals

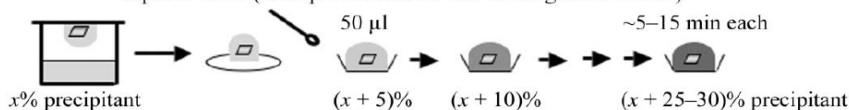
To improve diffraction quality of MAS crystals, the solvent content of the crystals was reduced in a controlled manner by implementing different step-wise dehydration strategies (Heras and Martin, 2005) (Fig. 2-1). After screening a large number of crystals, a dehydration protocol was designed wherein single well-formed crystals of MAS were transferred to 4 °C and slowly dehydrated by small step-wise increments in the concentrations of the precipitant PEG 3,350 and the cryoprotectant glycerol in the drop. Three solutions with different concentrations of PEG 3,350 and glycerol were mixed in different ratios to give final concentrations of PEG 3,350 and glycerol from 12-25% and 19-30%, respectively (Table 2-1). At each step, the reservoir buffer was replaced with 50-100 μ l of the dehydrating solution and the drop was allowed to equilibrate for 12-24 h between the steps. After the final step, the drop was left to equilibrate for an extra day against the stabilization buffer (0.1 M HEPES Na, pH 7.5, 10% (v/v) 2-propanol, 30% (w/v) PEG 3,350 and 25% (v/v) glycerol) in the reservoir.

In addition to the gradual step-wise dehydration method, a modified open air dehydration protocol was also tested for improvements in the diffraction of MAS crystals (Fig 2-1). In this extreme dehydration protocol, MAS crystals were first transferred to a stabilizing solution (0.1 M HEPES Na, pH 7.5, 8-10% (v/v) 2-propanol, 20% (w/v) PEG 3,350 and 15% (v/v) glycerol) and equilibrated over night against 100 μ l of this solution at 18 °C. After incubation, the sealing tape was removed and the stabilized crystals were air dehydrated until only a thin film of the stabilizing solution remained around the crystals and harvested by mounting on cryo-loops.

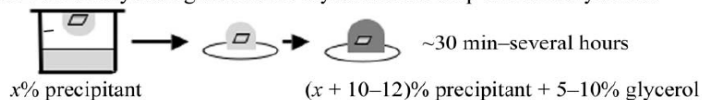
Table 2-1. Crystal dehydration protocol. The solutions used are: solution A (0.1 M HEPES Na, pH 7.5, 10% (v/v) 2-propanol, 15% (w/v) PEG 3,350 and 10% (v/v) glycerol); solution B (same as A, but with 25% (w/v) PEG 3,350 and 15% (v/v) glycerol); and solution C (as A, but with 30% (w/v) PEG 3,350 and 25% (v/v) glycerol).

Step	A : B : C (%)	Incubation time (h)	Volume added (μl)	[Peg 3,350] %	[Glycerol] %
1	60 : 40 : 0	12	2	19	12
2	20 : 80 : 0	12	4	23	14
3	0 : 80 : 20	24	8	26	17
4	0 : 50 : 50	24	10	27.5	20
5	0 : 30 : 70	12	10	28.5	22
6	0 : 10 : 90	12	10	29.5	24
7	0 : 0 : 100	12	10	30	25
8	0 : 0 : 100	24	10	30	25

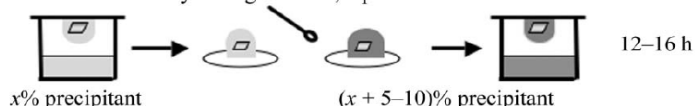
Method 1: Serial transfer to increasing concentration of precipitant; dehydrate either over reservoir or exposed to air (example increments and soaking times shown)



Method 2: Add dehydrating solution to crystallization drop and air dehydration



Method 3: Transfer to dehydrating solution, equilibrate over reservoir



Method 4: Transfer cover slip to reservoirs containing serial increase of dehydrating solution

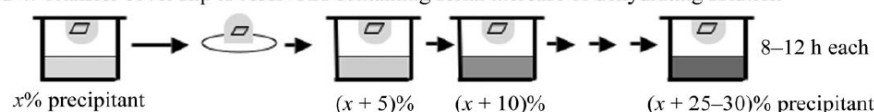


Figure 2-1. Crystal dehydration methods. In these methods the crystal is either transferred to the dehydrating solution, or the dehydrating agent is added to the crystallization drop. The dehydrating solutions are made by adding increasing concentrations of the precipitant to the mother liquor, or by supplementing with increasing concentrations of the cryoprotectant. The concentration of the dehydrating solution is increased in small steps of 0.5-5% up to a final concentration of 5-30% (w/v) in the drop, and the soaking time ranges from few minutes to days. The methods outlined here are easily adapted to the sitting drop vapor diffusion experiments. Figure reproduced from Heras, B. and Martin, J. L., *Acta Cryst.* (2005). D61, 1173-1180, with permission of © [2005] IUCr. <http://dx.doi.org/10.1107/S0907444905019451>

Crystallographic screening and data collection

The crystals were mounted using cryo-loops (Hampton Research) and flash-frozen by plunging in liquid nitrogen for screening and data collection. Prior to freezing, MAS crystals obtained after initial optimizations were briefly incubated (0.5-5 min) in a drop of mother liquor containing 15-20% glycerol as a cryoprotectant. For the crystals that were dehydrated, the final stabilization conditions provided sufficient

cryoprotection, so they were directly frozen from the drops. The crystals were screened for diffraction at cryogenic temperatures on beamlines 19-ID, 23ID-B and 23ID-D of the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL). For all the screened crystals, the diffraction data was analyzed by the *HKL2000* suite (Otwinowski and Minor, 1997). After screening hundreds of crystals, diffraction data was collected for a crystal that diffracted to better than 10 Å on the beamline 19-ID at the Advanced Photon Source, Argonne National Laboratory. A complete dataset was collected at 100 K using an ADSC Q315 CCD detector at a wavelength of 0.97929 Å and a detector distance of 800 mm. For data collection, the crystal was exposed 12 sec per image at 0.3° oscillation per frame and a total of 320 images were recorded. The diffraction images were indexed, integrated and scaled using *HKL2000*. The scaled dataset was analyzed by *XPREP* and *Xtriage* to determine the space group of the crystal (Sheldrick, 2008a). Solvent-content calculations (Matthews, 1968) to determine the number of molecules in the asymmetric unit were done in the *CCP4* suit.

RESULTS AND DISCUSSION

Cloning and purification of *Mtb* MAS

An important prerequisite for the protein crystallization studies is the availability of sufficient amounts of pure and homogenous form of the protein under study. This is frequently achieved by overexpression of the protein of interest in a suitable heterologous host, followed by purification using an engineered affinity tag. A similar

approach was adopted for the production and purification of *Mtb* MAS protein from *E. coli*, a commonly used host for the recombinant expression of bacterial proteins. However, due to the inherent difficulties in the cloning and heterologous overexpression of large macromolecular complexes, standard cloning protocols were modified for successful cloning of the *mas* gene from *Mtb* H37Rv genomic DNA. Screening of different commercially available PCR reaction mixtures and DNA polymerases led to the identification of a combination of failsafe mix D and the high fidelity Phusion DNA polymerase that amplified the full length *mas* gene (6336 bp) successfully without amplifying non-specific fragments from the *Mtb* genome (Fig. 2-2a). It was also observed that keeping the number of extension cycles in the PCR to below 20 prevented mutations in the amplicons. For cloning the PCR product, a chimeric vector pET28_30b was created by incorporating the MCS of pET30b into the MCS of pET28b vector to create the restriction sites for *NheI* and *KpnI*, restriction enzymes that were found not to cut the *mas* gene. The restriction digestion of the PCR product followed by its ligation into the empty vector produced the pET28_30b-MAS expression construct that incorporated an N-terminal His₆-tag for aiding the purification of the recombinantly expressed MAS protein by Ni²⁺ affinity chromatography.

The MAS protein expressed in the *Mtb* cells undergoes a posttranslational modification (PTM) that covalently attaches a 4'-phosphopantetheine prosthetic group derived from CoA, to a conserved Ser residue on its carrier domain (ACP). In *Mtb*, this PTM is performed by the phosphopantetheinyl transferase (PPTase) PptT (Rv2974c) (Chalut et al., 2006). However, the PTM of heterologously expressed MAS is not

efficient due to the non-specificity of the *E. coli* PPTase towards the ACP domain of the MAS protein. As a result, a mixed population of *apo*- and *holo*-MAS is produced in the *E. coli* cells. The homogeneity of the protein sample is a critical parameter that can influence crystallization and diffraction properties of the crystals (McPherson, 2004). Therefore, to improve the homogeneity of MAS, *Mtb* PptT was cloned in a pCDFDuet-1 vector and co-expressed with MAS in the *E. coli* cells. The vector pCDFDuet-1 was chosen as it provided plasmid compatibility and antibiotic selectivity with the pET28_30b-MAS expression plasmid, and allowed construction of a tag-less PptT expression system to facilitate its easy separation from MAS on a Ni²⁺ affinity purification system. The integrity of the cloned constructs was verified by double-digestions (not shown) and DNA sequencing.

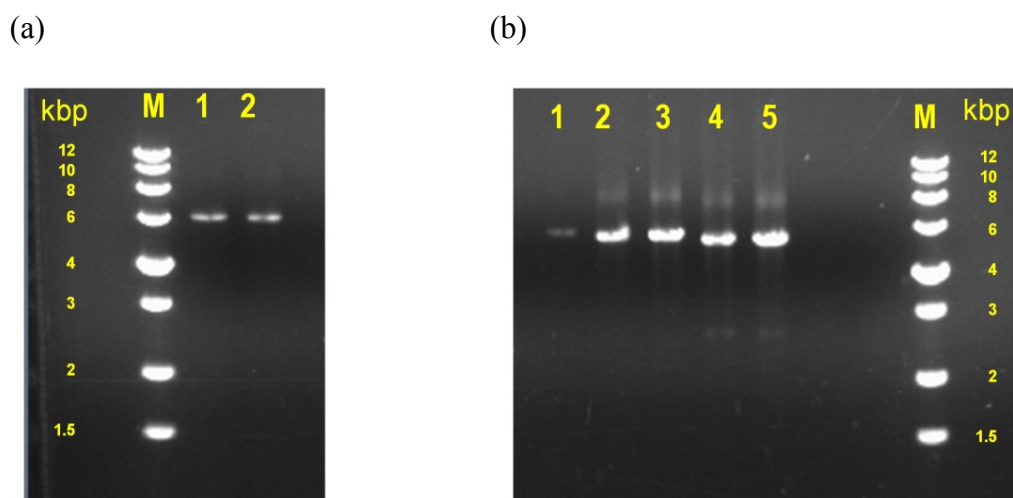
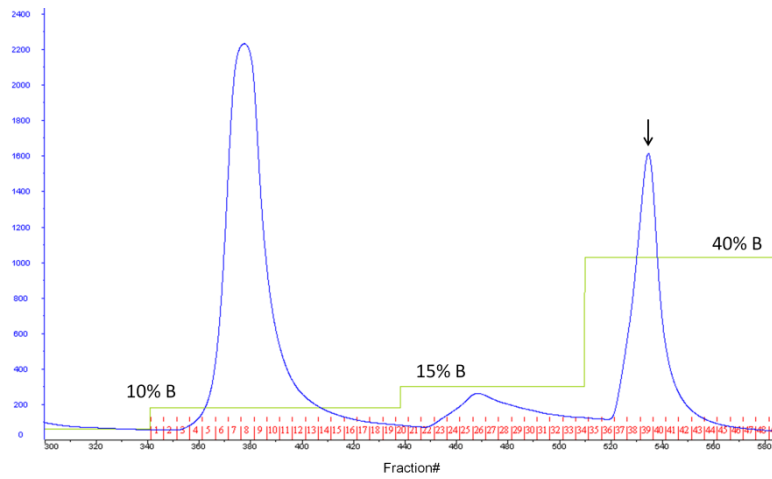


Figure 2-2. PCR amplification of MAS. DNA electrophoresis on 0.8% agarose gel shows PCR amplified fragments of MAS from the H37Rv genomic DNA. (a) Full length (6336 bp) MAS product shown in lanes 1 and 2. (b) Amplified C-terminal ACP domain truncations of MAS, T₁-T₅ shown in lanes 1-5, respectively. For reference, DNA molecular weight markers (0.5–12 kbp) (Novagen) are shown in lane marked M.

The cloned constructs pET28_30b-MAS and pCDFDuet-1-Rv2794c were co-transformed into different expression strains of *E. coli* and tested for the overexpression of soluble MAS protein. After screening, the *E. coli* host strain Rosetta 2(DE3)pLysS was selected for large scale production of MAS since it not only enhances the expression of rare codon containing proteins, but also prevents leaky expression of the recombinant proteins through the suppression of basal T7 RNA polymerase production by T7 lysozyme (Studier, 1991). It was observed that optimum MAS expression was obtained by induction with 0.5 mM IPTG followed by overnight growth of the cells at 18 °C. The overexpressed His₆-tagged MAS protein was initially purified by Ni²⁺ affinity chromatography by a step-gradient protocol that was designed to significantly reduce the amount of co-eluted impurities (Figs. 2-3a and 2-3b).

After further purification using the protocols described in the methods section, the MAS protein was successfully purified to homogeneity and >95% purity as judged by the SDS-PAGE (Fig. 2-4b). The purified recombinant MAS protein showed a band at 225 kDa consistent with the predicted molecular weight based on the gene sequence (Fig. 2-4b).

(a)



(b)

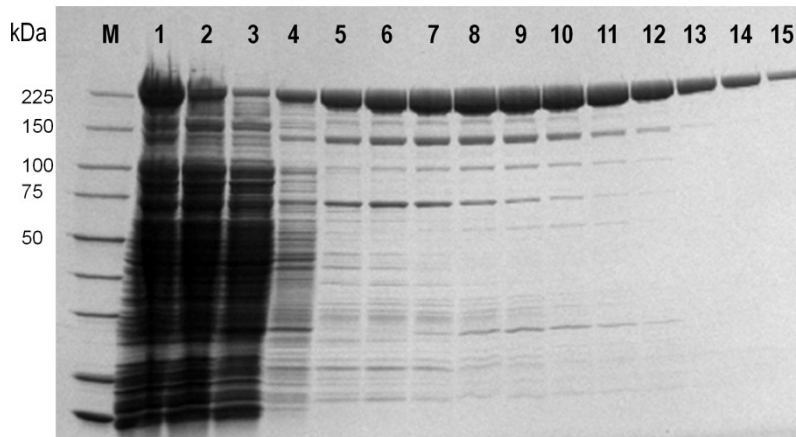
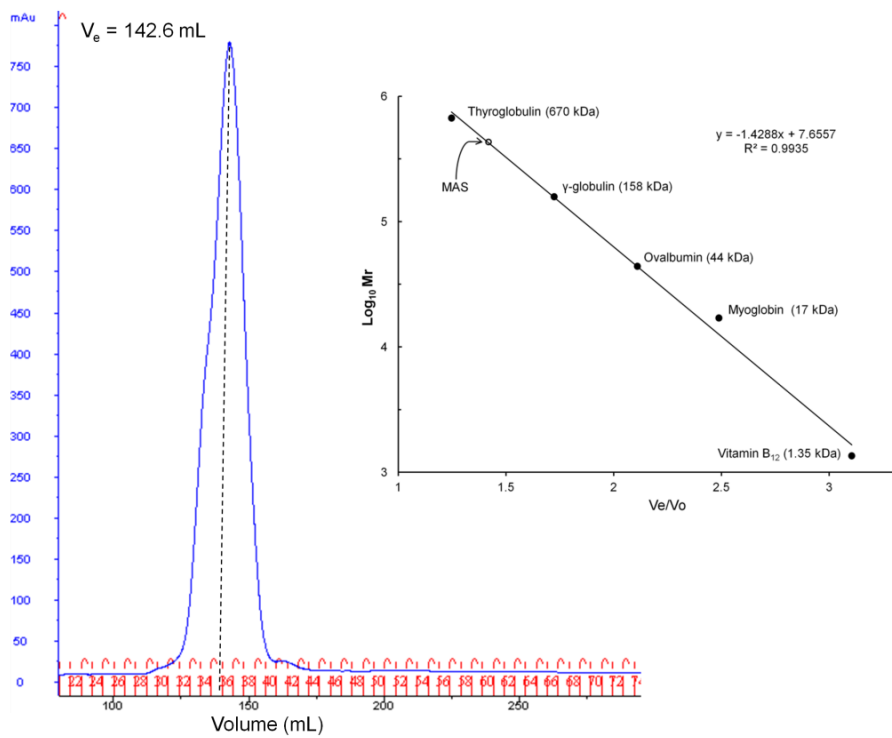


Figure 2-3. Purification of MAS by Ni^{2+} affinity chromatography. (a) The elution profile of MAS obtained from the step-gradient purification protocol. The blue trace represents the absorbance at 280 nm and shows absorbance peak (black arrow) under which MAS elutes from the column. (b) A 4-20% gradient gel showing protein content of the fractions eluted using different concentrations of buffer B in (a) and visualized by coomassie staining. Most impurities eluted in 10% and 15% B wash (lanes 2-4) and purified MAS eluted with 40% B (lanes 5-15). For reference, Perfect protein markers were run on the gel (M) and their molecular weights are listed alongside the gel in kDa.

(a)



(b)

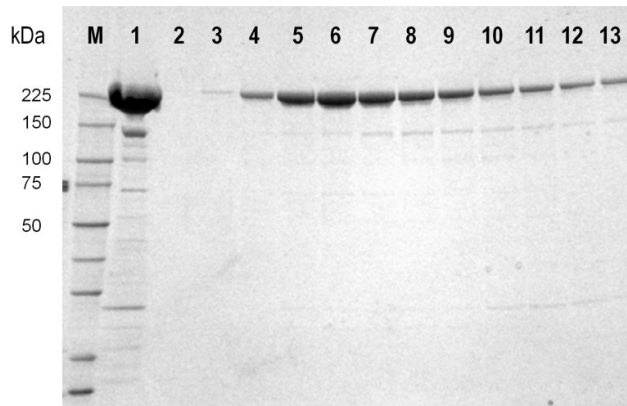


Figure 2-4. Gel filtration of MAS. (a) MAS eluted as a single peak from Suprose 6 gel filtration column. Inset shows the calibration curve determined using BioRad gel filtration molecular weight standards (filled circles). MAS is shown on the curve (open circle) corresponding to a MW of ~ 446 kDa, consistent with MW of a dimer. (b) Fractions of MAS obtained after gel filtration analyzed on 4-20% gradient SDS-PAGE gel and visualized by coomassie staining. MAS ran as a single band at ~ 225 kDa on the gel. For reference, Perfect protein markers were run on the gel (M) and their molecular weights are listed alongside the gel in kDa.

It has been observed that the large multi-domain proteins have higher structural and conformational heterogeneity compared to the single domain proteins (Leibundgut et al., 2008). Such heterogeneity can either significantly lower the likelihood of crystallization, or it can lead to poorly formed crystals with internal disorder and lattice imperfections (Mueller et al., 2007). This suggested that a reduction in the flexibility of multi-domain MAS protein could facilitate its crystallization. The ACP domains of FAS I and PKS systems have inherent flexibility due to their functional requirement to deliver the tethered substrate to different active sites at different locations on these proteins. This is exemplified by the lack of electron density for the ACP domain in the high resolution mammalian FAS I structure that was attributed to its flexible nature (Maier et al., 2008). Therefore, to increase the likelihood of obtaining crystals of MAS, five different ACP domain truncations were designed, cloned and tested for expression. These truncations were designed based on the secondary structure and domain predictions. All five truncation were successfully PCR amplified from *Mtb* H37Rv genomic DNA using the same reaction mix that was used for amplifying the full length *mas* gene (Fig. 2-2b). These amplified PCR fragments were subsequently cloned into the pET28_30b vector as was done for the full length MAS, and tested for protein expression in *E. coli* Rosetta 2(DE3)pLysS cells. Analysis of the expression test by SDS-PAGE revealed that only two truncations of MAS, T₁ and T₂ expressed as soluble proteins (Fig. 2-5). These two truncations were subsequently expressed on a larger scale and purified to homogeneity using the same purification protocols that were used to purify the full length MAS protein.

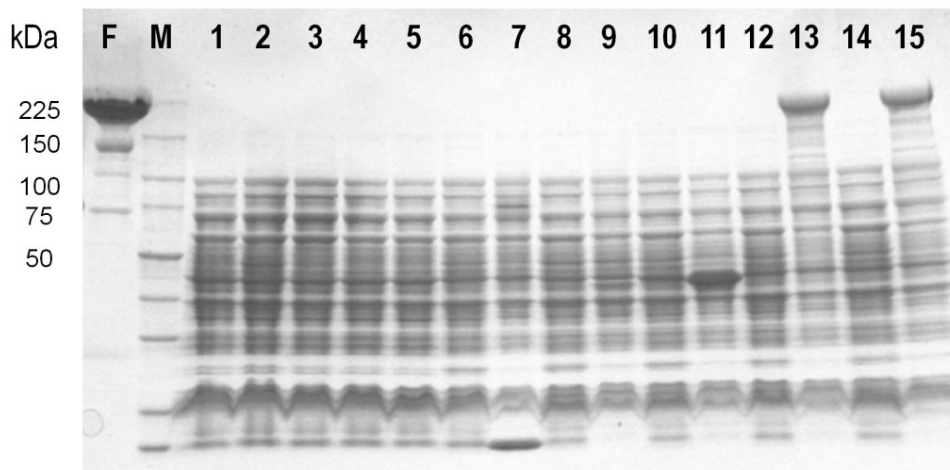


Figure 2-5. Expression analysis of the ACP domain truncation constructs of MAS. *E. coli* Rosetta 2(DE3)pLysS cells transformed with five different truncation constructs T₁-T₅ were tested for soluble protein expression. Empty vector controls (lanes 1-5) were cultured the same way as the cells that were transformed with the expression constructs. Aliquots of the induced cultures containing equal number of cells (corresponding to OD₆₀₀ = 0.3) were lysed and centrifuged at 13,000g to remove insoluble material from the soluble fraction in the supernatant. After transferring supernatant to another tube, the pellet from the lysate was re-suspended in a small volume of buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl). Equal amounts of proteins from the supernatants (lanes 7, 9, 11, 13, 15 for T₅, T₄, T₃, T₂ and T₁ respectively) and the re-suspended pellets (lanes 6, 8, 10, 12, 14 for T₅, T₄, T₃, T₂ and T₁ respectively) were loaded on a 4-20% gradient SDS-PAGE gel after protein concentrations were determined using the Bradford assay kit (BioRad). Protein bands on the gel were visualized by staining with coomassie blue. It is seen in the figure that only MAS truncations T₁ (lane 15) and T₂ (lane 13) expressed as soluble proteins. For reference, purified full length MAS (F) and Perfect protein markers were run on the gel (M). The molecular weights of the standards are listed alongside the gel in kDa.

Oligomeric state of *Mtb* MAS

The MAS protein that was recombinantly expressed and purified from *E. coli* was analyzed by analytical size-exclusion chromatography on a Superose 6 column to determine its oligomeric state. To calibrate the size-exclusion column, molecular weight standards were also run on the column using the same buffer system that was used for

MAS. The elution volumes of the standards were plotted and a calibration curve was obtained by linear regression (Fig. 2-4a). The elution profile of MAS contained a single peak and based on the elution volume, the molecular weight of MAS in solution was calculated to be ~446 kDa (Fig. 2-4a). This result indicated the oligomeric state of MAS in solution to be a dimer. This finding was in agreement with earlier studies which reported that the MAS purified from *Mtb* strain BCG was a dimer in solution (Rainwater and Kolattukudy, 1985). Thus, the recombinantly expressed MAS protein purified from *E coli* was determined to have the same oligomeric state as the natively expressed MAS.

Enhancement of 4'-PPT modification in recombinantly expressed MAS

The carrier domains of FAS, PKS and NRPS systems are converted to their *holo*-forms *via* covalent attachment of a 4'-PPT modification by their cognate PPTase in the host organism (Lambalot et al., 1996). Thus, non-specificity of the PPTase of the expression host towards the carrier domain of a heterologously expressed PKS would lead to production of an inactive or partially active protein. This would lead to a stoichiometric heterogeneity in the protein sample that can significantly reduce the likelihood of crystallization. A convenient strategy for overcoming this limitation in expression host is to co-express the PKS with its cognate PPTase. The *Mtb* genome encodes two PPTases, AcpS and PptT, however, since the information regarding their ability to transfer 4'-PPT to the ACP domain of MAS was not known, a preliminary co-expression study was conducted to evaluate their PPTase activity towards MAS. Both the PPTases of *Mtb* were cloned from the H37Rv genomic DNA and separately co-

transformed with MAS in *E. coli* Rosetta 2(DE3)pLysS cells. The MAS protein from both sets of co-expressions was purified separately as detailed in the methods section. The efficiency of 4'-PPT transfer to the ACP domain was then determined for both of the PPTases using an *in vitro* fluorescence labeling assay. This assay utilized the ability of the promiscuous PPTase Sfp from *B. subtilis* to transfer a fluorescently labeled 4'-PPT from BODIPY-CoA to the ACP domain of the MAS protein *in vitro*. After the assay, samples were run on a SDS-PAGE gel and the extent of fluorescent labeling of the ACP domain of MAS was determined by visualization under UV-light. Analysis of the fluorescent bands showed that the MAS protein co-expressed with PPTase PptT had the lowest fluorescence labeling of the protein (Fig. 2-6). This indicated an efficient *in vivo* 4'-PPT transferase activity of PptT toward MAS ACP domain, and suggested it to be the cognate *holo*-ACP synthase for the *Mtb* MAS protein. This finding was later confirmed in a published report that identified PptT as the cognate PPTase for all the PKS systems in *Mtb* (Chalut et al., 2006). Therefore, PptT was selected for co-expression with MAS to produce *holo*-MAS for crystallization studies.

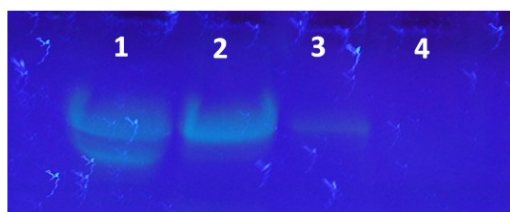


Figure 2-6. Analysis of 4'-PPT modification in recombinantly expressed MAS. The MAS protein co-expressed with *Mtb* PPTase AcpS (lane 2) or PptT (lane 3) was incubated *in vitro* with *B. subtilis* Sfp and BODIPY-CoA to assess their 4'-PPT transferase activity towards MAS ACP domain. Equal amount of protein was loaded in each lane, and the controls included MAS expressed in *E. coli* without *Mtb* PPTase (lane 1) and a C-terminal ACP domain truncation of MAS (lane 4).

Crystallization of *Mtb* MAS

Large macromolecular complexes generally crystallize under a narrow range of parameters and only from a specific combination of precipitants and chemical additives in the crystallization cocktail (Maier et al., 2010). Since each protein represents a unique chemical entity, it is not possible to predict conditions under which a protein of unknown structure will crystallize. Therefore, large scale screening trials are performed to identify initial crystallization conditions. A similar high throughput screening approach was used to identify conditions favorable for growing MAS crystals. In this approach, a number of crystallization trials of MAS were set up using commercial, as well as custom made crystallization screens in a 96-well format by a nanoliter screening robot. Commercial screens contain precipitants at concentrations that can be too high for large macromolecules. So, MAS was also screened with conditions diluted in different ratios. In addition, the substrates of MAS were also added during crystallization trials in order to facilitate crystal formation. Thus, the combination of different crystallization parameters resulted in >2,000 unique crystallization experiments per sample of MAS protein. However, despite extensive screening, initial attempts to obtain the MAS crystals were unsuccessful.

In large multifunctional proteins confirmation heterogeneity can occur due the domain movements, which can significantly lower the likelihood of crystal formation. In PKS and FAS I systems, the ACP domain displays inherent flexibility due to its functional requirement to deliver covalently tethered reaction intermediates for processing to different active sites on the protein (Smith and Tsai, 2007). The mobile

nature of the ACP domain was exemplified by the missing electron density for the ACP domain in both the fungal and the mammalian FAS I crystal structures (Jenni et al., 2007; Maier et al., 2008). Therefore, to facilitate MAS crystallization, two ACP domain truncations of MAS were made and screened for crystallization as was done for the full length MAS. Furthermore, *in situ* proteolysis was also performed in the crystallization trials of MAS to further increase chances of obtaining crystals. Limited proteolysis was successfully used to separate more flexible TE domain from the core domains of animal FAS I (Mattick et al., 1983). Also, many studies have reported application of *in situ* proteolysis for obtaining crystals of proteins that had failed to yield crystals by conventional screening (Dong et al., 2007). Analysis of these proteolysis experiments indicated that the protease had modified the proteins by cleaving off flexible linkers, N- or C- termini, or surface residues from the original protein resulting in the formation of more stable domains or fragments that were more likely to crystallize (Gao et al., 2005). The full length MAS was tested in the *in situ* proteolysis screening trials using trypsin and chymotrypsin proteases since these were the most commonly used and successful proteases in such studies.

Extensive screening of the MAS truncations and the full length MAS under these conditions also failed to yield any crystals. Investigation of the proteolytic cleavage products of MAS resulting from the treatment with trypsin and chymotrypsin by SDS-PAGE revealed a number of fragments with no significant production of any major bands (Fig. 2-7). This indicated that the cleavage sites on MAS were easily accessible to a range of proteases, presumably due to the inherent domain flexibility in MAS. This

suggested that the MAS protein was not suitable for crystallization screening by the method of *in situ* proteolysis.

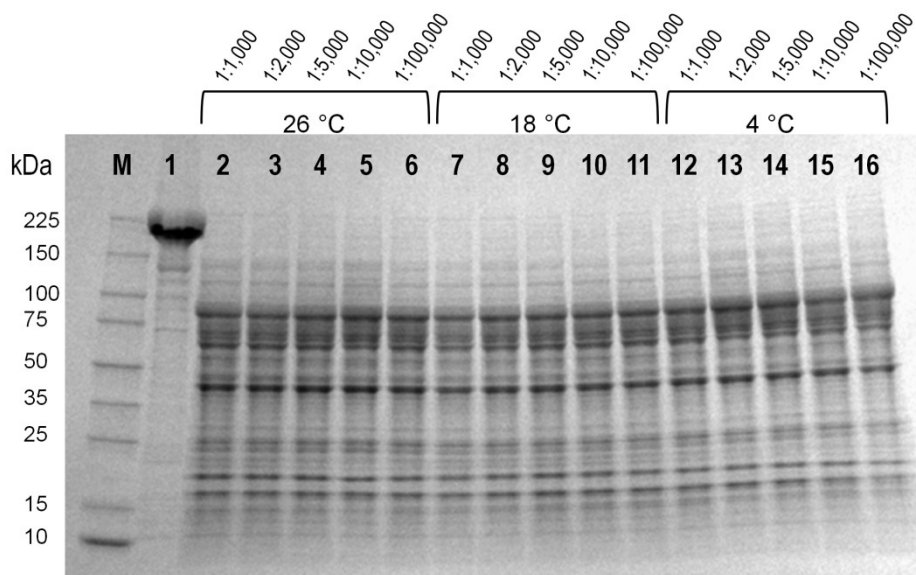
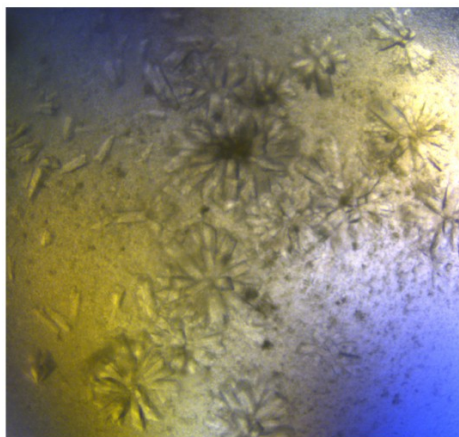


Figure 2-7. Analysis of MAS proteolysis by chymotrypsin. MAS was incubated with different ratios (*w/w*) of chymotrypsin at three temperatures for 5-30 min. After the incubation period, reaction was stopped by addition of the gel loading dye and heating the samples at 95 °C for 5 min. Samples were run on a 4-20% SDS-PAGE gel and the cleavage products were visualized by staining with coomassie blue. MAS cleavage with trypsin was analyzed in the same manner (not shown). For reference, purified full length MAS (lane 1) and Perfect protein markers were run on the gel (M). The molecular weights of the standards are listed alongside the gel in kDa.

For successful crystal formation, an essential step is the formation of a stable crystal nucleus under the supersaturating conditions in the crystallization drop (McPherson, 2004). Crystal nucleation is an energy driven event that can be adversely affected by various parameters in the crystallization experiment. One of the ways to increase the likelihood of finding conditions suitable for stable nucleus formation is to systematically introduce heterogeneous nucleating agents in the crystallization trials. The heterogeneous nucleants provide a wide range of pores, cavities and other surface structural features that can confine protein molecules thereby raising their local concentration leading to increased instances of nuclei formation (Chayen et al., 2006; Thakur et al., 2007). Thus, they can help initiate crystal formation even in solutions which do not contain supersaturating concentrations of the protein of interest. Therefore, a similar crystallization screen was performed for MAS by addition of a 1:1 (w/w) mixture of seaweed and hydroxyapatite as heterogeneous nucleating agents to the screening drops. This screen led to the successful identification of a single condition employing PEG 4,000 and 2-propanol as precipitants, in which small clusters of MAS crystals were formed (Fig. 2-8a).

(a)



(b)

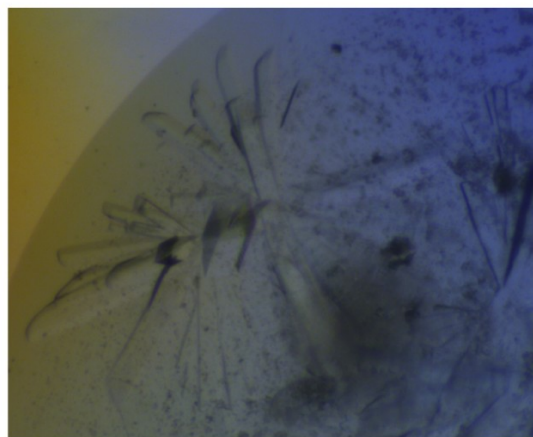
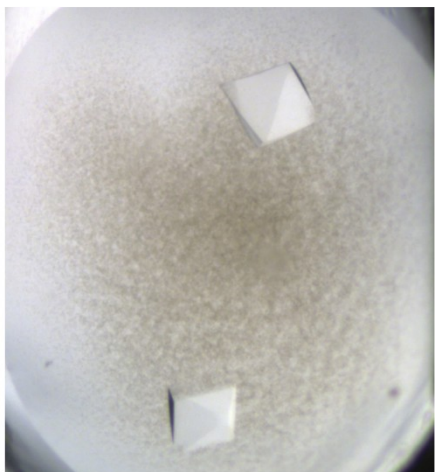


Figure 2-8. Initial crystallization hit of MAS. (a) First crystals of MAS obtained by screening with a 1:1 (*w/w*) mixture of seaweed and hydroxyapatite in the crystallization drop. MAS crystals grew as small clusters in a single condition consisting of 20% (*w/v*) Polyethylene glycol 4,000 and 10% (*v/v*) 2-Propanol in 100 mM HEPES sodium (pH 7.5) buffer. (b) Initial optimization of the hit condition gave larger crystals, but still required presence of heterogeneous nucleating agents for crystal formation.

