



MONASH University

**Structural Characterization of Three Novel Proteins
from *Mycobacterium species* Vital for their Survival**

By

Adam Elias Shahine (B.Sc)

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

Department of Biochemistry & Molecular Biology
Faculty of Medicine, Nursing and Health Sciences

Monash University

2016

Copyright notice

© Adam Elias Shahine 2016. Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

Mycobacterium tuberculosis, the etiological agent of tuberculosis (TB), is one of the deadliest human pathogens, with one third of the world's population presently infected, resulting in 1.3 million deaths in 2012. The emergence of multiple drug resistant strains of *M. tuberculosis* has resulted in an inability to control and treat this pathogen. Accordingly, there is an urgent need for the development of new anti-TB therapeutics.

Structural genomics has played a pivotal role in the characterization of the *M. tuberculosis* proteome to gain functional insight into pathogenesis, with applications for future rational drug design. While TB structural genomics has generated data on a number of mycobacterial targets, structural data is missing for up to 90% of the *M. tuberculosis* proteome. To gain a global understanding of mycobacterial pathogenesis, an in-depth structural characterization of the *M. tuberculosis* proteome is required.

My work described here explores the structural characteristics of three novel proteins essential for mycobacterial growth and host survival. Utilizing applications of x-ray crystallography, crystal structures of each target were determined, and have been utilized to characterise drug inhibition mechanisms, as well as provide functional insight into the biological role of the targets.

The key first line anti-TB drug, isoniazid, exhibits inhibition against multiple *M. tuberculosis* targets, where resistance is readily developed. Rv2971, an essential aldo-keto reductase, is a recently identified target of isoniazid. To gain structural insight into isoniazid inhibition mechanisms, the crystal structure of Rv2971 was determined to 1.6 Å, revealing the structural architecture of the isoniazid and NADPH binding sites, paving the way for future structural characterization of inhibition.

A recently characterized drug, tetrahydrolipstatin (THL), has been found to strongly inhibit Rv3802c, an essential lipase involved in mycolic acid biosynthesis. To characterise mechanisms of THL inhibition, the crystal structure of Rv3802c in complex with THL was determined to 2.9 Å. The crystal structure reveals the binding mechanisms of THL against an essential mycobacterial protein. Identification and structural characterization of additional THL drug targets will allow for future THL based drug design.

A high percentage of the *M. tuberculosis* proteome is annotated as “hypothetical” proteins. To truly understand the mechanism of mycobacterial pathogenesis, further characterisation of hypothetical proteins is required. Rv0807, and its *M. smegmatis* orthologue, MSMEG_5817, have recently been identified as essential for mycobacterial survival within macrophages. The gene is annotated as a hypothetical protein, with little known on its functionality. The crystal structure of MSMEG_5817 was, via Se-MAD approach, successfully determined to 1.7Å. The structure shares similarities with sterol carrier proteins (SCP), which bind and transport biologically relevant lipids. The lipid binding capabilities of MSMEG_5817 was probed, revealing binding to a number of phospholipids in a differing binding mechanism to the SCPs. The crystal structure determined structurally represents this new class of mycobacterial proteins, providing further insight into mechanisms of host survival.

The three crystal structures provide insight into the drug binding mechanisms and functional characteristics of each target, allowing for further characterization of the novel mechanisms of mycobacterial pathogenesis, with potential for development of new anti-TB therapeutics.

Declaration of Authenticity

I hereby declare that the content of this thesis contains no material that has been accepted for the award of any other degree of diploma at any university or equivalent institution. To the best of my knowledge, this thesis contains no material previously published or written by another person, except where the appropriate reference has been made.

In addition, I declare that the text of this thesis, including figures and publications, is less than 100,000 words.

Signed:



Date: 21/10/15

Publications

The research conducted as part of this study has previously been reviewed and published in the following papers:

Shahine, A. E., Chan, P. Y., Littler, D., Vivian, J., Brammananth, R., Crellin, P. K., Rossjohn, J., Beddoe, T., (2013). “Cloning, expression, purification and preliminary X-ray diffraction studies of a mycobacterial protein implicated in bacterial survival in macrophages.” *Acta Cryst (2013)*. F69, 566-569

Shahine, A., Prasetyoputri, A., Rossjohn, J., Beddoe, T., (2014) “A structural characterization of the isoniazid Mycobacterium tuberculosis drug target, Rv2971, in its unliganded form.” *Acta Cryst (2014)*. F70, 572-577

Shahine, A., Littler, D., Brammananth, R., Chan, P., Crellin, P., Coppel, R., Rossjohn, J., Beddoe, T., (2014) “A structural and functional investigation of a protein from *Mycobacterium smegmatis* implemented in mycobacterial macrophage survivability.” *Acta Cryst (2014)*. D70, 2264-2276

Thesis including published work General Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three original papers published in peer-reviewed journals and one unpublished publication. The core theme of the thesis is mycobacterial structural genomics. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Biochemistry and Molecular Biology, Faculty of Medicine, Nursing and Health Sciences under the supervision of Professor Jamie Rossjohn.

(The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.)

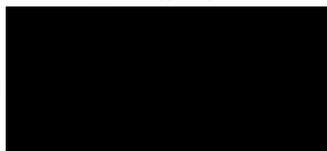
In the case of Chapters 2, 4, 5 and 6 my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status	Nature and extent (%) of students contribution
2	A structural characterization of the isoniazid Mycobacterium tuberculosis drug target, Rv2971, in its unliganded form	Published 2014	Expression, purification and x-ray data collection of Rv2971. Data analysis, figure and manuscript preparation
4	Cloning, expression, purification and preliminary X-ray diffraction studies of a mycobacterial protein implicated in bacterial survival in macrophages	Published 2013	Expression, purification, and crystallisation of native and Se-Met derivative MSMEG_5817. Data analysis, figure and manuscript preparation
5	A structural and functional investigation of a protein from <i>Mycobacterium smegmatis</i> implemented in mycobacterial macrophage survivability	Published 2014	Structural determination of MSMEG_5817. Cloning, expression, purification of hSCP. PIP-Strip and ELISA analysis of MSMEG_5817 compared to hSCP. Alanine-scanning mutagenesis generation of nine mutants, expression and purification of six mutants. Circular dichroism and thermal melting of MSMEG_5817 and 6 mutants. Data analysis, figure and manuscript preparation.

6	The high resolution structure of a novel uncharacterized <i>M. smegmatis</i> gene product, MSMEG_5817, expressed within its native host	Submitted 2015	Expression and purification of MSMEG_5817 in <i>M. smegmatis</i> . Crystallisation, data collection and structural determination. Data analysis, figure and manuscript preparation
---	---	----------------	--

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature:

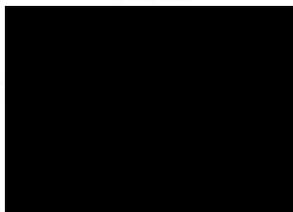


Date:

21/10/15

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student and co-authors' contributions to this work.

Main Supervisor signature:



Date:

21.10.15

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Jamie Rossjohn, for all the guidance and expertise he has provided since I joined the laboratory in 2009. This thesis would not have come into fruition without his continued drive and resolve throughout the course of this PhD. And, despite our clashing musical tastes, I am truly grateful and honoured to have been granted the opportunity to complete a PhD within his laboratory. Secondly, I would also like to thank my co-supervisor, Associate Professor Travis Beddoe, for his continued support, experimental advice and patience over the last six years. From his wealth of knowledge in experimental design and optimisation to his persistence when progression seemed dire, he has inspired me in becoming a better research scientist.

Many members of the laboratory, both past and present, have contributed greatly to this thesis through sharing their expertise, as well as making the lab a positive place to work. In particular, thank you to Dr Jérôme LeNours, who originally taught me the intricacies of protein x-ray crystallography, and for assisting in many early Australian Synchrotron visits. I would also like to thank my original laboratory mentors, Dr Adrian McAlister and Ms Sally Troy, who taught me the workings molecular biology and recombinant protein production. I'd also like to thank all the members of the Rossjohn laboratory, both past and present, including fellow PhD students, for making this a positive working environment and for the many laughs we've had over the last six years.

I would like to acknowledge Dr Paul Crellin and Ms Rajini Brammananth from Monash University Microbiology Department for their continued collaboration and for demystifying the world of mycobacterial microbiology. My thanks also go out to Ms Irene Hatzinisiriou from the Imaging Facility, as well as Dr Danuta Maksel and Dr Robyn Gray from the Monash Macromolecular Crystallisation Facility for their assistance with usage of their respective platforms.

I would also like to thank my family and friends, especially to my father, mother and sister, who have always remained a positive force throughout this undertaking. Finally, to my beautiful daughter Aaliyah and partner Natalie, you've made the completion of the last six years worthwhile, and now that this PhD is finally over, I can't wait to begin the next chapter in our lives.