Since the crystals in the initial hit were too small for mounting on the x-ray diffractometer, systematic grid screens were set up for optimization of the initial hit condition. Larger sized crystals were obtained from the optimization trials; however, their formation still required addition of the mixture of seaweed and hydroxyapatite (Fig. 2-8*b*). Subsequent optimization screens using alternate precipitants and additives identified 15-20 % (*w/v*) PEG 3,350 as a suitable precipitant and 5-10% (*v/v*) glycerol as an additive that led to growth of well formed, large single crystals of MAS. The optimized MAS crystals were either bipyramidal or rectangular rod like in morphology with dimensions of 0.1-0.3 x 0.4-0.5 x 0.6-0.9 mm³ (Figs. 2-9*a* and 2-9*b*).

(a)



(b)

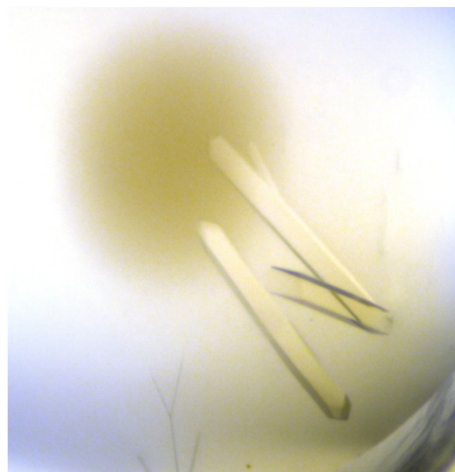


Figure 2-9. Optimization of MAS crystals. Refinement of crystallization conditions coupled with microseeding yielded well formed single crystals of MAS with (a) bipyramidal, or (b) rectangular rod like morphology. Crystals were obtained using 15-20 % (w/v) PEG 3,350 and 8-10% (v/v) 2-propanol as precipitants, with 5-10% (v/v) glycerol as an additive.

Initial diffraction images from the screening of the two crystal forms of MAS on the home source Bruker Daltonics diffractometer (Bruker Inc) showed diffraction spots in low resolution region (40 – 34 Å) indicating that these were indeed protein crystals (Figs. 2-10a and 2-10b). The bipyramidal form crystals gave much lower resolution diffraction (>40 Å) whereas the rectangular rod shaped crystals diffracted up to ~33 Å.

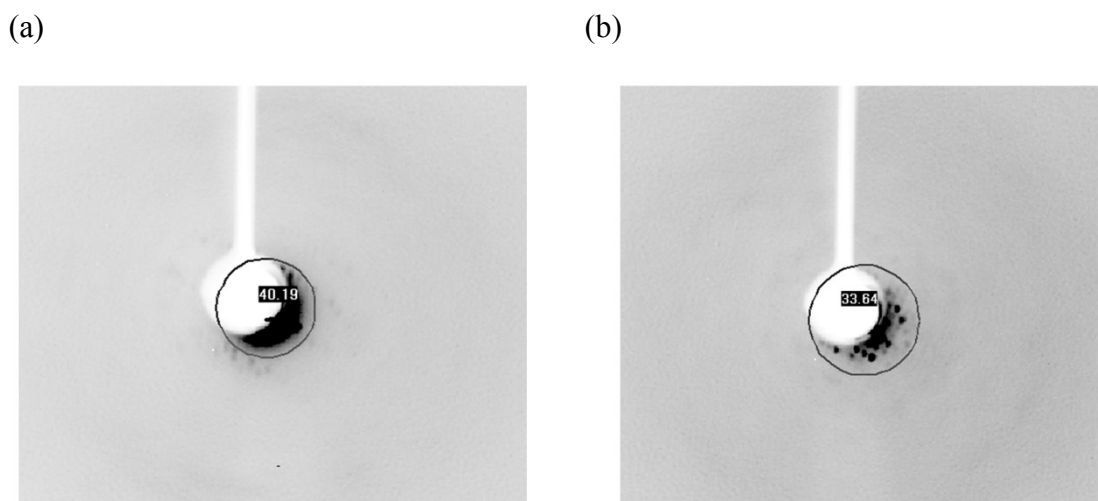


Figure 2-10. Initial diffraction from MAS crystals. Representative diffraction patterns from the initial screening of two different (a) bipyramidal, and (b) rectangular rod shaped, crystal forms of MAS, obtained using in-house Bruker X-ray diffractometer. The images were recorded on a CCD detector placed at a distance of 150 mm from the crystal that was exposed for 10 min at 1° oscillation per frame. The data was analyzed by proteum-II software and the resolution limits are indicated by black circles at ~40 Å for (a) and ~34 Å for (b).

In general, crystals of large macromolecules exhibit large asymmetric unit and unit cell dimensions. Thus, they diffract weakly and the diffraction spots are closely spaced on the detector as a consequence of long unit cell axes (Mueller et al., 2007). Due to the limitations of the lower x-ray intensities and the small detector area of in-house diffractometers, synchrotron radiation is the method of choice for screening and data collection from crystals of large macromolecular complexes. The synchrotron beamlines offer very high intensity x-rays that can compensate for weak diffraction from the crystals, and they provide CCD detectors with large surface areas which allow crystal to

detector distance to be adjusted to reduce the problem of overlapping spots (Mueller et al., 2007). Therefore, MAS crystals were also screened for diffraction on the beamlines 23-ID B/D and 19-ID at the Advanced Photon Source (APS), Argonne National Laboratory. Indeed, high intensity beams resulted in more diffraction spots with the resolution extending up to ~ 27 Å for the same rectangular rod shaped MAS crystals that were screened at the home source, and the larger crystal to detector distances of 400 – 800 mm also resulted in better spot separation (Fig. 2-11a).

Due to the better diffraction and crystal handling properties of the rectangular rod shaped crystals of MAS, they were selected for further optimization by post-crystallization treatments to improve their diffraction qualities. Crystal dehydration is a particularly suited technique for controlled removal of solvent from the crystals of large macromolecules. Solvent reduction in crystal by dehydration decreases unit cell volume that can alter the crystal contacts resulting in a better packing and ordering of the macromolecule leading to better diffracting crystals (Heras and Martin, 2005). Successful application of dehydration for improving the diffraction properties of the crystals of fungal FAS (improved from 8 Å to < 4 Å) (Jenni et al., 2007); yeast RNA polymerase II (from 6 Å to < 2.6 Å) (Cramer et al., 2000) and DsbG (10 Å to 2 Å) (Heras et al., 2003), suggested that this method could be used for improvements in the diffraction properties of MAS crystals. Initial dehydration of MAS crystals was done at 18 °C by sequential transfer of the crystals to solutions of increasing concentrations of PEG 3,350 (up to 30% in 5% increments) in the mother liquor with a 15-30 min equilibration between each transfer step. Screening of the dehydrated crystals at APS

showed slight improvements in the diffraction properties of the crystals with the resolution extending up to 20 Å (Fig. 2-11*b*). Subsequent screening with more crystals did not show further improvement in the diffraction. Therefore, two different dehydration protocols were designed and employed as detailed in the methods section. In the slow step-wise dehydration method, the crystals were transferred to 4 °C and dehydrated sequentially in the solutions containing small increments in concentrations of PEG 3,350 (up to 30%) and glycerol (up to 25%) over many days. In the second method, the crystals were transferred to a stabilizing solution and then equilibrated against air until only a thin film of the stabilizing solution remained around the crystals (Fig. 2-12*a*). The crystals dehydrated using the gradual step-wise protocol showed marked improvement in their diffraction qualities and they diffracted up to ~9-10 Å at the APS (Fig. 2-11*c*), whereas the air dehydration method resulted in the complete loss of diffraction, even though the crystals remained intact after air dehydration (not shown).

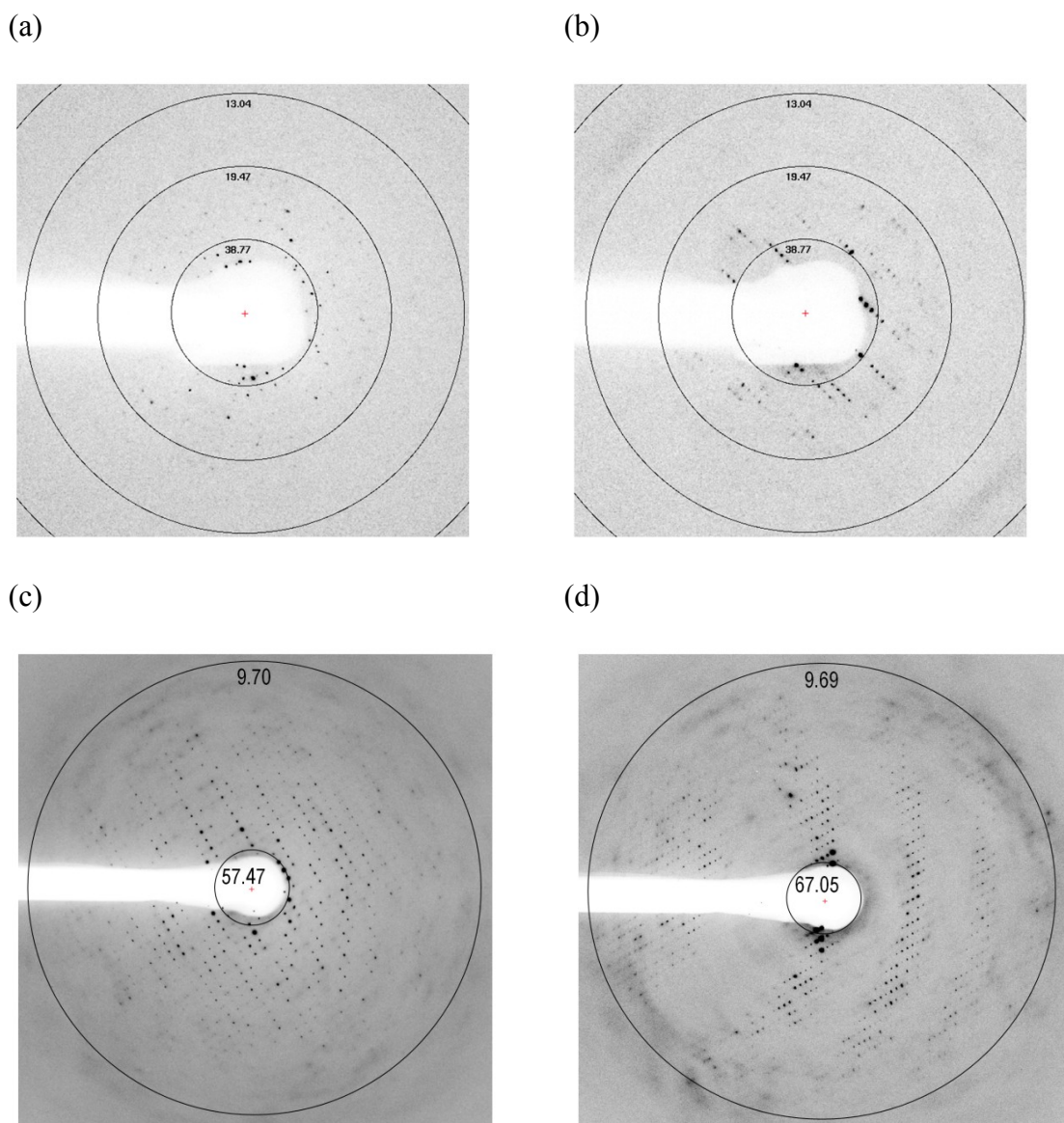


Figure 2-11. Improvement in the diffraction properties of MAS crystals. Diffraction pattern of the rectangular rod shaped MAS crystals collected on the beamlines 23-ID B/D and 19-ID at APS at wavelength of $1 - 0.979 \text{ \AA}$ and a detector distance of $600 - 800 \text{ mm}$. (a) Diffraction at APS from the crystal that was also screened at home source (see Fig. 2-11(b)). (b) Diffraction pattern extending up to $\sim 20 \text{ \AA}$ from the initial dehydration of MAS crystals at $18 \text{ }^\circ\text{C}$. (c) Improved diffraction from the slowly dehydrated crystals at $4 \text{ }^\circ\text{C}$ extending up to $\sim 9 \text{ \AA}$. (d) Diffraction from a dehydrated crystal grown in agarose gel also extending up to $\sim 9 \text{ \AA}$. The resolution rings are shown as black circles.

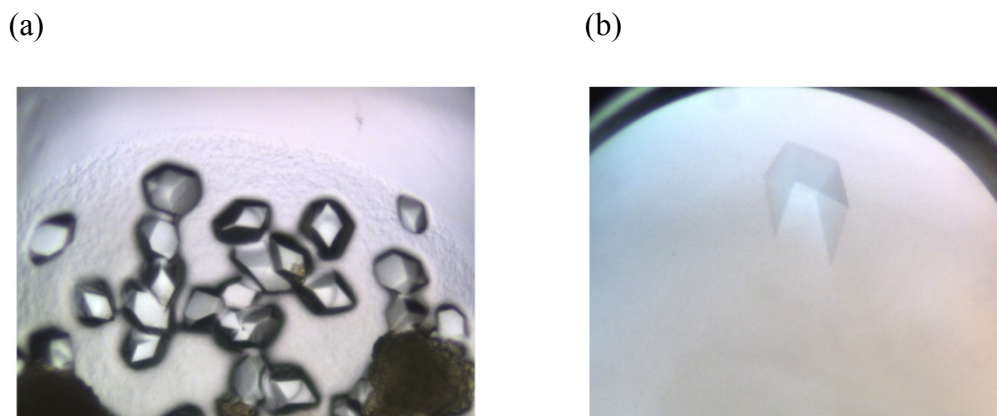


Figure 2-12. Post-crystallization treatment of MAS crystals. (a) Air dehydrated crystals of MAS. (b) Dehydration of MAS crystal grown in low melting agarose gel. The diffraction properties of this crystal were similar to the dehydrated crystals grown by sitting drop vapor diffusion method (see Fig. 2-12(d)). So crystals from the sitting drops were used for data collection due to ease of handling.

A complete diffraction dataset was collected from a slowly dehydrated crystal that diffracted to 9 Å on the beamline 19-ID at the APS using an ADSC Q315 CCD detector at a wavelength of 0.97929 Å and a detector distance of 800 mm. The diffraction images were indexed in the tetragonal Bravais lattice, integrated and scaled using *HKL2000*. Analysis of the integrated and scaled dataset by *XPREP* and *Xtriage* indicated that the crystal belonged to the tetragonal space group $P4_32_12$ or $P4_12_12$, with unit-cell parameters $a = b = 232.5$ Å, $c = 387.5$ Å, $\alpha = \beta = \gamma = 90^\circ$. Solvent-content calculations (Matthews, 1968) based on the molecular weight of MAS (MW ~225,000 Da) predicted either four or five molecules of MAS in the asymmetric unit. The Matthews coefficient for four molecules, $V_M = 2.74$ Å³ Da⁻¹, corresponded to a solvent content of 55.2% with a probability $P_{(\text{tot})}$ of 0.35, whereas for five molecules the V_M

($2.19 \text{ \AA}^3 \text{ Da}^{-1}$) corresponded to a solvent content of 44% with a probability $P_{(\text{tot})}$ of 0.54. The crystallographic data-collection statistics are summarized in Table 2-2.

To calculate electron density maps from the collected 9 \AA dataset, an attempt was made to obtain the initial phases by molecular replacement (MR) using crystal structures of the mammalian FAS I (PDB codes 2cf2, 2vz9) and the KS-AT didomain of module 5 from DEBS (PDB code 2hg4). MR trials using the full length FAS I or the KS-AT didomain did not give a solution. So another MR search with an ensemble of separated domains was performed and it was also unsuccessful in finding a solution. It has been reported that phasing of large macromolecules by MR can be difficult, particularly in the absence of a full length homologous structure, even though the template structures for individual domains may be available. MR is challenging in such cases because the domains represent only a small fraction of the total scattering mass thereby making it difficult to correctly place the component in the asymmetric unit (Maier et al., 2010). Similarly, the method of single heavy atom derivatization of native crystals for phasing is generally not suitable for large proteins due to a large number of potential binding sites in the asymmetric unit, which can lead to noisy Patterson maps, and thus decrease the likelihood of obtaining a solution (Mueller et al., 2007). Therefore, to overcome these limitations, heavy atom clusters have been used for initial phasing at intermediate ($\sim 10\text{-}4.5 \text{ \AA}$) resolutions (Johansson et al., 2008; Maier et al., 2010). These clusters bind to fewer sites on the proteins leading to limited heavy atom sites in the asymmetric unit, and they appear as single super-scatterers at intermediate resolutions

enabling easy interpretation of the Patterson maps for the identification of correct sites (Dauter, 2005).

Among the heavy atom clusters used for solving the structures of large macromolecular complexes, tantalum cluster ($\text{Ta}_6\text{Br}_{12}^{2+}$) has been the most widely used since it can be easily synthesized in many different ways. The $\text{Ta}_6\text{Br}_{12}$ cluster was successfully used for phasing structures of 50S and 30S ribosomal subunits (Ban et al., 2000); RNA polymerase II (Cramer et al., 2000); and the structures of *T. lanuginosus* FAS I and yeast FAS I (Jenni et al., 2006; Johansson et al., 2008). Therefore, a commercial $\text{Ta}_6\text{Br}_{12}$ cluster kit was used for preparing the $\text{Ta}_6\text{Br}_{12}$ -derivatized crystals of MAS. The heavy atom derivatization was performed either by transferring the MAS crystals to a stabilizing solution containing 1 mM $\text{Ta}_6\text{Br}_{12}$ cluster, or by adding few flakes of the dry powder directly to the crystal drop, and incubating for 1-2 days. After incubation crystals were inspected under microscope for successful derivatization as indicated by the green color of crystals, and screened for diffraction. However, the soaked crystals diffracted to only very low resolution, indicating that the crystal soaking had adversely affected crystal lattice and hence, no datasets could be collected. Further screening with other polynuclear metal clusters (for example, tungsten or platinum clusters) and refinement of back-soaking protocols will be needed for the successful determination of initial phases for the MAS structure determination.

Table 2-2. Data-collection and processing statistics for MAS crystals. Values in parentheses are for the highest resolution shell.

	<i>Mtb</i> MAS
Beamline	19-ID
Wavelength (Å)	0.97929
No. of crystals	1
Temperature (K)	100
Crystal-to-detector distance (mm)	800
Exposure time (s)	12
No. of images	320
Oscillation range (°)	0.3
Space group	<i>P4₁2₁2/P4₃2₁2</i>
Unit-cell parameters (Å, °)	$a = b = 232.5, c = 387.5, \alpha, \beta, \gamma = 90$
Solvent content (%)	55.2 (4 molecules/asu)/44 (5 molecules/asu)
Resolution range (Å)	50–8.96 (9.16–8.96)
No. of observations	13540 (291)
No. of unique reflections	7900 (281)
Completeness (%)	92.8 (58.2)
Multiplicity	3.2 (1.4)
R_{merge} (%)	12.8 (69.0)
$\langle I/\sigma(I) \rangle$	7.3 (0.5)

SUMMARY AND FUTURE DIRECTIONS

In this chapter, the development of the protocols for cloning, purification and crystallization of the 225 kDa full length MAS protein, a type I PKS and an essential enzyme of PDIM biosynthesis in *Mtb*, are described in detail. Furthermore, this is the first report, till date, of the successful crystallization and intermediate resolution data collection from the crystals of a recombinant full length type I PKS from *Mtb*.

The full length *mas* gene was successfully cloned from the *Mtb* genomic DNA by using a combination of reaction conditions and a high-fidelity DNA polymerase from two different suppliers. A chimeric expression vector was constructed for the heterologous expression of His-tagged MAS protein in *E. coli*. Further, it was determined that the co-expression of MAS with the *Mtb* PPTase Rv2794c was essential for the production of phosphopantetheinylated *holo*-MAS in *E. coli*. The full length *holo*-MAS was successfully purified using a step-gradient protocol and it was determined to be a dimer in solution, similar to the MAS purified from *M. bovis* BCG strain (Rainwater and Kolattukudy, 1985). Extensive screening for the crystallization of MAS led to the identification of a single condition in which MAS crystallized in the presence of heterogeneous nucleating agents. Further optimization of the initial hit condition produced large single diffractable crystals of MAS. The diffraction properties of the MAS crystals were significantly improved by the application of post-crystallization treatments. Crystal dehydration protocols were developed and employed for improving the diffraction of MAS crystals, and an intermediate resolution dataset

was successfully collected at the synchrotron from a dehydrated MAS crystal. Our initial attempts for obtaining the electron density maps at the intermediate resolution by using tantalum clusters for phasing were unsuccessful, suggesting that a wider screening of the heavy atom clusters and soaking conditions would be needed for the successful phasing of the MAS crystals. Furthermore, significant improvements in the diffraction properties of the mammalian FAS I crystals (Maier et al., 2010) due to fine refinements and optimizations of the crystal growth and stabilization conditions suggests that such an approach can also help to improve the diffraction properties of the MAS crystals. Additionally, the availability of the helium cryostat and the free-standing mounting systems on the modern synchrotron beamlines (Mueller et al., 2007) can help in obtaining the best quality data from the MAS crystals, which should further improve their resolution.

The methods and protocols described in this chapter were developed after a lot of troubleshooting of the technical challenges associated with the cloning, purification and crystallization of MAS, in *E. coli* as the heterologous host. The methods reported here should prove useful in devising strategies for initiating crystallographic studies on other PKS I systems in *Mtb*. It is hoped that such studies will provide atomic-level insights into the structure-function relationships of the core domains, as well as unravel the structural features required for their interaction with other proteins of the complex lipid biosynthesis pathways in *Mtb*. Finally, this information can also guide the development of structure-based intervention strategies to target PKS I systems for developing novel inhibitors against *Mtb*.

CHAPTER III

Mycobacterium tuberculosis* ACYL CARRIER PROTEIN SYNTHASE ADOPTS TWO DIFFERENT pH-DEPENDENT STRUCTURAL CONFORMATIONS

SYNOPSIS

The crystal structures of acyl carrier protein synthase (AcpS) from *Mycobacterium tuberculosis* (*Mtb*) and *Corynebacterium ammoniagenes* determined at pH 5.3 and pH 6.5, respectively, are reported. Comparison of the *Mtb* apo-AcpS structure with the recently reported structure of the *Mtb* AcpS–ADP complex revealed that AcpS adopts two different conformations: the orthorhombic and trigonal space-group structures show structural differences in the $\alpha 2$ helix and in the conformation of the $\alpha 3$ – $\alpha 4$ connecting loop, which is in a closed conformation. The apo-AcpS structure shows electron density for the entire model and was obtained at lower pH values (4.4–6.0). In contrast, at a higher pH value (6.5) AcpS undergoes significant conformational changes, resulting in disordered regions that show no electron density in the AcpS model. The solved structures also reveal that *C. ammoniagenes* AcpS undergoes structural rearrangement in two regions, similar to the recently reported *Mtb* AcpS–ADP

* Reprinted with permission from “*Mycobacterium tuberculosis* acyl carrier protein synthase adopts two different pH-dependent structural conformations” by Kuppan G., Aggarwal, A., Shipman, L., Besra, G.S. and Sacchettini, J.C., Acta Cryst. (2011). D67, 657–669. Copyright 2011 International Union of Crystallography. DOI:10.1107/S0907444911020221.

complex structure. *In vitro* reconstitution experiments show that AcpS has a higher post-translational modification activity between pH 4.4 and 6.0 than at pH values above 6.5, where the activity drops owing to the change in conformation. The results show that apo-AcpS and AcpS–ADP adopt different conformations depending upon the pH conditions of the crystallization solution.

INTRODUCTION

The fatty-acid synthase type I (FAS-I) pathway is widely present in eukaryotes (Smith, 1994) and consists of a single large polypeptide containing multiple active centers. The fatty-acid synthase type II (FAS-II) pathway is present in all bacteria and plants and comprises a system in which discrete monofunctional enzymes encoded by specific genes carry out each reaction in the pathway (Magnuson et al., 1993). Fatty-acid biosynthesis is more complicated in *Mycobacteria* and *Corynebacteria* species compared with other bacteria owing to the presence of both the FAS-I and FAS-II pathways.

Lipid biosynthesis is essential to the survival of *M. tuberculosis* (*Mtb*) (Taylor et al., 2010). There are two important components in fatty-acid biosynthesis, CoAs and acyl carrier proteins (ACPs), both of which are essential for priming and extending the growing acyl chain. Apo-ACP is post-translationally modified by the transfer of a 4'-phosphopantetheine (4'-ppt) prosthetic group from coenzyme A to holo-ACP by acyl carrier protein synthase (AcpS). In both the FAS-I and FAS-II pathways the fatty acids

are synthesized by repeated cycles of reactions, namely the transacylation, condensation, dehydration and reduction reactions (Rock and Cronan, 1996).

The AcpS enzyme was initially identified and characterized in *Escherichia coli*, in which it is an essential enzyme for bacterial growth (Lambalot and Walsh, 1995). The *Mtb* genome contains two phosphopantetheine transferases: AcpS and 4'-phosphopantetheinyl transferase (PptT) (Cole et al., 1998). Both enzymes are essential for bacterial survival and synthesis of virulence factors (Chalut et al., 2006). AcpS is predominantly associated with the biosynthesis of fatty acids in both FAS systems, but it is also involved in polyketide synthesis (Shen et al., 1992), nonribosomal peptide synthesis (Baldwin et al., 1991), transacylation of oligosaccharides (Geiger et al., 1991) and protein synthesis (Issartel et al., 1991). A recent study showed that AcpS is predominately associated with post-translational modification of FAS-I and the AcpM subunit of the FAS-II system (Chalut et al., 2006). In contrast, the PptT enzyme is involved in post-translational modification of type I polyketide synthase enzymes. The *Mtb* genome contains more than 18 type I polyketide synthase enzymes (Cole et al., 1998), which catalyze the biosynthesis of phenolic lipids (Dubey et al., 2003) in virulent strains. In the type II synthase, the hydrophobic fatty-acyl intermediates sequentially shuttle between enzymes through an ACP. All acyl chains undergoing biosynthesis by FAS are attached to the terminal cysteine thiol of 4'-ppt via a labile thioester bond, making AcpS an essential enzyme. In addition, the 4'-ppt prosthetic group has a long flexible arm that can reach into the active site of enzymes, where the sulfhydryl group is available for the attachment of an acyl group. The divalent Mg^{2+} ion is required for the

post-translational modification reaction and is essential for the biosynthesis of all bacterial fatty acids, several bacterial lipopolysaccharides and membrane lipids. Recently, it has been reported that the AcpS gene product is essential for the growth of *Streptococcus pneumoniae* (McAllister et al., 2000). The nature of FAS-II in bacteria is quite different from the mammalian lipid-biosynthetic pathway and provides various specific sites for selective inhibition by antibacterial agents. Furthermore, AcpS is widely present in *Mycoplasma*, Gram-positive and Gram-negative bacteria; thus, an AcpS inhibitor could alter the availability of holo-ACP for the acylation process and the availability of components for fatty-acid biosynthesis, making such inhibitors possible broad-spectrum antibacterial agents. In the past ten years, structures of AcpS have been solved for several pathogenic bacteria (Table 3-1), including *Streptomyces coelicolor* (PDB entry 2wdo; P. Dall'Aglio, C. Arthur, M. P. Crump, J. Crosby & A. T. Hadfield, unpublished work), *Streptococcus pneumoniae* (PDB entry 1fte) (Chirgadze et al., 2000), *Bacillus subtilis* (PDB entry 1f80) (Parris et al., 2000), *B. anthracis* (PDB entry 3hyk; Center for Structural Genomics of Infectious Diseases, un-published work) and *Mycobacterium smegmatis* (PDB entry 3gwm; C. Poulsen, M. Wilmanns & Y. H. Song, unpublished work), revealing the role of this enzyme in post-translational modification.

We report here the crystal structures of *C. ammoniagenes* AcpS (with and without CoA) and *Mtb* apo-AcpS. *Mtb* apo-AcpS has an overall structure similar to those of the recently reported *Mtb* AcpS–ADP complex and the *C. ammoniagenes* AcpS structure described here. Despite their overall similarity, the two structures exhibit significant conformational changes in the α -carbon backbone. *Mtb* AcpS was

crystallized both at low pH values (4.4–6.0) and higher pH values (6.5 and 7.5); interestingly, AcpS showed higher post-translational modification activity at the lower pH. The *Mtb* and *C. ammoniagenes* AcpS structures suggest that pH may play a significant role in conformational change but that CoA binding has no effect on the structural rearrangement. The transfer of the 4'-ppt moiety from CoA to the Asp-Ser-Leu (DSL) motif of the mycobacterial ACP (AcpM) is likely to require the association of AcpS and apo-AcpM. To understand the interaction between these two proteins, the *Mtb* AcpS and *Mtb* ACP NMR structures (PDB entry 1klp) (Wong et al., 2002) were superimposed on the *B. subtilis* AcpS–ACP complex. The superimposed model of *Mtb* AcpS–ACP shows that the conserved DSL motif of AcpM occupies a dimer interface of AcpS and is in close enough proximity to interact with either 3',5'-ADP or CoA.

Table 3-1. List of ACPS structures from several pathogenic bacteria and their sequence similarity to *Mtb* AcpS.

Bacterium	PDB entry	Sequence similarity to <i>Mtb</i> AcpS (%)	R.m.s.d. (Å)
<i>Streptomyces coelicolor</i>	2wdo	39	1.58
<i>Mycobacterium smegmatis</i>	3gwm	84	1.07
<i>Bacillus anthracis</i>	3hyk	32	2.07
<i>Staphylococcus aureus</i>	3f09	26	2.12
<i>Streptococcus pneumoniae</i>	1fte	27	2.18
<i>Bacillus subtilis</i>	1f80	29	1.98

MATERIAL AND METHODS

Cloning, expression and purification of *C. ammoniagenes* AcpS and *Mtb* AcpS

A 0.45 kb DNA fragment containing the AcpS gene was amplified by PCR using *C. ammoniagenes* genomic DNA as the template (forward primer, 5'-CATATGCTCGACAACCGTGAAGCG; reverse primer, 5'-AAGCTTTTACCGCAG-GTACCGCAG). The amplified DNA fragment was digested with *Nde*I and *Hind*III and subcloned into the corresponding restriction sites in the pET28a vector to yield N-terminally His₆-tagged AcpS. *E. coli* B834 (DE3) Met cells were transformed with the AcpS-pET28a/His vector. For the production of SeMet-incorporated protein the cells were grown in M9 minimal medium supplemented with all 19 standard amino acids and selenomethionine. Expression of AcpS was induced with 1 mM IPTG and cells were harvested after 6 h of growth at 289 K.

A 0.4 kb DNA fragment containing the AcpS gene (Rv2523c; Swiss-Prot accession No. O53228) was amplified by PCR using *Mtb* H37Rv genomic DNA as the template (forward primer, 5'-ATGGGCATCGTCGGTGTGGG; reverse primer, 5'-TCACGGGGCCTCCAGGATGGCGAC). The amplified DNA fragment was digested with *Nco*I and *Bam*HI and subcloned into the corresponding restriction sites in the pET28a vector to yield C-terminally His₆-tagged AcpS. The AcpS-pET28a vector was transformed into *E. coli* BL21 (DE3) cells by heat-shock transformation. The transformed cells were grown to mid-log exponential phase at 310 K in Terrific Broth

medium containing kanamycin. Expression of AcpS was induced with 1 mM IPTG and cells were harvested after 6 h of growth at 289 K.

The harvested cell pellet was resuspended in buffer *A* (20 mM Tris–HCl pH 8.0, 100 mM imidazole, 0.5 M NaCl) with 1 mM PMSF and Complete EDTA-free protease-inhibitor cocktail (Roche). The cell lysate was disrupted by three passages through a cooled French pressure cell and repeatedly sonicated at 277 K with 30 s pulses. The resulting cell extract was centrifuged at 15 000g at 277 K for 1 h. The cleared supernatant was loaded onto a Hi-Trap Ni²⁺-chelating column (Pharmacia Biosciences) and washed with 300 ml buffer *A*. His-tagged AcpS was eluted with a 100 ml linear gradient of 100–500 mM imidazole in 20 mM Tris–HCl pH 8.0 and 0.5 M NaCl. After purification to near-homogeneity by size-exclusion chromatography on a Superdex G-75 gel-filtration column (Pharmacia Biosciences), AcpS was dialyzed against 20 mM Tris–HCl pH 8.0 buffer, concentrated to 8 mg ml⁻¹ and stored in 20 mM Tris–HCl pH 8.0 at 193 K.

Cloning, expression and purification of *Mtb* ACP (AcpM)

A 0.35 kb DNA fragment containing the ACP gene (Rv2244; Uni-Prot accession No. POA4W6) was amplified by PCR from *Mtb* H37Rv genomic DNA (forward primer, 5'-AAGAAGAGGCGGCCGCATGCCTGTCACTCAGGAGCCTGTCACTCAGG-AAGA; reverse primer, 5'-AAGAAGAGCCATGCCTTGGACTCGGCCTCAAGC-CTC).

The amplified DNA fragment was digested with *NotI* and *NcoI* and subcloned into the corresponding restriction sites in the pET30b vector to yield C-terminally His₆-tagged AcpM. The AcpM-pET30b vector was transformed into *E. coli* BL21 (DE3) cells. The transformed cells were grown to mid-log phase at 310 K in Terrific Broth containing kanamycin. Expression of AcpM was induced with 1 mM IPTG and cells were harvested after 6 h of growth at 289 K.

The harvested cell pellet was resuspended in buffer *A* (20 mM Tris-HCl pH 8.0, 10 mM imidazole, 0.5 M NaCl) with 1 mM PMSF and Complete EDTA-free protease-inhibitor cocktail (Roche). The cell lysate was disrupted by three passages through a cooled French pressure cell and repeatedly sonicated at 277 K with 30 s pulses. The resulting cell extract was centrifuged at 15 000g at 277 K for 1 h. The cleared supernatant was loaded onto a Hi-Trap Ni²⁺-chelating column (Pharmacia Biosciences) and washed with 300 ml buffer *A*. His-tagged AcpM was eluted with a 200 ml linear gradient of 10–500 mM imidazole in 20 mM Tris-HCl pH 8.0 and 0.5 M NaCl. The peak fractions were pooled and dialyzed against 20 mM Tris buffer pH 8.0. The concentrated protein was stored at 193 K until use.

AcpS post-translational modification fluorescence assay

The 4'-ppt transferase activity of AcpS was assessed using BODIPY-labeled CoA. For this purpose, BODIPY-S-CoA was synthesized using CoA (Sigma) and BODIPY FL N-(2-aminoethyl)maleimide (Invitrogen) (La Clair et al., 2004). In the assay, 50 μM AcpM was incubated with either 0.2 or 2 μM AcpS for 1 h at room

temperature in the dark in 20 mM HEPES pH 7.5 buffer containing 10 μ l BODIPY-S-CoA (60 μ M final concentration) and 10 mM MgCl₂. As a control, AcpM was also pre-incubated with AcpS and 300 μ M CoA for 1 h at room temperature in the dark before the addition of BODIPY-S-CoA to the reaction mixture. PKS11 was used as a negative control as it is not a substrate of AcpS *in vivo*. PKS11 was incubated with BODIPY-S-CoA in the presence of AcpS, as performed for AcpM. After incubation, the reactions were stopped by the addition of 20% TCA to the mixture. The samples were incubated on ice for 15 min and then spun at 13 000g for 10 min. The supernatant was decanted and the pellets were washed twice with 200 μ l ice-cold acetone. The pellets were dried in a speed vacuum, dissolved in a small amount of buffer and analyzed by SDS-PAGE. The resulting bands were analyzed under a UV-light scanner. The 4'-ppt transferase activity of AcpS was also measured at different pH values (3.4–8.0) using BODIPY-labeled CoA.

Crystallization and data collection

Initial crystallization conditions were screened by the hanging-drop method using a sparse-matrix kit (Crystal Screen and Crystal Screen 2, Hampton Research, California, USA; McPherson, 1982). *C. ammoniagenes* AcpS alone produced diffraction-quality crystals (space group $P2_12_12_1$) in 4–8 d at 291 K from 4 μ l hanging drops consisting of a 1:1 ratio of AcpS (8 mg ml⁻¹ concentration) and crystallization buffer consisting of 100 mM MES buffer pH 6.5, 0.2 M ammonium sulfate and 30%(w/v) PEG MME 5K. AcpS-CoA crystals were also obtained by the hanging-drop method. Protein solution (10 mg

ml⁻¹) was mixed with an equal volume of precipitant solution to create 4 µl drops and incubated at 291 K. For protein crystals complexed with ligands, 0.5–2 mM CoA and 10 mM magnesium chloride were pre-incubated with AcpS for 10 min at room temperature. The precipitant solution for AcpS–CoA crystals consisted of 25%(w/v) PEG MME 5K, 0.1 M MES buffer pH 6.5 and 200 mM ammonium sulfate. The *C. ammoniagenes* AcpS–CoA complex crystallized in space group *P1* with two trimers in the asymmetric unit.

The crystallization screen for *Mtb* AcpS was set up in hanging drops using the factorial screening method. Initial crystallization screening was carried out with AcpS alone in 10 mM MgCl₂ and 14 mM KCl (24 h at 277 K). Diffraction-quality crystals were obtained at 291 K when drops (1:1 ratio of protein:reservoir solution) were equilibrated against 500 µl well solution consisting of 5.6%(w/v) PEG 4K, 0.07 M sodium acetate pH 4.6–5.5, 50 mM ammonium sulfate and 30%(v/v) glycerol. The crystal belonged to space group *I222*, with unit-cell parameters a = 83.12, b = 104.87, c = 105.19 Å, and contained one homotrimer per asymmetric unit. Diffraction-quality crystals were obtained after 3–4 d. Crystals belonging to space group *R3* were obtained by the same method using a well solution consisting of 5.6% PEG 4K, 0.07 M sodium acetate pH 5.0–5.6, 50 mM ammonium acetate, 15% glycerol and 100 mM sodium iodide. The *R3* crystals contained one molecule in the asymmetric unit.

A multi-wavelength anomalous dispersion data set for the *C. ammoniagenes* SeMet apo-AcpS crystal was collected on beamline 14-BM-D at the Advanced Photon Source (APS), Argonne National Laboratory using an Area Detector Systems

Corporation Q4 area detector. High-resolution data sets for *C. ammoniagenes* AcpS–CoA and native AcpS crystals were collected on beamline 14-BM-C at APS. PEG 400 was used as a cryoprotectant for both apo-AcpS and the AcpS–CoA complex. The crystals were flash-frozen in a liquid-nitrogen stream at 120 K. *Mtb* apo-AcpS crystals (space groups *I*222 or *I*₂₁₂₁ and *R*3) were captured in cryoloops and flash-cooled in an N₂ stream after brief soaks in mother liquor containing 20% ethylene glycol as a cryoprotectant. The data were reduced using *DENZO* (Otwinowski and Minor, 1997) and intensities were scaled with *SCALEPACK* (Otwinowski and Minor, 1997). Integrated and scaled data were analyzed by *XPREP* and indicated that the *Mtb* apo-AcpS crystals belonged to space group *I*222 or *I*₂₁₂₁. Solvent-content calculations (Matthews, 1968) indicated the presence of a homotrimer ($V_M = 2.55 \text{ \AA}^3 \text{ Da}^{-1}$, $V_S = 51.76\%$) in the asymmetric unit. Molecular replacement (MR) in space group *I*222 gave a higher correlation coefficient (74.4%) and lower *R* factor (39.3%) compared with space group *I*₂₁₂₁ (correlation coefficient of 34.2% and *R* factor of 58.9%). Therefore, *I*222 was the correct space group. The crystallographic data-collection statistics and refinement parameters are summarized in Tables 3-2 and 3-3.

Table 3-2. Anomalous data-collection (a) and phasing statistics (b) for *C. ammoniagenes* apo-AcpS. Values in parentheses are for the highest resolution bin.

(a)

	Inflection (maximum f')	Peak (maximum f'')	High remote
Wavelength (Å)	0.979	0.979	0.963
Resolution range (Å)	99.0-2.48	99.0-2.48	99.0-2.48
Completeness (%)	100 (100)	100 (100)	100 (100)
Completeness, 2 σ (%)	86 (67)	83 (56)	94 (85)
$\langle I/\sigma(I) \rangle$	13.7 (2.8)	15.9 (2.6)	23.1 (7.2)
R_{merge}	0.10 (0.31)	0.10 (0.32)	0.09 (0.35)
f' (e $^-$)	-11.03	-8.49	-4.93
f'' (e $^-$)	3.20	5.28	3.44

(b)

Resolution bin (lower limit) (Å)	8.53	5.42	4.25	3.61	3.19	2.89	2.66	2.48
FOM initial (<i>SOLVE</i>)	0.61	0.40	0.36	0.31	0.23	0.17	0.12	0.10
FOM final (<i>SHARP</i> , DM NCS averaging)	0.79	0.49	0.42	0.29	0.17	0.12	0.10	0.70

Table 3-3. Data-collection, refinement and geometry statistics. Values in parentheses are for the highest resolution bin.

	<i>Mtb</i> -Apo-AcpS		<i>C. ammoniagenes</i>	
			Apo-AcpS	AcpS-CoA
Space group	<i>I</i> 222	<i>R</i> 3	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 1
Unit cell parameters				
a (Å)	83.12	67.89	51.40	54.78
b (Å)	104.87	67.89	59.00	56.02
c (Å)	105.19	83.94	153.56	88.28
α	90.00	90.00	90.00	80.88
β	90.00	90.00	90.00	76.23
γ	90.00	120.00	90.00	60.92
No. of crystals	1	1	2	1
Resolution (Å)	30-2.51 (2.60-2.51)	30-1.90 (1.95-1.90)	30-2.48 (2.48-2.50)	30-1.89 (1.95-1.89)
Unique reflections	15471 (1254)	6051 (855)	16881 (2038)	52264 (5869)
Completeness (%)	95.7 (75.5)	100 (100)	94.7 (78.6)	73 (51.0)
R_{merge}	0.03 (0.14)	0.03 (0.19)	0.05 (0.15)	0.05 (0.18)
$\langle I/\sigma(I) \rangle$	13.83 (1.22)	54.00 (4.00)	17.35 (7.01)	58.10 (7.30)
Free R value (random 5%)	0.30 (0.35)	0.27 (0.38)	0.27 (0.28)	0.25 (0.29)
R value	0.23 (0.36)	0.24 (0.39)	0.20 (0.21)	0.21 (0.21)
No. of protein residues	387	129	399	804
B Factor (Å ²)	74.14	43.75	30.14	24.91
Wave length (Å)	1.54	0.97	0.97	0.97
No. of ADPs	0	0	0	6
No. of sulfates	4	0	4	0

Table 3-3. Continued.

	Mtb-Apo-AcpS		<i>C. ammoniagenes</i>	
			Apo-AcpS	AcpS-CoA
R.m.s.d. bond lengths (Å)	0.03	0.03	0.02	0.031
R.m.s.d. bond angles (°)	1.48	1.66	1.64	2.036
Overall coordinate error (Å)	0.30	0.28	0.23	0.26
Real- space CC (<i>REFMAC5</i>)	0.94	0.94	0.92	0.93
Ramachandran plot, residues in				
Most favored region	332 (87.30)	338 (88.70)	387 (96.99)	774 (96.27)
Allowed region	45 (12.0)	41 (10.7)	7 (1.75)	17 (2.11)
Generously allowed region	3 (0.7)	2 (0.3)	0 (0)	0
Disallowed region	0	1 (0.3)	5 (1.25)	13 (1.62)

Structure determination

The experimental phases for the *C. ammoniagenes* apo-AcpS structure were obtained by multi-wavelength anomalous dispersion phasing (Hendrickson and Ogata, 1997). *SHELXD* (Sheldrick, 2008a) located three selenium sites in the asymmetric unit consistent with a homotrimer in the asymmetric unit and *SOLVE* (Terwilliger and Berendzen, 1999) was used to refine the sites and calculate the initial protein phase, resulting in an overall figure of merit of 0.7 (Table 3-2) for data in the resolution range 100–2.48 Å. The heavy-atom parameters were then refined and the phase was calculated in *SHARP* (de La Fortelle and Bricogne, 1997). Further phase improvement with solvent flattening in *autoSHARP* (Bricogne et al., 2002) resulted in high-quality density-modified maps showing clear electron density for the correct hand for three molecules of protein in the asymmetric unit. The initial protein model was built using *XFIT* (McRee, 1999) and a final high-quality model was produced after several cycles of manual model building and NCS-restrained maximum-likelihood refinement with *REFMAC5* (Murshudov et al., 2011) against the high-remote data set (Table 3-3). Water molecules were manually added during iterative cycles of building and refinement.

The structure of the *C. ammoniagenes* AcpS–CoA binary complex was solved by molecular replacement with *EPMR* (Kissinger et al., 1999) using the final model of *C. ammoniagenes* apo-AcpS as the search model for data extending from 25 to 3.5 Å resolution. After model building and fitting, bias-minimized electron-density maps were obtained using the *Shake&wARP* protocol (Kantardjieff et al., 2002). Clear electron density for CoA molecules was visible in the *Shake&wARP* map prior to any model

building. After repeated cycles of refinement and manual model building, water molecules were added to the structure using *XFIT* (McRee, 1999) (Table 3-3).

The structure of *Mtb* apo-AcpS was solved by molecular replacement with *EPMR*, giving a correlation coefficient of 0.45, using the final *C. ammoniagenes* apo-AcpS model as the search model for data extending from 25 to 3.5 Å resolution. The initial molecular-replacement model was mutated according to the *Mtb* AcpS sequence and manually rebuilt using *XFIT* into bias-minimized multiple-averaged electron-density maps obtained from the *Shake&wARP* map server. The initial electron-density map allowed the construction of a model with 129 residues (residues 2–130) in each subunit. Rigid-body refinement was carried out using *REFMAC5*. After repeated cycles of refinement and manual building, water molecules were added to the structure using *XFIT* (McRee, 1999). The final model has a crystallographic *R* factor of 23% and an R_{free} of 30% to a resolution of 2.51 Å with good stereochemistry (Table 3-3). *Mtb* apo-AcpS also crystallized in space group *R3* with one molecule in the asymmetric unit. The *R3* crystal structure was solved by molecular replacement using a single chain from the final model of *Mtb* apo-AcpS solved in space group *I222*. Several cycles of manual model building and likelihood refinement in *REFMAC5* yielded a final 1.9 Å resolution model for the *R3* crystal (Table 3-3). Water molecules were manually added during iterative cycles of building and refinement.

RESULTS AND DISCUSSION

The initial phases for the *C. ammoniagenes* apo-AcpS crystal diffraction data were obtained by MAD (Hendrickson, 1991) using crystals of selenomethionyl protein in space group $P2_12_12_1$. The unit-cell parameters and scaling with diffraction amplitudes showed that the AcpS–CoA binary complex crystallized in space group $P1$ with two trimers in the asymmetric unit. The initial map allowed the construction of a model with 136 residues in each subunit. The crystal structure of AcpS from *C. ammoniagenes* revealed that it assembles as a tightly packed homotrimer. The final structure consists of residues 5–47 and 61–153 (Fig. 3-1a) in three subunits designated *A*, *B* and *C*. The three polypeptide chains can be superimposed with r.m.s.d.s of 0.43 Å between chains *A* and *B* and of 0.53 Å between chains *A* and *C*. In an experimental map and in an unbiased electron-density map of native *C. ammoniagenes* AcpS the structure shows no electron density for the four N-terminal residues (1–4) or for residues 48–60 (the $\alpha 2$ – $\alpha 3$ connecting loop). The *C. ammoniagenes* apo-AcpS structure has an α/β -fold (Fig. 3-1a) similar to those of the *B. subtilis* and *S. pneumoniae* AcpS structures (Chirgadze et al., 2000; Parris et al., 2000). The main core of the protein is the long $\alpha 3$ helix. One side of the $\alpha 3$ helix is covered by antiparallel β -sheets (consisting of $\beta 1$, $\beta 4$ and $\beta 5$), while the other side of the helix is covered by a β -sheet composed of strands $\beta 2$, $\beta 3$ and $\beta 4$ together with $\alpha 1$ and $\alpha 2$.

The *C. ammoniagenes* AcpS–CoA binary complex structure was solved by molecular replacement using the coordinates of the *C. ammoniagenes* apo-AcpS

structure as a starting model. The crystallographic data-collection statistics and refinement parameters are summarized in Table 3-3. To obtain the molecular-replacement solution for *Mtb* AcpS, we used the coordinates of *C. ammoniagenes* AcpS as an initial search model. The structural data reveal that native *Mtb* AcpS (Fig. 3-1b) assembles as a tightly packed homotrimer and each subunit consists of residues 2–130.

AcpS forms a homotrimer in which the polypeptide chain is folded into four α -helices and five β -strands and trimerization is mediated through the residues of the antiparallel β -strands (β 1, β 4 and β 5) of all three subunits. The three antiparallel β -strands of each subunit are arranged together and form a tunnel-like structure at the center of the trimer that runs through the entire length of the protein (Figs. 3-1a and 3-1b). The tunnel contains mainly hydrophobic residues that stabilize the biological unit, with intersubunit interactions occurring predominantly between the β 1 strand of one subunit and the β 4 strand of another subunit. The active site of AcpS is formed at the intermolecular interface between two subunits, specifically where the β -sheets of two AcpS molecules intersect in the trimer. The amino acids from one subunit surround one half of the active site and those from the second subunit surround the other half. Examination of the active site of the *C. ammoniagenes* native AcpS structure reveals a bowl-shaped depression (Figs. 3-1a and 3-1b) that is surrounded by charged residues and bound ordered water molecules. In the *C. ammoniagenes* and *Mtb* AcpS native structures the active sites are occupied by sulfate ions (Figs. 3-1c and 3-1d). In addition, hydrophobic residues create a hydrophobic tunnel for the 4'-ppt moiety.

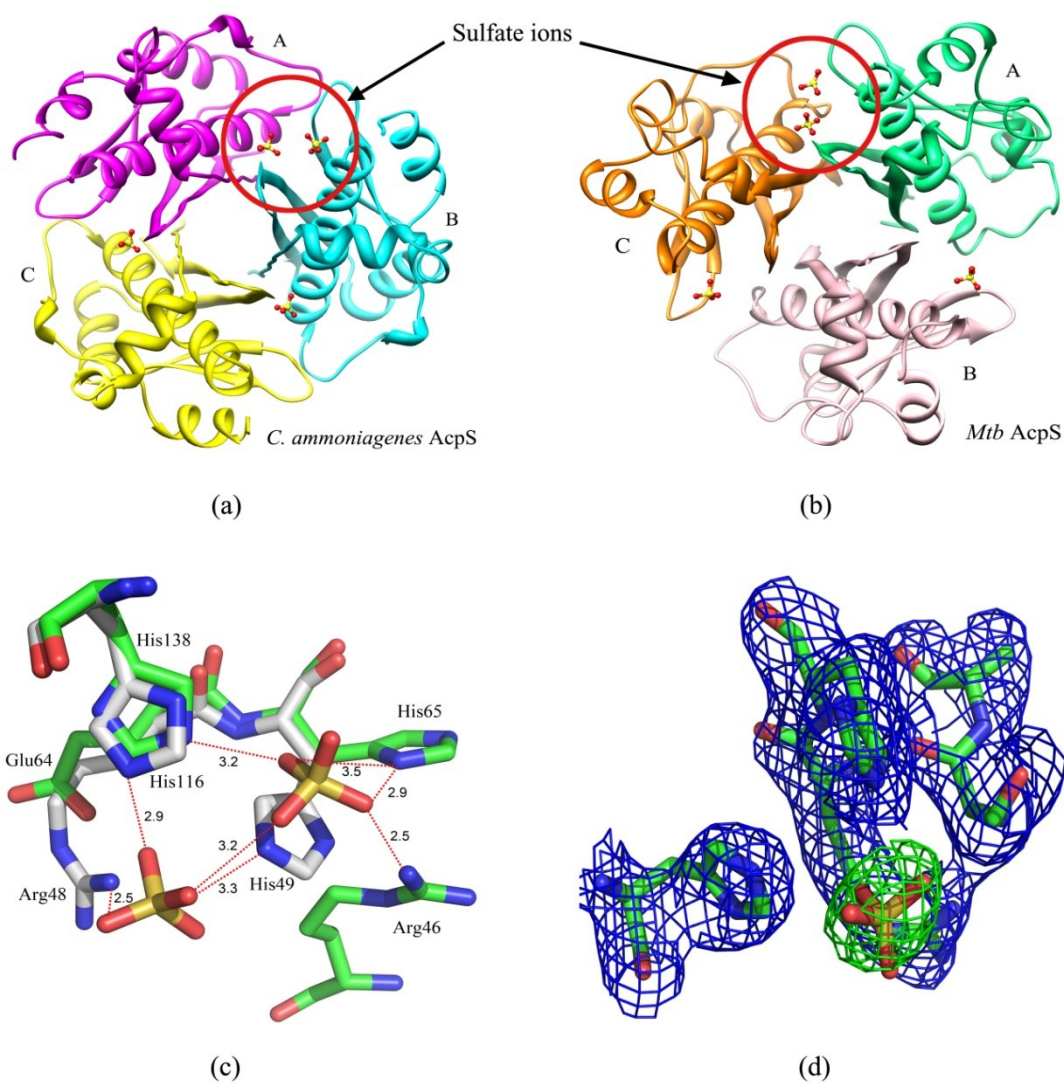


Figure 3-1. Overview of the *C. ammoniagenes* and *Mtb* AcpS structures with ribbon representation of the homotrimer. In the final model of the *C. ammoniagenes* AcpS structure each monomer contains residues 5–47 and 61–153. In the *Mtb* AcpS structure each monomer contains residues 2–130. The view is from the top of the trimer. (a) Ribbon diagram of the *C. ammoniagenes* AcpS homotrimeric structure: subunit A (magenta), subunit B (cyan) and subunit C (yellow). (b) Ribbon diagram of the *Mtb* AcpS homotrimer structure: subunit A (sea green), subunit B (plum) and subunit C (orange). The sulfate ion (shown in ball-and-stick representation) binds at the dimer interface, specifically in the cleft region that is formed by the two monomers. (c) Structural superposition of the *Mtb* and *C. ammoniagenes* AcpS active-site residues with sulfate ions. The bound sulfate ions are shown in stick representation. In *Mtb* AcpS His49 swings 101° compared with His65 of *C. ammoniagenes* AcpS. (d) Electron-density map of the substrate-binding pocket of *Mtb* AcpS. The carve feature in PyMOL was used to limit display of the electron density of the model. A $2F_o - F_c$ map contoured at 1.4σ (blue) and an $F_o - F_c$ map contoured at 4σ (green) are shown.

Sequence and structural comparison of *Mtb* AcpS and *C. ammoniagenes* AcpS with other bacterial AcpSs

Multiple sequence alignments of known AcpS sequences are presented in Fig. 3-2 and show that *Mtb* AcpS has a lower sequence identity to *B. subtilis* AcpS and *S. pneumoniae* AcpS (29 and 27%, respectively) than to *C. ammoniagenes* AcpS, which displays 34% sequence identity to the homologous enzyme. The KEAXXKA motif, containing the Lys62 residue, is within hydrogen-bonding distance of the 5'-phosphate and is conserved in the prokaryotic AcpS enzymes; the SXXXH motif interacts with either the 3'-phosphate of CoA or with ADP. Other conserved motifs include GXD (involved in chelating the Mg²⁺ ion), and RW and GRP, which are part of the CoA-binding regions. The RRXA and SXXXH sequence motifs are not conserved in the prokaryotic AcpS and surfactin synthetase enzymes. The SXXXH motif is highly conserved in AcpS enzymes from *Mycobacterium* and *Corynebacterium* species and is located in the vicinity of the CoA-binding pocket. In the *Mtb* apo-AcpS structure His49 (His65 in *C. ammoniagenes* AcpS) interacts with the sulfate ion. The present structural study also shows the participation of a second histidine in the active site. In *B. subtilis* AcpS the hydrophobic Phe49 residue occupies the place of His49 and does not contribute to active-site formation. The sequence alignment also shows that *C. ammoniagenes* AcpS has an insertion of 13 amino acids in the $\alpha 2$ – $\alpha 3$ connecting loop.

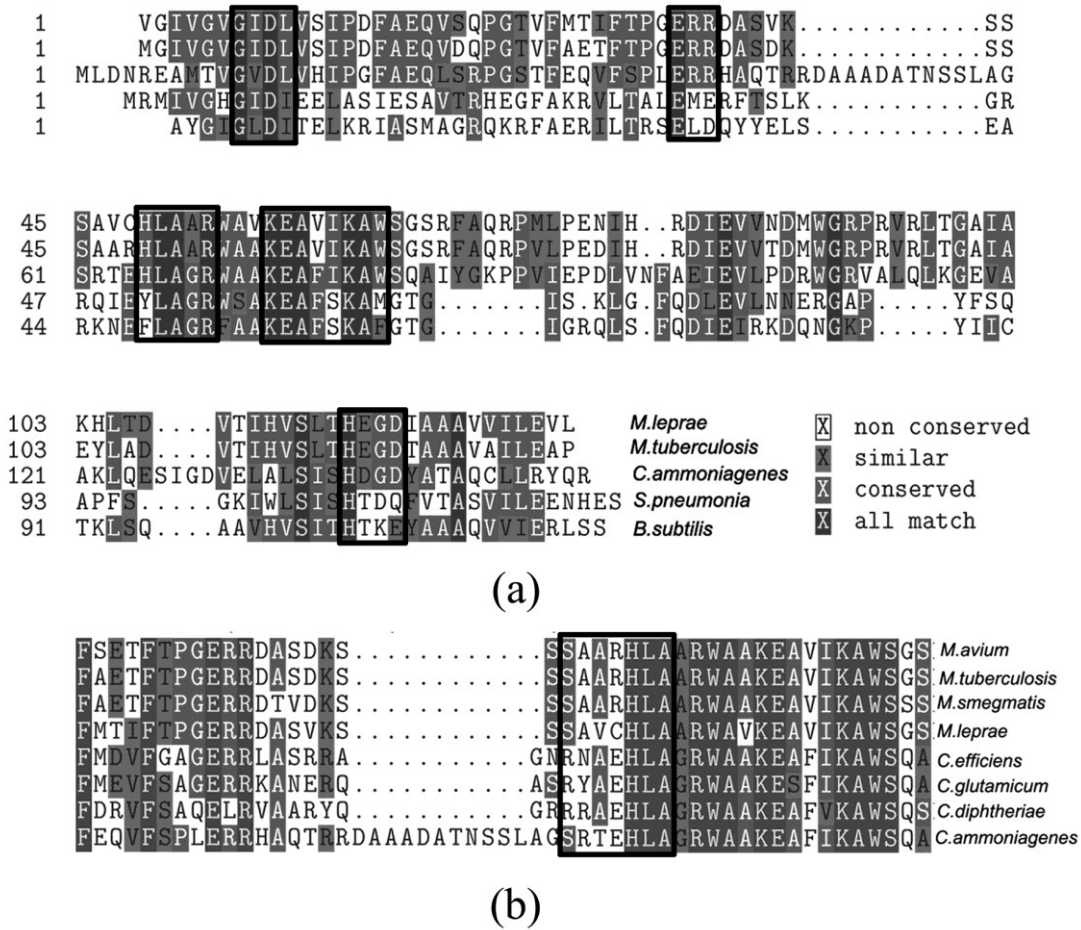


Figure 3-2. Sequence alignment of AcpS. (a) Multiple sequence alignment of AcpS enzymes. The sequence alignment shows that *C. ammoniagenes* AcpS has an additional 12-residue (RDAADATNSSL) insertion sequence between the $\alpha 2$ and $\alpha 3$ helices which is disordered in the structure. (b) Sequence alignment of AcpS enzymes from *Mycobacterium* and *Corynebacterium* species. The conserved SXXXHLA motif is shown in a box.

Least squares superposition of the structures indicate the close resemblance of *Mtb* AcpS and *C. ammoniagenes* AcpS to the *B. subtilis* and *S. pneumoniae* enzymes (r.m.s.d for 112 C^α carbon atoms ~1.67 and 1.8 Å respectively; (Chirgadze et al., 2000; Parris et al., 2000); Figs. 3-3a-3-3c). In *B. subtilis* AcpS and *S. pneumoniae* AcpS three short α -helices (α_1 , α_2 , and α_3) are located between the β_1 strand and the long α_4 helix (Fig. 3-3b). In contrast, in *C. ammoniagenes* AcpS and *Mtb* AcpS the two short α -helices (α_1 and α_2) are located between the β_1 strand and the long α_3 helix (Fig. 3-3b). Structural alignment shows that the β_1 strand and the α_3 helix are connected by 30 residues in known AcpS structures. In the *Mtb* and *C. ammoniagenes* AcpS structures these 30 residues form two short α -helices (α_1 and α_2). In contrast, in the *S. pneumoniae* and *B. subtilis* AcpS enzymes the equivalent 30 residues (12-43) form three well ordered short α -helices (α_1 , α_2 , and α_3). Because of the absence of one short α -helix (equivalent to α_2 in *B. subtilis*), the connecting loop between the α_1 and α_2 helices is much larger and more flexible in the *Mtb* and *C. ammoniagenes* AcpS proteins. Furthermore, in *C. ammoniagenes* AcpS and *Mtb* AcpS the β_3 and β_4 strands are connected by a well defined α_4 helix (residues 115-129 and 98-105, respectively; Fig. 3-3c), while in *B. subtilis* and *S. pneumoniae* the β_3 and β_4 strands are not connected by an α -helix but by a short loop (91-96). The topological rearrangements of the α -helix in *Mtb* AcpS and *C. ammoniagenes* AcpS may be required for their interactions with ACP, as *Mtb* AcpS and *C. ammoniagenes* AcpS are unusual members of the ACP family owing to their extended C-termini.

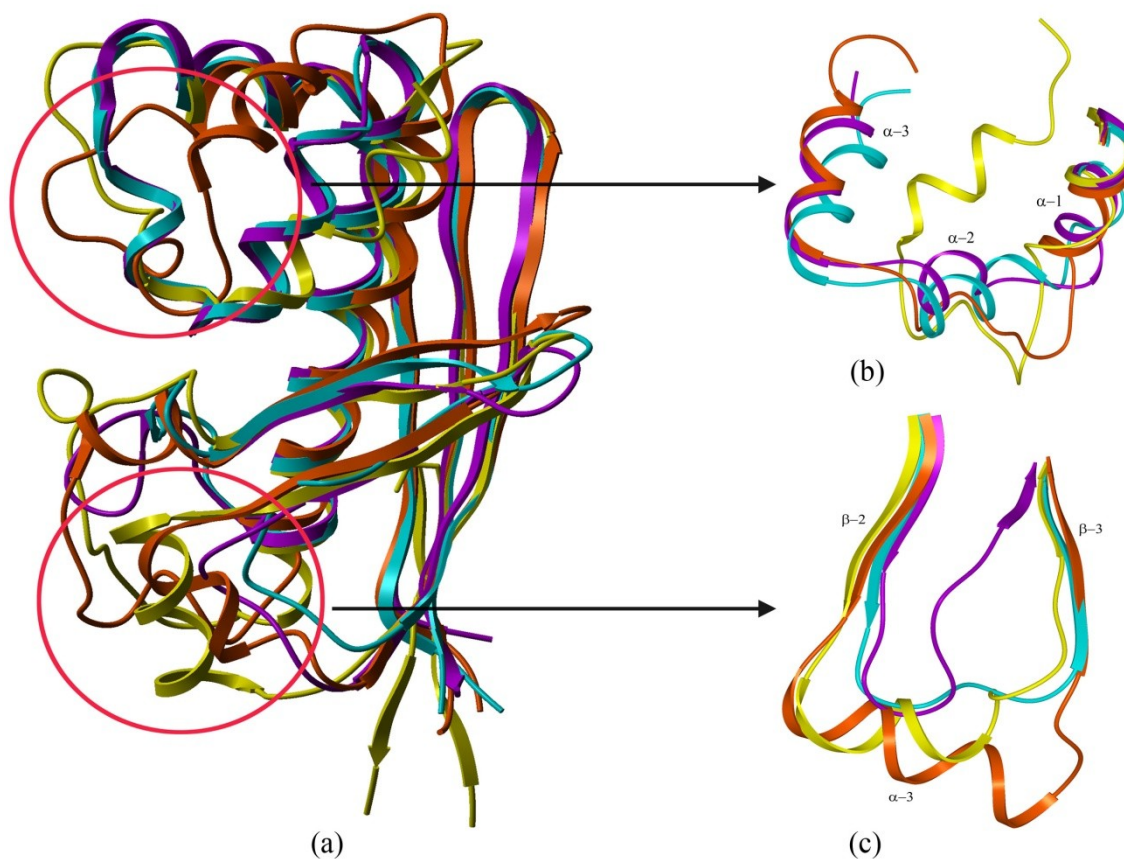


Figure 3-3. Structural superposition of the AcpS enzymes. (a) Backbone superposition of the AcpS enzymes. Color key: orange, *Mtb*; yellow, *C. ammoniagenes*; cyan, *B. subtilis*; purple, *S. pneumoniae*. (b) The structural superposition reveals that the *B. subtilis* and *S. pneumoniae* AcpS structures have three short helices ($\alpha 1$, $\alpha 2$ and $\alpha 3$) between the $\beta 1$ strand and the $\alpha 3$ 'core' helix. The *Mtb* and *C. ammoniagenes* AcpS structures contain two short helices between $\beta 1$ and the core helix. (c) The $\beta 2$ and $\beta 3$ strands of the *Mtb* and *C. ammoniagenes* AcpS structures are connected by a short helix which is absent in *B. subtilis* and *S. pneumoniae* AcpS.

Structural differences between the open and closed structural forms of *Mtb* AcpS

A crystal structure of *Mtb* AcpS has recently been reported in which AcpS was crystallized in space group *P23* and contained one molecule in the asymmetric unit (Dym et al., 2009). The reported AcpS–ADP structure is disordered in three regions (residues 22–30, 41–44 and 75–78) accounting for 13% of the AcpS model. *Mtb* apo-AcpS crystallized in space groups *I222* and *R3*. Most notable was that in the *I222* and *R3* space groups AcpS had clear density for the entire model. AcpS crystallized as a homotrimer in space group *I222*, but in space group *R3* it crystallized with a monomer in the asymmetric unit. The trimeric structure was maintained by the symmetry. The monomeric form was superimposed individually on chains *A*, *B* and *C* of the trimeric apo-AcpS structure with r.m.s.d.s of 0.91, 1.02 and 0.85 Å, respectively. Structural superposition of C α atoms revealed that AcpS–ADP crystallized in space group *P23* retained an overall fold analogous to those of previously reported AcpS structures (Chirgadze et al., 2000; Parris et al., 2000). However, superposition of the C α atoms reveals significant conformational changes of the C α atoms in two regions, the α 2 helix (residues 30–40; r.m.s.d. of 12 Å) and the connecting loop located between the α 3 and α 4 helices (residues 67–73; r.m.s.d. of 9 Å), when compared with the *Mtb* AcpS structure crystallized in space groups other than *P23* (Figs. 3-4a and 3-4b). In contrast, the closed conformation was not observed in the *C. ammoniagenes* AcpS structure. The crystallographic data for *C. ammoniagenes* AcpS suggest that both the apo-AcpS and the AcpS–CoA complex structures adopt an open conformation. It is possible that in some cases crystal packing can contribute to the structural and conformational changes in the

proteins. However, our current structural data show that *Mtb* AcpS crystallizes in three different space groups at lower pH and no structural changes were observed between them. In contrast, the structure of a crystals obtained at higher pH (unreported low-resolution data for apo-AcpS crystallized in space group *P23*, unit-cell parameters $a = b = c = 73.74 \text{ \AA}$, $\alpha = \beta = \gamma = 90.0^\circ$) superposes very well with the recently reported AcpS-ADP complex (Dym et al., 2009). The crystals obtained at higher pH show three disordered regions and also structural movement in two regions. Taken together, these results suggest that the pH of the crystallization condition plays a significant role in the conformational changes of *Mtb* AcpS.

The effect of pH on the structure of the $\alpha 2$ helix

At lower pH values (between 4.4 and 6.0), the $\alpha 1$ - $\alpha 2$ connecting loop is well ordered in the *Mtb* apo-AcpS structure. Residues from this loop make hydrogen-bonding interactions with residues in the $\alpha 2$ helix that result in close packing of the $\alpha 1$ and $\alpha 2$ helices on one side of the $\alpha 3$ or 'core' helix. Likewise, residues from helix $\alpha 2$ interact with residues from helix $\alpha 3$. In particular, the distance between the hydroxyl group of Thr25 and the main-chain N atom of Phe27 is 3.06 \AA and the guanidinium group (NH1) of Arg53 from the $\alpha 3$ helix interacts with the acidic side chain of Glu35 (OE2) from the $\alpha 2$ helix (3.19 \AA) at low pH. At higher pH values the AcpS structure undergoes a significant conformational change in which the $\alpha 2$ helix moves 12 \AA away from the $\alpha 1$ helix and occupies a position close to the antiparallel $\beta 2$ and $\beta 3$ strands (Fig. 3-4a).