Table of Contents

Copyright notice	ii
Abstract	iii
Declaration of Authenticity	v
Publications	vi
Thesis including published work General Declaration	vii
Acknowledgements	ix
Table of Contents	x
List of Figures	xii
Abbreviations	xiii
Chapter 1: Literature Review	1
1.1 <i>Mycobacterium tuberculosis</i>	1
1.2 Current prevention and therapeutic treatments	4
1.2.1 <i>Isoniazid</i>	4
1.2.1.1 <i>Rv2971: An essential oxidoreductase</i>	6
1.2.2 <i>Rifampicin</i>	8
1.2.3 <i>Ethambutol</i>	8
1.2.4 <i>Pyrazinamide</i>	9
1.2.5 <i>Characterization and treatment of drug resistant M. tuberculosis strains</i>	9
1.3 The Cell Wall of <i>Mycobacterium tuberculosis</i>	12
1.3.1 <i>Mycolic Acids</i>	14
1.3.2 <i>Enzymes involved in mycolic acid biosynthesis</i>	15
1.4 Rv3802c – A proposed mycobacterial lipase/thioesterase	20
1.4.1 <i>Tetrahydrolipstatin (THL) inhibition of Rv3802c</i>	25
1.5 <i>M. tuberculosis</i> survival within host macrophages	29
1.5.1 <i>Mechanisms of M. tuberculosis within host macrophages</i>	32
1.6 MSMEG_5817 – A novel drug target implemented in macrophage survival	36
1.7 Aims of this study	40
Chapter 2: Structural investigation into INH drug target, Rv2971	44
2.1 Introduction	44
Declaration of Thesis Chapter 2	47
Chapter 3: Structural determination of Rv3802c in complex with THL	54
3.1 Introduction	54

3.2	Materials and Methods	56
3.2.1	<i>Cloning of Rv3802c^{Δ1-67}</i>	56
3.2.2	<i>Expression, Refolding and Purification of Rv3802c^{Δ1-67}</i>	57
3.2.3	<i>Circular Dichroism and Thermal Melting</i>	58
3.2.4	<i>Enzyme Kinetic and Inhibition Assays</i>	59
3.2.5	<i>Cocomplexation and Crystallization of Rv3802c^{Δ1-67}-THL</i>	60
3.2.6	<i>X-ray Data Collection and Structural Determination of Rv3802c^{Δ1-67}-THL</i>	60
3.3	Results	61
3.3.1	<i>Expression and purification of Rv3802c^{Δ1-67}</i>	61
3.3.2	<i>Circular Dichroism and Thermal Melt Analysis of Rv3802c</i>	65
3.3.3	<i>Enzyme Kinetic and Inhibition assays of Rv3802c^{Δ1-67} against THL</i>	67
3.3.4	<i>Crystallization and Structural Determination of Rv3802c^{Δ1-67}-THL</i>	70
3.3.5	<i>Overview Structural Analysis of Rv3802c^{Δ1-67}-THL</i>	76
3.3.6	<i>Comparison of Rv3802c^{Δ1-67}-THL with Apo-MSMEG_6394</i>	78
3.3.7	<i>Structural Analysis of the THL Binding Site</i>	80
3.4	Discussion	86
	Chapter 4: Crystallization and X-Ray Data Collection of MSMEG_5817	91
4.1	Introduction	91
	Declaration of Thesis Chapter 4	93
	Chapter 5: Structural and Functional characterization of MSMEG_5817	98
5.1	Introduction	98
	Declaration of Thesis Chapter 5	100
	Chapter 6: Crystal Structure of MSMEG_5817 produced in <i>M. smegmatis</i>	114
6.1	Introduction	114
	Declaration of Thesis Chapter 6	116
	Chapter 7: Discussion	133
7.1	Structural genomics and TB	133
7.2	Drug discovery and structure based rational drug design against TB	135
7.3	Hypothetical mycobacterial proteins	142
7.4	Overcoming bottlenecks in TB structural biology	146
7.5	Conclusions and Future Directions	148
	Chapter 8: References	151
	Chapter 9: Appendix	164

List of Figures

- Figure 1.1: *M. tuberculosis* detection and XDR-TB incidence map
- Figure 1.2: TB First-line drug mechanisms of action against TB
- Figure 1.3: Crystal structures of INH drug targets
- Figure 1.4: Schematic representation of the mycobacterial cell wall
- Figure 1.5: Overview of the mycolic acid biosynthetic pathway
- Figure 1.6: Crystal structures of the *F. solani* cutinase and the essential *M. smegmatis* lipase MSMEG_6394
- Figure 1.7: Structural basis for THL inhibition against the human Fatty Acid Synthase II Thioesterase (FASIITE) domain
- Figure 1.8: Overview of macrophage phagolysosome maturation
- Figure 1.9: Phagosomal maturation in the presence of pathogenic and non-pathogenic mycobacteria
- Figure 1.10: *in vivo* growth curves of the *M. smegmatis* transposon mutant Myco132
- Figure 2.1: Determination of INH drug targets in *M. tuberculosis* by INH-NAD(P) affinity chromatography
- Figure 3.1: Expression of Rv3802c and Rv3802c^{Δ1-67}
- Figure 3.2: Purification of Rv3802c and Rv3802c^{Δ1-67}
- Figure 3.3: Circular Dichroism analysis of Rv3802c and Rv3802c^{Δ1-67}
- Figure 3.4: Enzyme kinetic activity and inhibition analysis of Rv3802c and Rv3802c^{Δ1-67}
- Figure 3.5: Crystals of Rv3802c^{Δ1-67} in its Apo form and in complex with THL
- Figure 3.6: Modeling of THL
- Figure 3.7: Overall structure of Rv3802c^{Δ1-67}-THL
- Figure 3.8: Rv3802c^{Δ1-67} active site cavity and structural overlay with MSMEG_6394
- Figure 3.9: Solvent accessibility between “open” and “closed” conformations
- Figure 3.10: Structural characterizations of THL bound within the Rv3802c enzyme active site
- Figure 3.11: Surface electrostatic characterisation of Rv3802
- Figure 3.12: Chemical structure of THL and modified THL compounds
- Figure 7.1: Lead inhibitor clusters identified via high-throughput screening
- Figure 7.2: Graphical representation of annotated ORF function in the *M. tuberculosis* H37Rv genome

Abbreviations

Å	Angstrom
AKR	Aldo-keto reductase
ATP	Adenosine tri-phosphate
CD	Circular Dichroism
CoA	Coenzyme-A
DNA	Deoxyribonucleic Acid
°C	Degrees Celsius
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme-Linked Immunosorbent Assay
FAS	Fatty Acid Synthase
FASTE	Fatty Acid Synthase: Thioesterase Domain
HCl	Hydrochloric Acid
HIV	Human Immunodeficiency Virus
INH	Isoniazid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
K	Kelvin
kDa	Kilo-Dalton
LAM	Lipoarabinnomannan
LB	Luria Broth
LIC	Ligation-Independent Cloning
M	Molar
MAD	Multiple wavelength anomalous dispersion
mAGP	mycolyl arabinogalactan-peptidoglycan complex
MDR-TB	Multiple drug resistant tuberculosis
MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligrams
MW	Molecular Weight
NADPH	Nicotinamide adenine dinucleotide phosphate
nm	nanometers
NF- κ B	Nuclear Factor kappa B
OD	Optical Density

ORF	Open Reading Frame
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
pH	Power of Hydrogen
PI3P	phosphatidylinositol 3-phosphate
PMSF	Phenylmethanesulfonylfluoride
RMSD	Root Mean Square Deviation
SCP	Sterol Carrier Protein
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
TB	Tuberculosis
TBS	Tris-buffered Saline
TE	Thioesterase
THL	Tetrahydrolipstatin
TIM	Triosephosphate isomerase
T _m	Thermal Melting Temperature
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
WHO	World Health Organization
WT	Wild Type
XDR-TB	Extensively drug resistant tuberculosis

Chapter 1: Literature Review

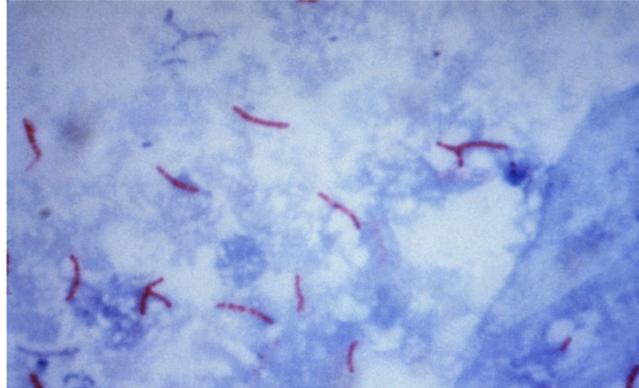
1.1: *Mycobacterium tuberculosis*

Mycobacterium tuberculosis, the key etiologic agent of human tuberculosis (TB), is a major cause of death and disease globally. It is estimated that approximately 2 billion people around the world, one third of the world's population, is currently infected with *M. tuberculosis*, with approximately 1.5 million deaths annually (WHO, 2015). TB is ranked the second highest cause of death by a single pathogenic agent, second only to Human Immunodeficiency Virus (HIV). It is reported that there were approximately 6.1 million new cases of TB in 2013, with 13% of cases involving patients co-infected with HIV (WHO, 2015). *M. tuberculosis* is also one of the oldest known human pathogens, with earliest known evidence supporting its role as a human pathogen dating back 9,000 years (Herskovitz, *et al.*, 2008). Despite its devastating effects and its long influence as a human pathogen, difficulties still remain in its treatment and prevention.