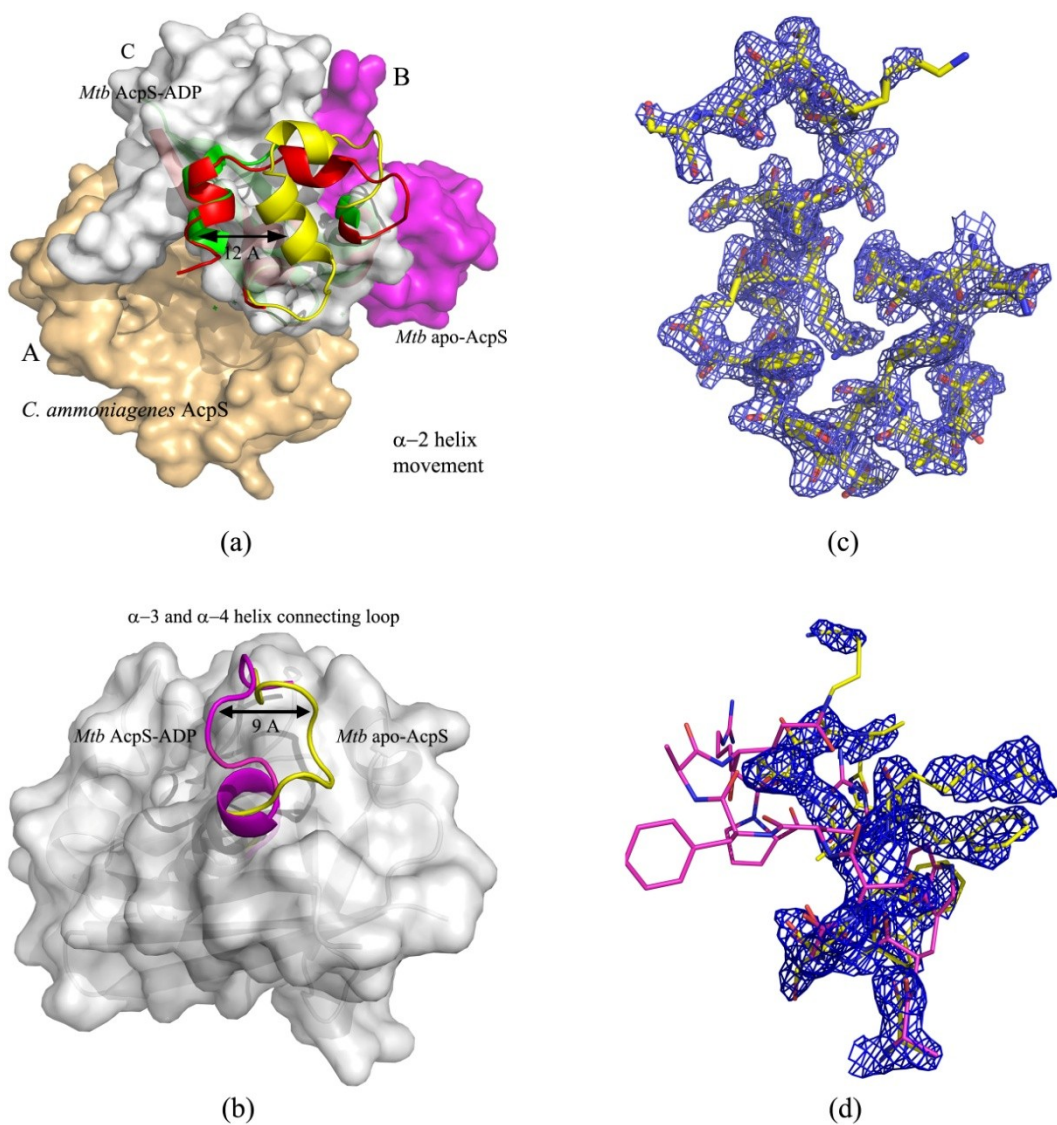


Figure 3-4. Overview of the $\alpha 2$ -helix movement and the open and closed conformations of the $\alpha 3$ - $\alpha 4$ connecting loop. (a) Structural superposition of helices $\alpha 1$, $\alpha 2$ and the connecting loop of the *Mtb* AcpS-ADP complex (green), *Mtb* apo-AcpS (yellow) and the *C. ammoniagenes* AcpS-CoA complex (red) represented by ribbon diagrams. *Mtb* apo-AcpS is shown as a surface representation; the surface of subunit C is shown partially in order to view the movement of the $\alpha 2$ helix at higher pH. (b) The $\alpha 3$ - $\alpha 4$ connecting loop is shown in both the open and closed conformations. At higher pH values the $\alpha 3$ - $\alpha 4$ connecting loop of the *Mtb* AcpS-ADP complex adopts an open conformation (purple). In the apo-AcpS structure (yellow) this same loop adopts a closed conformation. (c) Electron-density map of the *Mtb* apo-AcpS structure contoured at 1.4 σ . At lower pH values the $\alpha 1$ and $\alpha 2$ helices are packed together; electron density is shown for residues 18-44. (d) Electron density for the $\alpha 4$ helix contoured at 1.4 σ .

The rearrangement of the $\alpha 2$ helix also leads to the destabilization of the connecting loops at both ends, resulting in disordered electron density for residues 21–30 and 41–44. This movement increases the flexibility of the connecting loop, thereby altering the hydrogen bonds and hydrophobic interactions with the $\alpha 2$ helix as well as the ionic interaction between the $\alpha 2$ and $\alpha 3$ helices. The ionic interaction between the Glu35 and Arg53 residues is lost; instead, the side chain of Arg53 interacts with the phosphate group of ADP. Superposition of the higher pH structure of *Mtb* apo-AcpS with that of *C. ammoniagenes* AcpS–CoA shows that the Arg69 residue and the *Mtb* Arg53 residue overlay each other. It is clear that the ionic interaction is lost as a consequence of the change in pH and not the ligand binding.

Open and closed conformations of the connecting loop between helices $\alpha 3$ and $\alpha 4$

Structure-based sequence alignment revealed that the connecting loop between helices $\alpha 3$ and $\alpha 4$ of *Mtb* AcpS contains six additional residues compared with the known AcpS structure, resulting in the loop adopting an open conformation (Fig. 3-4b) (Dym et al., 2009). In our structure, the $\alpha 3$ – $\alpha 4$ connecting loop adopts a closed conformation despite the presence of the additional six residues (Fig. 3-4b). The pH of the conditions used for crystal growth is the only difference between these two structures. At lower pH values (4.4–6.0) the $\alpha 3$ – $\alpha 4$ connecting loop adopts a closed conformation. Structural superposition reveals that the closed conformation of *Mtb* AcpS is similar to those in the structures of AcpS from *B. subtilis* (Parris et al., 2000) and *S. pneumoniae* (Chirgadze et al., 2000). These results indicate that the $\alpha 3$ – $\alpha 4$ connecting

loop adopts two different conformations based on the pH of the crystallization conditions. The r.m.s.d. value of the individual residues of the $\alpha 3$ – $\alpha 4$ connecting loop of AcpS at pH 6.5 compared with the low-pH structure is approximately 6 Å. At lower pH values, the $\alpha 3$ – $\alpha 4$ connecting loop points towards the connecting loop between the $\alpha 3$ helix and the $\beta 2$ strand of the same subunit. The Phe69 side chain is buried in the hydrophobic pocket formed by Val4, Leu105, Val108 and Ala129. At higher pH values, residues 68–72 flip 150° in the opposite direction and move towards the connecting loop between the $\beta 2$ and $\beta 3$ strands of the neighboring subunit. In the open conformation, the Phe69 residue shows the most deviation (the r.m.s.d. of the C^α atom is 9.6 Å), followed by Ala70 (the r.m.s.d. of the C^α atom is 9.2 Å), when compared with the closed conformation (Figs. 3-4*b* and 3-4*d*).

Dym and coworkers postulated that in *C. ammoniagenes* AcpS the $\alpha 3$ – $\alpha 4$ connecting loop will adopt an open conformation as in the *Mtb* AcpS–ADP complex structure owing to their sequence conservation and identity (Dym et al., 2009). Here, we report the crystal structure of *C. ammoniagenes* AcpS with and without CoA. Both apo-AcpS and AcpS–CoA crystallized as homotrimers and superimposed very well (r.m.s.d. of 0.45 Å). Superposition of the *Mtb* apo-AcpS structure with the *C. ammoniagenes* AcpS structure (apo-AcpS and AcpS–CoA) shows that the $\alpha 2$ helix (residues 33–43) moves 12 Å towards the $\beta 2$ and $\beta 3$ strands and the $\alpha 3$ – $\alpha 4$ connecting loop (residue 82–88) undergoes a conformational change resembling the open structure of the *Mtb* AcpS–ADP structure. The open and closed conformations of AcpS cannot be correlated with complex formation but rather with changes in pH.

Another unexpected observation is that a portion of the $\alpha 4$ helix (residues 73–78) is disordered in the *Mtb* AcpS–ADP complex. It has been hypothesized that this loop (74–78) may be stabilized by the interaction of AcpS with AcpM. In contrast, the pH 5.3 crystal structure of *Mtb* AcpS and the pH 6.5 crystal structure of *C. ammoniagenes* AcpS exhibit electron density for all of the backbone and side-chain residues of the $\alpha 4$ helix (Fig. 3-4*d*). This suggests that the interaction between AcpM and AcpS is not required for stabilization of the $\alpha 4$ helix of AcpS. In the lower-pH AcpS structure the connecting loops and the $\alpha 4$ -helix region have higher B factors in comparison with the remaining region, indicating that these regions are highly flexible and are beginning to undergo conformational change. This pH-induced conformational change in AcpS may provide insight into a possible mechanism of protein–protein interaction. At higher pH values the Arg48 and His49 residues of the $\alpha 3$ helix do not participate in the CoA-binding mechanism; these residues are highly conserved in AcpS enzymes from *Mycobacterium* species. In contrast, at low pH these two residues actively participate in sulfate binding in the apo structure of *Mtb* AcpS (Figs. 3-1*c* and 3-1*d*). Similarly, in *C. ammoniagenes* the equivalent residues (Glu64 and His65) participate in both sulfate and CoA binding. The His65 residue is highly conserved in *Corynebacterium* AcpS enzymes; it is located in the active site and forms a hydrogen bond to the phosphate group of CoA (Fig. 3-2*c*). AcpS crystallized in space group *P*23 does not show an interaction with the second histidine in the sulfate/CoA-binding mechanism.

Active site of the apo-AcpS structure

The active sites of *Mtb* AcpS and *C. ammoniagenes* AcpS are formed at the interface of the trimer. Electron density for a sulfate ion is visible in the unbiased electron-density maps of both the *Mtb* and *C. ammoniagenes* AcpS native structures (Fig. 3-1*d*). The locations of the identified sulfate ions are adjacent to a series of well conserved residues, specifically Arg48, His49 and His116 in *Mtb* AcpS, and Arg46, Glu64, His65 and His138 in *C. ammoniagenes* AcpS. These sulfate ions are associated with residues found in the CoA-binding pocket identified in the *B. subtilis* AcpS structure (Parris et al., 2000). The sulfate ion occupies the same position as the 3'-phosphate of CoA. Previous studies have shown that the active sites of *S. pneumoniae* AcpS are occupied by sulfate ions (Chirgadze et al., 2000). Of particular interest are the two O atoms (O1 and O2) of the sulfate that form hydrogen bonds to Arg48 NH1 (distances of 2.48 and 2.89 Å, respectively) of subunit *A*. The guanidinium group of Arg48 is at a favorable distance to form a hydrogen bond to the carbonyl groups of the highly conserved Ser43 (3.01 Å; Fig. 3-1*d*). O atoms O1 and O3 of the sulfate ion also form hydrogen bonds to NE2 of His49 and His116 (2.82 Å), respectively. Of the three active sites, two contained one sulfate ion and one was occupied by two sulfate ions. A similar result was also observed in the *C. ammoniagenes* apo-AcpS structure (Figs. 3-1*a* and 3-1*b*). In *C. ammoniagenes* AcpS the Glu64 residue is replaced by Arg48, which occupies an identical location as in the *Mtb* AcpS structure. However, the location and orientation of His65 and the sulfate ion in the active site of the *C. ammoniagenes* AcpS structure are not identical to their positions in *Mtb*. This can be explained in two ways.

Firstly, in *C. ammoniagenes* AcpS the His65 and His138 residues are surrounded by Glu64 on one side and Arg46 on the other. Owing to the negative charge of Glu69 the sulfate ion is located in a favorable position to interact with the basic residue Arg46. Secondly, owing to the long side chain of Arg48 the sulfate ions occupy a different location in *Mtb* AcpS, making the side chain of His49 swing approximately 101° compared with His65 of *C. ammoniagenes* AcpS and facilitating interaction with the sulfate ion (Fig. 3-1c). Nevertheless, the structural data clearly indicate that in the absence of CoA the active sites are occupied by sulfate ions.

The binary complex of AcpS and CoA

The CoA-bound *C. ammoniagenes* AcpS structure was solved by molecular replacement. This binary complex crystallized in space group *P1* and contained six monomers arranged as two groups of trimers. In the AcpS–CoA complex density for ADP was very clear in each binding site and allowed unambiguous positioning of ADP. Superposition of the carbon backbone of native AcpS with the ADP-bound binary complex reveals that binding of CoA does not alter the structural conformation of the biological unit. The adenine ring of ADP fits between the $\beta 2$ and $\beta 3$ loops, consisting of residues 108–112 of subunit *A* and residues 81–86 of subunit *B* of the AcpS trimer. The adenine base is immobilized by a network of interactions created by Ser81, Trp108 and Gly109 (Fig. 3-5). The ribose is present in the 3'-*endo* conformation, with axial orientation of the 2'-hydroxyl group and equatorial orientation of the 3'-phosphate group. The 3'-phosphate group of ADP is bound tightly to the protein through a number of

interactions, specifically with the imidazole rings of His65 and His138, and His46 and Arg69 make contacts with the 3'-phosphate through a water molecule. The 5'- α -phosphate is packed against the β 1 strand and the α 3 helix of the second molecule and is positioned within hydrogen-bonding distance of the main chain of His105, the N^ε group of Lys78 and the hydroxyl group of Ser137. The 5'- β -phosphate group of CoA does not interact with any of the protein residues. The 4'-ppt arm and the β -mercaptoethanol-binding pocket extend away from the adenine base and are surrounded by mostly hydrophobic residues, specifically Ile77B, Pro93B, Phe98B and Trp108A. The involvement of the imidazole rings of His65 and His138 in CoA binding probably contributes to the strict pH-dependence of AcpS activity.

***Mtb* AcpS can selectively transfer BODIPY-labeled CoA to *Mtb* apo-ACP**

The activity of *Mtb* AcpS in post-translational modification of apo-ACP was assessed. We incubated recombinant *Mtb* AcpS with BODIPY-labeled CoA (a fluorescent dye is attached to the sulfhydryl group of CoA) and apo-ACP and an aliquot of the reaction was analyzed by SDS-PAGE. When the SDS gel was viewed under a UV illuminator, the post-translationally modified holo-ACP was visualized as a fluorescent band with the expected molecular size. The fluorescent SDS gel clearly demonstrated that *Mtb* AcpS was able to transfer the 4'-phosphopantetheine arm to ACP upon incubation with BODIPY-CoA and *Mtb* apo-ACP. We used two different concentrations of AcpS, 0.2 and 2 μ M, and the fluorescence intensity of holo-ACP clearly showed that AcpS effectively transfers the 4'-phosphopantetheine arm to apo-ACP even at 0.2 μ M

(Fig. 3-6a). Similarly, AcpS was pre-incubated with unlabeled CoA and then assessed for post-translational modification activity in the presence of BODIPY-CoA and apo-ACP. The result shows that the fluorescence intensity decreases upon pre-incubation with CoA, as part of the holo-ACP contains nonfluorescent 4'-phosphopantetheine and part contains fluorescent BODIPY-4'-phosphopantetheine (Fig. 3-6a). Owing to the heterogeneous population, diminished fluorescence intensity was observed in the CoA pre-incubation experiment.

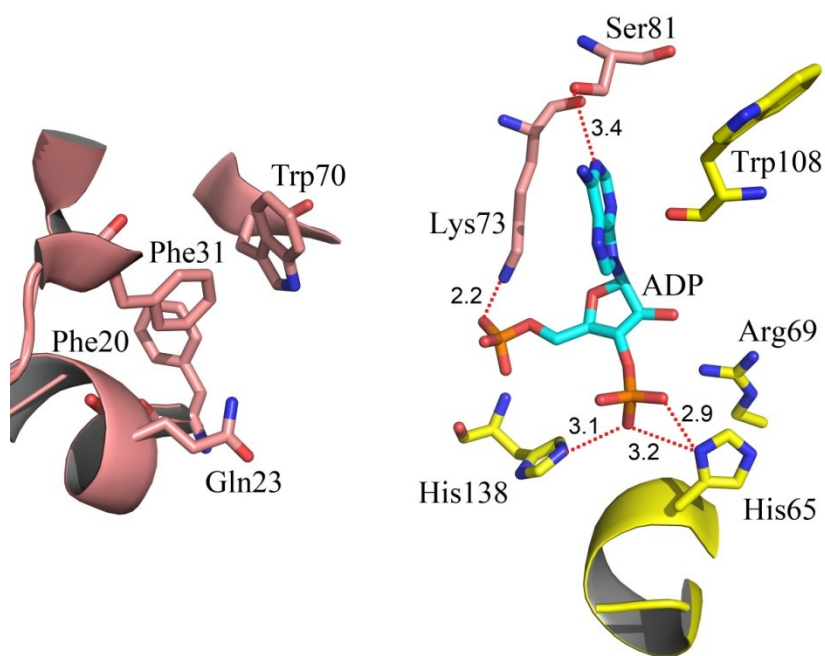


Figure 3-5. ADP binds in the dimer interface between two subunits. The active-site residues of one subunit are colored yellow, those of the second subunit are colored salmon and ADP is colored cyan.

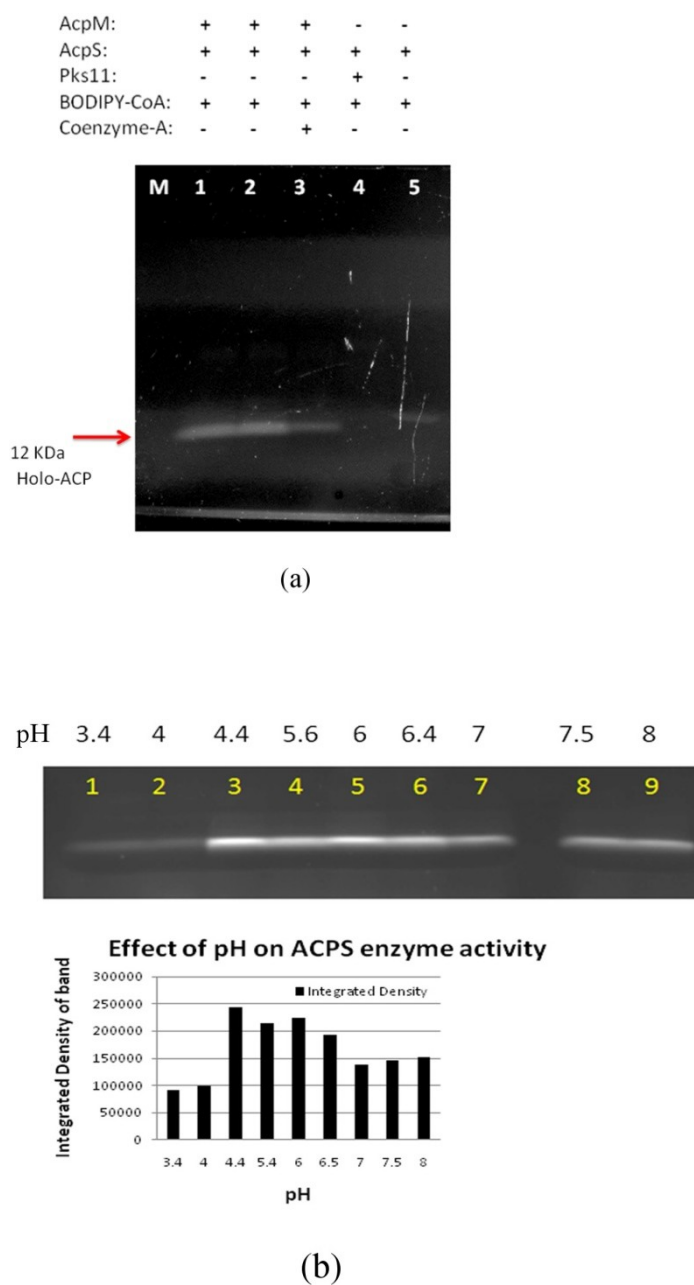


Figure 3-6. AcpS post-translational modification activity assay. (a) Lane M, molecular-weight markers; lane 1, 0.2 μ M AcpS; lane 2, 2 μ M AcpS; lane 3, pre-incubation of AcpS with CoA; lane 4, incubation with PKS11; lane 5, incubation of AcpS and BODIPY-CoA alone. (b) *Mtb* AcpS activity was assessed from pH 3.4 to pH 8.0. The intensities are shown in the bar diagram.

Next, we assessed the specificity of the AcpS post-translational modification activity. PKS11 was used as a negative control since it is not a substrate of AcpS *in vivo*. In this experiment, we incubated *Mtb* AcpS with BODIPY-labeled CoA and *Mtb* PKS11 in order to determine whether AcpS is able to transfer 4'-ppt to PKS11 (as this protein has a malonyl-CoA-binding tunnel). *Mtb* AcpS was unable to transfer the 4'-ppt moiety to PKS11 despite the presence of a CoA-binding tunnel, which demonstrates the specificity of AcpS. This experiment shows that *Mtb* AcpS is functionally active and is able to convert apo-ACP to holo-ACP. The crystal structure reveals that *Mtb* AcpS adopts one conformation at low pH and a different conformation at higher pH. This led us to assess the post-translational modification activity of AcpS at different pH values from 3.4 to 8.0. The fluorescence SDS gel clearly shows that AcpS is unable to transfer the 4'-ppt arm to ACP at lower pH values (3.4–4.0), but it is able to transfer the 4'-ppt arm to ACP at pH 4.4–8.0. However, the relative fluorescence intensity was higher between pH 4.4 and 6.0 compared with pH 6.6–8.0 (Fig. 3-6b). Recently, the *Mtb* FAS-I enzyme has been expressed and purified from *E. coli* and a reconstitution experiment showed that the FAS-I enzyme was inactive. In contrast, co-expression of the *Mtb* FAS-I enzyme with *Mtb* AcpS (in a duet vector) resulted in a FAS-I enzyme which was biochemically active (Qingan & Sacchettini, unpublished work). During co-expression, *Mtb* AcpS post-translationally modifies the ACP domain of the FAS-I enzyme *via* a CoA-derived 4'-ppt group. It is evident that AcpS modifies the ACP domain of FAS-I as well as AcpM.

Comparison of the *B. subtilis* AcpS–ACP complex with a proposed model of the *Mtb* AcpS–AcpM complex

ACP is a key component in both fatty-acid biosynthesis and polyketide biosynthesis. Initially, ACP is post-translationally modified by AcpS. During this process, AcpS recognizes the conserved Asp-Ser-Leu motif in apo-ACP and transfers the 4'-ppt moiety from CoA to a serine residue. The crystal structure of the *B. subtilis* AcpS–holo-ACP complex has been reported (Parris et al., 2000). In *B. subtilis* the Arg14, Arg21, Arg22 and Arg28 residues of AcpS interact with the Asp38 and Glu46 acids of holo-ACP. Multiple sequence alignments indicate that *Mtb* AcpS does not contain an Arg residue in the N-terminal region (residues 14, 21, 22 and 28); rather, it contains Asp, Gln and Glu in place of Arg (Fig. 3-7a). *C. ammoniagenes* AcpS has similar residues to those of *Mtb* AcpS, with the exception of the Arg22 residue. In contrast, the ACP amino-acid sequence alignment indicates that the EKS, DED and IPDED motifs and the 4'-ppt-binding motif (DSL) are highly conserved in all bacteria. In *B. subtilis*, the conserved residues of ACP interact with AcpS. In order to understand the interaction between AcpS and AcpM in *Mtb*, we superimposed the *Mtb* AcpS structure (the current model) and the AcpM NMR structure (PDB entry 1klp; (Wong et al., 2002); residues 1–90) with the *B. subtilis* AcpS–holo-ACP complex structure (Fig. 3-7a). The superimposed model of the AcpS–holo-ACP complex was energy-minimized using the *Discovery* III suite of the *INSIGHT* II software. The complex shows that the location and orientation of *Mtb* ACP are very similar to those of *B. subtilis* ACP; however, the interacting residues are different. Specifically, Glu19, Gln22, Ser45, Arg48, His49,

Glu77 and Arg81 of AcpS are responsible for *Mtb* ACP binding. In particular, AcpS residues Ser45, Arg48 and His49 are close enough to interact with Asp38 of ACP (Fig. 3-7b). The conserved IPDED motif of ACP is sufficiently near to form a hydrogen-bonding network with Glu77 and Arg81 of AcpS. Another conserved residue, Asp46, of ACP is close enough to interact with Glu19 and Gln22 of AcpS. The minimized AcpS–AcpM model shows that Arg48 and His49 of AcpS are at a favorable distance to interact with AcpM residues in addition to binding CoA or ADP. This model contains residues 1–90 of ACP but does not address the role of the C-terminal region of ACP and its interaction with AcpS. We are currently working to produce an ACP–AcpS complex to answer these questions.

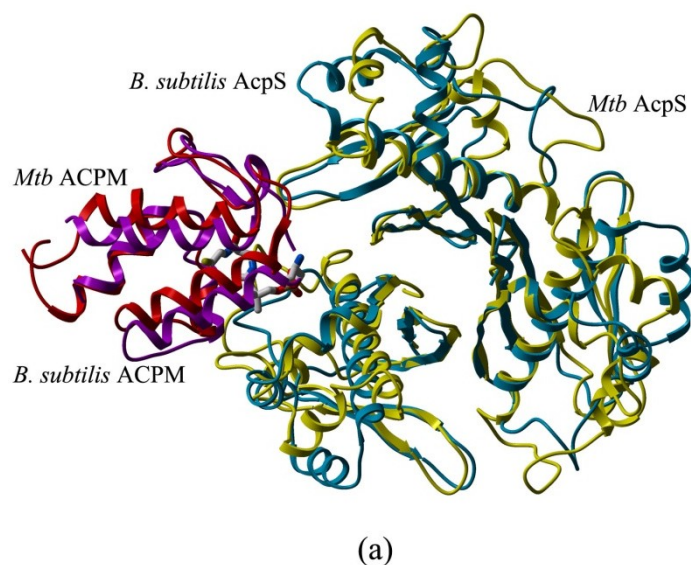
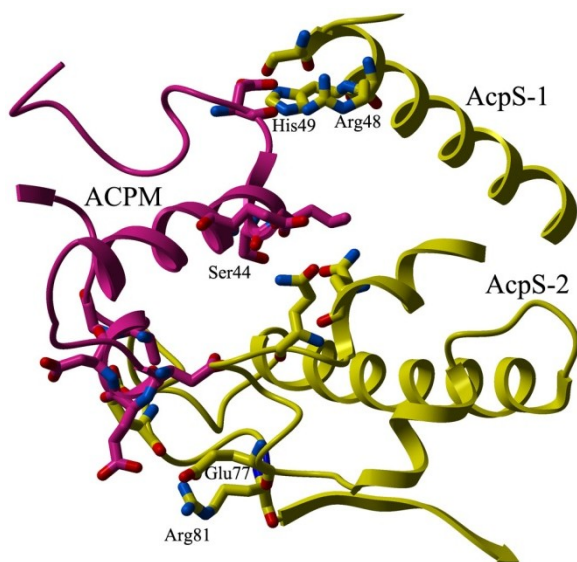


Figure 3-7. Comparison of AcpS-AcpM complex. (a) Structural superposition of AcpS–holo-ACP from *B. subtilis* with *Mtb* AcpS and the solution structure of AcpM. (b) The interacting residues of *Mtb* AcpS and AcpM are shown as stick models. ACP occupies the space between the dimer interface and SXXXH in a favorable position to interact with ACP.



(b)

Figure 3-7. Continued

CONCLUSIONS

We have determined the structures of both apo-AcpS and AcpS in complex with CoA from *Mtb* and *C. ammoniagenes* and have found that AcpS adopts different pH-dependent conformations. The active sites of the apo-AcpS and CoA-bound AcpS structures show that His49 and His116 (His65 and His136 in *C. ammoniagenes*) participate in either sulfate binding or CoA binding. The fluorescence assay demonstrates the functionality of AcpS, which converts apo-ACP to holo-ACP. The structural differences observed between *Mtb* AcpS and *C. ammoniagenes* AcpS may be essential for interaction with ACP. Finally, the structural basis of AcpS–ACP interactions in the *Corynebacterium* subgroup is fundamentally different from the electrostatic complementarity found in other bacterial systems.

ADDITIONAL UNPUBLISHED RESULTS

Mtb AcpS represents an attractive target for inhibitor discovery as it is critical for the activation of FAS-I and FAS-II systems in *Mtb* (Chalut et al., 2006). To identify novel chemical scaffolds that can be developed as leads we have applied a fragment-based ligand discovery (FBLD) approach. The FBLD paradigm requires relatively smaller number of compounds (hundreds to few thousands) in contrast to conventional high-throughput screening, to sample large chemical space enabling rapid and efficient identification of small molecular-weight ($M_r \sim 120-250$) chemical fragments that can serve as starting points for lead generation and drug development (Rees et al., 2004). Differential scanning fluorimetry (DSF) is particularly suited for fragment screening in a high-throughput manner as it is an inexpensive and efficient technique for identifying small ligands that bind to the target protein (Niesen et al., 2007). In DSF, thermal unfolding of proteins in presence of ligands is monitored using fluorescent dyes that fluoresce highly upon interaction with the exposed hydrophobic sites on the unfolded protein, by using a real-time PCR instrument (Fig. 3-8). DSF enables efficient detection of ligands that bind and stabilize a protein as they cause a rightward shift in the thermal unfolding curve of the protein with a concomitant increase in its melting temperature (T_m) (Figs. 3-9a and 3-9b).

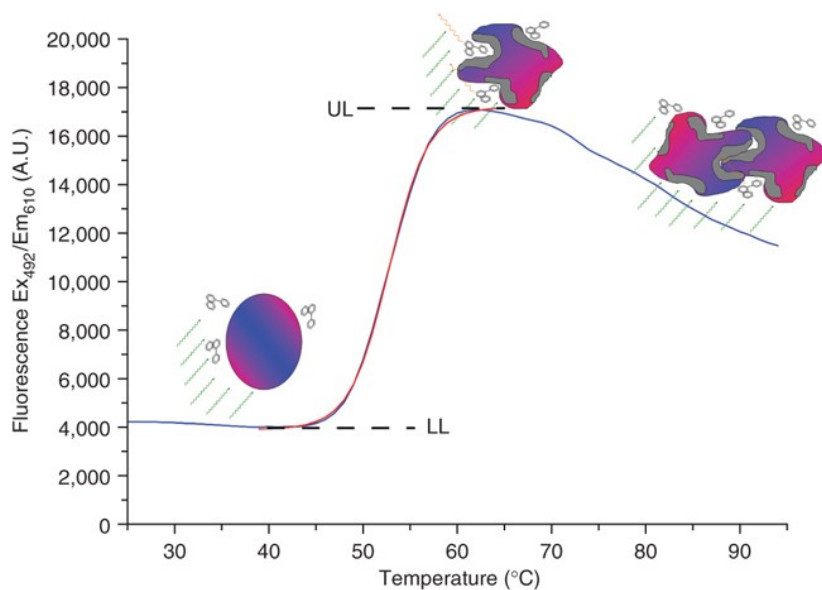


Figure 3-8. Schematic representation of DSF experiment. SYPRO orange (three ring molecule) has low fluorescence when the globular protein (spherical shape) is folded at the room temperature. The protein begins to unfold as the temperature increases, exposing its hydrophobic patches (grey) which interact with SYPRO orange causing an increase in fluorescence intensity. As the protein continues to unfold, the fluorescence also increases until it reaches a peak, and then it declines due to aggregation and precipitation of the denatured protein. Figure adapted from Niesen et al., *Nat Protoc.* (2007) 2(9):2212-21, by permission of © [2007] Macmillan Publishers Ltd. DOI: 10.1038/nprot.2007.321

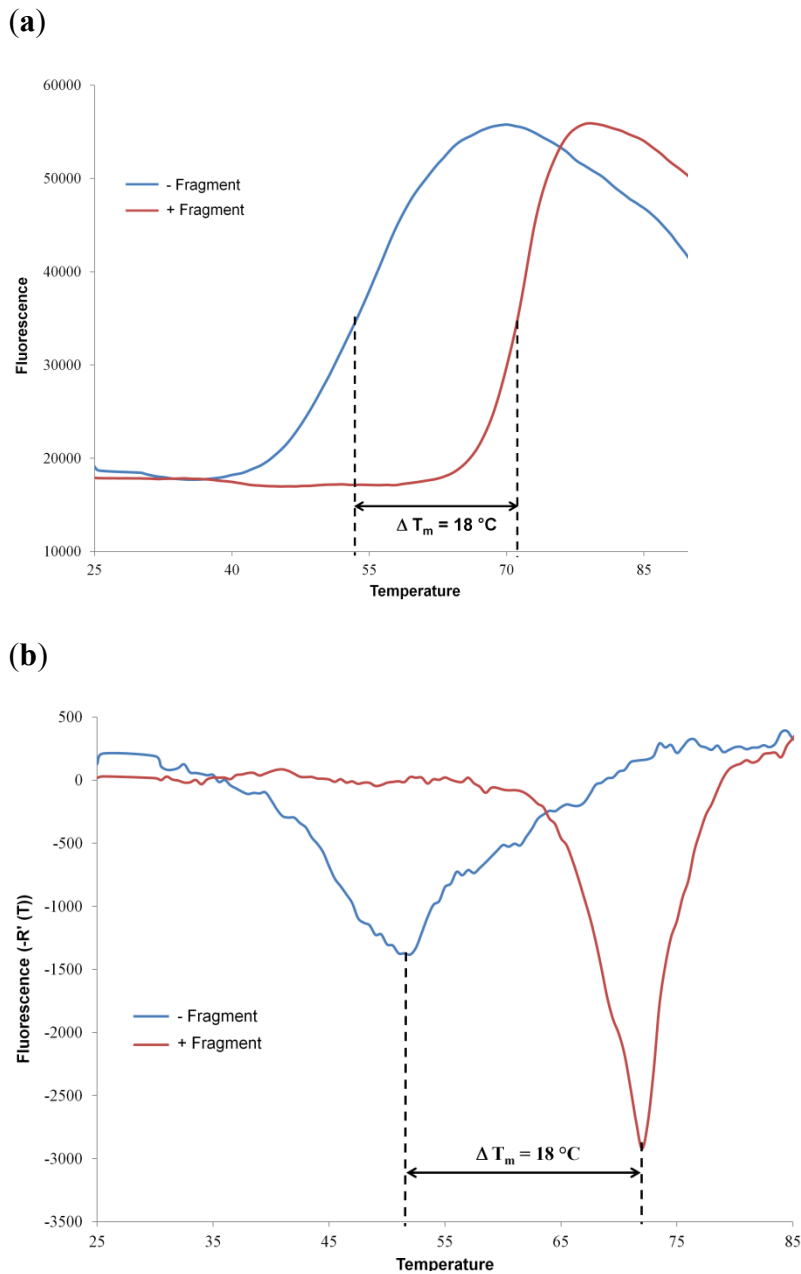


Figure 3-9. Typical fluorescence intensity plots from DSF. **(a)** Fluorescence intensity plotted as a function of temperature from raw amplification data gives a sigmoidal curve. The melting temperature (T_m) is calculated by fitting the data to Boltzmann equation. **(b)** Another representation of the fluorescence intensity data from DSF plotted as a first derivative plot. This simple method is implemented in most RT-PCR data analysis software packages in which T_m is calculated by determining maximum of the first derivative which is indicated by the tip of the peak on the curves.

Fragment based screening for ligands of *Mtb* AcpS

To establish a hit selection criteria, AcpS was characterized by DSF to determine its melting temperature (T_m) both without and with its ligand CoA. AcpS alone had an average T_m of 50 °C, whereas in the presence of CoA the average T_m increased to 68 °C (ΔT_m of 18 °C) (Figs. 3-9a and 3-9b). For fragment screening, a total of 780 compounds from an in-house fragment library (obtained from Chris Abell lab at University of Cambridge) were screened for binding to AcpS by DSF. The primary screen yielded 24 fragments that caused the melting temperature of AcpS to increase by more than 3 °C (Table 3-4). To confirm the binding of the hits from the primary screen, 6 out of the 24 hit compounds were repurchased based on their commercial availability, and tested by re-screening AcpS at different concentrations of the compounds using DSF. This titration assay confirmed 4 of the compounds as valid and reproducible hits (Fig 3-10). These fragments will serve as initial scaffolds for the design of more potent leads against *Mtb* AcpS with structure guided rational chemical modifications. Further, as more fragments become available, they will also be validated by titration assay and put into the lead development pipeline.

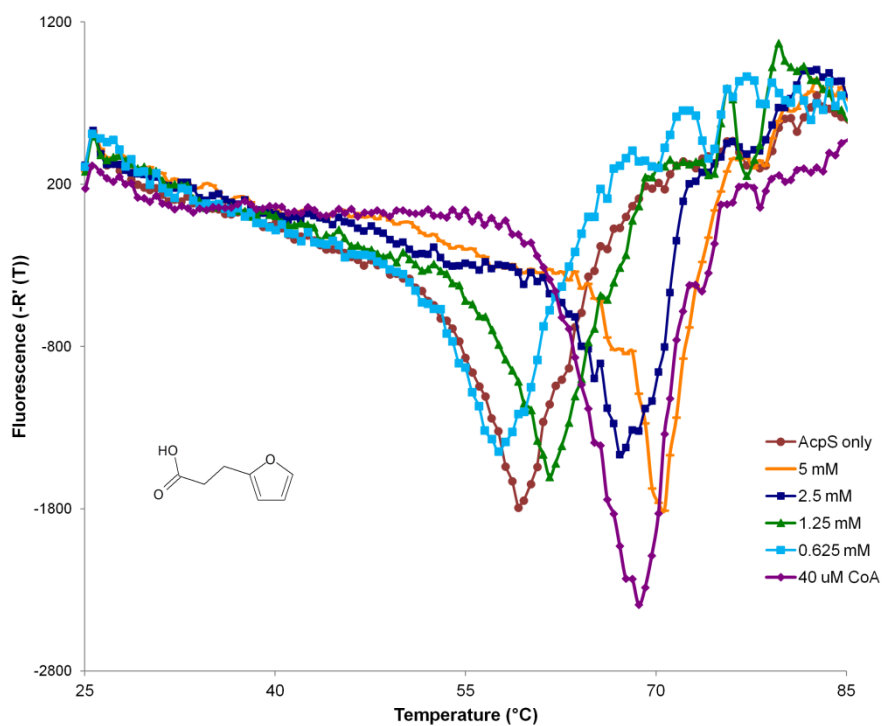


Figure 3-10. A representative plot of the compound titration assay. Concentration-dependent increase in the thermal stability of AcpS confirms the compound 1 (shown in line diagram) as a valid hit.

Table 3-4. Fragment hits from DSF screening of *Mtb* AcpS.

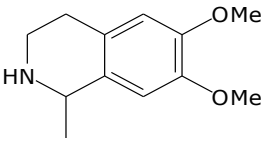
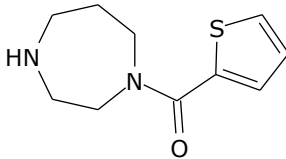
Compound number	Hit fragment identified by DSF	ΔT_m (°C)
1		6.5
2		7.5

Table 3-4. Continued.

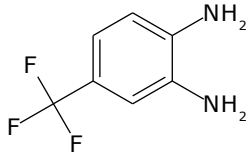
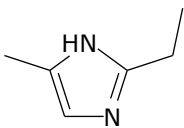
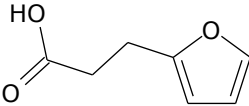
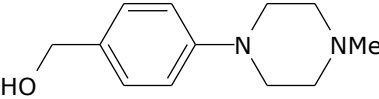
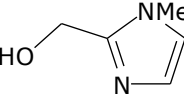
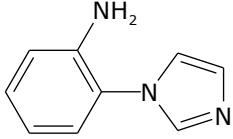
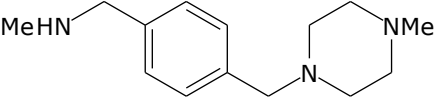
Compound number	Hit fragment identified by DSF	ΔT_m (°C)
3		4.5
4		4
5		18
6		5
7		3.1
8		8.5
9		6

Table 3-4. Continued.

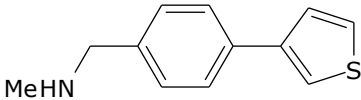
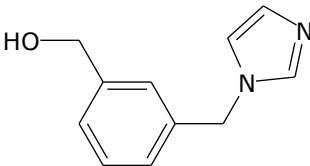
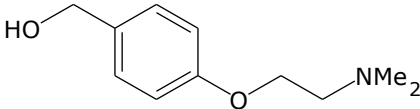
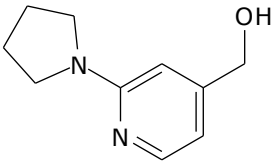
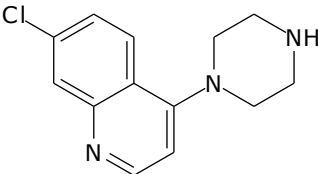
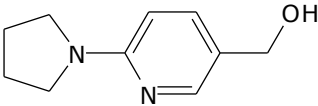
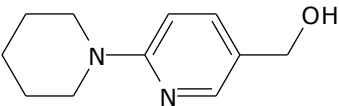
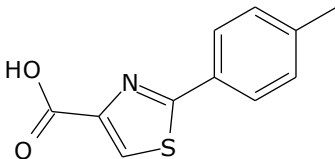
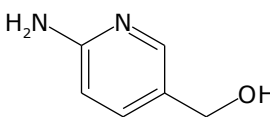
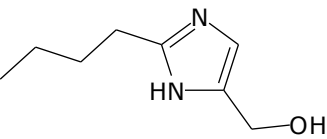
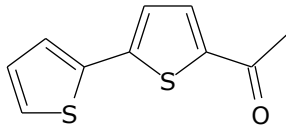
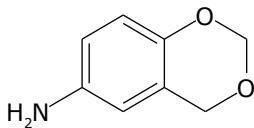
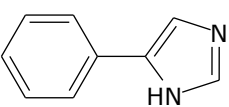
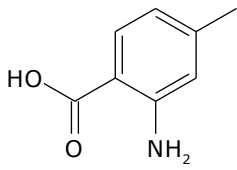
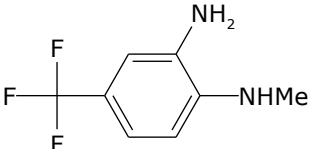
Compound number	Hit fragment identified by DSF	ΔT_m (°C)
10		6.5
11		5.5
12		7
13		7
14		9
15		6.6
16		8.5

Table 3-4. Continued.

Compound number	Hit fragment identified by DSF	ΔT_m (°C)
17		13.6
18		7
19		8
20		5
21		5.5
22		9.5
23		7
24		7

ADDITIONAL UNPUBLISHED EXPERIMENTAL METHODS

Differential scanning fluorimetry (DSF) of *Mtb* AcpS

The DSF experiments were performed in 96-well PCR plates using Stratagene Mx3005P RT-PCR system. For fragment screening, 19 μL of AcpS master-mix containing 4 μM AcpS in 0.2 M MES, pH 5.2 buffer with 4 mM MgCl_2 , and 2.5x SYPRO orange (Life Technologies) was dispensed in 96-well assay plates. To this 1 μL of compounds from the fragment library were added and mixed well. Then the plates were sealed and spun at 300g for 1 min to make sure no bubbles were trapped and no solution was sticking to the walls. As a positive control, 40 μM CoA was included in each assay and 1 μL DMSO was added to all the control wells since the fragment compounds were dissolved in 100% DMSO. The data from the assay was analyzed using MxPro v4.10 software that came pre-installed with the instrument.

The compound titration assays for validating the primary screen hits were done in 96-well PCR plates as was done for the initial screen. The reaction mix contained 4 μM AcpS in 0.2 M MES, pH 5.2 buffer, 2.5x SYPRO orange dye and varying concentrations of the selected hit compound in a total volume of 20 μL . The compounds were dissolved in 100% DMSO and the final concentration of DMSO in the reaction mix was 5%. Two types of controls were also setup for the assay- one without CoA and the other with 40 μM CoA, and both of these control reactions also contained 5% DMSO. All the reactions were setup in duplicates and a total of 6 compounds were tested at 4 different concentrations (5, 2.5, 1.25 and 0.625 mM).

FUTURE DIRECTIONS

Prospects for PPTases as targets for inhibitor discovery

In microorganisms, PPTases are essential enzymes that activate carrier domains on proteins involved in primary and secondary metabolic pathways that produce a wide variety of pathogenesis related virulence factors and toxins (Lambalot et al., 1996). Some of the representative virulence factors from pathogenic organisms include complex lipids PDIM and mycolic acids of *Mtb* cell wall, toxin mycolactone from *M. ulcerans*, and siderophores such as mycobactin in *Mtb* or pyoverdine in *P. aeruginosa* (Kosa et al., 2014). Thus, as a result of their involvement in host viability and virulence, PPTases have started to attract attention as promising new targets for drug development (Joseph-McCarthy et al., 2005), especially against the pathogens that have developed significant clinical drug resistance and threaten health-care efforts worldwide. For example, *Mtb* is among the top threats to global health for which there is an urgent need to identify drugs that can shorten treatment duration, and are also active against MDR and XDR strains. In *Mtb*, PPTases participate in activation of carrier domains in >20 proteins belonging to FAS I & II, PKS and NRPS systems (Cole et al., 1998). These enzyme systems synthesize a diverse array of complex lipids that are essential for the viability, virulence and pathogenicity of *Mtb* (Neyrolles and Guilhot, 2011). Therefore, discovery and development of inhibitors of the two non-redundant PPTases in *Mtb* offers attractive prospects for the development of novel broad-spectrum anti-tubercular drugs as their inhibition would disrupt multiple biosynthetic pathways in *Mtb*. In addition,

other mycobacterial pathogens such as *M. leprae*, *M. avium* and *M. ulcerans* also have orthologs of the two *Mtb* PPTases, suggesting that the inhibitor discovery efforts can be extended to include these pathogens too.

However, despite the knowledge of the essential nature of this family of enzymes, only a limited number of studies have been done for inhibitor screening and development targeting these enzymes. The major reason for this has been lack of a simple and robust PPTase assay that is adaptable to automated high-throughput screening of compound libraries. The traditionally used assays for the PPTase activity have utilized technically complex HPLC or radiolabelling and involved cumbersome treatment steps (Chalut et al., 2006). To overcome these limitations, recently many spectrophotometry based assays that are suitable for HTS have become available for screening PPTases. Two studies have reported development of fluorescence based assays that use fluorescence resonance energy transfer (FRET) (Yasgar et al., 2010) and fluorescence polarization (FP) techniques (Duckworth and Aldrich, 2010) to spectrophotometrically monitor the PPTase activity. These assays utilized fluorescent CoA conjugates for HTS of PPTase Sfp from *B. subtilis*, and the FRET-based method was successfully used to screen a LOPAC¹²⁸⁰ library of bioactives for inhibitors of Sfp (Yasgar et al., 2010). Furthermore, another study reported the use of blue pigment indigoidine synthesizing, single module NRPS BpsA as a reporter for assessment of PPTase activity in a HTS (Owen et al., 2011). These developments have enabled rapid screening of promiscuous Sfp-type PPTases, however, they have not been adopted for more specific AcpS-type PPTases. Though these assays did not enable evaluation of

natural substrates as they required fluorophore labeled substrates, modifications of these methods have been reported that allow direct and continuous assay measurements (Kosa et al., 2014).

More recently, a scintillation proximity assay (SPA) using radiolabeled CoA for HTS was reported for *Mtb* PPTase PptT (Leblanc et al., 2012). Moreover, we have used DFS to screen a fragment library in a high-throughput manner to identify compounds that bind *Mtb* AcpS, suggesting that DSF is also amenable for HTS of compound libraries. In addition, we have successfully demonstrated transfer of a fluorescent label on acyl carrier protein AcpM, the natural substrate of AcpS in *Mtb*, so this *in vitro* assay can be also used for the HTS of inhibitors, or to validate hits from the DSF screens. Availability of these simple and robust assays for HTS will facilitate anti-*Mtb* inhibitor discovery, and coupled with the structural information reported in this chapter, the identified inhibitors can be improved through structure-guided rational development.

CHAPTER IV

TARGETING *Mtb* PKS13 THIOESTERASE DOMAIN FOR ANTITUBERCULAR DRUG DISCOVERY

SYNOPSIS

Tuberculosis (TB) continues to be a major health problem worldwide. Emergence of multidrug-resistant *Mycobacterium tuberculosis* (*Mtb*), the etiologic agent for TB, has created an urgent need for the discovery and development of new antitubercular drugs that are effective against the drug-resistant bacteria. We report here biochemical and structural studies detailing mechanism of inhibition by a novel benzofuran class inhibitor that was found to inhibit *Mtb* growth. Our studies establish that the inhibitor has a unique mode of inhibition whereby it targets the thioesterase activity of the essential mycolic acid condensing enzyme Pks13 in *Mtb*. Structure of the Pks13 thioesterase domain in complex with the inhibitor revealed important interactions, and studies with analogs led to the identification of structural features required for effective inhibitors. Two mutations, D1607N and D1644G, conferred resistance to the thioesterase domain against inhibition by the benzofuran compound, further confirming it as the target of the compound. Structure of the D1607N mutant allowed us to determine molecular basis for the mechanism of resistance. This study identifies a novel mode of inhibition of *Mtb* Pks13 and further validates it as a critical target for drug

discovery. The structural studies reported here provide a platform for structure-based design and development of better inhibitors.

INTRODUCTION

The mycolic acid biosynthesis pathway is well-established to be an essential pathway for survival of *M. tuberculosis*. A number of clinically-used drugs target genes on this pathway, including the first-line drug isoniazid, as well as ethionamide. *M. tuberculosis* is the causative agent of tuberculosis, which causes more than 1.5 million deaths annually, and for which drug resistance, including multi-drug resistance, has become an increasing problem worldwide. Like other mycobacteria, *M. tuberculosis* produces a waxy outer cell-wall layer of very-long-chain fatty acids called mycolic acids, and disruption of mycolic acid synthesis is lethal. Isoniazid targets the enoyl-acyl-ACP reductase (InhA) enzyme that is one of several proteins of the FAS II system participating in the elongation of long-chain (C40-C60) fatty acid components of mycolic acids (after being activated by catalase-peroxidase (KatG) and forming an adduct with NADH) (Banerjee et al., 1994). However, resistance to isoniazid is increasingly observed and is currently estimated at around 5% worldwide (including both monoresistance as well as MDR/XDR-TB).

We recently reported the identification of a benzofuran-based compound from a whole-cell screen that targets polyketide synthase Pks13, a different enzyme in mycolic acid synthesis pathway (Ioerger et al., 2013). Whole-genome sequencing of resistant

mutants revealed resistance mutations in the thioesterase domain of Pks13. Pks13 is a large, multi-modular polyketide synthase that performs the final condensation of a long-chain (C40-C60) fatty acid (synthesized by the FAS II system) with a ~C26 fatty acid (synthesized by the FAS I system) to form an alpha-hydroxy meromycolate. Pks13 is essential *in vitro*, but is not targeted by any existing drugs in clinical use, making this a promising route for development of new drugs that interfere with the sensitive pathway of mycolic acid synthesis.

Pks13 is comprised of five domains that harbor all the activities required for the single round of condensation of the two long fatty-acyl chains. Pks13 has two acyl carrier protein domains, ACP_N and ACP_C, a β -ketoacyl-synthase (KS), an acyltransferase (AT) and a C-terminal thioesterase (TE) domain. The two ACP domains accept distinct substrates – the ACP_N accepts activated meroacyl-AMP from FadD32 and transfers it to the KS domain, and the ACP_C is loaded with 2-carboxyacyl-CoA chain by the AT domain. The KS domain then performs the Claisen-type condensation to produce α -alkyl β -ketoester attached to the ACP_C which is then released by the TE domain for subsequent modification reactions (Takayama et al., 2005). Recently, a pyrrole-based compound identified through another high-throughput screen, with an MIC of 0.5 μ M against H37Rv, was proposed to inhibit the N-ACP domain of Pks13 (Wilson et al., 2013), adding evidence that chemical inhibition of this enzyme, which is a critical step in the mycolic acid biosynthesis pathway, is lethal to *M. tuberculosis*.

Herein, we report the biochemical and structural validation of the *Mtb* Pks13 thioesterase (TE) domain as the target of benzofuran-based whole-cell active inhibitors

based on compound **1** (Fig. 4-1). We confirmed the inhibition of TE activity by **1** biochemically, and we determined the binding mode by solving the crystal structure of the Pks13 TE domain in complex with **1**. We performed a preliminary structure-activity relationship (SAR) study by testing several structural analogs of **1** for enzyme inhibition and whole cell activity. Furthermore, we solved the structures of four of these analogs in complex with the TE domain to identify structural requirements for active inhibitors against the Pks13 TE domain. In addition, to elucidate the mechanism of resistance, we also solved the structure of mutant forms of the TE domain. Our structural analysis of Pks13 and the sensitivity of mycobacterial cells to its chemical inhibition show that it has the potential to make a good platform for development of new inhibitors of the well-established essential mycolic acid pathway.

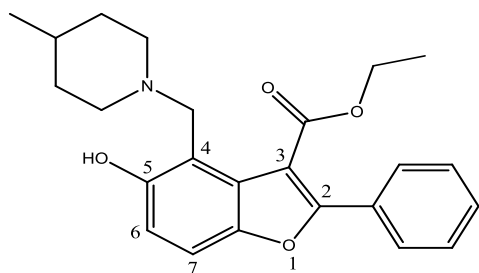


Figure 4-1. Chemical structure of benzofuran **1**.

EXPERIMENTAL METHODS

Cloning and overexpression of *Mtb* Pks13 TE domain constructs

The TE domain constructs corresponding to the predicted TE domain in *Mtb* Pks13 gene (Rv3800c) were made by PCR from the *Mtb* H37Rv genomic DNA as the template. The amplified DNA fragments were incorporated into the pMCSG-19b vector by ligation independent cloning (LIC) to yield TEV protease cleavable N-terminal His₆-tagged TE domain constructs (Donnelly et al., 2006). The TE-pMCSG-19b vectors were transformed into *E. coli* BL21(DE3)pLysS cells (Novagen) and the transformed cells were grown at 37 °C in LB media containing carbenicillin (100 µg/ml) and chloramphenicol (34 µg/ml) to an OD₆₀₀ of 0.6. Expression of TE constructs was induced with 0.5 mM IPTG, and cells were harvested after 16 hours of growth at 20 °C. The D1607N and D1644G mutants of Pks13 TE domain were constructed using the QuikChange site-directed mutagenesis kit (Stratagene). The mutations were confirmed by DNA sequencing. Mutant plasmids were transformed into *E. coli* BL21(DE3)pLysS cells, and mutant proteins were expressed by induction with 0.5 mM IPTG at 20 °C for 18 h.

Purification of Pks13 TE domain

The harvested cells were resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10% (v/v) glycerol, 1 mM β-mercaptoethanol (BME) and DNase) and lysed by French press. The resulting cell extract was clarified by centrifugation (15,000

x g at 4 °C) for 1 hour. The cleared supernatant was loaded onto a Ni-affinity column and the His-tagged TE domain constructs were eluted with a linear gradient of 10-250 mM imidazole in 20 mM Tris-HCl, pH 8.0 and 0.5 M NaCl. The peak fractions were pooled and the His-tag was cleaved by overnight incubation with TEV protease in dialysis buffer (20 mM Tris-HCl pH 8.0, 10% (v/v) glycerol and 1 mM DTT). The TEV cleaved protein was passed through Ni-column to remove any uncleaved His-tagged protein using 20 mM Tris-HCl (pH 8.0) with 100 mM NaCl and 1 mM BME. His-tag cleaved protein eluted in the flow-through and was concentrated for loading onto a Superdex-200 gel filtration column (GE Healthcare). The TE domain constructs eluted under a single peak as a monomer (~32 kDa) from the gel filtration column and were >95% pure as observed by SDS-PAGE. The purified protein was concentrated to 20-25 mg/ml, flash-frozen and stored at -80 °C. The TE domain mutants were purified using the same protocol as for the wild-type TE domain constructs. Both the mutants and the wt TE domain protein constructs have the amino acids SNA from the TEV cleavage site appended to the N-terminus.

Crystallization and soaking with ligands

Initial screening for crystallization conditions for the soluble TE domain constructs was done by sitting drop method using 1 µl of purified protein (15-20 mg/ml) and 1 µl of crystallization buffer from the well solution. After extensive screening, crystals were obtained for only the 283 residue long construct of the TE domain starting from residue 1451 in full length Pks13 (hereafter referred to as TE1451). The TE1451

crystals were obtained in crystallization buffer containing 0.1 M Tris-HCl, pH 8.5 and 2.0-1.8 M ammonium sulfate as precipitant. The crystals were further optimized by using polypropylene glycol P-400 as an additive at 2%-5% (v/v) in the original condition. To obtain TE1451-inhibitor complex crystals, soaking of the inhibitors was done by transferring apo-TE1451 crystals into a drop consisting of 0.1 M Tris-HCl, pH 8.5 and 2-2.2 M ammonium sulfate with 1-2.5 mM inhibitor added from a DMSO stock keeping the final DMSO concentration at <5%, and incubated at 18 °C and 4 °C for 4-20 hours.

Crystals of the TE11451:D1607N mutant were obtained by sitting drop method at 18 °C. The crystallization drops contained an equal volume of the protein solution (15-20 mg/ml) and mother liquor (0.1 M HEPES, pH 7.5, 2%-4% (v/v) PEG 400, and 1.8-2 M ammonium sulfate), and the diffraction quality crystals were obtained within 2 weeks.

Data collection and processing

For diffraction data collection the crystals were cryo-protected using either Fomblin (Sigma) or 2.4 M malonate (Hampton Research) and flash frozen in liquid nitrogen. High resolution data was collected at a wavelength of 0.98 Å on the beamlines 19-ID and 23-ID at the Advanced Photon Source (APS) of the Argonne National Laboratory. All the data sets were processed and scaled with HKL2000 (Otwinowski and Minor, 1997). Analysis of the integrated and scaled data by Xprep indicated that TE1451 crystallized in P2₁2₁2 space group (Sheldrick, 2008b). Solvent content analysis

indicated the presence of two molecules (V_M 2.16, V_S 43.2%) in the asymmetric unit (Matthews, 1968).

Determination of TE1451 structures and model refinement

The structure of the TE domain was solved by molecular replacement method (MR) using *E. coli* EntF (PDB entry 3TEJ), as search model. A single MR solution was obtained using Phenix AutoMR (Adams et al., 2010) which was input into the AutoBuild wizard to generate the initial model for apo-TE1451. The initial model was improved by further manual rebuilding in COOT (Emsley and Cowtan, 2004). The final model was obtained after iterative cycles of model building and Phenix refinement with simulated annealing yielding a 1.72 Å resolution apo-TE1451 model with R_{cryst} of 17% and an R_{free} of 20% with good stereochemistry (Table 4-1). The final refined apo-model has two chains, designated A and B, and 388 water molecules in the asymmetric unit.

To solve the TE1451-inhibitor complex structures, as well as the D1607N mutant structure, only the protein atoms from chain A of the apo-TE1451 structure were used as search model in the initial rigid body refinement of the isomorphous $P2_12_12$ crystals in the Phenix Refine module. Inspection of electron density maps showed clear $|F_o - F_c|$ positive difference density for the ligands which were fit into the density using Ligandfit routine in Phenix (Terwilliger et al., 2006). The ligand model and geometry restraint files were created in ELBOW BUILDER of the Phenix suite (Moriarty et al., 2009).

Table 4-1. Data collection and refinement statistics for TE1451.

	ApoTE1451	TE1451:1	TE1451(D1607N)
Data collection			
Space group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	88.5, 106.7, 57.7	89.2, 109.5, 57	88.7, 108.9, 58.1
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	39.19-1.72 (1.78-1.72)*	50-1.94 (1.97-1.94)	44.36-1.88 (1.95-1.88)
<i>R</i> _{merge}	0.089 (0.508)	0.172 (1.78)	0.141 (1.44)
<i>I</i> / σ <i>I</i>	23.81 (4.33)	18.8 (2.3)	8.64 (0.53)
Completeness (%)	99.34 (96.09)	94.67 (93.29)	92.47 (59.25)
Redundancy	7.6 (6.7)	6.6 (5.2)	4.9 (2.9)
Refinement			
Resolution (Å)	1.72	1.94	1.88
No. reflections	58442	40142	42956
<i>R</i> _{work} / <i>R</i> _{free}	0.17/0.20	0.19/0.23	0.18/0.22
No. atoms			
Protein	4277	4217	4257
Ligand/ion	27	63	44
Water	388	210	309
<i>B</i> -factors			
Protein	22	47	35
Ligand/ion	31	48	54
Water	28	45	40
R.m.s. deviations			
Bond lengths (Å)	0.011	0.007	0.014
Bond angles (°)	1.35	1.17	1.44

Data from 1 crystal was used to solve each structure.

*Values in parentheses are for highest-resolution shell.

Iterative cycles of model building and NCS-restrained maximum likelihood refinement with simulated annealing in Phenix refine yielded high quality models for TE1451-inhibitor complexes (Table 4-2). Since the thioesterase domains have many loops with intrinsic flexibility, some of the residues in TE1451 structures could not be

built into the model due to the missing electron density and some residues are modeled as alanine due to ambiguous side chain electron density. In all of the structures >98% of residues are placed in the favoured region of the Ramachandran plot as determined by MolProbity(Chen et al., 2010) validation tool in Phenix.

Enzyme assay

Activity of TE1451 was assessed using 4-methylumbelliferyl heptanoate (4-MUH, Sigma) as a fluorogenic substrate in a 96-well plate format (Richardson and Smith, 2007). To make initial velocity measurements, TE1451 (1 μ M) in 0.1 M Tris-HCl, pH 7 buffer was incubated with different concentrations of 4-MUH (2-150 μ M in DMSO) in a 100 μ l reaction volume, and the fluorescence of the hydrolyzed product 4-methylumbelliferone was read (Ex at 355 nm and Em at 460 nm) using PolarStar Omega plate reader (BMG Labtech) at 5-10 min intervals over 60-70 min. The reaction rate was observed to be linear in the measured range. 4-MUH in buffer alone was included as a control to quantify its background hydrolysis. Data points were plotted as an average of triplicates and each experiment was repeated 2-3 times independently. The initial velocity data was curve fit to Michaelis-Menten equation by nonlinear regression using Prism software (Graphpad) to determine the kinetic parameters K_m and V_{max} . The assay and data analysis for TE1451 mutants was done the same way as that for the wild-type protein with the 4-MUH concentration varying from 2 to 300 μ M.

Table 4-2. Data collection and refinement statistics for analogs of **1**.

	TE1451:3	TE1451:4	TE1451:5	TE1451:6
Data collection				
Space group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	89.4, 109.4, 56.9	89, 110.2, 57.5	88.9, 109.7, 57.4	88, 109.4, 57
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	33.76-1.99 (2.06-1.99)*	48.33-2.04 (2.12-2.04)	48.2-2.05 (2.12-2.05)	46.45-1.99 (2.06-1.99)
<i>R</i> _{merge}	0.114 (1.15)	0.158 (2.01)	0.153 (1.04)	0.124 (0.936)
<i>I</i> / σ <i>I</i>	14.1 (1.5)	9.0 (0.8)	11.5 (1.6)	9.9 (1.3)
Completeness (%)	87.8 (84.4)	87.8 (84.9)	94.0 (91.6)	99.0 (93.9)
Redundancy	5.2 (4.5)	4.3 (3.7)	5.5 (4.6)	4.3 (4.1)
Refinement				
Resolution (Å)	1.99	2.04	2.05	1.99
No. reflections	34387	32201	33776	38083
<i>R</i> _{work} / <i>R</i> _{free}	0.20/0.24	0.21/0.24	0.20/0.24	0.19/0.21
No. atoms				
Protein	4232	4238	4203	4235
Ligand/ion	54	34	28	50
Water	156	102	218	390
<i>B</i> -factors				
Protein	44	56	34	34
Ligand/ion	33	79	33	29
Water	45	53	34	38
R.m.s. deviations				
Bond lengths (Å)	0.009	0.005	0.007	0.008
Bond angles (°)	1.27	0.95	1.10	1.12

Data from 1 crystal was used to solve each structure.

*Values in parentheses are for highest-resolution shell.

IC₅₀ determination

To determine the potency of **1** and its analogs against wt TE1451, the compounds were tested at concentrations ranging from 0.012 to 20 μ M in a 96-well plate format. The reaction mix contained 1 μ M TE1451 in 0.1 M Tris-HCl, pH 7 buffer with 1 μ l of each dilution of the compound or DMSO in a total volume of 99 μ l. The reaction

was initiated by addition of 1 μ l of 2 mM 4-MUH in DMSO (20 μ M final, equal to its K_m for TE1451) to the reaction mix. Initial velocity data was obtained by monitoring increase in the fluorescence due to hydrolysis of the substrate using PolarStar Omega plate reader at 10 min intervals over 60 min. The data points were collected in triplicate and the averaged value was used to generate concentration-response plots for **1** and its analogs. The IC₅₀ value for each compound was obtained by nonlinear regression curve fitting of a four-parameter variable slope equation to the dose-response data using Prism software. The IC₅₀ values of **1** for TE1451 mutants were determined in the same way as that for wt TE1451, however, the testing concentration of **1** ranged from 0.04 to 40 μ M and the substrate 4-MUH was used at 20 μ M for ease of comparison.

Whole cell and cytotoxicity testing

Whole cell testing for determining MIC was done using Alamarblue assay in 96-well plates (Franzblau et al., 1998). *Mtb mc*²-7000 strain cells (Sambandamurthy et al., 2006) were grown in 7H9 media supplemented with OADC (Middlebrook), 0.05% Tyloxapol (Sigma), and 25 mg/ml pantothenate to an OD₆₀₀ of 1–2. The cells were then diluted into testing media (7H9 media with 0.2% dextrose, 0.085% NaCl, 0.05% Tyloxapol, and 25 mg/ml pantothenate) to an OD₆₀₀ of 0.01 and dispensed into testing plates at 200 μ l per well. Then the compounds were added as a 2-fold serial dilution in DMSO (2% DMSO final in each well). The test plates also had a DMSO only control and a Rifampicin control. The plates were incubated with shaking at 37 °C for 6 days and then stained with resazurin (Sigma) for an additional 2 days at 37 °C. After staining

the fluorescence of reduced resazurin was read (Ex 544 nm, Em 590 nm) using PolarStar Omega plate reader. The fluorescence data were plotted as percent growth inhibition against the compound concentration and curve fitting was done by nonlinear regression using Prism software. Minimum inhibitory concentration (MIC) values were determined from the fitted curves.

Compounds were tested for toxicity by the Human Dermal Fibroblast (HDF) cytotoxicity assay. HDF cells were purchased from ATCC (Manassas, VA). The cells were cultured in DMEM (Lonza) media supplemented with 10% fetal bovine serum (Lonza) and penicillin/streptomycin (Lonza). For setting the cytotoxicity assay, compound stocks were serially diluted in phosphate buffered saline (PBS) plus 10% DMSO. On the day of assay, HDF cells were trypsinized, counted and resuspended at a concentration of 64,000 cells/ml in the media. Cells were plated, overlaid with the compound serial dilutions and incubated at 37 °C. After 48 h, resazurin dye was added and the assay plates were cultured for another 24 h. The next day the absorbance of the resazurin was measured on a microplate reader (BMG Labtech) to assess cell death. Cytotoxicity was determined as a percent of dead cells versus living.

RESULTS

Discovery of benzofuran **1** and identification of Pks13 as its target using forward chemical genetics

The benzofuran **1** was identified from a whole-cell screen for growth inhibition of *Mtb* strain H37Rv (Ioerger et al., 2013). The minimum inhibitory concentration (MIC) of **1** was determined to be 2.0 μ M against H37Rv. To identify the molecular target of **1**, spontaneous resistant mutants of *Mtb* were selected on agar plates with the concentration of **1** at 8x the liquid MIC (16 μ M). Sequencing of the genome of four resistant colonies identified only non-synonymous single nucleotide polymorphisms (SNPs) in the *pks13* gene. These SNPs result in two different amino acid substitutions in the TE domain region of the Pks13 protein: D1644G (one isolate) and D1607 (three isolates). Recombineering (oligo-mediated transfer of SNPs into the parental strain) was used to confirm that each SNP alone was sufficient to confer *Mtb* six- to eight-fold increased resistance to **1** (Ioerger et al., 2013).

Three different recombinant expression constructs encompassing different start and stop sites for the TE domain of *Mtb* Pks13 were designed based on the sequence alignments and the secondary structure from homologs. These were cloned into *E. coli* recombinant vector and tested for expression. One of the three protein constructs encompassing the residues 1400-1733 of Pks13 expressed His₆-tagged protein that was soluble even after removal of the tag. This construct containing residues 1451-1733 of Pks13 with three additional residues vector derived from the TEV cleavage site

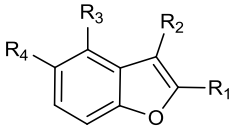

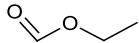
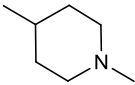

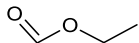
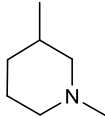

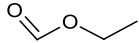


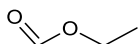
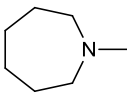

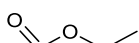
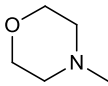

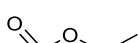


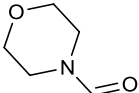
appended to its N-terminus produced single diffraction quality crystals. An enzyme assay was developed using this purified recombinant TE domain (hereafter called TE1451), with a fluorescence fatty acid, 4-Methylumbelliferyl heptanoate (4-MUH) as a substrate (Richardson and Smith, 2007). TE1451 showed good esterase activity using 4-MUH as the substrate with a $K_m \sim 20 \mu\text{M}$, and $k_{cat}/K_m \sim 7.2 \times 10^{-4} \mu\text{M}^{-1} \text{min}^{-1}$ (Table 4-3). The IC_{50} of **1** against TE1451 was determined to be $0.26 \mu\text{M}$ (Table 4-4). Next, in order to confirm the two mutations observed in isogenic resistant isolates, we expressed and purified the mutant proteins using an identical recombinant expression system as was used for the wt TE domain construct.

Table 4-3. Kinetic parameters of wild-type TE1451 and D1607N, D1644G mutants.

Protein	K_m (μM)*	$k_{cat} \text{ min}^{-1}$	$k_{cat}/K_m \mu\text{M}^{-1} \text{ min}^{-1}$	Relative activity
Wt.	19.5 ± 3.4	$140 \pm 9 \times 10^{-4}$	$7.2 \pm 1.3 \times 10^{-4}$	1
D1607N	8.5 ± 0.9	$105 \pm 3 \times 10^{-4}$	$12.3 \pm 1.4 \times 10^{-4}$	1.7
D1644G	6.9 ± 1.1	$166 \pm 6 \times 10^{-4}$	$24 \pm 3.8 \times 10^{-4}$	3.3

*Kinetic values were obtained by fitting the raw data to Michaelis-Menten equation. All assays were carried out in triplicates and the results are presented as mean \pm s.d.