Disease by *M. tuberculosis* predominantly occurs within the lungs (pulmonary TB); however *M. tuberculosis* is known to result in secondary sequelae affecting the central nervous system, the circulatory system, the lymphatic system and the skeletal system (Grosset, 2003; Herrmann, *et al.*, 2005; Houben, *et al.*, 2006). The mycobacterium itself is spread through the air and, whilst *M. tuberculosis* is highly infectious, only 3-15% of infected people display signs of active disease, with the majority of the people asymptotically infected with *M. tuberculosis* comprising a latent TB infection (LTBI) (Herskovitz, *et al.*, 2008; Parker, *et al.*, 2007). Infection with *M. tuberculosis* is predominantly diagnosed through sputum smear microscopy via Zeihl-Neelson staining techniques (Figure 1.1a), chest x-rays, culture based methods and mantoux testing.

Two decades ago, it was considered that the spread and treatment of tuberculosis was under control, and it was predicted that this disease would be eradicated in the near future (WHO, 2015). However due to the emergence of drug resistant strains of *M. tuberculosis* and the increasing HIV epidemic, this is no longer the case (WHO, 2015). The lack of drug compliance due to the long-term treatment of *M. tuberculosis* in diseased patients has resulted in the emergence and the spread of multiple drug resistant strains of tuberculosis (MDR-TB), and more recently, extremely drug resistant strains of tuberculosis (XDR-TB) (Figure 1.1b). Immuno-compromised patients diagnosed with HIV have higher susceptibility to *M. tuberculosis*, with deaths usually arising from TB (Corbett, *et al.*, 2003). The World Health Organization (WHO) declared TB to be a global public health emergency in 1993, and has since adopted The Stop TB Strategy, which aims to establish an internationally agreed strategy for TB control, and to halt and reduce the TB epidemic by 2015. While the program is on track to reducing mortality rates by 50% by 2015 compared to 1990, the global TB burden remains substantial. The emergence and prevalence of drug resistant strains of TB have resulted in difficulties in our ability to combat and protect against the spread of the disease, increasing the burden on global health of TB. There is now an urgent need for new therapies to treat TB.

A



B

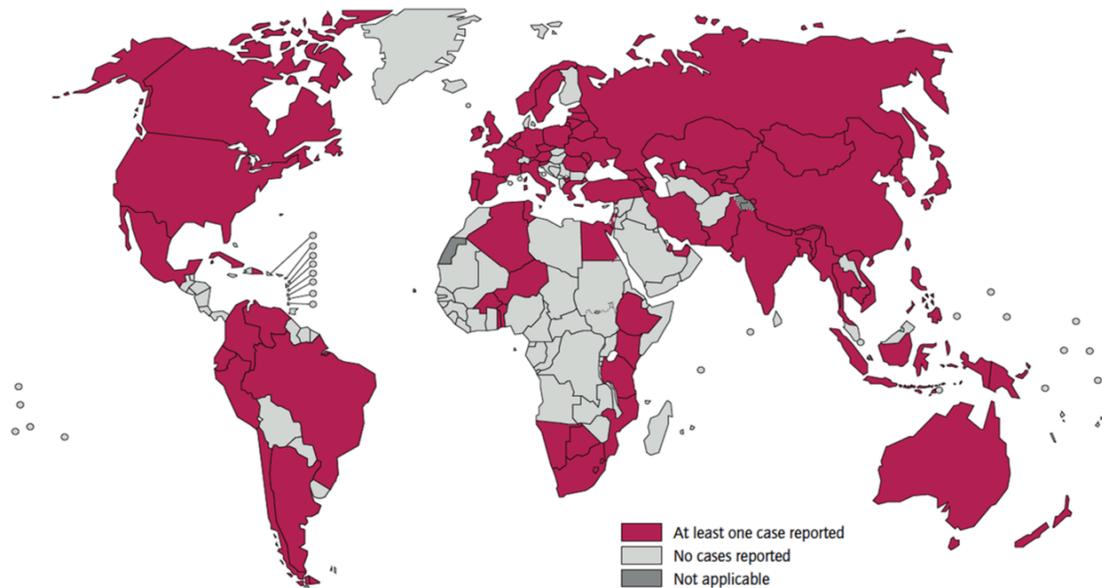


Figure 1.1: *M. tuberculosis* detection and XDR-TB incidence map

- (A) Zeihl-Neelson stain of *Mycobacterium tuberculosis*. Mycobacterium are stained in pink, with negative background tissue stained in blue (CDC, 1979)
- (B) Cases of XDR-TB reported by 2013. Countries indicated in dark red have reported at least one case of XDR-TB by 2013, with countries in light grey reporting no cases of XDR-TB (WHO, 2015)

1.2: Current prevention and therapeutic treatments

New patients diagnosed with TB undergo an intricate course of anti-mycobacterial treatment lasting six to nine months utilizing a variety of drugs. The general course of anti-mycobacterial treatment employs four first line drugs, taken in various combinations for optimal efficiency in eradicating the bacterium, and as a means of preventing drug resistance. The current regimen, as outlined by the WHO, recommends the use of the four first line orally administered drugs; Isoniazid (INH), Rifampicin, Ethambutol (EMB) and Pyrazinamide, each to be administered daily for two months in an intensive course, with a further continuous course of four months of daily administration of Isoniazid and Rifampicin (WHO, 2015). This regimen is known as directly observed treatment, short-course, or DOTS, and is recommended by the WHO as the most cost efficient means to preventing the spread of TB via treatment (WHO, 2015)

1.2.1: Isoniazid

Each of the four first line drugs act on different targets on *M. tuberculosis*, where the combined actions of each drug acts to eliminate the mycobacteria from the patient (Figure 1.2). INH is a hydrazide that requires activation by the mycobacterium itself, and acts to inhibit mycolic acid synthesis, whereupon mycolic acid represents a key component of the mycobacterial cell wall. Upon passive entry into mycobacterial cells, the INH prodrug is activated by the mycobacterial enzyme KatG: a multipurpose catalase-peroxidase. KatG activates INH by peroxidation that generates reactive species that form high levels of adducts with NAD⁺ and NAD(P), which in turn act to damage mycolic acid and nucleic acid biosynthetic enzymes (Timmins, *et al.*, 2006). This INH-NAD adduct is a potent inhibitor of InhA, an enoyl-acyl carrier protein reductase which plays a key role in mycolic acid synthesis (Figure 1.3a) (Quémard *et al.*, 1995; Dessen, *et al.*, 1995).

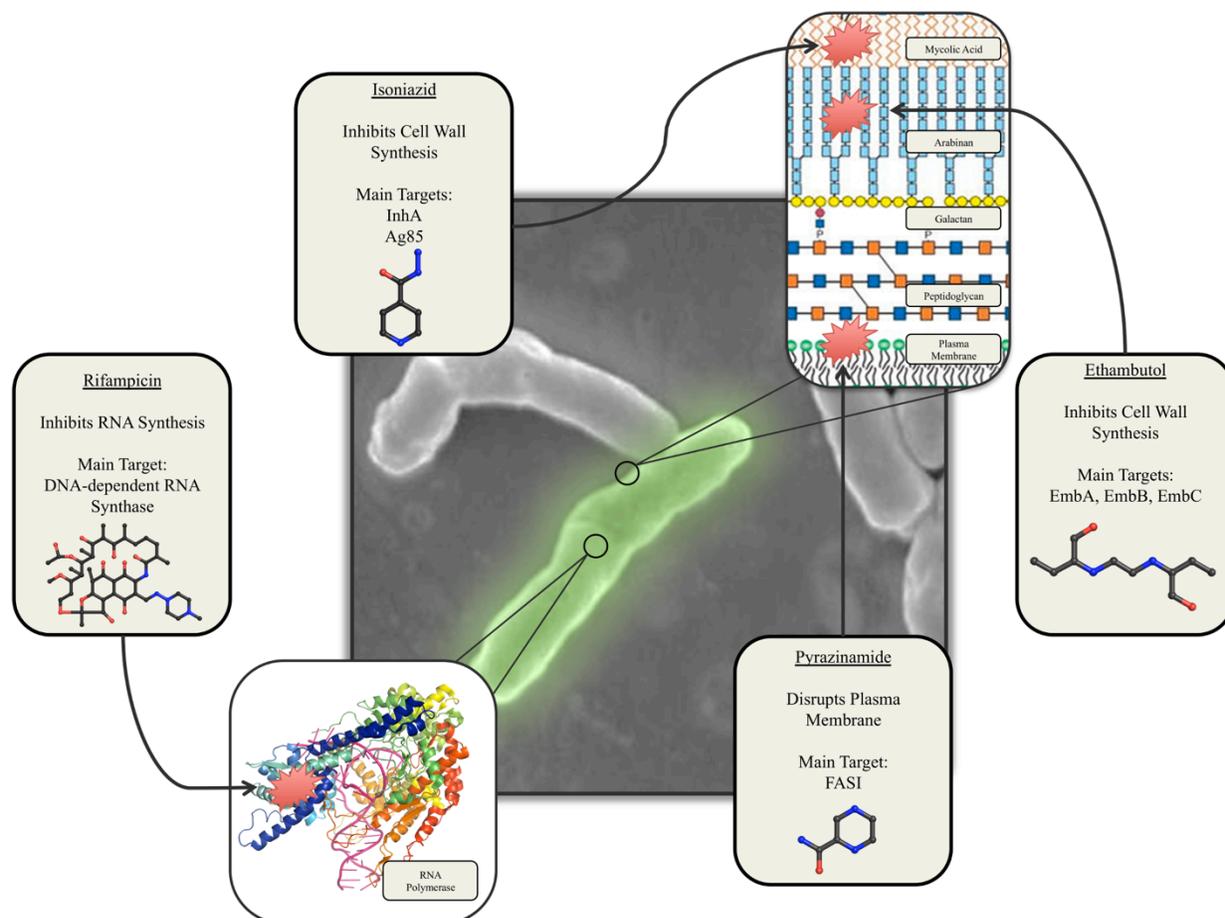


Figure 1.2: TB First-line drug mechanisms of action against TB.

The first four line drugs, Isoniazid, Rifampicin, Ethambutol and Pyrazinamide targets and mechanisms of actions are indicated, including structural representation of each drug obtained from the DrugBank (www.drugbank.ca). Isoniazid, Ethambutol and Pyrazinamide each act on targets within the mycobacterial cell wall and plasma membrane, indicated in the top right insert, with the target of Rifampicin indicated in the bottom left insert. Crystal structure of Bacteriophage T7 RNA Polymerase used as a model structure (PDB ID: 1MSW)

Resistance to INH is quickly established by the mycobacterium, so the drug is always used in combination with another drug to prevent resistance from occurring (Timmins & Deretic, 2006). Resistance to INH occurs predominantly as a result of mutation or deletion of the *KatG* gene, which prevents the activation of INH to play its role as an anti-mycobacterial (Stoeckle, *et al.*, 1993; Rozwarski *et al.*, 1998). INH may also cause a variety of side effects such as hepatitis (liver inflammation), rash formation, and nerve damage resulting in peripheral neuropathy due to the production of these INH derived oxidative free radicals.

It was previously considered that the sole target of INH was InhA, however this is no longer the case. Recently, a total of 17 INH drug targets have been identified, with targets including a range of enzymes essential for mycobacterial survival, and include a number of oxidoreductases, nucleosidases and universal stress proteins (Argyrou, Jin, *et al.*, 2006). Despite a number of targets identified, the mechanism of INH inhibition is presently unknown, with the exact targets still unclear. For instance, inhibition against the dihydrofolate reductase encoded by *DfrA* involved in nucleic acid synthesis by INH-NAD has been reported (Argyrou, Vetting, *et al.*, 2006). However a further study has revealed that DfrA is not directly inhibited by INH (Wang, *et al.*, 2010). Further characterization of these additional INH targets will provide a further understanding of the mechanisms of INH inhibition and a basis for INH resistance.

1.2.1.1: *Rv2971: An essential oxidoreductase*

One target of INH inhibition is the essential aldo-keto reductase (AKR) AKR5H1, encoded by the gene *Rv2971*. *Rv2971* has previously been characterised as an essential gene for growth and survival in *M. tuberculosis* (Sasseti, *et al.*, 2003) with a potential role in the detoxification of toxic metabolites

(Grimshaw, 1992; Grant, *et al.*, 2003; Penning, Drury, 2007). The AKR protein family catalyses the NADPH-dependant oxidation & reduction of a broad range of substrates, including aldehydes, ketones and steroid hormones. While the exact function of Rv2971 is presently unknown, it has previously been characterised to hydrolyse a number of aldehydes in an NADPH-dependant manner (Scoble, *et al.*, 2010).

To gain insight into the function of Rv2971, the crystal structure of its *M. smegmatis* orthologue, MSMEG_2407 (65% sequence identity) has been determined (Figure 1.3b) (Scoble, *et al.*, 2010). The crystal structure revealed an overall triosephosphate (TIM) barrel fold, comprised of eight alternating α -helices and eight β -sheets that form a barrel like structure, characteristic of gluconic acid reductases. The crystal structure of MSMEG_2407 was determined in its unliganded form (APO) and in its holoenzyme form in the presence of the NADPH cofactor (Figure 1.3b). The structure of the holoenzyme reveals the architecture of the NADPH binding pocket and key residues involved in direct contact with the cofactor. In addition, this provides insight into the mechanisms of inhibition by the INH-NADPH adduct.

Despite the high sequence identity between MSMEG_2407 and Rv2971, differences in substrate specificities and INH inhibition levels exist. Namely, INH has previously been reported to inhibit MSMEG_2407 approximately 5 times tighter than Rv2971, with K_i values of 6 μ M and 31 μ M reported respectively (Scoble, *et al.*, 2010). This is presumably a result of subtle differences in the substrate and cofactor binding pockets between MSMEG_2407 and Rv2971, with 10 amino-acid differences identified. The crystal structure of Rv2971 will allow for a more in depth analysis of the substrate and cofactor-binding pocket, and to allow for further characterization into its biological function.

1.2.2: Rifampicin

Rifampicin is primarily a mycobacterial antibiotic and acts to inhibit DNA-dependant RNA polymerase, inhibiting RNA transcription within the mycobacterium. Rifampicin binds directly to DNA-dependant RNA polymerase, forming a stable complex. This inhibits the RNA polymerase, preventing RNA translation and peptide formation (Hartmann, *et al.*, 1967). This drug may also cause a variety of side effects, such as hepatitis and liver damage, along with flu like symptoms and fever. Rifampicin and INH are the two most potent first line TB drugs (Gagneux, 2009). In the case of treatment against drug resistant strains of TB, rifabutin, a semi-synthetic derivative of rifampicin, is used as a replacement first line drug; however due to the greater availability and reduced cost of rifampicin, it is not widely administered (Bass Jr, *et al.*, 1994).

1.2.3 Ethambutol

Ethambutol is a bacteriostatic drug that acts to inhibit arabinogalactan, a key component of the mycobacterial cell wall that is synthesised in actively growing mycobacteria (Rallis, *et al.*, 2009). The drug was shown to act during the synthesis of arabinogalactan, and while the exact mode of action is unknown, a variety of targets have been identified. The current known targets are all encoded by the *emb* cluster of genes, which consists of the *embA*, *embB* and *embC* genes, each encoding products involved in the synthesis of the mycobacterial cell wall. EmbA and EmbB are both involved in arabinogalactan synthesis (Escuyer, *et al.*, 2001) and are predicted to work together as a heterodimer (Amin, *et al.*, 2008), whereas EmbC is involved in lipoarabinomannan (LAM) synthesis (Goude, *et al.*, 2009). The inhibition of these three Emb proteins by ethambutol decreases arabinogalactan content within the mycobacterial cell wall, which leads to an increase in cell wall permeability and decreased cell viability (Lee, *et al.*, 1995). Resistance may also arise against ethambutol as a result of mutations

in the *embB* gene, reducing ethambutol specificity against embB (Alcaide, *et al.*, 1997). As with INH and rifampicin, ethambutol may also cause a variety of side effects, such as optic nerve inflammation (optic neuritis) and red/green colour blindness.

1.2.4 Pyrazinamide

Pyrazinamide, a nicotinamide analogue prodrug, is administered within the first two months of intensive treatment as a means of reducing overall treatment time to six months. *M. tuberculosis* is the only mycobacterial species that pyrazinamide is active against, where alternate species of mycobacteria such as *M. leprae*, the causative agent of leprosy, is innately immune to the drug's effects. Upon passive entry into the mycobacterial cell, the pyrazinamide prodrug is hydrolysed by the enzyme pyrazinamidase, releasing pyrazinoic acid (Konno, *et al.*, 1967). The pyrazinoic acid product acts to inhibit the activity of the Fatty Acid Synthase I system (Zimhony, *et al.*, 2000), involved in biosynthesis of C₁₆ to C_{24/26} fatty acids from acetyl-CoA (Brindley, *et al.*, 1969), which are major constituents cell plasma membrane. While the mechanisms of inhibition are known, the direct target of the drug is presently unknown, with the FAS-I enzyme not being directly inhibited by the drug (Boshoff, *et al.*, 2002). Resistance may also arise against pyrazinamide as a result of mutations of the gene encoding the pyrazinamidase enzyme, abolishing its hydrolytic activity (Scorpio, *et al.*, 1996).

1.2.5 Characterization and treatment of drug resistant *M. tuberculosis* strains

For TB cases exhibiting infection by a drug resistant strain of *M. tuberculosis*, which accounts for approximately 15% of new TB cases, the treatment regimen greatly differs. Multiple drug resistance is defined by *M. tuberculosis* strains exhibiting resistance to at least isoniazid and rifampicin, the two key first line drugs. Treatment against drug resistant strains of TB employs the use of a number of

injectable antibiotics, including kanamycin and streptomycin, and a number of second line drugs, including a range of fluoroquinolones and bacteriostatic agents. The second line drugs are less cost effective compared to the first line drugs, and can result in serious side effects. Due to the inability in the usage of a number of first line drugs, treatment regimens against drug resistant strains of TB are extended, and can last as long as 24 months. Due to the extended treatment of drug resistant strains, as well as mismanagement of drug resistance patient treatment, XDR-TB strains can quickly become established. XDR-TB is defined as exhibiting resistance to isoniazid and rifampicin, as well as resistance to injectable drugs and fluoroquinolones, which greatly limits treatment options to the less effective drugs (Raviglione and Smith, 2007).

With the emergence of these new drug resistant strains, new TB treatment regimens are required, employing the use of new anti-mycobacterial drugs. For this, new drug targets are required for the development of new therapeutics for the treatment of TB. Due to the success of isoniazid and ethambutol, which both act on targets involved in mycobacterial cell wall synthesis, alternate components and enzymes located in the cell wall, could provide a multitude of novel targets for development of new anti-tuberculosis drugs.