Table 4-4. Preliminary SAR of **1** and its analogs.

						
Compound	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)*	MIC (μM)*
1				OH	0.26 ± 0.03	2.3 ± 0.2
2				OH	0.12 ± 0.02	4.4 ± 0.2
3				OH	0.24 ± 0.02	4.1 ± 0.1
4				OH	0.28 ± 0.03	4.6 ± 0.3
5				OH	0.71 ± 0.05	13.3 ± 1.6
6				OH	1.57 ± 0.15	7.3 ± 0.3
7			H	OH	No binding	ND

IC₅₀ values were determined using the *Mtb* Pks13 TE domain as described in the methods section. MIC values were determined for mc²7000 in liquid medium in 96-well plates. ND, not determined. *Values are shown as mean ± s.d. of triplicates.

Biochemical consequences of TE domain mutations D1607N and D1644G

The mutants (TE1451:D1607N and TE1451:D1644G) were characterized by determining the basic kinetic parameters (Table 4-3). Interestingly, both mutations had increased k_{cat}/K_m with the D1607N mutant showing an increase of ~1.7-fold and the D1644G >3-fold over the wt-TE1451. The D1644G mutation increased the IC_{50} of **1** by more than 66-fold from 0.26 μ M to 17.4 μ M against the TE domain, whereas the increase was more modest (0.26 to ~0.76 μ M) for the D1607N mutant (Fig. 4-2).

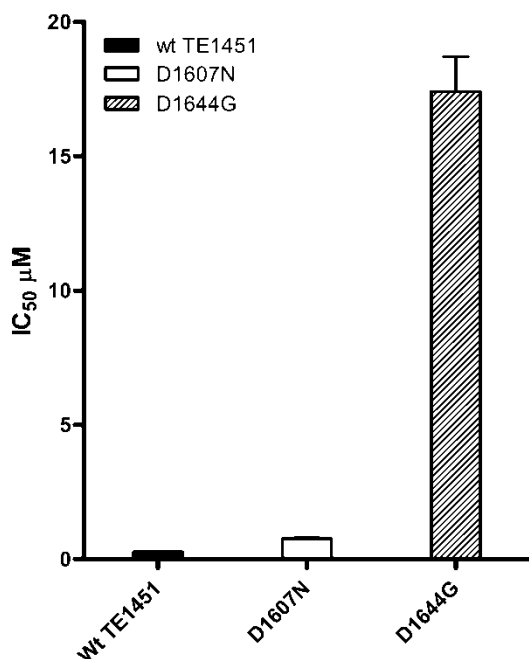


Figure 4-2. Comparison of IC_{50} values of **1** against wt TE1451 and D1607N, D1644G mutants. IC_{50} values were determined for the wt TE1451 and the mutants as described in the experimental methods section. The results are shown as the mean \pm s.d of triplicate experiments.

Preliminary SAR by inhibition studies with analogs of **1**

To establish a preliminary SAR for **1** based inhibition, a pharmacophore search was conducted and a set of six commercially available structural analogs of **1** were identified and evaluated for inhibition of TE1451 activity *in vitro*. These analogs differed primarily in replacements for the piperidine ring at R₃ of the benzofuran core by other saturated heterocycles (Table 4-4).

The IC₅₀ values ranged from 0.12 – 1.57 μM against the TE domain (Table 4-4). IC₅₀ data shows that only **2** was more potent (0.12 μM) compared to **1** (0.26 μM) against TE1451. The only difference between these two positional isomeric compounds is that the methyl group on the piperidine ring is in *para* position in **1** and in *meta* position in **2**. Compounds **3** and **4** have five- and seven-membered rings respectively at R₃, and they exhibit IC₅₀ similar to **1**. However, **5** which has a six-membered morpholine ring with a polar oxygen atom has ~3-fold higher IC₅₀ against TE1451 compared to **1** (Table 4-4). In contrast to the compounds with the cyclic amine substituents at R₃, when we tested the compound **6** with acyclic dimethyl amine at the R₃, there was >6-fold decrease in the inhibitory potency compared to **1**. In comparison to **1** and compounds **2-6**, **7** which lacks a group at R₃ and has a morpholine ring at R₂ in place of the ethyl ester, did not show any appreciable binding to TE1451 even at a concentration of 30 μM. Taken together, this SAR data suggests that the presence of a hydrophobic cyclic amine group at the R₃ position on the benzofuran core is important for potency of the compound, compared to the presence of a cyclic polar or an acyclic group at this position.

We next tested the structural analogs of **1** for anti-*Mtb* activity in a whole-cell growth inhibition assay using the *Mtb* strain *mc*²7000 (Sambandamurthy et al., 2006). All the compounds which were active against TE1451 *in vitro* also inhibited the growth of bacteria to varying extents (Table 4-4). Inhibitor **1** showed the most potent inhibitory effect on bacterial growth, with a minimum inhibitory concentration (MIC) of 2.3 μ M. Compounds **2**, **3** and **4** had IC₅₀ values comparable to **1** against TE1451, but they exhibited ~2-fold higher MICs against the bacteria. In comparison, **5** and **6** exhibited 6- and 3-fold higher MICs compared to **1**. Interestingly, **5** had an IC₅₀ value >2-fold lower than **6** against TE1451 *in vitro*, but in whole cell testing its MIC was ~2-fold higher than **6**. Since **5** is more polar, it is possible that this compound had reduced cell penetration resulting in an increase in its MIC. However, other effects like efflux or metabolism of **5** that can lead to an increase in its MIC cannot be ruled out.

To assess the potential of **1** and analogs that demonstrate whole cell activity (**2-5**) as suitable candidates for further development, we tested them for cytotoxicity against human dermal fibroblast (HDF) cells using resazurin dye assay (O'Brien et al., 2000). Compound **1** and analogs **2**, **4** and **5** did not show any growth inhibition in HDF cells up to 40 μ M, whereas **3** exhibited a half-maximum inhibitory concentration of >15 μ M (Fig. 4-3). Overall, the cytotoxicity testing data suggest that these compounds have good selectivity.

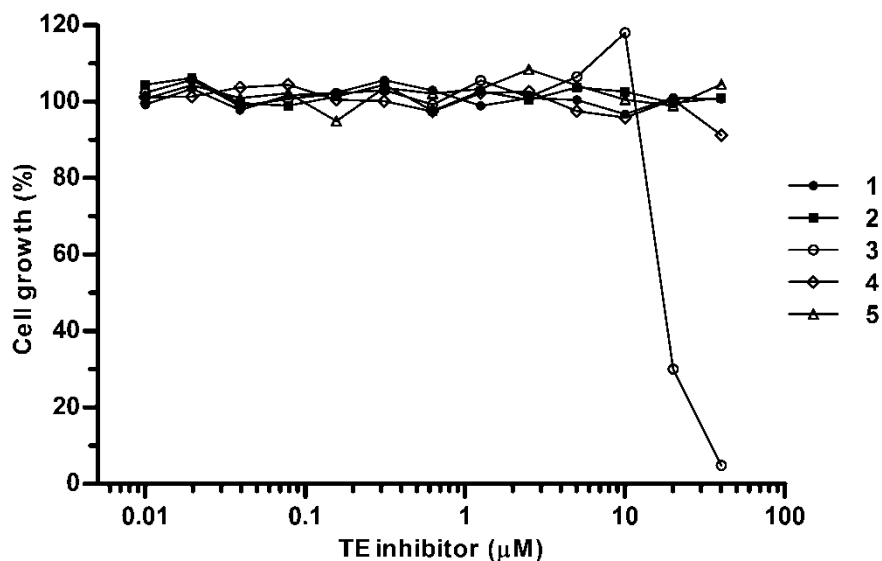


Figure 4-3. Cytotoxicity profile of **1** and its analogs. For cytotoxicity profile, compounds were evaluated by incubating with human dermal fibroblast cells as described in experimental methods section. The compounds were tested in a two-fold dilution series starting at a concentration of 40 μM . Cell growth is plotted as the percent of the DMSO only control.

Structural basis of inhibition of TE1451 by **1**

In order to understand the molecular basis of the interaction of TE1451 with inhibitors, we determined the crystal structures of TE1451 in apo and inhibitor complexed forms.

Thin plate like crystals of apo TE1451 were obtained by vapor-diffusion method in a condition with ammonium sulfate as precipitant at pH 8.5. The crystals diffracted to a maximum resolution of 1.7 \AA and they belonged to the space group $P2_12_12$ with two molecules in the asymmetric unit (designated A and B). The phase solution for the apo-

TE1451 structure was determined by molecular replacement method using the crystal structure of the *E. coli* EntF (PDB entry 3TEJ) as the search model. The final apo-structure contained molecules A and B with 278 and 272 residues built out of 283 residues, respectively, and was refined to a resolution of 1.72 Å ($R_{\text{work}} = 17\%$ and $R_{\text{free}} = 20\%$) with good stereochemistry (Table 4-1).

Since the two protein molecules in the asymmetric unit show a fairly similar overall structure (RMSD of 1.6 Å over 264 C_{α} atoms) with minor differences only in the mobile smaller mobile lid domains (Fig. 4-4), structure of the more complete molecule A is hereafter used to describe the overall structure of the TE domain. The TE1451 structure is divided into two domains (lid and core), with the larger core domain (residues 1451-1570, 1646-1660 and 1680-1733) possessing the canonical α/β -hydrolase fold comprised of a central seven-stranded β -sheet ($\beta 1$ - $\beta 7$) flanked by four α -helices ($\alpha 1$ - $\alpha 3$ and $\alpha 11$) with the N-terminal $\beta 1$ strand anti-parallel to other strands (Cantu et al., 2010; Chakravarty et al., 2004; Nardini and Dijkstra, 1999) (Fig. 4-5). The smaller lid domain (residues 1575-1645) is inserted between strands $\beta 5$ and $\beta 6$ of the core domain and consists of four α helices $\alpha 4$ - $\alpha 7$ adjacent to the core domain. Two small helices $\alpha 8$ and $\alpha 9$ (residues 1665-1675) that are present on a long loop between strands $\beta 6$ and $\beta 7$ of the core domain are intermediate between the core and the lid domain. Analysis of the TE1451 structure with VAST server (Gibrat et al., 1996) uncovered >100 structures belonging to thioesterase, hydrolase and lipase class enzymes with structural alignment scores well above the threshold of significance (VAST $-\log(p) > 10$). The top hit (VAST $-\log(p) > 15$) in the structural alignment was the thioesterase of fengycin synthesis non-

ribosomal peptide synthetase (PDB entry 2CB9) from *B. subtilis*. However, regardless of the enzyme class, the structural similarity was within the α/β -hydrolase core domain, the lid domain did not show significant structural conservation among the hits. In a VAST search using only the coordinates of the lid domain (1575-1645), only three hits were returned with alignment scores well below the significance threshold (VAST $-\log(p) < 4$) which had no structural or functional relationship with the lid domain. This finding is consistent with the role of lid domains in substrate recognition and binding in a variety of α/β -hydrolase enzymes, which accounts for their highly variable shapes and substrate specificities (Tsai and Ames, 2009).

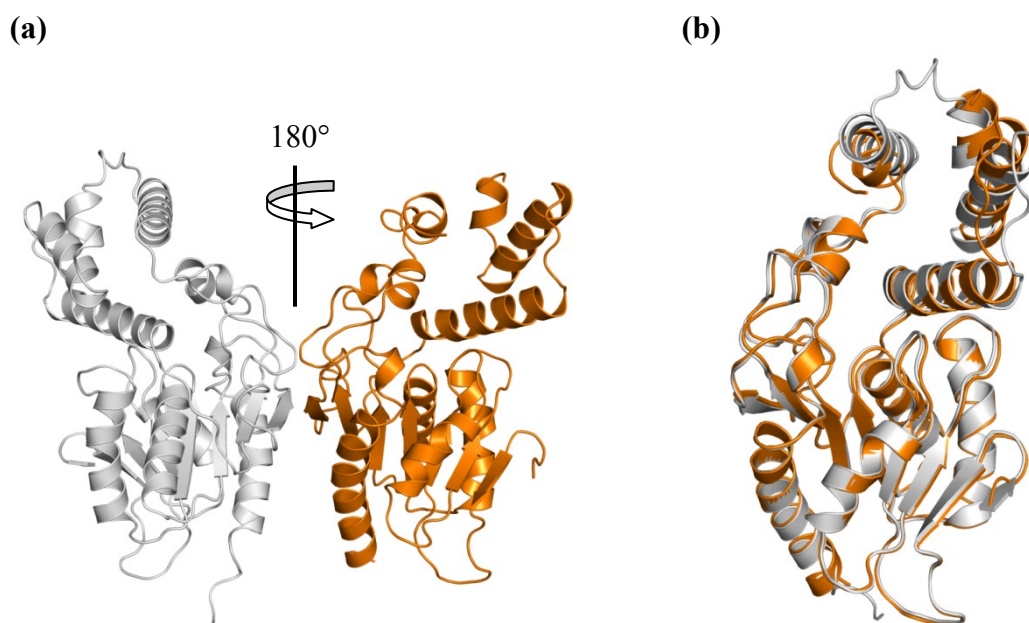


Figure 4-4. Overall structure of Pks13 TE domain. **(a)** Two molecules in the asymmetric unit (asu) are shown as cartoons with one monomer in white and the other in orange color. The two monomers are related by a two-fold rotation axis. **(b)** Superimposed structures of the two monomers in the asu showing slight differences in the flexible lid domain.

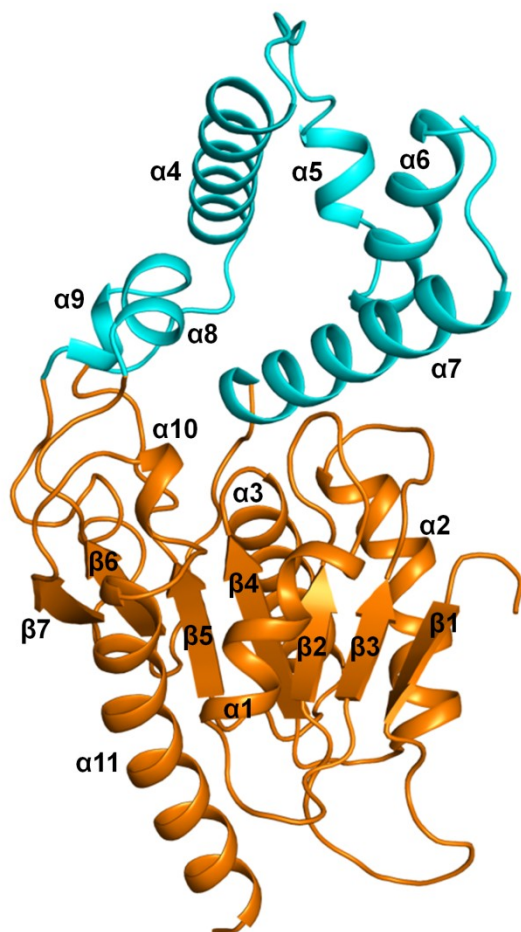


Figure 4-5. Overall fold and structural elements of the TE domain. The monomer structure is composed of two domains – lid domain (shown in cyan, comprised of helices $\alpha 4$ - $\alpha 9$) and the core domain (shown in orange, comprised of a seven-stranded β -sheet and helices $\alpha 1$ - $\alpha 3$, $\alpha 11$). The topology of the TE domain in the TE1451-1 complex structure is very similar to the apo-enzyme structure (RMSD of 0.94 Å over 272 paired C_{α} atoms).

Active site and binding mode of 1

Active sites of all α/β -hydrolase thioesterases contain a canonical Ser-His-Asp catalytic triad in a conserved topology (Nardini and Dijkstra, 1999). Superimposition of TE1451 structure on *E. coli* EntF structure (PDB entry 3TEJ) helped us to identify Ser1533 as the active site nucleophile, and Asp1560 and His1699 as the other two members of the catalytic triad (Fig. 4-6).

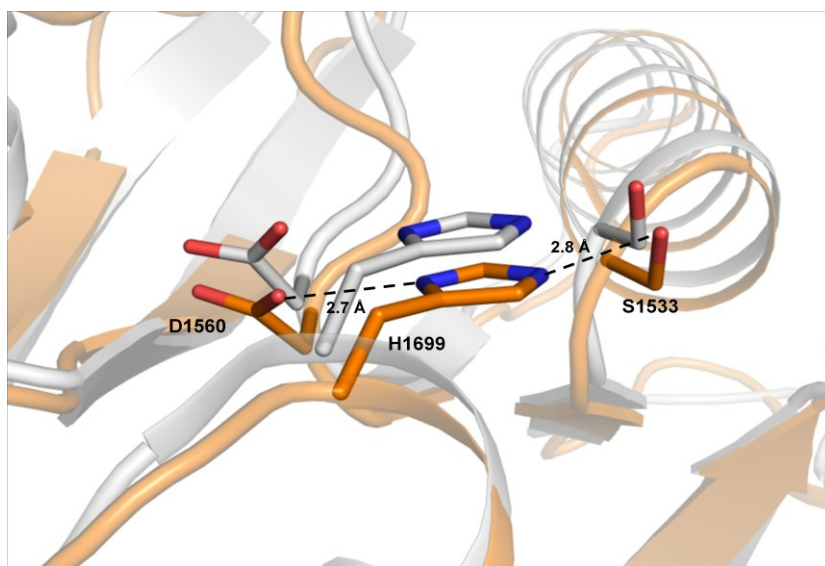


Figure 4-6. Identification of the catalytic residues of TE1451. Structural alignment of TE domain (in orange) with the *E. coli* EntF (white, PDB code 3TEJ) structure showing conserved catalytic residues Ser1533, Aps1560 and His1699. The hydrogen bonding network between the catalytic residues is shown by black dashed lines. The nitrogen and the oxygen atoms are shown in blue and red colors, respectively.

The active site pocket is formed at the interface of the two domains and it is situated at the proximal end of a long surface groove on the lid domain. This groove spans the full length of the lid domain (~ 30 Å) with a total surface area ~ 1290 Å², as calculated using the CASTp server (Dundas et al., 2006) (Fig. 4-7). The residues forming this hydrophobic groove are primarily contributed by the structural elements of the lid domain. The bottom of the groove is lined by residues from helices $\alpha 4$, $\alpha 5$, $\alpha 7$, $\alpha 8$, and the loop connecting $\alpha 4$ to $\alpha 5$, namely Trp1579, Tyr1582, Ile1597, Leu1602, Tyr1637 and Phe1670. Residues that form the wall of the groove are contributed by helices $\alpha 4$ (Phe1585, Ala1586, Phe1590), $\alpha 6$ (Val1614, Val1618) and $\alpha 8$ (Met1669) along with Ile1594 from the loop connecting end of $\alpha 4$ to $\alpha 5$. A similar hydrophobic surface groove (~ 20 Å long) leading to the active site serine was also observed on the all α -helical lid domain in the crystal structure of bovine palmitoyl-protein thioesterase 1 bound to palmitate (Bellizzi et al., 2000).

In the apo-TE domain structure, we observed a fragment of the additive polypropylene glycol (C₁₂O₅H₂₅, Fig. 4-8) bound in the active site. Thus, the structure strongly suggests that the surface groove presents the substrate binding site that can accommodate the C₂₆ α -chain of mycolic acid and leads into the catalytic pocket formed at the domain-domain interface.

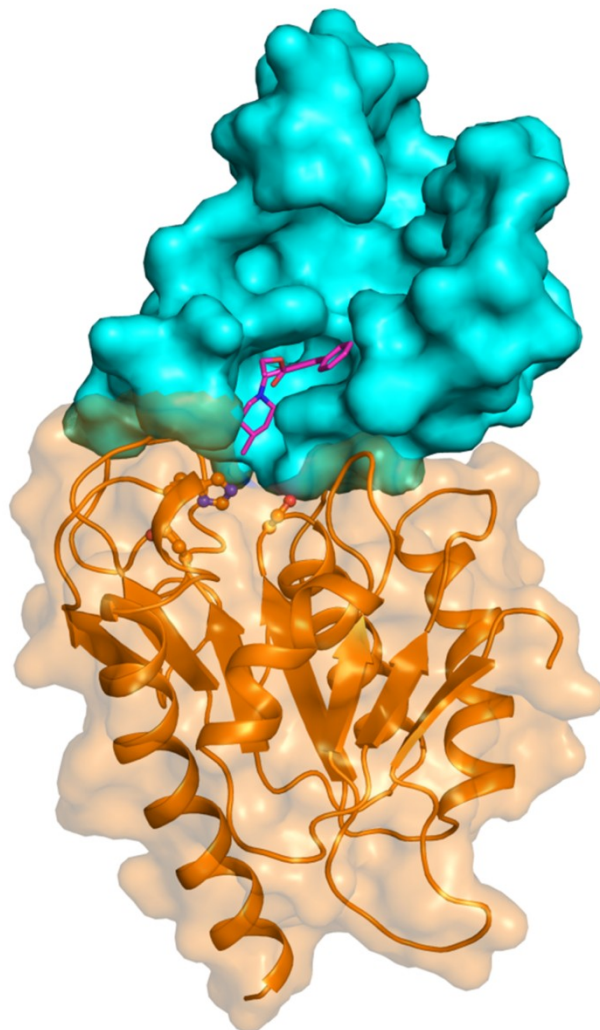


Figure 4-7. Surface groove on the lid domain in TE1451 structure. TE domain in the TE1451-1 complex structure is shown in molecular surface representation depicting the surface groove on the lid domain. The colors reflect the domain division as in Fig. 4-5. Compound **1** is shown as sticks in magenta color. The active site residues His1699 and Ser1533 at the interface of the lid and the core domains are shown as ball and stick. The nitrogen and the oxygen atoms are colored blue and red, respectively.

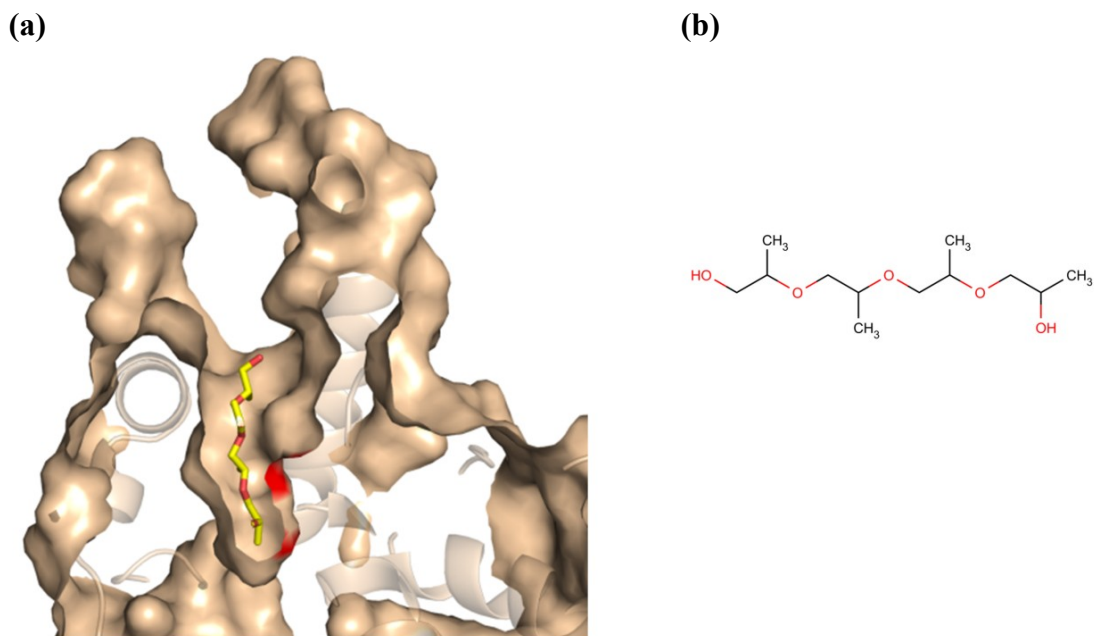


Figure 4-8. Polypropylene glycol binding in the TE1451 active site. (a) Vertical cross-sectional view of the active site pocket and substrate binding groove on the lid domain of TE1451 showing a fragment of the additive polypropylene glycol (yellow sticks) bound near the catalytic Ser1533. The oxygen atoms are shown in red color. (b) Chemical structure of the bound PPG fragment.

Crystals of TE1451 complexed to **1** were obtained by soaking the apo-TE1451 crystals with **1**. The complex structure was determined by MR to a resolution of 2 Å (space group P2₁2₁2 and unit cell similar to apo-TE1451) with 273 and 274 residues built out of a total of 283 residues for molecules A and B, respectively, and refined to a final R_{work} = 20% and R_{free} = 24% (Table 4-1). **1** binds at the mouth of the substrate binding groove close to the opening to the catalytic site, and thus, binding of the inhibitor effectively blocks access of the substrate to the active site (Fig. 4-7). The four different side groups attached to the benzofuran core of **1** interact mainly with the residues from helix α7 of the lid domain (Gln1633, Ser1636, Tyr1637, Asn1640, Arg1641, Ile1643 and Asp1644) and the two supporting helices α8-α9 along with the loop that connects them to strand β6 of the core domain (Tyr1663, Ala1667, Phe1670, Glu1671 and Tyr1674) (Table 4-5). There are several changes that occur upon binding of **1**, most notably the flipping of Phe1670 located at the end of helix α8 to form a planar stacking interaction (3.6-3.7 Å) with the five membered ring of the benzofuran and hydrophobic interactions with the core phenyl ring (Fig. 4-9). Phe1670 side chain also forms van der Waals interactions with the R₁ phenyl ring and the ethyl group of the ethyl ester attached to the R₃ position of **1**, respectively (~ 4 Å). The side chains of Tyr1582 and Tyr1637 also participate in van der Waals interactions with the phenyl ring of the benzofuran core (3.7-3.8 Å).

Table 4-5. Interactions of compound **1** with the residues of TE domain.

TE domain residue	Type of interaction	Interacting group of 1
Phe1670	Planar stacking, hydrophobic	Benzofuran core
Asp1644	Hydrogen bond (2.4 Å)	-OH at C-5
Asn1640	Hydrogen bond (3 Å)	Basic N of piperidine at C-4
Gln1633, Ser1636	van der Waals	Phenyl at C-2
Tyr 1674, Tyr1663, Ala1667	van der Waals	Piperidine ring at C-4
Glu1671	van der Waals	-OH at C-5
Tyr1582, Tyr1637	van der Waals	Phenyl ring of core
Arg1641, Ile1643	Hydrophobic	Phenyl ring of core and methyl of piperidine

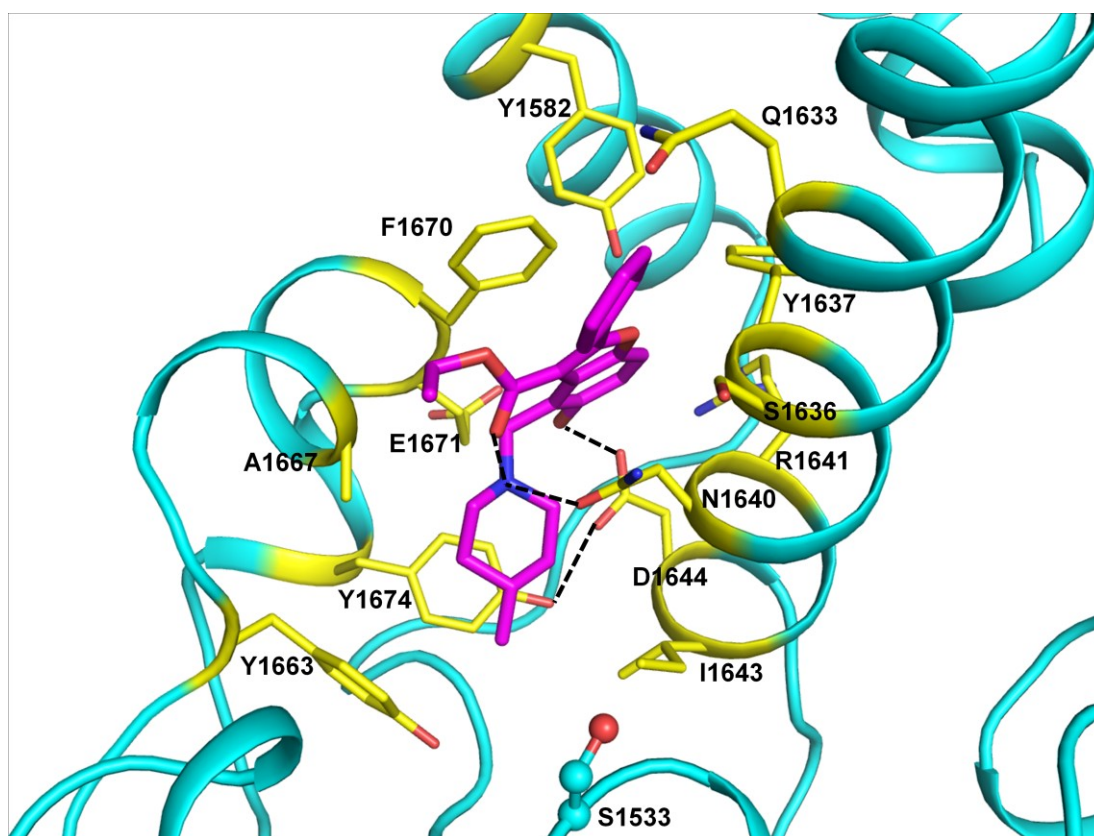


Figure 4-9. Cartoon representation of the binding interactions of **1** with the TE lid domain. **1** is shown as sticks with carbons in magenta color and the residues of the lid domain that interact with **1** are shown as sticks with carbons in yellow color (other atom colors are: red, oxygen; blue, nitrogen). Hydrogen bond interactions are shown as black dashed lines with the distances ranging from 2.4 to 3.3 Å. The active Ser1533 is shown as ball and stick.

One of the key interactions for **1** binding involves Asp1644 which is located at the end of helix $\alpha 7$ of the lid domain. The carbonyl oxygen of Asp1644 forms a strong hydrogen bond (Jeffrey, 1997) (2.4 Å) with the R₄ hydroxyl of **1**. In addition to the direct interactions with **1**, a carboxylate oxygen at Asp1644 also forms hydrogen bond interaction with the side chain hydroxyl of Tyr1674 (2.7 Å), helping to orient it to form a stacking van der Waals interaction with the R₃ piperidine ring (Fig. 4-9). The basic N of piperidine acts as a bifurcated donor making one hydrogen bond with the side chain oxygen of Asn1640 (3 Å) and another with carbonyl oxygen of R₂ ethyl ester (3.3 Å). Other residues that participate in the van der Waals interactions with **1** are Gln1633 and Ser1636 with the R₁ phenyl ring; Tyr1663 and Ala1667 with the R₃ piperidine; and Glu1671 with the R₄ hydroxyl. Arg1641 and Ile1643 form hydrophobic interactions with the phenyl ring of the benzofuran core and the methyl group of R₃ piperidine, respectively (Fig. 4-9).

Structural basis for resistance of D1607N and D1644G mutants to **1**

The effect of D1644G mutation is direct, in that it removes the interaction with R₄ hydroxyl of **1**. The effect of D1607N mutation is less apparent, because it does not make any direct interactions with **1**. Inspection of the interactions of Asp1607 in the mutant structure (refinement statistics in Table 4-1) revealed that in the wtTE1451- **1** complex structure the side chain carbonyl oxygen atoms of Asp1607 form bidentate hydrogen bond interactions with the side chain amide group of Arg1641 (3.2 Å and 3.3 Å) which helps to anchor the C-terminal end of the helix $\alpha 7$ in a position that facilitates

Asp1644 to form hydrogen bond interaction with R₄ hydroxyl of **1**. In contrast, the D1607N mutation disrupts this hydrogen bond interaction with Arg1641 allowing it to adopt a different rotamer conformation (Fig. 4-10). This results in the loss of anchoring effect by Arg1641 causing the C-terminal end of helix α 7 to move away from the substrate binding groove. The resulting flexibility also causes Asp1644 to move away from the substrate binding groove by 3 Å, altering its hydrogen bond interactions. The movement of Asp1644 disrupts its interaction with **1**, in addition it also gets positioned within hydrogen bonding distance to Arg1641 (2.7 Å) (Fig. 4-10). Thus, the D1607N mutation introduces subtle changes in structural interactions, in contrast to the D1644G mutation, to disrupt TE domain interaction with **1** and impart resistance to the bacterium.

Structural basis for inhibition of TE1451 by analogs of **1**

The structures of TE1451-analog binary complexes were refined using the TE1451-apo structure. As expected, there was clear positive $|F_o|-|F_c|$ difference electron density for inhibitors **3-6** that was in a very similar position in the substrate binding groove. Ligands were fit into the electron density, and the structures were built and refined to a resolution of 2 Å with good stereochemistry (Fig. 4-11, Table 4-2).

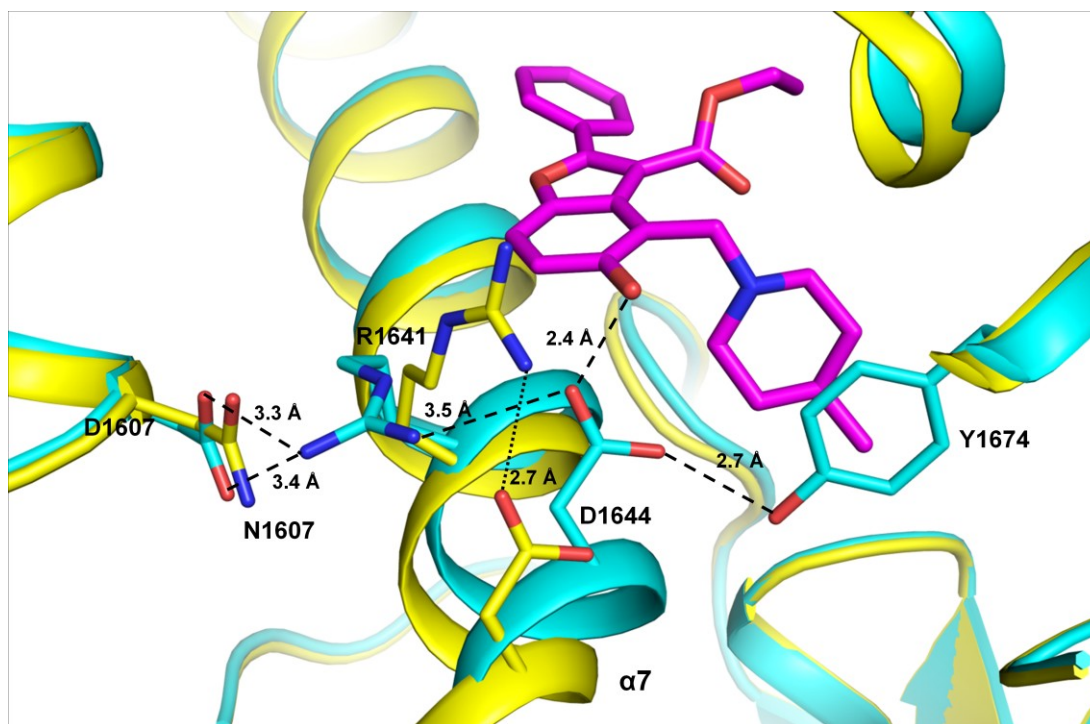


Figure 4-10. Structural basis for resistance of the TE1451:D1607N mutant to **1**. Structure of TE domain-**1** complex (TE in cyan color, **1** as magenta sticks) superimposed with the structure of D1607 mutant (yellow color) showing altered conformation of Arg1641 in the mutant. This results in an altered hydrogen bonding of Asp1644 to Arg1641 in the D1607N mutant (black dotted line). In the wt-TE structure, Asp1607 hydrogen bonds to Arg1641 that anchors C-terminal end of helix $\alpha 7$ thereby allowing Asp1644 to form hydrogen bond interaction with **1** and Tyr1674 (black dashed lines). The mutation D1607N results in a loss of hydrogen bond mediated anchoring of helix $\alpha 7$ that causes Asp1644 to move away by $\sim 3 \text{ \AA}$ which consequently disrupts its strong interaction with **1**.