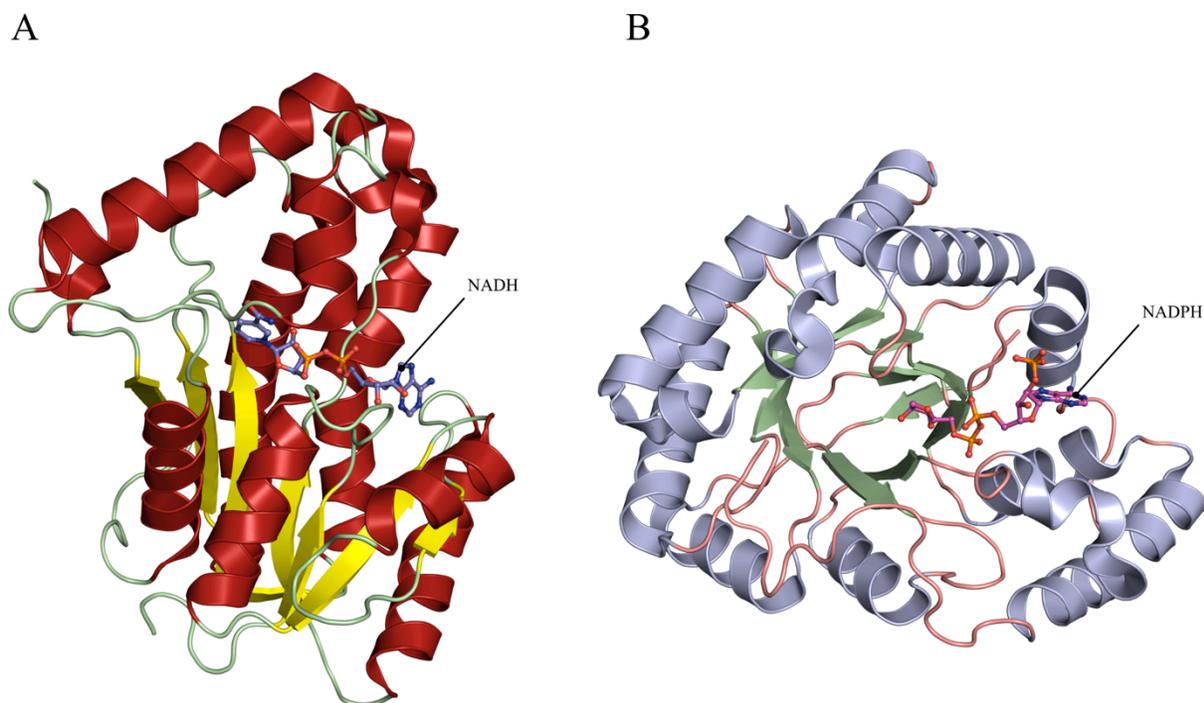


Figure 1.3: Crystal structures of INH drug targets.

(A): Cartoon representation of the crystal structure of InhA (Rv1484); the primary target of the drug isoniazid. The structure is colour coded based on secondary structure, with the bound NADH cofactor indicated. (PDB ID: 3OEW)

(B): Cartoon representation of the crystal structure of ARK5H1 (MSMEG_2407) from *M. smegmatis*: The orthologue of Rv2971 from *M. tuberculosis*. Rv2971 is an essential aldo-keto reductase recently identified as an INH drug target. Crystal structure is colour coded based on secondary structure, with the bound NADPH cofactor indicated. (PDB ID: 2WZM)

1.3: The Cell Wall of *Mycobacterium tuberculosis*

The cell wall of *M. tuberculosis* is the key element that allows for its persistence and success within its host, allowing for the evasion of the immune system and internal survival within macrophages: the host reservoir of pathogenic mycobacteria. The cell wall is categorised as having a complex, lipid-rich cell wall composed of two key segments: The lower segment and upper segment (Brennan, 2003). The lower segment is termed the cell wall “core”, or the mycolyl arabinogalactan-peptidoglycan (mAGP) complex. Beyond the plasma membrane of the mycobacterium, the peptidoglycan layer is covalently attached via a phosphodiester bridge to arabinogalactan, which in turn is linked to mycolic acids. These mycolic acids are composed of a long meromycolate and a short α -chain attached to a proximal cyclopropane ring (Crick, *et al.*, 2001; Brennan, 2003; Takayama, *et al.*, 2005). The upper segment is more associated with the mycobacterial capsule, and is primarily composed of free lipids and cell wall proteins. The free lipids contained within this segment of the cell wall either have long fatty acid chains complementing short α -chains, or short fatty acid chains complementing long α -chains, which displays a diverse array of free lipids present within the mycobacterial cell wall (Brennan, 2003). The components of the upper layer are categorised into a variety of groups. These groups include cord factor/dimycolyltrehalose (TDM), sulfolipids, phthiocerol dimycocerosate (DIM) and the group containing free lipids such as phosphatidylinositol mannosides (PIMs), lipomannan (LM), and lipoarabinomannan (LAM) (Brennan, 2003; Kremer, *et al.*, 2005). Through cell wall disruption studies, these free lipids and cell wall proteins are described as signalling and effector molecules involved in the disease process, whereas the mAGP complex is involved in forming a stable scaffold for the outer segment of the cell wall and increasing viability of the mycobacterial structure (Brennan, 2003) (Figure 1.4).

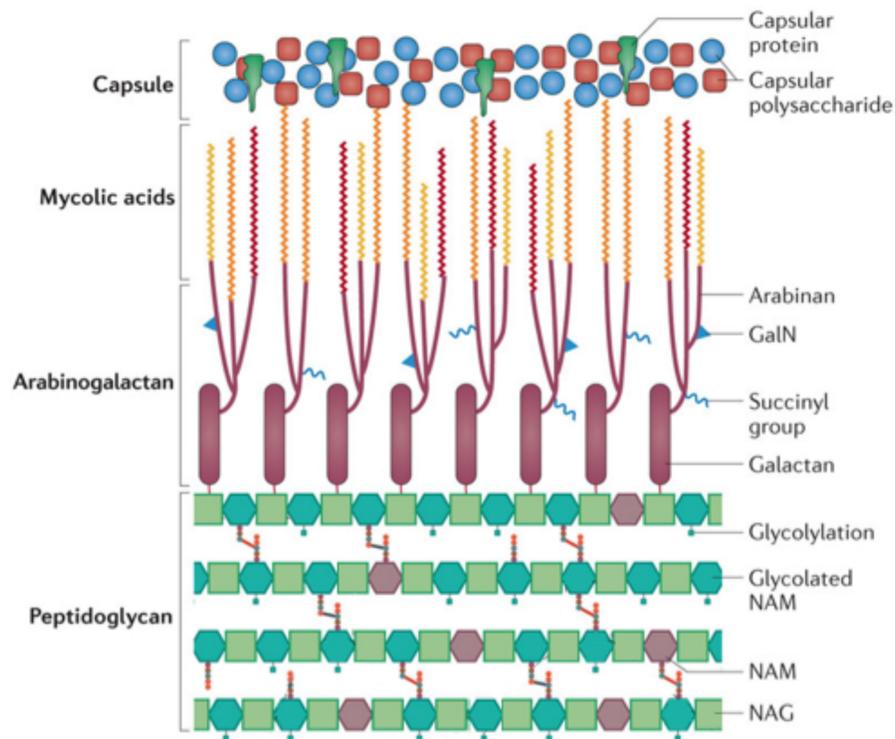


Figure 1.4: Schematic representation of the mycobacterial cell wall.

Key components of the mycolyl arabinogalactan-peptidoglycan core (mAGP core), made up by the peptidoglycan, arabinogalactan and mycolic acid layers, are emphasised. Sugar groups GalN (acetylated galactosamine), NAM (*N*-acetylmuramic acid) and NAG (*N*-acetylglucosamine) within the mAGP core are highlighted. (Adapted from Keiser and Rubin, 2014)

Members of the groups mentioned above play a large number of important roles in pathogenicity and mycobacterial structure viability. Cord factor/TDM's high toxicity, as observed in mice, is its defining characteristic, and has been hypothesised that it stimulates host NADase activity, which decreases levels of NAD within host cells, ultimately reducing activity of a multitude of NAD-dependant enzymes (Brennan 2003). This has devastating effects on the host cell, for example disrupting electron flow through host cell mitochondria that in turn negatively impacts on cellular respiration. Interestingly, TDMs are actually free mycolic acids present in the upper segment of the mycobacterial cell wall, further reinforcing the importance of the presence of mycolic acids (Besra, *et al.*, 1994). Mycobacterial sulfolipids are also attributed to pathogenicity, and are hypothesised to be involved in phagosome-lysosome fusion, preventing *M. tuberculosis* containing phagosomes from fusing to lysosomes within macrophages (Brennan 2003). Phthiocerol dimycocerosates are a major lipid in the mycobacterial cell wall. They are highly apolar, wax-like lipids and are the cause for the waxy characteristics of the mycobacterial cell wall. Finally, the group containing phosphatidylinositol mannosides, lipomannans, and lipoarabinomannans play a variety of roles in pathogenicity (Brennan 2003). For example, LAMs have short mannose-containing oligosaccharide caps that allow it to bind to mannose receptors on macrophages, thus facilitating mycobacterial entry into macrophages (Chatterjee & Khoo, 1998). They also have the ability to bind to toll like receptors (TLRs) and can physically insert themselves into macrophage membranes.

1.3.1 Mycolic Acids

A major component of the mycobacterial cell wall are the mycolic acids, which play a key role in the structural integrity of the cell wall and upper segment, along with protecting the mycobacterium from environmental and host immune oxidative stresses. By definition, mycolic acids are high molecular mass (C₆₀ – C₉₀) β-hydroxy fatty acids with a long α-alkyl side chain, and in *M. tuberculosis*, they are

categorised into three structural groups (Takayama, *et al.*, 2005). Mycolic acids within the bacterium exist either as α -mycolic acids, methoxy-mycolic acids, and keto-mycolic acids, where α -mycolic acids are found to be the most abundant form (>70%) (Takayama, *et al.*, 2005). Mycolic acids are aligned perpendicular to the cell membrane in order to interact with the upper segment of the cell wall, increasing structural stability of the upper segment and serving as a low permeability monolayer (Figure 1.4). Mycolic acids also play a role in *M. tuberculosis* virulence. Deletion studies in all three mycolic acid categories have shown significant attenuation of growth of *M. tuberculosis* in both mouse models and within macrophages (Takayama, *et al.*, 2005). A full understanding of the mycolic acid biosynthesis pathway will allow us to fully understand their roles in the mycobacterium and allow us to characterise the enzymes involved in their synthesis, providing us with a multitude of viable drug targets (Brennan & Crick, 2007).

1.3.2: Enzymes involved in mycolic acid biosynthesis

Extensive research has been conducted over the last 25 years to characterize the mycolic acid biosynthetic pathway, with particular emphasis on identifying the roles of each enzyme within the pathway. An essential gene cluster, spanning 30 genes between *Rv3779c* to *Rv3809c*, comprises open reading frames (ORFs) that code for enzymes involved in the biogenesis of the major components of the *M. tuberculosis* cell wall, including mycolic acids, arabinogalactan and LAMs (Kaur, *et al.*, 2009). A key component of this segment of the *M. tuberculosis* genome is a specific gene cluster containing open reading frames involved in mycolic acid synthesis located from *Rv3799c* to at least *Rv3804c* (Parker, *et al.*, 2009). The gene cluster contains five genes, each coding for enzymes involved in various stages of mycolic acid synthesis. The functions of each of the enzymes produced by these five genes are well understood (Lea-Smith, *et al.*, 2007; Jackson, *et al.*, 1999; Seidel, *et al.*, 2007; Mikusová, *et al.*, 2000; Kremer, *et al.*, 2001), except for the gene product of *Rv3802c*.

Figure 1.5 outlines the current mycolic acid biosynthetic pathway, which involves Fatty Acid Synthase I (FAS-I) and Fatty Acid Synthase II (FAS-II) systems in the production of preliminary mycolyl products. The FAS-I enzyme, encoded by a single gene denoted as *fab* (*Rv2524c*), is a multifunctional enzyme exhibiting all functions for de novo fatty acid synthesis (Smith *et al.*, 2003). The FAS-I multifunctional enzyme produces the C₂₆-CoA fatty acid product that ultimately contributes to the production of the short α -alkyl chain of the final mycolic acid product. The FAS-I system elongates acetyl groups using Acetyl-CoA and Malonyl-CoA as its substrates in five stages to produce the final product. Acetyl and malonyl transacylation converts Acetyl-CoA and Malonyl-CoA into Acetyl-S-CoA and Malonyl-S-CoA respectively, which are processed through a condensation reaction to produce β -ketoacyl-C₄-S-enzyme complex. Via repeated cycles of β -ketoacyl reduction, dehydration and enoyl reduction, the final C₂₆-S-CoA product is produced. The FAS-I system also produces a shorter C₂₀-S-CoA product that provides a starting point for the FAS-II System to produce the long-chain β -hydroxy fatty acids.

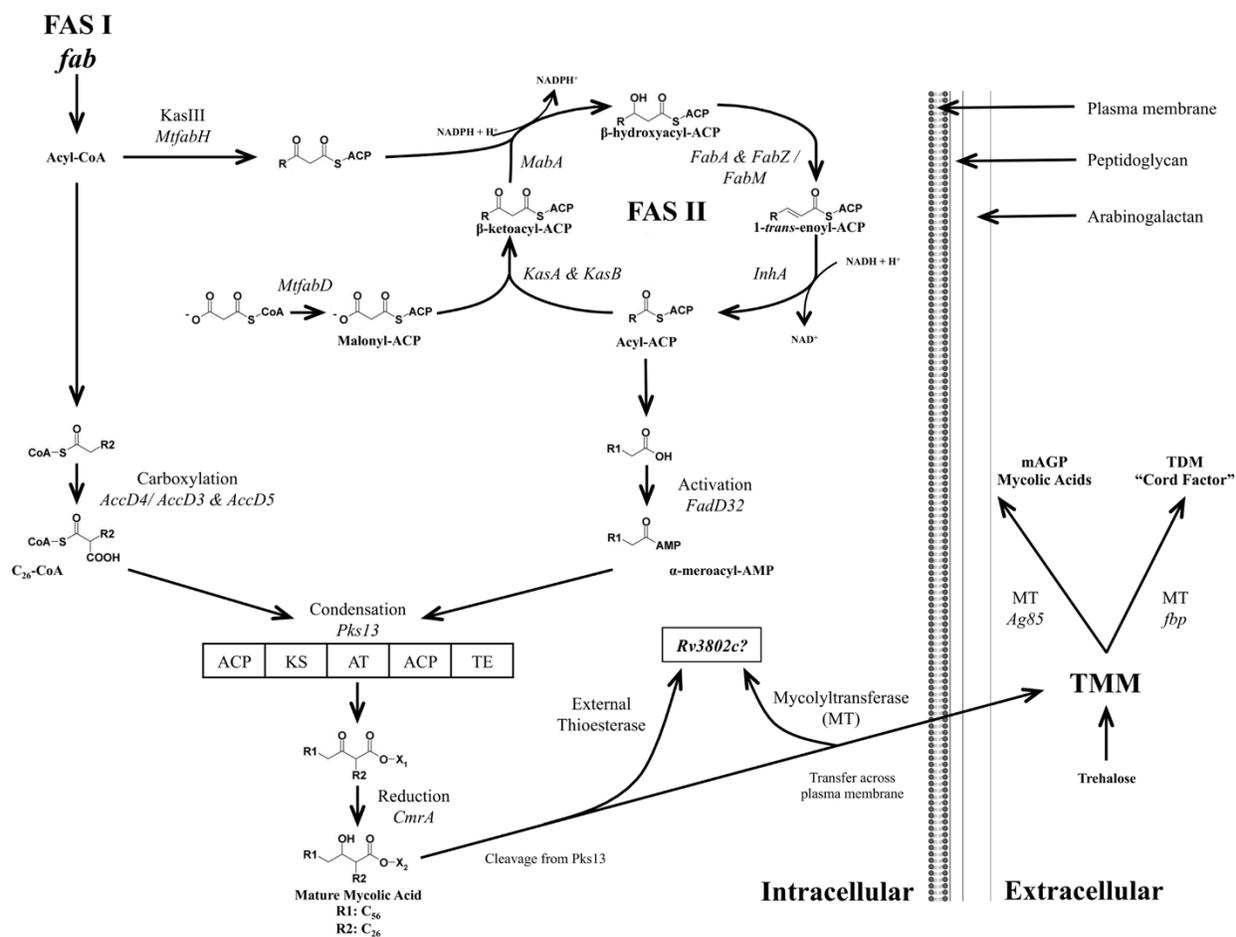


Figure 1.5: Overview of the Mycolic Acid Biosynthetic Pathway.

The schematic diagram highlights the key enzymes and products produced throughout the biosynthesis of mycolic acids, with emphasis on the Fatty Acid Synthase I and II pathways (FASI & FAS II). The final stages of mycolic acid synthesis driven by *Pks13* are indicated. A brief overview of the mycolic acid post-production stages, including attachment to the mAGP core and production of TDM “Cord factor” are additionally indicated. Speculation on the potential role of the essential lipase *Rv3802c* as either an external thioesterase or mycolyltransferase is presented.

The primary focus of the FAS-II system is the synthesis of meroacid for the production of the long-chain β -hydroxy fatty acid tail in mycolic acids. The final FAS-I system product requires transition into the first stage of FAS-II processing utilizing β -ketoacyl-ACP synthase-III (KasIII) encoded by the *MtfabH* gene. KasIII initiates elongation of the C_{20} -S-CoA product to produce β -ketoacyl-S-ACP, which is the starting product of the FAS-II system (Choi, *et al.*, 2000). An additional source of β -ketoacyl-S-ACP is also produced by mtFabD, a malonyl-CoA:ACP transacylase that converts malonyl-CoA to malonyl-S-ACP (Kremer, *et al.*, 2001), which in turn undergoes rounds of acyl extension via the activity of the β -ketoacyl-ACP synthases KasA and KasB (Schaeffer, *et al.*, 2001). The accumulated β -ketoacyl-S-ACP undergoes an NADPH dependant reduction to β -hydroxyacyl-ACP, facilitated by the β -ketoacyl-ACP reductase MabA, encoded by the gene *FabG1* (Marrakchi, *et al.*, 2002). The product undergoes a dehydrase reaction, facilitated by the β -hydroxyacyl-ACP dehydrases FabA and FabZ (Mohan, *et al.*, 1994), yielding *trans*-2-enoyl-ACP, which in turn is isomerized by the 1-*trans*-enoyl-ACP isomerase FabM (Marrakchi, Choi, *et al.*, 2002). The final elongation stage in the cyclic FAS-II system is catalysed by the 2-*trans*-enoyl-ACP reductase, encoded by the gene *InhA*: the key target for the first line anti-TB drug isoniazid (Dessen, *et al.*, 1995; Bianchard, *et al.*, 1995). The reduction of the 2-*trans*-enoyl chains yields the saturated meroacyl-S-ACP, which in turn is processed further to form the β -hydroxy fatty acid component of the mycolic acid (Figure 1.5).