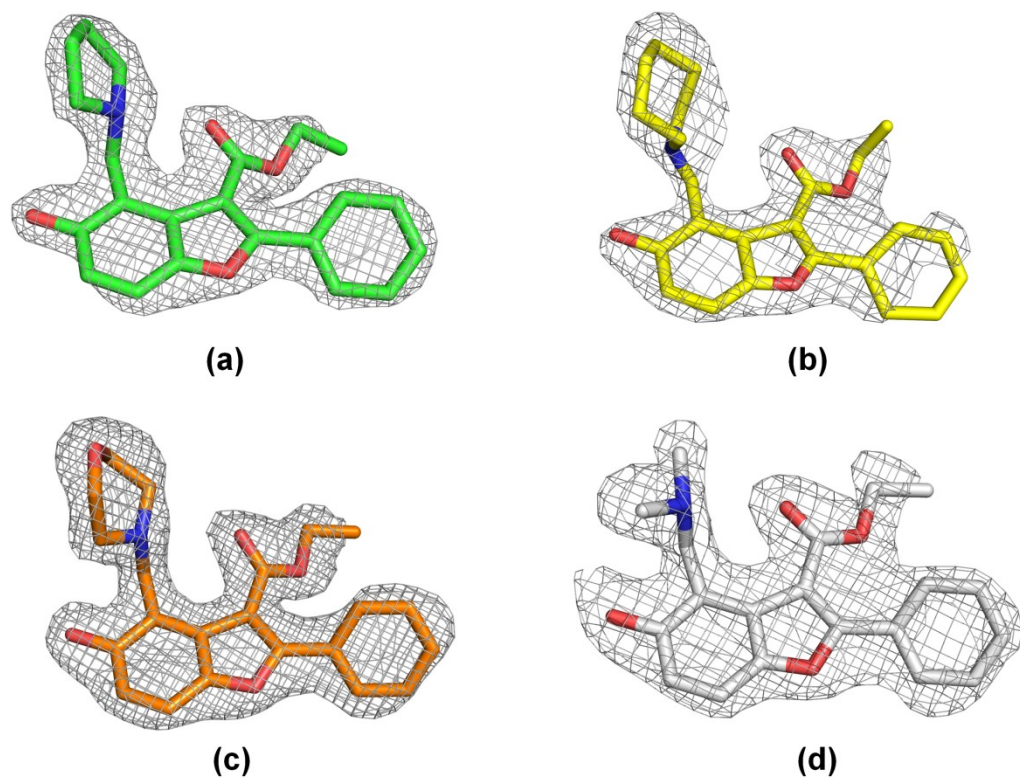


Figure 4-11. Fitting of the inhibitors in the TE1451-analog complex structures. The panels show the $2mF_o - DF_c$ maximum-likelihood omit maps contoured at 2σ (grey mesh) for the TE1451 complex structures with the following ligands: (a) analog **3**, (b) analog **4**, (c) analog **5**, and (d) analog **6**. Maps were calculated for the refined models with the ligand atoms omitted. The carbon atoms are colored green for **3**, yellow for **4**, orange for **5**, and white for **6**. The oxygen and nitrogen atoms are colored red and blue, respectively, in all the analog structures.

The protein component of all of the complexes exhibit a nearly identical overall structure as compared to the TE1451-**1** complex (RMSD of 0.4 Å for **4**, 0.7 Å for **3**, 0.9 Å for **5** and **6** vs. **1** complex over 264 paired C_α atoms). In the structures, the variation in the tertiary amine group at R₃ of the analogs did not affect their mode of binding; however, based on our biochemical data it did contribute to the variations in the efficacy of inhibition by different analogs (Table 4-4). The crystal structures of TE1451-**3** and TE1451-**4** complexes revealed that the five- and seven-membered cyclic amine groups at R₃ of **3** and **4**, respectively, formed stacking interaction with Tyr1674 in a manner similar to **1**. In addition, the non-polar rings at R₃ of **3** and **4** also maintained multiple van der Waals and hydrophobic interactions with the side chains of Ile1643 and Tyr1663 (3.5-4.1 Å), that were also formed by the piperidine ring of **1**. The conservation of interactions observed in the structures of TE1451 in complex with **1**, **3** and **4** suggests similar inhibitory potency for these compounds. Indeed, the IC₅₀ values of **3** and **4** were determined to be similar to the IC₅₀ of **1** (Table 4-4). The small differences in the IC₅₀ values among **1**, **3** and **4** can be attributed to variations in the strength of stacking interaction of the planar side chain of Tyr1674 with the variably puckered rings at R₃, similar to the sugar-ring Tyr interactions in the carbohydrate/sugar binding proteins (Weis and Drickamer, 1996) (Fig. 4-12). In the TE1451-**5** complex structure, we observed that the polar cyclic amine at R₃ also had a stacking interaction with Tyr1674, however, it formed only van der Waals interactions with side chains of Ile1643 and Tyr1663 (4-4.3 Å). Moreover, contrary to our expectations the oxygen atom of the R₃ morpholine ring of **5** did not form a hydrogen bond interaction with the hydroxyl of

Tyr1663, as its rotamer conformation positioned the hydroxyl at a distance of 4.2 Å. The effect of the reduced interactions of **5** on its inhibitory activity was reflected in ~3-fold increase in its IC₅₀ (0.71 μM) compared to **1** (IC₅₀ 0.26 μM). In contrast to the cyclic groups at R₃, the presence of an acyclic dimethyl amine group at R₃ in **6** is expected to abolish the stabilizing stacking interaction with Tyr1674, and it was confirmed in the TE1451-**6** complex structure. The structure also revealed that the smaller size of the side group at R₃ led to reduced van der Waals and hydrophobic interactions with Tyr1674, Tyr 1663 and Ile1643. As a consequence of this loss of interactions there was >6-fold decrease in the inhibitory activity of **6** against TE1451 (Table 4-4).

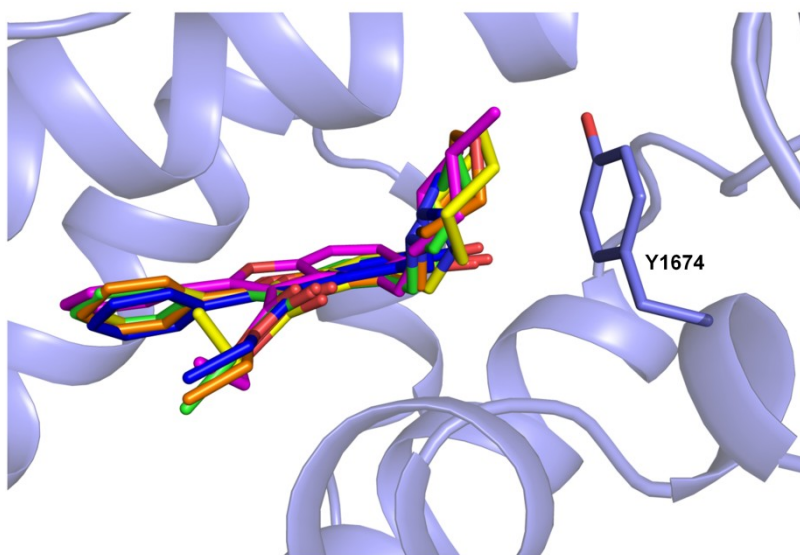


Figure 4-12. Binding mode of analogs of **1** to TE domain. Overlay of the structure of **1** with analogs **3-6** (in stick representation with **1** in magenta, **3** in green, **4** in yellow, **5** in orange and **6** in blue color) bound to TE domain as seen in the complex structures, reveals a binding mode similar to **1**. The cyclic amine substituent groups at C-4 of the benzofuran in **1** and its analogs form a stacking interaction with Tyr1674 of the TE domain.

DISCUSSION

This is the first benzofuran based inhibitor reported to be active against *M. tuberculosis*. Molecule **1** has been validated biochemically and structurally to target the TE domain of *Mtb* Pks13. Inhibition of Pks13 represents an attractive alternative target for TB therapy as targeting it would inhibit cell wall biosynthesis leading to cell death, especially in the INH resistant strains (Stanley et al., 2013). In addition, it was recently reported that a combination of Pks13 inhibitors with other inhibitors of the mycolic acid biosynthesis (for e.g. INH) can potentially eradicate persisters as well as reduce the development of resistance (Wilson et al., 2013). Our preliminary SAR using structural analogs of **1** in an inhibition assay was conducted to determine the effect of different substituent groups at R₃ of the benzofuran core on inhibitory activity of analogs. The MIC data from whole cell testing also correlated well with the IC₅₀ data regarding the effect of R₃ substitutions (Table 4-4). Notably, these compounds exhibited low cytotoxicity profiles suggesting that these could be used for structure guided development to further improve their activity profile and generate lead candidates for antitubercular drug development.

Therefore, in order to gain insights into the structural features of TE domain-**1** interaction, we determined the first crystal structure of the Pks13 TE domain -**1** complex. **1** bound non-covalently at the entrance of the catalytic chamber of the TE domain, effectively blocking access of the substrate to the active site Ser (Fig. 4-7). In contrast, inhibitors of human FAS TE domain were found to exert their inhibitory effects by

forming covalent adduct that mimics the acyl-enzyme intermediate (Pemble et al., 2007). Analysis of the binding mode of **1** and analogs suggested that the benzofuran core is critical for binding and orientation at the entrance of the active site. These structures also revealed the structural basis for our findings from the preliminary SAR studies, which indicated that the non-polar nature and planarity of the ring substituents at R₃ are the key factors for potency of the compounds. From the structures, it can be seen that a ring with more planar conformation at R₃ could form a stronger π - π interaction, rather than just van der Waals or stacking interaction with Tyr1663. We are currently pursuing design and synthesis of analogs based on these structural and the SAR studies.

In order to design inhibitors which have a lower potential for resistance development, it is also important to understand the resistance mechanisms bacteria may deploy to render the inhibitor ineffective. Structural studies on the two Pks13 TE domain mutations (D1607N and D1644G) that were identified to confer resistance to *Mtb* against **1**, revealed that both the mutations disrupted the interaction between Asp1644 and **1**, but interestingly the mechanism of disruption was found to be indirect and more subtle in the D1607N mutant. It is plausible that by increasing the flexibility of helix α 7 in D1607N mutant, the bacterium is able to reduce the interaction and as a result the affinity for **1** just enough for the substrate to have a preferential binding, without distorting structural features of the substrate binding groove and hence preserving the substrate affinity. In contrast, the D1644G mutation has a direct effect on α 7, which not only forms a part of the substrate binding groove, but may also interact with the phosphopantetheine arm (pPant) of the ACP domain for positioning the

substrate in the groove. In other PKS systems, TE domain interactions with the pPant arm of ACP domain have been reported to facilitate recognition and proper positioning of the substrate for catalysis (Xu et al., 2013). Although the *in vitro* activity and especially IC₅₀ for the D1644G mutant was much enhanced in comparison to the wild type TE, remarkably it did not translate into a corresponding *in vivo* advantage for the bacterium, as indicated by the 2-fold higher MIC observed for the D1607N mutant strains compared to the D1644G mutants (Ioerger et al., 2013).

In conclusion, our work presents the first report of elucidation of the structural basis for the mechanism of inhibition of Pks13, an essential type 1 PKS from *Mtb* that presents a promising candidate for anti-*Mtb* drug discovery (Wilson et al., 2013). The structural and biochemical analysis performed in this study validates the TE domain of Pks13 as the target of a novel inhibitor **1**. Further, analysis of the structures of the TE domain in complex with **1** and its analogs has revealed important details on structural interactions that will aid in the design and development of more potent inhibitors. In addition, we have also determined the structural basis for the mechanism of resistance towards **1**. Our work provides a foundation for initiating structure-based studies and development of inhibitors that target a novel enzyme in the essential mycolic acid pathway.

ADDITIONAL UNPUBLISHED RESULTS

Extended SAR with the newly synthesized analogs of **1***

The results from our preliminary SAR studies with the analogs of **1** (Table 4-4) had suggested that the replacement of the R₃ piperidine ring with a planar phenyl ring could lead to a stronger π - π stacking interaction with Tyr1674 in the binding pocket of TE domain (Fig. 4-12). To test this hypothesis, we synthesized an analog of **1** (MKP-116L) with a phenyl ring in place of the R₃ piperidine ring, and evaluated its inhibitory activity. Contrary to our expectations, the incorporation resulted in loss of activity of MKP-116L by >40-fold compared to **1** (Table 4-6). Since the inhibitors with puckered tertiary amine rings at R₃ position showed good activity we have synthesized and tested an analog with a R₃ cyclohexyl group to investigate the effect of ring N-atom. The cyclohexane ring of this analog (MKP-126) has a puckered chair conformation similar to the piperidine of **1**. However, this analog showed a higher IC₅₀ (20 μ M) compared to the phenyl ring containing analog MKP-116L (IC₅₀ 12 μ M), resulting in a loss of activity by >70-fold when compared to the original hit compound **1** (Tables 4-4 and 4-6). These results indicate that the presence of a non-tertiary amine ring either affects the conformation of the inhibitor, or results in an alteration in the interactions of the inhibitor with the binding pocket residues of the protein. This suggested that intramolecular hydrogen bond between the basic N-atom of piperidine and carbonyl oxygen of R₂ ethyl ester might be important for maintaining the conformation, and

* The analogs were synthesized by Maloy Parai, a post-doctoral research associate in the Sacchettini lab at TAMU.

hence, the mode of binding of the ring at R₃. Indeed, when the 4-methyl piperidine ring was replaced at R₃ in the analog MKP-120, the analog showed an IC₅₀ (0.3 μM) similar to the original hit compound **1** (IC₅₀ 0.26 μM), confirming that the tertiary amine is critical for the potency of the compounds. Interestingly, MKP-120 displayed ~5-fold lower MIC in the whole cell assays compared to **1** (Table 4-6). Since MKP-120 does not have the *para*-methyl group on the piperidine ring as **1**, these results indicate that the methyl group does not make a significant contribution to the potency; however, its removal enhanced the whole cell activity. This finding has been further supported by the IC₅₀ data from the analog AA-7 which contains an N-methyl piperazine group at R₃ (Table 4-6). The R₃ substituent of AA-7 shares a similar puckered conformation with the piperidine of **1**, but has a more polar ring. The IC₅₀ of AA-7 was found to be 2.7 μM, which is ~10-fold higher than the IC₅₀ of **1** (Tables 4-4 and 4-6). This finding is consistent with the results obtained in the preliminary SAR from the testing of analog **5** which had a morpholine ring at R₃ and displayed a ~3-fold higher IC₅₀ compared to **1** (Table 4-4). Overall, the results from these SAR studies of the R₃ substituents suggest that the piperidine ring at R₃ is critical for the potency of the compounds.

Table 4-6. Pks13 TE domain SAR of the synthesized benzofuran analogs of **1**.

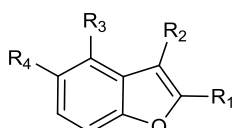

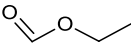


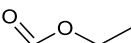
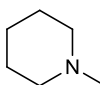

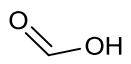
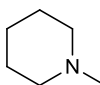

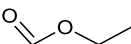
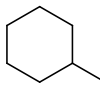

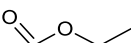
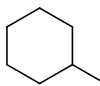

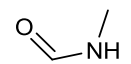
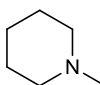

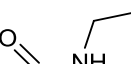
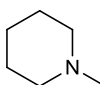
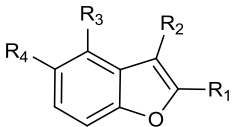

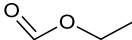
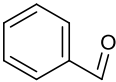

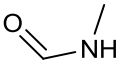
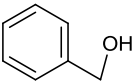
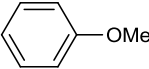
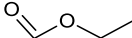
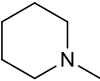
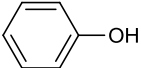
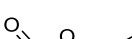
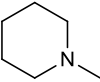

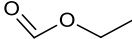
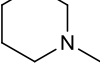

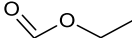
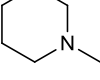

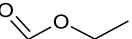
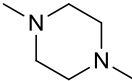
						
Compound	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)*	MIC (μM)
MKP-116L				OH	11.9 ± 2.3	>40
MKP-120				OH	0.26 ± 0.04	0.4
MKP-121				OH	6.6 ± 0.7	NI
MKP-125				MeO	NI	ND
MKP-126				OH	19.6 ± 1.4	5.2
MKP-128				OH	0.29 ± 0.01	0.2
MKP-130				OH	0.34 ± 0.02	14

Table 4-6. Continued.

						
Compound	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)*	MIC (μM)
MKP-133				MeO	>50	ND
MKP-138B				MeO	>50	ND
MKP-148				OH	0.28 ± 0.01	2.6
MKP-149				OH	0.17 ± 0.02	1.2
MKP-150				MeO	35.8 ± 2.2	ND
MKP-152				H	2.0 ± 0.1	
AA-7				OH	2.7 ± 0.1	1.4

IC₅₀ values were determined using the *Mtb* Pks13 TE domain as described in the methods section. MIC values were determined for mc²7000 in liquid medium in 96-well plates. MeO, methoxy; ND, not determined; NI, no inhibition. *Values are shown as mean ± s.d. of triplicates.

We next performed SAR on the potentially labile R₂ ethyl ester of the inhibitor **1**. Since the ethyl ester can be hydrolyzed to corresponding acid by the non-specific esterases inside the mycobacterial cells, it is plausible that the acid form of **1** is the active compound. This would suggest that the ester represents the prodrug form of **1** which upon cellular uptake gets converted to an active acid form. We tested this hypothesis by synthesizing the acid analog (MKP-121) of MKP-120 and evaluating it for enzyme inhibition and whole cell activity. MKP-121 had an IC₅₀ of 6.6 μM, a ~20-fold loss in activity compared to MKP-120; moreover its whole cell activity was completely abolished (Table 4-6). The MIC data are consistent with the reported impermeability of the *Mtb* cell wall for the acidic compounds (Brennan and Nikaido, 1995). Also, in a study with the phenyl-diketo acid inhibitors of the *Mtb* malate synthase, it was shown that masking of the acid group via esterification was a good strategy to enhance the whole cell uptake of the inhibitors (Krieger et al., 2012). However, even though the lack of whole cell activity for MKP-121 could be attributed to its reduced permeability into the cells, the IC₅₀ data clearly established that MKP-121 was not a potent inhibitor (Table 4-6). Since the R₂ ethyl ester could also be subjected to hydrolysis by the serum esterases (Tsaion and Kates, 2011), it still represented a metabolic liability that could lead to the production of the less active acid form of the inhibitor. Therefore, we synthesized analogs of MKP-120 that contain methyl- (MKP-128) or ethyl-amide (MKP-130) functionality at R₂ (Table 4-6). Both methyl- and ethyl-amide groups are metabolically more stable and they provide a similar isoelectronic structure as the ethyl ester. Both MKP-128 and MKP-130 showed similar enzyme inhibitory activity (IC₅₀

0.3-0.4 μM) as the ethyl ester containing MKP-120 (IC_{50} 0.3 μM), which indicated that methyl- and ethyl-amide groups maintain the inhibitor potency, while at the same time they provided the desired metabolic stability at R_2 . We determined their mode of interaction with the TE domain by solving the crystal structure of TE1451 in complex with MKP-128. The crystal structure of TE1451-MKP-128 complex was nearly identical to the TE1451-**1** complex structure, and it showed that MKP-128 bound to TE1451 in a fashion similar to **1** (Fig. 4-13).

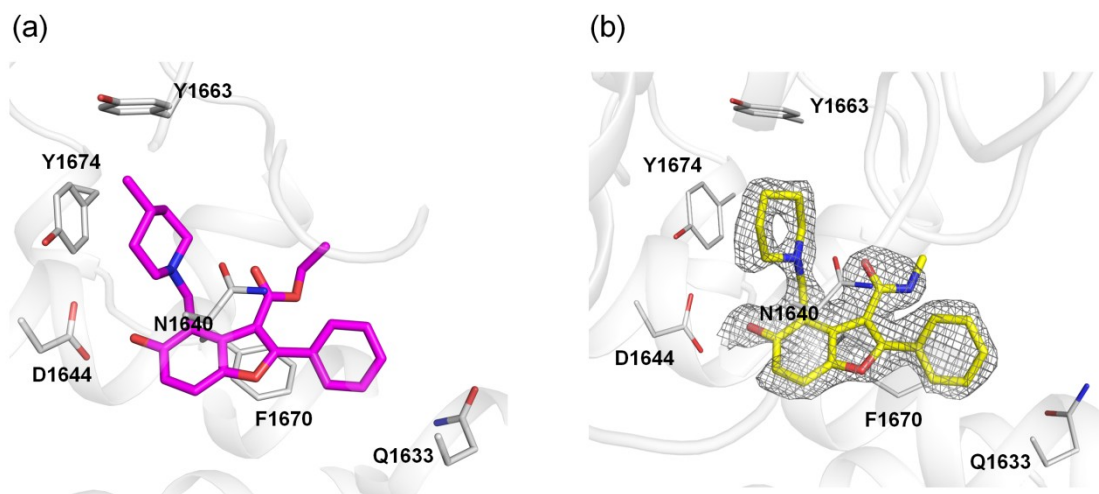


Figure 4-13. Binding mode of methylamide analog MKP-128. (a) Binding mode of **1** (shown as magenta sticks). (b) Binding mode of MKP-128 (shown as yellow sticks) to the Pks13 TE domain. The grey mesh represents the $2mF_o - DF_c$ maximum-likelihood omit map contoured at 2σ . Map was calculated for the refined model with the ligand atoms omitted. The carbon atoms are colored magenta for **1** and yellow for MKP-128. The oxygen and nitrogen atoms are colored red and blue, respectively, in both the structures.

All the major interactions of MKP-128 with the TE domain are very similar to inhibitor **1** complex structure including the H-bonding interactions with Asp1644 and Asn1640, and the stacking interactions with Tyr1674 and Phe1670, which provided the structural basis for the similar *in vitro* potency. When MKP-128 was tested for the whole cell activity, its MIC (0.2 μ M) was found to be >10-fold lower than that of **1** (2.3 μ M). In addition, it was also 2-fold lower than the MIC of MKP-120, suggesting that substitution of the ester with the amide at R₂ had also enhanced whole cell activity (Table 4-6). Therefore, the R₂ SAR data suggests that the replacement of the more labile ethyl-ester with the more stable amide group not only preserved the potency of the parent compound but also improved its whole cell activity.

The phenolic OH at the R₄ of the benzofuran core was found to make a strong hydrogen bond to the side chain carbonyl of Asp1644 (distance 2.4-2.5 Å) of the TE domain in the crystals structures of **1** and its analogs. Since this R₄ OH has a potential for oxidation, we performed a SAR with different substituents at this position to determine its role in inhibitor binding. Two analogs of MKP-120 were synthesized in which the R₄ OH was substituted with either a methoxy group (MKP-150) or a hydrogen atom (MKP-152) (Table 4-6). The IC₅₀ for MKP-150 was determined to be ~36 μ M, whereas it was 2 μ M for MKP-152. Therefore, removal of the hydrogen bonding capability in MKP-150 caused a >100-fold decrease in its inhibitory activity. Some contribution to this loss of activity in MKP-150 could also be attributed to the presence of the bulky methoxy group at R₄, since it can have some steric clash with the side chains of Asp1644 and Glu1671 (distances 2.4 and 3.3 Å, respectively). This inference is


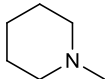
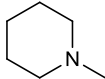
further supported by the inhibition data on MKP-152, which showed only a modest ~7-fold decrease in activity due to substitution of the R₄ OH with just a hydrogen atom. Thus, the SAR data from the R₄ substitutions suggests that a hydrogen bond donor is required at that position for the potency of the inhibitor. In addition, analysis of the TE1451 complex structures with inhibitors suggests that the OH can be moved to the C-6 position on the benzofuran core which could preserve the current hydrogen bonding pattern while eliminating the oxidation liability associated with the R₄ OH. We are currently pursuing design and synthesis of new analogs based on these findings.

In the TE1451-1 complex structure, we observed that the R₁ phenyl substituent is exposed to the solvent and does not make any interactions with the protein. However, the structure also indicated that the side chain carbonyl and amide groups of the Gln1633 residue on the helix α 7 are positioned at a distance of 3.3 and 4.4 Å, respectively, from the C-4 of the R₁ phenyl substituent (Fig. 4-9). This suggested that incorporation of either a hydrogen bond donor or a hydrogen bond acceptor group at the C-4 of the R₁ phenyl would allow an additional hydrogen bond interaction of the inhibitor with Gln1633, which could improve the potency of the inhibitor. To address this point, we synthesized two analogs of MKP-120 containing a methoxy (MKP-148) and an OH (MKP-149) group at the C-4 of the R₁ phenyl ring (Table 4-6). The IC₅₀ of the methoxy substituent (MKP-148) was found to be similar to MKP-120 (Table 4-6), but the OH substituted compound (MKP-149) exhibited a slightly better IC₅₀ (0.19 μ M) compared to the non-OH substituted MKP-120 (IC₅₀ 0.3 μ M). We are currently pursuing synthesis of

compounds with other H-bond donor and acceptor substituents at the C-4 of the R₁ phenyl group. These will then be evaluated to identify the most suitable substitution.

In addition to the structure-guided development of the analogs of the benzofuran **1**, we also performed a pharmacophore search to identify other structural mimetic of the benzofuran core. We identified and tested a benzimidazole derivative, AA-8 with a phenyl, a piperidine and an OH group substituted at the R₁, R₃ and R₄, respectively, similar to the benzofuran **1** (Table 4-7). However, AA-8 did not have any substituent at R₂. Interestingly, AA-8 exhibited moderate inhibition of TE1451 with an IC₅₀ of ~6.8 μM. We elucidated the structural basis for its inhibitory activity by solving the crystal structure of the TE1451-AA-8 complex. The structure showed that AA-8 bound to TE1451 in a pose similar to **1** (Fig. 4-14). The benzimidazole core of AA-8 is co-planar with R₁ phenyl ring, and forms a stacking interaction with Phe1670 similar to the benzofuran core of **1**. In addition, the structure showed that similar to **1**, AA-8 also formed hydrogen bonding interactions with Asp1644 and Asn1640 through its R₄ OH and the R₃ piperidine, respectively. Furthermore, the piperidine also formed the stacking interaction with Tyr1674, similar to the piperidine of **1** (Fig. 4-14). Thus, the structure showed that the benzimidazole AA-8 has a binding mode similar to **1**. We also tested an analog of AA-8 that contains a methyl group substituted at the R₁ position (AA-9). This analog showed an IC₅₀ of >100 μM, suggesting that the R₁ phenyl is critical for inhibitor binding. Thus, the AA-8 structure will be used as a base for designing future analogs.

Table 4-7. Pks13 TE domain SAR of the benzimidazole analogs of **1**.

Comp- ound	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)*	MIC (μM)
AA-8				OH	6.8 ± 0.9	~22
AA-9	CH ₃			OH	>100	ND

IC₅₀ values were determined using the *Mtb* Pks13 TE domain as described in the methods section. MIC values were determined for mc²7000 in liquid medium in 96-well plates. ND, not determined. *Values are shown as mean ± s.d. of triplicates.

Screening of the whole cell active compounds for identification of novel inhibitors

In addition to our structure-guided inhibitor development, we also screened TE1451 against a collection of 560 compounds that were identified from an in-house library of 51K compounds to be whole cell active against *Mtb* strain mc²7000, to identify inhibitors with novel chemical scaffolds. The screening was performed as detailed in the Methods section, and the initial screening identified nine compounds that showed a >40% inhibition of the enzyme activity compared to the DMSO only control (Table 4-8).

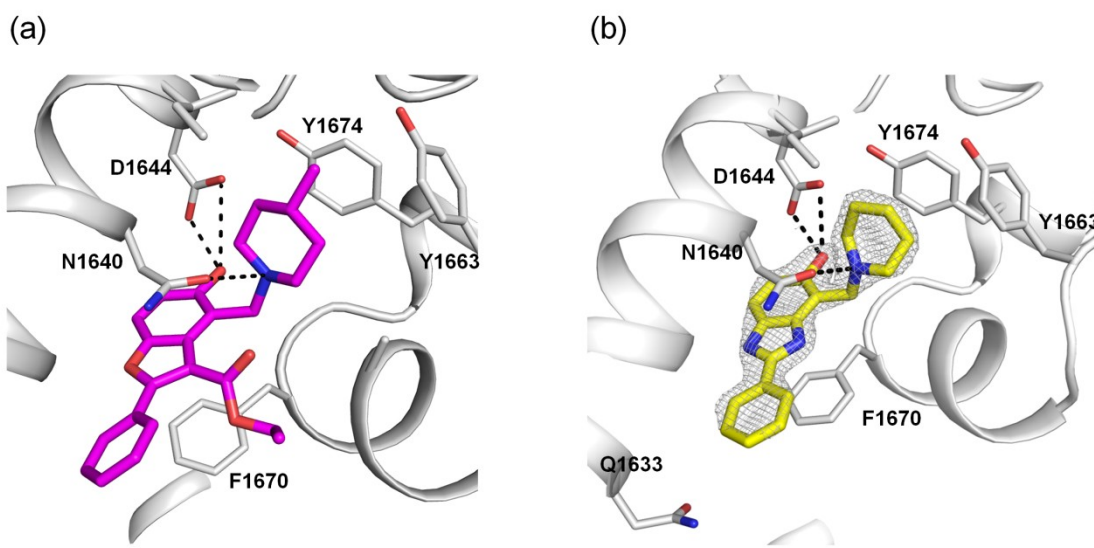
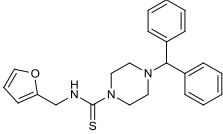
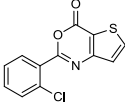
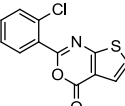
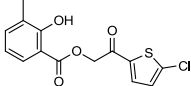
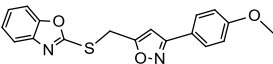
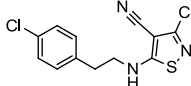
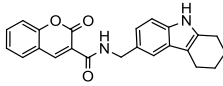
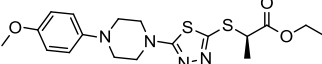
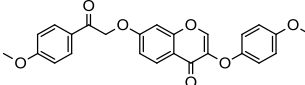


Figure 4-14. Binding mode of benzimidazole analog AA-8. (a) Binding mode of **1** (shown as magenta sticks). (b) Binding mode of AA-8 (shown as yellow sticks) to the Pks13 TE domain. The grey mesh represents the $2mF_o - DF_c$ maximum-likelihood omit map contoured at 2σ . Map was calculated for the refined model with the ligand atoms omitted. The carbon atoms are colored magenta for **1** and yellow for AA-8. The oxygen and nitrogen atoms are colored red and blue, respectively, in both the structures. The hydrogen bonding interactions are shown as black dashed lines.

These initial nine hits were then evaluated in a dose-response assay to determine their IC_{50} values (Table 4-8). Some of the compounds showed good inhibitor activity with the IC_{50} values ranging from 3-1.6 μ M (Table 4-8), and these will be used for co-crystallization trails to determine the structural basis of inhibition.

Table 4-8. Hits from the whole cell actives screening of TE1451.

Compound ID	Structure	IC ₅₀ (μM)*
1:G11		34.2 ± 1.7
3:F8		13 ± 1
3:F9		2.1 ± 0.2
3:H3		>100
4:E11		30 ± 2
5:E8		1.6 ± 0.2
5:H5		>50
6:F9		3 ± 0.3
6:E10		>25

IC₅₀ values were determined using the *Mtb* Pks13 TE domain as described in the methods section. *Values are shown as mean ± s.d. of triplicates.

SUMMARY

The mycolic acid biosynthesis pathway is well-established to be an essential pathway for survival of *M. tuberculosis*. Many of the clinically-used antitubercular drugs in the current TB therapy target enzymes in this pathway, including the first-line drug isoniazid. However, resistance to isoniazid is increasingly observed and is currently estimated at around 5% worldwide (including both monoresistance as well as MDR/XDR-TB). Therefore, it is critical to identify novel druggable targets in this validated essential biosynthetic pathway. In this study, we have performed biochemical and structural validation of the *Mtb* Pks13 thioesterase (TE) domain as the target of benzofuran-based whole-cell active inhibitors. We determined the binding mode of these inhibitors by solving the crystal structures of the Pks13 TE domain in complex with benzofuran **1** and its analogs. These preliminary SAR studies identified structural requirements for active inhibitors against the Pks13 TE domain. In addition, we also determined the structural basis for the mechanism of resistance toward **1** by solving the structures of the mutant forms of the TE domain. To develop lead compounds from the initial hits, we synthesized a number of analogs of **1** which were substituted with different functionalities at the R₁, R₂, R₃ and R₄ positions on the benzofuran core, and performed extensive SAR studies with these analogs. These studies revealed that for the potent inhibitor activity a cyclic tertiary amine group is required at the R₃ position which participates in a stacking interaction with Tyr1674. The R₄ OH is another important substituent that has been found to be critical for the potency of the

compound as it participates in a strong H-bond interaction with Asp1644. All of the findings from our extensive SAR are summarized in Figure 4-15. Based on the information obtained from the SAR studies, we have developed two lead compounds, MKP-128 and MKP-149, which exhibit 8- to 12-fold better whole cell activity compared to the original hit compound **1**, while maintaining a similar *in vitro* potency. These will be subjected to pharmacokinetic and animal model studies for evaluation as potential drug development leads.

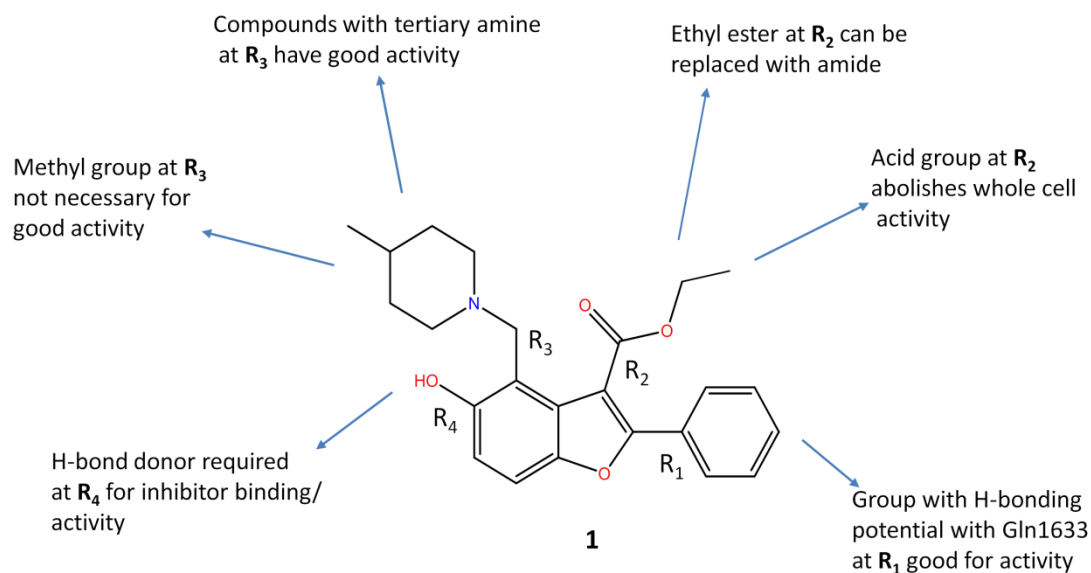


Figure 4-15. Summary of the SAR studies from the analogs of **1**.