The final stages of mycolic acid biosynthesis involves the processing of the products generated via the FAS-I and FAS-II systems to form mature mycolic acids, which is primarily driven by the gene cluster spanning *Rv2799c* – *Rv3804c*. Before the two products are linked to form the mature mycolic acid, the α -meroacyl-S-ACP and C_{26} -S-CoA products must first be processed into the required substrates for the next stages of biosynthesis. The fatty acyl-AMP ligase FadD32 (*Rv3801c*) catalyses the activation in deriving α -meroacyl-AMP from α -meroacyl-S-ACP, a preliminary product produced from the FAS-II

system. The acyl-CoA carboxylase AccD4 (Rv3799c), along with AccA3 and AccD5 (Rv3285 and Rv3280; not encoded by this gene cluster) are involved in the carboxylation process in deriving 2-carboxyl- C_{26} -S-CoA from C_{26} -S-CoA, a preliminary product from the Fatty Acid Synthase I (FAS-I) system. These two fatty acid chains are then the substrates required to form the mature mycolic acid, and are joined together via thioester bonds and attached to acyl carrier domains onto the polyketide synthase Pks13 (Rv3800c) (Bergeret, *et al.*, 2012). Pks13 is a multifunctional enzyme comprised of five enzymatic domains that catalyses the final stages of mycolic acid biosynthesis via a Claisen-type condensation reaction (Portevin, *et al.*, 2004). Initially, the two substrates are covalently loaded on Pks13, which undergoes a condensation of the 3-oxo group of the meroacyl group transferred from the N-terminal ACP domain to the KS condensing domain (Portevin, *et al.*, 2004). The final stage of mature mycolic acid production is facilitated by the short chain dehydrogenase CmrA, which catalyses the reduction of the 3-oxo group, thus yielding the mature mycolic acid (Bhatt, *et al.*, 2008.)

The mature mycolic acid is cleaved from Pks13 by a presently unknown thioesterase, and transferred to a TMM molecule; a mycolic acid donor (Takayama, *et al.*, 2005). At this stage the mycolyltransferase Ag85 complex, an additional isoniazid drug target, acts to transfer the mature mycolic acid to arabinogalactan to form the mAGP complex (Parker, *et al.*, 2009). This just leaves the function of *Rv3802c* presently unknown. However, it is hypothesized that *Rv3802c* plays the role of either the external thioesterase that cleaves the mature mycolic acid product from Pks13, or as a mycolyltransferases that facilitates transfer of the cleaved mycolic acid through the plasma membrane to a TMM molecule that ultimately contributes to the formation of either the mAGP core component or as free TDM-cord factor.

1.4: Rv3802c – A proposed mycobacterial lipase/thioesterase

Rv3802c has been annotated as belong to a family of putative cutinases (Parker, *et al.*, 2007; Parker, *et al.*, 2009). Cutinases are serine esterases that primarily act on cutin, an insoluble polyester that is a major component of the exterior layer of plants, thereby contributing to their waxy texture (Martinez, *et al.*, 1994; Longhi, Cambillau, 1999; West, *et al.*, 2008). Cutinases are primarily expressed in plant-targeting pathogenic fungi, and are used to break down the cutin polymer into monomeric components, facilitating entry into plants for invasion (Takayama, *et al.*, 2005). However, since the human host of *M. tuberculosis* does not contain any cutin, this suggests that these enzymes in fact act on various other fatty acid containing substrates. Additionally, cutinases exhibit the ability to hydrolyse fatty acid esters and emulsified triglycerols as effectively as lipases. The three dimensional structure of a cutinase from *F. solani* has been solved to a resolution of 1.0 Å (Figure 1.6a), and was shown to be a member of the α/β -hydrolase fold family (Longhi, Cambillau, 1999). Cutinases are the smallest member of the α/β -hydrolase fold family, composed of a single domain with an approximate molecular weight of 22-25kDa (Longhi, Cambillau, 1999; Hotelier, *et al.*, 2004). The enzyme itself contains a hydrophobic core comprising of five parallel-stranded β sheets surrounded by four α helices with a preformed, positively charged oxyanion hole (Martinez, *et al.*, 1994). The oxyanion hole is formed between the Glycine and Serine catalytic residues that acts to stabilise the transient oxyanions (oxygen based anionic species) formed during the processes of acylation and deacylation. The active site of cutinase consists of the catalytic serine residue part of the catalytic triad, Ser120–Asp175–His188, most categorised in serine esterases (Parker, *et al.*, 2007).

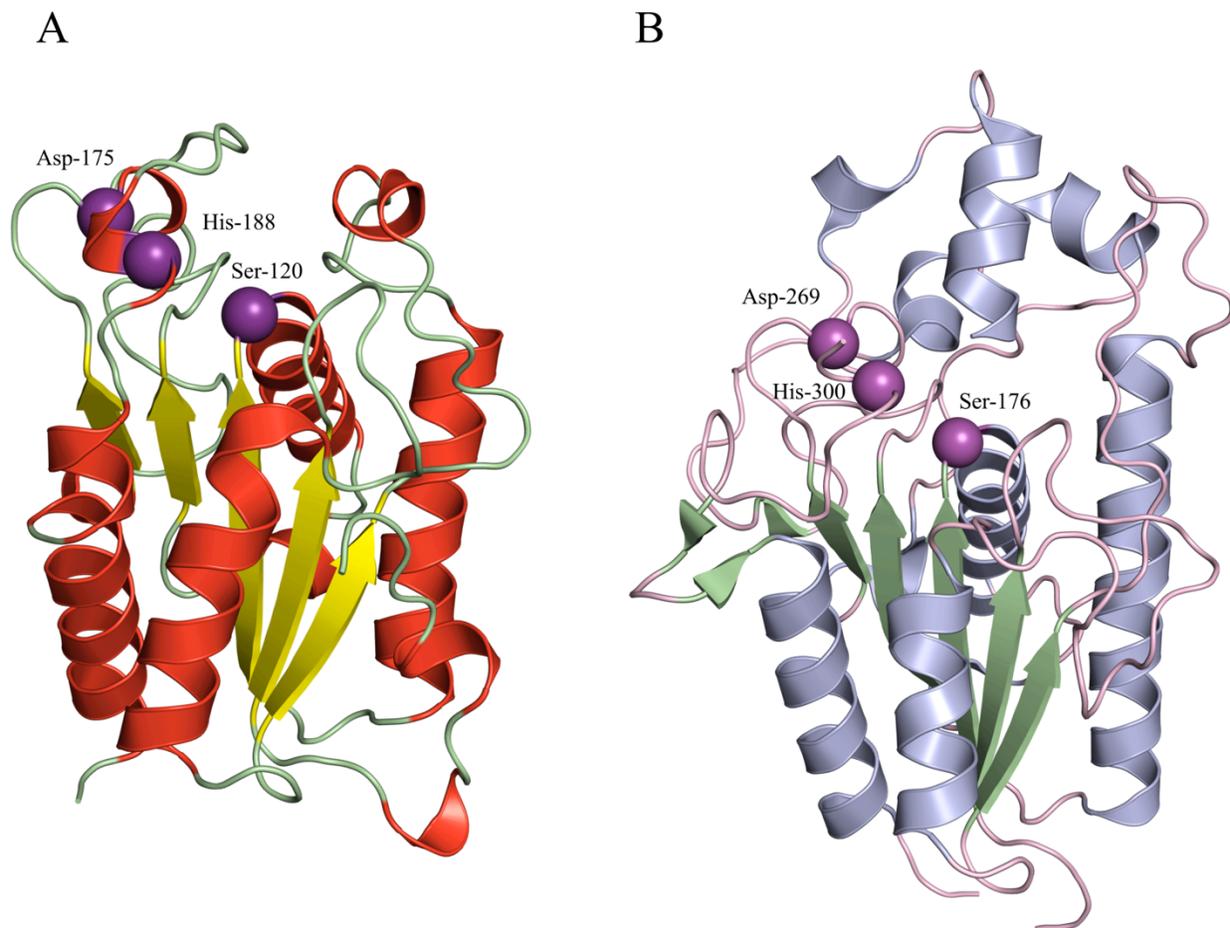


Figure 1.6: Crystal structures of (A) the *F. solani* cutinase (PDB ID: 1OXM) and (B) the essential *M. smegmatis* lipase MSMEG_6394 (PDB ID: 3AJA).

Cartoon representation of each structure highlights the secondary structure composition typical of the α/β -hydrolase superfamily fold, comprised of a single β -sheet flanked by 4-5 α -helices. The catalytic triad typical of serine esterases belonging to the α/β -hydrolase superfamily fold are represented in purple spheres.