CHAPTER IV

CONCLUSIONS

The status of *M. tuberculosis* (*Mtb*), which causes TB in humans, as one of the deadliest and most successful of all human pathogens prompted the World Health Organization to declare it a global health emergency in 1993. Even though a successful treatment regimen exists which can cure over 95% of the infected individuals, TB continues to be one of the highest burden infectious disease worldwide. In addition, emergence of multi-drug resistant and extensively drug resistant TB threatens to jeopardize the global efforts to contain the TB pandemic. Therefore, there is an urgent need for the development of new TB drugs that are effective against drug resistant *Mtb* and can help shorten the treatment duration. To discover new anti-tubercular drugs, it is critical to identify novel targets in either pre- or newly validated essential pathways in *Mtb*, and determine the molecular basis for their function. Structural studies of these proteins can provide important mechanistic insights into their function which can help guide structure-based inhibitor discovery against these targets. Here, we have employed various X-ray crystallographic techniques for the structural characterization of three different proteins that are involved in the complex lipid synthesis pathways in the *Mtb*.

Chapter I of this dissertation presented essential facts about TB and its etiological agent *Mtb*. The architecture of the cell envelope, which is essential for the survival and virulence of *Mtb*, was described in detail. Next, details on the distribution of different

mycolic acids, which are the most abundant lipid of the *Mtb* cell envelope, were presented and their roles in the mycobacterial physiology and pathogenesis were described. The essential pathway of mycolic acid biosynthesis with its novel drug targets was also described in detail. In addition, a general description of all of the polyketide synthases in *Mtb* was presented, along with the details on the different methyl-branched fatty acids containing lipids of the cell envelope. Also, the biosynthetic pathway of the phthiocerol dimycocerosate esters, the most abundant non-covalently attached lipids of the *Mtb* cell envelope, was presented in detail. Lastly, the mechanism for the activation of the acyl carrier protein domains of the PKS and the FAS systems in *Mtb* was described.

Chapter II described the purification and initial crystallographic characterization of the large 225 kDa multifunctional methyl-branched fatty acid synthesizing MAS enzyme. The development of the methods for the successful cloning, expression and purification of the phosphopantetheinylated *holo*-MAS from *E. coli* was described in detail. The recombinantly expressed MAS was determined to be a dimer in solution. A screening protocol that utilized heterogeneous nucleating agents for obtaining first crystallization hits of MAS was developed and described. To improve the diffraction properties of the MAS crystals, extensive post-crystallization treatment protocols were developed and described in detail. Initial crystallographic characterization indicated that MAS crystallized in the tetragonal space group with four to five molecules in the asymmetric unit. The results of this study suggest that the methods reported in this

chapter could prove useful for initiating crystallographic studies on other PKS I systems in *Mtb*.

In the Chapter III, structural studies on acyl carrier protein synthase (AcpS) from *Mtb* and *C. ammoniagenes* were presented. We determined the crystal structures of the AcpS at different pH which revealed that AcpS adopted two different conformations – open and closed. At a lower pH (4.0 to 6.0) the helix $\alpha 2$ of *Mtb* AcpS was found to be well ordered, whereas at higher pH values it underwent structural rearrangement that led to the destabilization of the connecting loops at its ends. The structure showed that a change in the pH disrupted ionic interactions of the helix $\alpha 2$ leading to its open conformation. Furthermore, using *in vitro* reconstitution assays we determined that the *Mtb* AcpS had a higher modification activity between pH 4.4 to 6.0 than at pH above 6.5, in agreement with the structural data. The structure also revealed the active site architecture and helped us to identify Arg48, His49 and His116 as the critical residues. This information will be useful for structure-based inhibitor discovery against AcpS which represents a good target for simultaneous disruption of several lipid biosynthesis pathways in *Mtb*. To this end, we have developed a differential scanning fluorimetry based assay for fragment-based ligand discovery to identify novel chemical scaffolds that can be developed as leads.

In Chapter IV, we described the biochemical and structural validation of the *Mtb* Pks13 thioesterase (TE) domain as the target of benzofuran-based whole-cell active inhibitors. Pks13 is the central condensing enzyme in the mycolic acid biosynthesis pathway in *Mtb* and represents an attractive alternate target in this essential pathway for

anti-*Mtb* drug discovery. We determined the binding mode of the inhibitors by solving the crystal structure of the Pks13 TE domain in complex with several benzofuran-based inhibitors. The crystal structure revealed that the TE domain belonged to the α/β -hydrolase family of enzymes. The structures allowed us to determine the binding mode of the inhibitors and elucidate the structural basis for their mechanism of inhibition. We determined that the critical interactions for the binding and the potency of the inhibitors included stacking interactions with Tyr1674 and Phe1670, and hydrogen bond interactions with Asp1644 and Asn1640. Furthermore, we also described the structural basis for the mechanism of resistance towards these inhibitors. Based on the results from our preliminary structure-activity relationship (SAR) studies, we designed and synthesized a number of analogs of the initial hit compound. Analysis of SAR data from these analogs identified two analogs, MKP-128 and MKP-149, that showed similar *in vitro* potency, but exhibited 8- to 12-fold better whole cell activity compared to the original hit compound. These will be further evaluated as potential drug development leads. Lastly, we identified a benzimidazole scaffold as a potential replacement for the benzofuran core, and determined that the benzimidazole analogs bind in a similar pose as the benzofurans by solving the complex structure.

REFERENCES

- 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.
- Arora, P., Goyal, A., Natarajan, V.T., Rajakumara, E., Verma, P., Gupta, R., Yousuf, M., Trivedi, O.A., Mohanty, D., Tyagi, A., *et al.* (2009). Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nat Chem Biol* 5, 166-173.
- Astarie-Dequeker, C., Le Guyader, L., Malaga, W., Seaphanh, F.K., Chalut, C., Lopez, A., and Guilhot, C. (2009). Phthiocerol dimycocerosates of *M. tuberculosis* participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS Pathog* 5, e1000289.
- Astarie-Dequeker, C., Nigou, J., Passemar, C., and Guilhot, C. (2010). The role of mycobacterial lipids in host pathogenesis. *Drug Discovery Today: Disease Mechanisms* 7, e33-e41.
- Azad, A.K., Sirakova, T.D., Fernandes, N.D., and Kolattukudy, P.E. (1997). Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem* 272, 16741-16745.
- Baldwin, J.E., Bird, J.W., Field, R.A., O'Callaghan, N.M., Schofield, C.J., and Willis, A.C. (1991). Isolation and partial characterisation of ACV synthetase from *Cephalosporium acremonium* and *Streptomyces clavuligerus*. Evidence for the presence of phosphopantothenate in ACV synthetase. *J Antibiot (Tokyo)* 44, 241-248.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.
- 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.
- Barry, C.E., 3rd, Lee, R.E., Mdluli, K., Sampson, A.E., Schroeder, B.G., Slayden, R.A., and Yuan, Y. (1998). Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* 37, 143-179.
- Belanger, A.E., Besra, G.S., Ford, M.E., Mikusova, K., Belisle, J.T., Brennan, P.J., and Inamine, J.M. (1996). The *embAB* genes of *Mycobacterium avium* encode an arabinosyl

- transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc Natl Acad Sci U S A* *93*, 11919-11924.
- Bellizzi, J.J., 3rd, Widom, J., Kemp, C., Lu, J.Y., Das, A.K., Hofmann, S.L., and Clardy, J. (2000). The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc Natl Acad Sci U S A* *97*, 4573-4578.
- Bergfors, T. (2003). Seeds to crystals. *J Struct Biol* *142*, 66-76.
- Boissier, F., Bardou, F., Guillet, V., Uttenweiler-Joseph, S., Daffe, M., Quemard, A., and Mourey, L. (2006). Further insight into S-adenosylmethionine-dependent methyltransferases: structural characterization of Hma, an enzyme essential for the biosynthesis of oxygenated mycolic acids in *Mycobacterium tuberculosis*. *J Biol Chem* *281*, 4434-4445.
- Brennan, P.J. (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* *83*, 91-97.
- Brennan, P.J., and Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* *64*, 29-63.
- Bricogne, G., Vonnrhein, C., Paciorek, W., Flensburg, C., Schiltz, M., Blanc, E., Roversi, P., Morris, R., and Evans, G. (2002). Enhancements in autoSHARP and SHARP, with applications to difficult phasing problems. *Acta Crystallographica Section A* *58*, c239.
- Buglino, J., Onwueme, K.C., Ferreras, J.A., Quadri, L.E., and Lima, C.D. (2004). Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from *Mycobacterium tuberculosis*. *J Biol Chem* *279*, 30634-30642.
- Camacho, L.R., Constant, P., Raynaud, C., Laneelle, M.A., Triccas, J.A., Gicquel, B., Daffe, M., and Guilhot, C. (2001). Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* *276*, 19845-19854.
- Camacho, L.R., Ensergueix, D., Perez, E., Gicquel, B., and Guilhot, C. (1999). Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* *34*, 257-267.
- Cantu, D.C., Chen, Y., and Reilly, P.J. (2010). Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci* *19*, 1281-1295.
- Chakravarty, B., Gu, Z., Chirala, S.S., Wakil, S.J., and Quioco, F.A. (2004). Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. *Proc Natl Acad Sci U S A* *101*, 15567-15572.

- Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C., and Guilhot, C. (2006). The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of *Mycobacteria*. *Proc Natl Acad Sci U S A* *103*, 8511-8516.
- Chayen, N.E., Saridakis, E., and Sear, R.P. (2006). Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. *Proc Natl Acad Sci U S A* *103*, 597-601.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* *66*, 12-21.
- Chirgadze, N.Y., Briggs, S.L., McAllister, K.A., Fischl, A.S., and Zhao, G. (2000). Crystal structure of *Streptococcus pneumoniae* acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis. *Embo J* *19*, 5281-5287.
- Chopra, T., Banerjee, S., Gupta, S., Yadav, G., Anand, S., Surolia, A., Roy, R.P., Mohanty, D., and Gokhale, R.S. (2008). Novel intermolecular iterative mechanism for biosynthesis of mycoketide catalyzed by a bimodular polyketide synthase. *PLoS Biol* *6*, e163.
- Chopra, T., and Gokhale, R.S. (2009). Polyketide versatility in the biosynthesis of complex mycobacterial cell wall lipids. *Methods Enzymol* *459*, 259-294.
- Cole, S.T. (1999). Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett* *452*, 7-10.
- 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.
- Constant, P., Perez, E., Malaga, W., Laneelle, M.A., Saurel, O., Daffe, M., and Guilhot, C. (2002). Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. *J Biol Chem* *277*, 38148-38158.
- Cox, J.S., Chen, B., McNeil, M., and Jacobs, W.R., Jr. (1999). Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* *402*, 79-83.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* *288*, 640-649.

- D'Arcy, A., Villard, F., and Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr* *63*, 550-554.
- Daffe, M., and Laneelle, M.A. (1988). Distribution of phthiocerol diester, phenolic mycosides and related compounds in mycobacteria. *J Gen Microbiol* *134*, 2049-2055.
- Dauter, Z. (2005). Use of polynuclear metal clusters in protein crystallography. *Comptes Rendus Chimie* *8*, 1808-1814.
- de La Fortelle, E., and Bricogne, G. (1997). [27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* *276*, 472-494.
- Demartreau-Ginsburg, H., Lederer, E., Ryhage, R., Stallberg-Stenhagen, S., and Stenhagen, E. (1959). Structure of phthiocerol. *Nature* *183*, 117-119.
- 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.
- Dong, A., Xu, X., Edwards, A.M., Midwest Center for Structural, G., Structural Genomics, C., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., *et al.* (2007). In situ proteolysis for protein crystallization and structure determination. *Nat Methods* *4*, 1019-1021.
- Donnelly, M.I., Zhou, M., Millard, C.S., Clancy, S., Stols, L., Eschenfeldt, W.H., Collart, F.R., and Joachimiak, A. (2006). An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr Purif* *47*, 446-454.
- Doublet, S. (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol Biol* *363*, 91-108.
- Dubey, V.S., Sirakova, T.D., Cynamon, M.H., and Kolattukudy, P.E. (2003). Biochemical function of *msl5* (*pks8* plus *pks17*) in *Mycobacterium tuberculosis* H37Rv: biosynthesis of monomethyl branched unsaturated fatty acids. *J Bacteriol* *185*, 4620-4625.
- Dubey, V.S., Sirakova, T.D., and Kolattukudy, P.E. (2002). Disruption of *msl3* abolishes the synthesis of mycolipanoic and mycolipenic acids required for polyacyltrehalose synthesis in *Mycobacterium tuberculosis* H37Rv and causes cell aggregation. *Mol Microbiol* *45*, 1451-1459.

- 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.
- Duckworth, B.P., and Aldrich, C.C. (2010). Development of a high-throughput fluorescence polarization assay for the discovery of phosphopantetheinyl transferase inhibitors. *Anal Biochem* 403, 13-19.
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34, W116-118.
- Dym, O., Albeck, S., Peleg, Y., Schwarz, A., Shakked, Z., Burstein, Y., and Zimhony, O. (2009). Structure-function analysis of the acyl carrier protein synthase (AcpS) from *Mycobacterium tuberculosis*. *J Mol Biol* 393, 937-950.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Favrot, L., and Ronning, D.R. (2012). Targeting the mycobacterial envelope for tuberculosis drug development. *Expert Rev Anti Infect Ther* 10, 1023-1036.
- Fitzmaurice, A.M., and Kolattukudy, P.E. (1998). An acyl-CoA synthase (acoas) gene adjacent to the mycocerosic acid synthase (mas) locus is necessary for mycocerosyl lipid synthesis in *Mycobacterium tuberculosis* var. *bovis* BCG. *J Biol Chem* 273, 8033-8039.
- Forrellad, M.A., Klepp, L.I., Gioffre, A., Sabio y Garcia, J., Morbidoni, H.R., de la Paz Santangelo, M., Cataldi, A.A., and Bigi, F. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3-66.
- Franzblau, S.G., Witzig, R.S., McLaughlin, J.C., Torres, P., Madico, G., Hernandez, A., Degnan, M.T., Cook, M.B., Quenzer, V.K., Ferguson, R.M., and Gilman, R.H. (1998). Rapid, low-technology MIC determination with clinical *Mycobacterium tuberculosis* isolates by using the microplate Alamar Blue assay. *J Clin Microbiol* 36, 362-366.
- Gao, X., Bain, K., Bonanno, J.B., Buchanan, M., Henderson, D., Lorimer, D., Marsh, C., Reynes, J.A., Sauder, J.M., Schwinn, K., *et al.* (2005). High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *J Struct Funct Genomics* 6, 129-134.
- Garcia-Ruiz, J.M., Gonzalez-Ramirez, L.A., Gavira, J.A., and Otalora, F. (2002). Granada Crystallisation Box: a new device for protein crystallisation by counter-diffusion techniques. *Acta Crystallogr D Biol Crystallogr* 58, 1638-1642.

- Gavalda, S., Leger, M., van der Rest, B., Stella, A., Bardou, F., Montrozier, H., Chalut, C., Burlet-Schiltz, O., Marrakchi, H., Daffé, M., and Quemard, A. (2009). The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in *Mycobacterium tuberculosis*. *J Biol Chem* 284, 19255-19264.
- Geiger, O., Spaink, H.P., and Kennedy, E.P. (1991). Isolation of the *Rhizobium leguminosarum* NodF nodulation protein: NodF carries a 4'-phosphopantetheine prosthetic group. *J Bacteriol* 173, 2872-2878.
- Gibrat, J.F., Madej, T., and Bryant, S.H. (1996). Surprising similarities in structure comparison. *Curr Opin Struct Biol* 6, 377-385.
- 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.
- Gokhale, R.S., Saxena, P., Chopra, T., and Mohanty, D. (2007). Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. *Nat Prod Rep* 24, 267-277.
- Goren, M.B., Brokl, O., and Das, B.C. (1976). Sulfatides of *Mycobacterium tuberculosis*: the structure of the principal sulfatide (SL-I). *Biochemistry* 15, 2728-2735.
- Guilhot, C., Chalut, C., and Daffé, M. (2008). Biosynthesis and Roles of Phenolic Glycolipids and Related Molecules in *Mycobacterium tuberculosis*. In *The Mycobacterial Cell Envelope*, M. Daffé, and J.M. Reyat, eds. (Washington, DC: ASM Press), p. 395.
- Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254, 51-58.
- Hendrickson, W.A., and Ogata, C.M. (1997). [28] Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol* 276, 494-523.
- Heras, B., Edeling, M.A., Byriel, K.A., Jones, A., Raina, S., and Martin, J.L. (2003). Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure* 11, 139-145.
- Heras, B., and Martin, J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* 61, 1173-1180.
- Ioerger, T.R., O'Malley, T., Liao, R., Guinn, K.M., Hickey, M.J., Mohaideen, N., Murphy, K.C., Boshoff, H.I.M., Mizrahi, V., Rubin, E.J., *et al.* (2013). Identification of

new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*. PLOS ONE 8, e75245.

Issartel, J.P., Koronakis, V., and Hughes, C. (1991). Activation of *Escherichia coli* prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation. Nature 351, 759-761.

Jackson, M., Stadthagen, G., and Gicquel, B. (2007). Long-chain multiple methyl-branched fatty acid-containing lipids of *Mycobacterium tuberculosis*: biosynthesis, transport, regulation and biological activities. Tuberculosis (Edinb) 87, 78-86.

Jain, M., and Cox, J.S. (2005). Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in *M. tuberculosis*. PLoS Pathog 1, e2.

Jain, M., Petzold, C.J., Schelle, M.W., Leavell, M.D., Mougous, J.D., Bertozzi, C.R., Leary, J.A., and Cox, J.S. (2007). Lipidomics reveals control of *Mycobacterium tuberculosis* virulence lipids via metabolic coupling. Proc Natl Acad Sci U S A 104, 5133-5138.

Jeffrey, G.A. (1997). An introduction to hydrogen bonding (New York: Oxford University Press).

Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolasek, B., and Ban, N. (2007). Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. Science 316, 254-261.

Jenni, S., Leibundgut, M., Maier, T., and Ban, N. (2006). Architecture of a fungal fatty acid synthase at 5 Å resolution. Science 311, 1263-1267.

Johansson, P., Wiltschi, B., Kumari, P., Kessler, B., Vonrhein, C., Vonck, J., Oesterhelt, D., and Gringer, M. (2008). Inhibition of the fungal fatty acid synthase type I multienzyme complex. Proc Natl Acad Sci U S A 105, 12803-12808.

Joseph-McCarthy, D., Parris, K., Huang, A., Failli, A., Quagliato, D., Dushin, E.G., Novikova, E., Severina, E., Tuckman, M., Petersen, P.J., *et al.* (2005). Use of structure-based drug design approaches to obtain novel anthranilic acid acyl carrier protein synthase inhibitors. J Med Chem 48, 7960-7969.

Kantardjieff, K.A., Hochtl, P., Segelke, B.W., Tao, F.M., and Rupp, B. (2002). Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility. Acta Crystallogr D Biol Crystallogr 58, 735-743.

- Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* 9, 522-526.
- Khosla, C., Gokhale, R.S., Jacobsen, J.R., and Cane, D.E. (1999). Tolerance and specificity of polyketide synthases. *Annu Rev Biochem* 68, 219-253.
- Kissinger, C.R., Gehlhaar, D.K., and Fogel, D.B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 2), 484-491.
- Kolattukudy, P.E., Fernandes, N.D., Azad, A.K., Fitzmaurice, A.M., and Sirakova, T.D. (1997). Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol* 24, 263-270.
- Kosa, N.M., Foley, T.L., and Burkart, M.D. (2014). Fluorescent techniques for discovery and characterization of phosphopantetheinyl transferase inhibitors. *J Antibiot (Tokyo)* 67, 113-120.
- Krieger, I.V., Freundlich, J.S., Gawandi, V.B., Roberts, J.P., Gawandi, V.B., Sun, Q., Owen, J.L., Fraile, M.T., Huss, S.I., Lavandera, J.L., *et al.* (2012). Structure-guided discovery of phenyl-diketo acids as potent inhibitors of *M. tuberculosis* malate synthase. *Chem Biol* 19, 1556-1567.
- La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. (2004). Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* 11, 195-201.
- Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* 3, 923-936.
- Lambalot, R.H., and Walsh, C.T. (1995). Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J Biol Chem* 270, 24658-24661.
- Leblanc, C., Prudhomme, T., Tabouret, G., Ray, A., Burbaud, S., Cabantous, S., Mourey, L., Guilhot, C., and Chalut, C. (2012). 4'-Phosphopantetheinyl transferase PptT, a new drug target required for *Mycobacterium tuberculosis* growth and persistence in vivo. *PLoS Pathog* 8, e1003097.
- Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008). The multienzyme architecture of eukaryotic fatty acid synthases. *Curr Opin Struct Biol* 18, 714-725.

- Liu, J., and Nikaido, H. (1999). A mutant of *Mycobacterium smegmatis* defective in the biosynthesis of mycolic acids accumulates meromycolates. *Proc Natl Acad Sci U S A* *96*, 4011-4016.
- Lorber, B., Sauter, C., Theobald-Dietrich, A., Moreno, A., Schellenberger, P., Robert, M.C., Capelle, B., Sanglier, S., Potier, N., and Giege, R. (2009). Crystal growth of proteins, nucleic acids, and viruses in gels. *Prog Biophys Mol Biol* *101*, 13-25.
- Magnuson, K., Jackowski, S., Rock, C.O., and Cronan, J.E., Jr. (1993). Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiol Rev* *57*, 522-542.
- Maier, T., Jenni, S., and Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* *311*, 1258-1262.
- Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science* *321*, 1315-1322.
- Maier, T., Leibundgut, M., Boehringer, D., and Ban, N. (2010). Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys* *43*, 373-422.
- Marks, G.S., and Polgar, N. (1955). Mycoceranic acid. Part II. *Journal of the Chemical Society (Resumed)*, 3851.
- Mathur, M., and Kolattukudy, P.E. (1992). Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. *bovis* *Bacillus Calmette-Guerin*. *J Biol Chem* *267*, 19388-19395.
- Matsunaga, I., Bhatt, A., Young, D.C., Cheng, T.Y., Eyles, S.J., Besra, G.S., Briken, V., Porcelli, S.A., Costello, C.E., Jacobs, W.R., Jr., and Moody, D.B. (2004). *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J Exp Med* *200*, 1559-1569.
- Matsunaga, I., and Sugita, M. (2012). Mycoketide: a CD1c-presented antigen with important implications in mycobacterial infection. *Clin Dev Immunol* *2012*, 981821.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J Mol Biol* *33*, 491-497.
- Mattick, J.S., Nickless, J., Mizugaki, M., Yang, C.Y., Uchiyama, S., and Wakil, S.J. (1983). The architecture of the animal fatty acid synthetase. II. Separation of the core and thioesterase functions and determination of the N-C orientation of the subunit. *J Biol Chem* *258*, 15300-15304.
- McAllister, K.A., Peery, R.B., Meier, T.I., Fischl, A.S., and Zhao, G. (2000). Biochemical and molecular analyses of the *Streptococcus pneumoniae* acyl carrier

- protein synthase, an enzyme essential for fatty acid biosynthesis. *J Biol Chem* 275, 30864-30872.
- McNeil, M., Daffe, M., and Brennan, P.J. (1990). Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. *J Biol Chem* 265, 18200-18206.
- 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.
- McPherson, A. (2004). Introduction to protein crystallization. *Methods* 34, 254-265.
- McRee, D.E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 125, 156-165.
- Meier, J.L., and Burkart, M.D. (2009). The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* 38, 2012-2045.
- Mohanty, D., Sankaranarayanan, R., and Gokhale, R.S. (2011). Fatty acyl-AMP ligases and polyketide synthases are unique enzymes of lipid biosynthetic machinery in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 91, 448-455.
- Moriarty, N.W., Grosse-Kunstleve, R.W., and Adams, P.D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* 65, 1074-1080.
- Mueller, M., Jenni, S., and Ban, N. (2007). Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr Opin Struct Biol* 17, 572-579.
- Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67, 355-367.
- Nardini, M., and Dijkstra, B.W. (1999). Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* 9, 732-737.
- Neyrolles, O., and Guilhot, C. (2011). Recent advances in deciphering the contribution of *Mycobacterium tuberculosis* lipids to pathogenesis. *Tuberculosis (Edinb)* 91, 187-195.
- Nguyen, L., Chinnapagari, S., and Thompson, C.J. (2005). FbpA-Dependent biosynthesis of trehalose dimycolate is required for the intrinsic multidrug resistance, cell wall structure, and colonial morphology of *Mycobacterium smegmatis*. *J Bacteriol* 187, 6603-6611.

- Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2, 2212-2221.
- O'Brien, J., Wilson, I., Orton, T., and Pognan, F. (2000). Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur J Biochem* 267, 5421-5426.
- Onwueme, K.C., Ferreras, J.A., Buglino, J., Lima, C.D., and Quadri, L.E. (2004). Mycobacterial polyketide-associated proteins are acyltransferases: proof of principle with *Mycobacterium tuberculosis* PapA5. *Proc Natl Acad Sci U S A* 101, 4608-4613.
- Onwueme, K.C., Vos, C.J., Zurita, J., Ferreras, J.A., and Quadri, L.E. (2005). The dimycocerosate ester polyketide virulence factors of mycobacteria. *Prog Lipid Res* 44, 259-302.
- Ortalo-Magne, A., Lemassu, A., Laneelle, M.A., Bardou, F., Silve, G., Gounon, P., Marchal, G., and Daffe, M. (1996). Identification of the surface-exposed lipids on the cell envelopes of *Mycobacterium tuberculosis* and other mycobacterial species. *J Bacteriol* 178, 456-461.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods of Enzymology* 276, 307-326.
- Owen, J.G., Copp, J.N., and Ackerley, D.F. (2011). Rapid and flexible biochemical assays for evaluating 4'-phosphopantetheinyl transferase activity. *Biochem J* 436, 709-717.
- Parris, K.D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C.C., Seehra, J., and Somers, W.S. (2000). Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure Fold Des* 8, 883-895.
- Pemble, C.W.t., Johnson, L.C., Kridel, S.J., and Lowther, W.T. (2007). Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat. *Nat Struct Mol Biol* 14, 704-709.
- Portevin, D., De Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffe, M., and Guilhot, C. (2004). A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A* 101, 314-319.
- Rainwater, D.L., and Kolattukudy, P.E. (1985). Fatty acid biosynthesis in *Mycobacterium tuberculosis* var. *bovis* *Bacillus Calmette-Guerin*. Purification and

characterization of a novel fatty acid synthase, mycocerosic acid synthase, which elongates n-fatty acyl-CoA with methylmalonyl-CoA. *J Biol Chem* 260, 616-623.

Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G., and Barry, C.E., 3rd (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 431, 84-87.

Rees, D.C., Congreve, M., Murray, C.W., and Carr, R. (2004). Fragment-based lead discovery. *Nat Rev Drug Discov* 3, 660-672.

Richardson, R.D., and Smith, J.W. (2007). Novel antagonists of the thioesterase domain of human fatty acid synthase. *Molecular Cancer Therapeutics* 6, 2120-2126.

Rock, C.O., and Cronan, J.E. (1996). *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim Biophys Acta* 1302, 1-16.

Rousseau, C., Neyrolles, O., Bordat, Y., Giroux, S., Sirakova, T.D., Prevost, M.C., Kolattukudy, P.E., Gicquel, B., and Jackson, M. (2003a). Deficiency in mycolipenate- and mycosanoate-derived acyltrehaloses enhances early interactions of *Mycobacterium tuberculosis* with host cells. *Cell Microbiol* 5, 405-415.

Rousseau, C., Sirakova, T.D., Dubey, V.S., Bordat, Y., Kolattukudy, P.E., Gicquel, B., and Jackson, M. (2003b). Virulence attenuation of two Mas-like polyketide synthase mutants of *Mycobacterium tuberculosis*. *Microbiology* 149, 1837-1847.

Rousseau, C., Winter, N., Pivert, E., Bordat, Y., Neyrolles, O., Ave, P., Huerre, M., Gicquel, B., and Jackson, M. (2004). Production of phthiocerol dimycocerosates protects *Mycobacterium tuberculosis* from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. *Cell Microbiol* 6, 277-287.

Russell, D.G. (2007). Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5, 39-47.

Sambandamurthy, V.K., Derrick, S.C., Hsu, T., Chen, B., Larsen, M.H., Jalapathy, K.V., Chen, M., Kim, J., Porcelli, S.A., Chan, J., *et al.* (2006). *Mycobacterium tuberculosis* DeltaRD1 DeltapanCD: a safe and limited replicating mutant strain that protects immunocompetent and immunocompromised mice against experimental tuberculosis. *Vaccine* 24, 6309-6320.

Saxena, P., Yadav, G., Mohanty, D., and Gokhale, R.S. (2003). A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. *J Biol Chem* 278, 44780-44790.

Schluger, N.W., and Rom, W.N. (1998). The host immune response to tuberculosis. *Am J Respir Crit Care Med* 157, 679-691.

Sheldrick, G. (2008a). A short history of SHELX. *Acta Crystallographica Section A* 64, 112-122.

Sheldrick, G.M. (2008b). A short history of SHELX. *Acta Crystallogr A* 64, 112-122.

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* 7, 285-295.

Shen, B., Summers, R.G., Gramajo, H., Bibb, M.J., and Hutchinson, C.R. (1992). Purification and characterization of the acyl carrier protein of the *Streptomyces glaucescens* tetracenomycin C polyketide synthase. *J Bacteriol* 174, 3818-3821.

Shenai, S., and Rodrigues, C. (2011). Mycobacterium. In *Molecular detection of human bacterial pathogens*, D. Liu, ed. (Boca Raton, FL: Taylor & Francis/CRC Press), pp. xxvii, 1251 p.

Simeone, R., Constant, P., Guilhot, C., Daffe, M., and Chalut, C. (2007). Identification of the missing trans-acting enoyl reductase required for phthiocerol dimycocerosate and phenolglycolipid biosynthesis in *Mycobacterium tuberculosis*. *J Bacteriol* 189, 4597-4602.

Sirakova, T.D., Dubey, V.S., Cynamon, M.H., and Kolattukudy, P.E. (2003a). Attenuation of *Mycobacterium tuberculosis* by disruption of a mas-like gene or a chalcone synthase-like gene, which causes deficiency in dimycocerosyl phthiocerol synthesis. *J Bacteriol* 185, 2999-3008.

Sirakova, T.D., Dubey, V.S., Kim, H.J., Cynamon, M.H., and Kolattukudy, P.E. (2003b). The largest open reading frame (pks12) in the *Mycobacterium tuberculosis* genome is involved in pathogenesis and dimycocerosyl phthiocerol synthesis. *Infect Immun* 71, 3794-3801.

Sirakova, T.D., Thirumala, A.K., Dubey, V.S., Sprecher, H., and Kolattukudy, P.E. (2001). The *Mycobacterium tuberculosis* pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J Biol Chem* 276, 16833-16839.

Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *Faseb J* 8, 1248-1259.

Smith, S., and Tsai, S.C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* 24, 1041-1072.

Stanley, S.A., Kawate, T., Iwase, N., Shimizu, M., Clatworthy, A.E., Kazyanskaya, E., Sacchettini, J.C., Ioerger, T.R., Siddiqi, N.A., Minami, S., *et al.* (2013). Diarylcoumarins

inhibit mycolic acid biosynthesis and kill *Mycobacterium tuberculosis* by targeting FadD32. Proc Natl Acad Sci U S A 110, 11565-11570.

Studier, F.W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol 219, 37-44.

Sulzenbacher, G., Canaan, S., Bordat, Y., Neyrolles, O., Stadthagen, G., Roig-Zamboni, V., Rauzier, J., Maurin, D., Laval, F., Daffe, M., *et al.* (2006). LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium tuberculosis*. Embo J 25, 1436-1444.

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.

Taylor, R.C., Brown, A.K., Singh, A., Bhatt, A., and Besra, G.S. (2010). Characterization of a beta-hydroxybutyryl-CoA dehydrogenase from *Mycobacterium tuberculosis*. Microbiology 156, 1975-1982.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallographica Section D 55, 849-861.

Terwilliger, T.C., Klei, H., Adams, P.D., Moriarty, N.W., and Cohn, J.D. (2006). Automated ligand fitting by core-fragment fitting and extension into density. Acta Crystallogr D Biol Crystallogr 62, 915-922.

Thakur, A.S., Robin, G., Guncar, G., Saunders, N.F., Newman, J., Martin, J.L., and Kobe, B. (2007). Improved success of sparse matrix protein crystallization screening with heterogeneous nucleating agents. PLOS ONE 2, e1091.

Trivedi, O.A., Arora, P., Sridharan, V., Tickoo, R., Mohanty, D., and Gokhale, R.S. (2004). Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature 428, 441-445.

Trivedi, O.A., Arora, P., Vats, A., Ansari, M.Z., Tickoo, R., Sridharan, V., Mohanty, D., and Gokhale, R.S. (2005). Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. Mol Cell 17, 631-643.

Tsai, S.C., and Ames, B.D. (2009). Structural enzymology of polyketide synthases. Methods Enzymol 459, 17-47.

Tsaioun, K., and Kates, S.A. (2011). ADMET for medicinal chemists : a practical guide (Hoboken, N.J.: John Wiley & Sons).

Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* 303, 1805-1810.

Weis, W.I., and Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* 65, 441-473.

Wilson, R., Kumar, P., Parashar, V., Vilcheze, C., Veyron-Churlet, R., Freundlich, J.S., Barnes, S.W., Walker, J.R., Szymonifka, M.J., Marchiano, E., *et al.* (2013). Antituberculosis thiophenes define a requirement for Pks13 in mycolic acid biosynthesis. *Nat Chem Biol* 9, 499-506.

Wong, H.C., Liu, G., Zhang, Y.M., Rock, C.O., and Zheng, J. (2002). The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*. *J Biol Chem* 277, 15874-15880.

Xu, W., Qiao, K., and Tang, Y. (2013). Structural analysis of protein-protein interactions in type I polyketide synthases. *Crit Rev Biochem Mol Biol* 48, 98-122.

Yasgar, A., Foley, T.L., Jadhav, A., Inglese, J., Burkart, M.D., and Simeonov, A. (2010). A strategy to discover inhibitors of *Bacillus subtilis* surfactin-type phosphopantetheinyl transferase. *Mol Biosyst* 6, 365-375.

Yuan, Y., Zhu, Y., Crane, D.D., and Barry, C.E., 3rd (1998). The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis*. *Mol Microbiol* 29, 1449-1458.

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.