Sequence alignment studies have shown that Rv3802c shares high amino acid sequence similarity to its orthologs in a number of mycobacterial species, along with the other annotated cutinases found in *M. tuberculosis*, displaying conservation within their cutinase motifs containing these putative catalytic residues. (Parker, *et al.*, 2007; Parker, *et al.*, 2009). The Rv3802c orthologue in *Corynebacterium glutamicum*, denoted as Ncgl2775, has been found to play a role in the regulation of outer lipid composition under heat stress conditions (Meniche, *et al.*, 2009)

MSMEG_1403, a protein annotated as a putative cutinase found in *M. smegmatis*, has also been characterised to exhibit Phospholipase A (PLA) activity (Parker, *et al.*, 2007). This is reinforced through the same study conducted showing that the proteins exhibiting cutinase motifs observed in *M. tuberculosis* and *M. smegmatis* both correlate with PLA activity. Since known cutinases exhibit PLA activity, it is reasonable to suggest that a protein purified for its PLA activity can be identified as a putative cutinase, which had not been previously observed. These previous studies have shown that both *M. tuberculosis* Rv3802c and *M. smegmatis* MSMEG_1403 both contain phospholipase A (PLA) activity (Parker, *et al.*, 2007). PLAs act on hydrolysing phospholipids into lysophosphatidylcholine and fatty acid subunits that are key precursors for signalling molecules used in a multitude of biological functions (Schaloske, Dennis, 2006). They are then further characterised in groups based on cellular locations (eg, secreted PLAs annotated as sPLA and cytosolic PLAs annotated as cPLA), along with other attributes, such as calcium dependence (eg, Ca²⁺ independent PLAs annotated as iPLA) (Sasseti, *et al.*, 2003; Schaloske, Dennis, 2006). PLAs play a key role in human inflammatory responses and pathogenesis, and are generally observed in snake or bee venom; however, PLAs are also characterised in humans, as well as virulence factors in a few cases of bacterial species (Sasseti, *et al.*, 2003). The functional consequence of Rv3802c exhibiting PLA activity is unclear; however, it may aid in pathogenicity of *M. tuberculosis* via the production of free fatty acids for the progression of inflammation or for the production of effector molecules. PLA activity may even be involved at an

early stage of cell wall synthesis. Purification studies and qualitative enzyme assays have shown that Rv3802c and MSMEG_1403 act on a variety of phospholipid substrates, demonstrating its PLA activity (Parker, *et al.*, 2007). While evidence for PLA activity is present, analysis of the *M. tuberculosis* amino acid sequence, along with BLAST and motif searches, shows no evidence of the presence of PLAs in its proteome (Parker, *et al.*, 2007; Parker, *et al.*, 2009). Despite this, studies on Rv3802c and MSMEG_1403 have shown that they exhibit associated activity with putative mycobacterial cutinases (Parker, *et al.*, 2007; Parker, *et al.*, 2009).

Recent studies have also noted that both Rv3802c and the MSMEG_1403 also exhibit thioesterase activity, and this is much more prominently observed in Rv3802c (Parker, *et al.*, 2009). Thioesterases are a group of enzymes that specifically hydrolyse thioesters, and are categorised into two classes. In general terms, integrated Type I thioesterases (TE I) domains catalyse a release stage from the “assembly line” of the final product where it is transported from one reaction centre to the next as a thioester linked to a cofactor covalently attached to the thioesterase domain (Devedjiev, *et al.*, 2000; Linne, *et al.*, 2004; Koglin, *et al.*, 2008). An example of this type of thioesterase domain is observed in pks13 (see section 1.4), which is involved in the formation of the mature mycolic acid. Type II thioesterases (TE II) are external stand-alone enzymes with proof reading and repair functions (Devedjiev, *et al.*, 2000; Parker, *et al.*, 2009). While it is unclear which type of thioesterase Rv3802c is, and what role it plays in mycolic acid synthesis, the enzyme has been demonstrated to hydrolyse thioesters through enzyme kinetic activity assays (Parker, *et al.*, 2007).

While there is evidence of Rv3802c displaying phospholipase/thioesterase activity, it is still not substantial enough to determine the exact role of the enzyme. While Rv3802c has only been categorised to exhibit PLA and thioesterase activity through various enzyme kinetic activity assays, it has been annotated as a putative cutinase and is a member of the α/β -hydrolase fold family, which

contains known PLAs, mycolyltransferases and thioesterases (Hotelier, *et al.*, 2004). The importance of Rv3802c is supported by the observation that the *Rv3802c* gene is conserved within the genome of *M. leprae*, which is considered important due to its reduced genome in comparison with *M. tuberculosis*, suggesting a requirement for survival (Vissa and Brennan, 2001; Mahapatra, *et al.*, 2008).

The essentiality of Rv3802c was recently demonstrated through work on MSMEG_6394, a close homologue of Rv3802c in *M. smegmatis* (~70% sequence identity). The creation of a knockout mutant of *M. smegmatis* lacking the *MSMEG_6394* was unsuccessful, implying that the enzyme might be essential to the viability of *M. smegmatis*. To counter this, a temperature conditional mutant was created, where the *MSMEG_6394* gene was disrupted within the mutant chromosome in the presence of a rescue plasmid carrying intact the *MSMEG_6394* gene. (Crellin, *et al.*, 2010). This conditional mutant could reproduce at a permissive temperature of 30°C, but could not reproduce at the non-permissive temperature of 42°C. It was found that the *M. smegmatis* mutant was viable at 30°C (at which the plasmid could replicate), but not at 42°C (where the plasmid could not replicate), indicating the essentiality of the *MSMEG_6394* gene. (Crellin, *et al.*, 2010). These findings were further aided by electron microscopy work, revealing that the loss of *MSMEG_6394* functionality decreased cell viability and increased cell death; with live mutant cells displaying an elongated morphology with rough surface relative to wild type *M. smegmatis* cells (Crellin, *et al.*, 2010). Transmission electron microscopy (TEM) was additionally applied to examine the cell wall and internal details of the *M. smegmatis* mutant. These results verified cell lysis in mutants lacking *MSMEG_6394* expression, with large electron transparent zones present in intact cells. It is considered that these electron transparent zones lead to cell lysis, and any attempts to stain these zones with fluorescent lipid dyes for the detection of lipid bodies were unsuccessful (Crellin, *et al.*, 2010). This recent work shows that while the function of *MSMEG_6394* is unknown, its essentiality has been defined, with lack of *MSMEG_6394* resulting in cell death and atypical cell morphology. These results directly correlate to

the role of Rv3802c in *M. tuberculosis*, with Rv3802c believed to play a fundamental and essential lipase-associated role involved in maintaining mycobacterial cell integrity.

To gain function insight into the role of Rv3802c, the crystal structure of MSMEG_6394 has been recently determined (Crellin, *et al.*, 2010). The crystal structure reveals an α/β hydrolase fold typical of serine esterases, comprised of a single 6-stranded parallel β -sheet flanked by 5 large α -helices (Figure 1.6b). The crystal structure allowed for the characterization of the enzyme substrate-binding pocket, with a focus on the conserved Ser176-Asp269-His300 catalytic triad previously determined to be essential for enzymatic activity (West, *et al.*, 2009). Enzyme kinetic activity assays comparing Rv3802c and MSMEG_6394 have demonstrated a difference in substrate specificities and kinetic activity rates between the two enzymes, due to subtle differences in the enzyme active site architecture.

1.4.1: Tetrahydrolipstatin (THL) inhibition of Rv3802c

Enzymatic inhibition of Rv3802c occurs with the use of a mycolic acid synthesis inhibitor known as tetrahydrolipstatin (THL). The compound itself has been implemented as an inhibitor of pancreatic lipases, which results in the inability of the body to absorb long chain fatty acids, and is sold over the counter as a weight loss agent known as AlliTM (Parker, *et al.*, 2009). THL has also been noted to selectively inhibit thioesterase function of human FAS in tumour cells without impacting on normal, non-cancerous cells. FAS activity is unregulated in prostate cancer cells and within various other types of tumours, allowing for greater cell proliferation and tumour formation. However with the implementation of THL, the action of FAS in cancerous cells has been interfered with, resulting in the induction of apoptosis in cancer cells. With this evidence, THL has become a valid oncology target, and various studies are underway for its use in anti-cancer drug development (Kridel, *et al.*, 2004).

It was shown that inhibition by THL only occurs for Rv3802c and not for the MSMEG_1403 (Parker, *et al.*, 2009). Inhibition of Rv3802c by THL occurs at micromolar concentrations *in vitro*, and demonstrates an inhibition of *M. tuberculosis* growth at <30µg/mL, however no inhibition of MSMEG_1403 occurred even at a 500 M excess (Crellin, *et al.*, 2010; Parker, *et al.*, 2009). THL inhibits Rv3802c greater than the *M. smegmatis* orthologue MSMEG_6394, with K_i values of 0.87 µM and 2.59 µM reported respectively (Crellin, *et al.*, 2010). The exact mechanism of Rv3802c inhibition is unknown, but it is hypothesised that inhibition occurs through similar methods of THL inhibition of Fatty Acid Synthetase Thioesterase (FASTE) domain. Inhibition of FASTE occurs competitively, with the THL palmitic core fitting into the hydrophobic substrate pocket with the hexanoyl tail binding to the catalytic histidine residue (Figure 1.7) (Pemble, *et al.*, 2007). This mode of inhibition significantly reduced hydrolysis, but this process is reversible, as THL is eventually hydrolysed. While it is shown in assays that THL inhibits Rv3802c, we cannot assume that it is its sole target within *M. tuberculosis*, as it is likely that THL inhibits other enzymes present. However the prior association between THL inhibition of Rv3802c provides a starting point of studying inhibition processes of Rv3802c and other *M. tuberculosis* targets of THL.

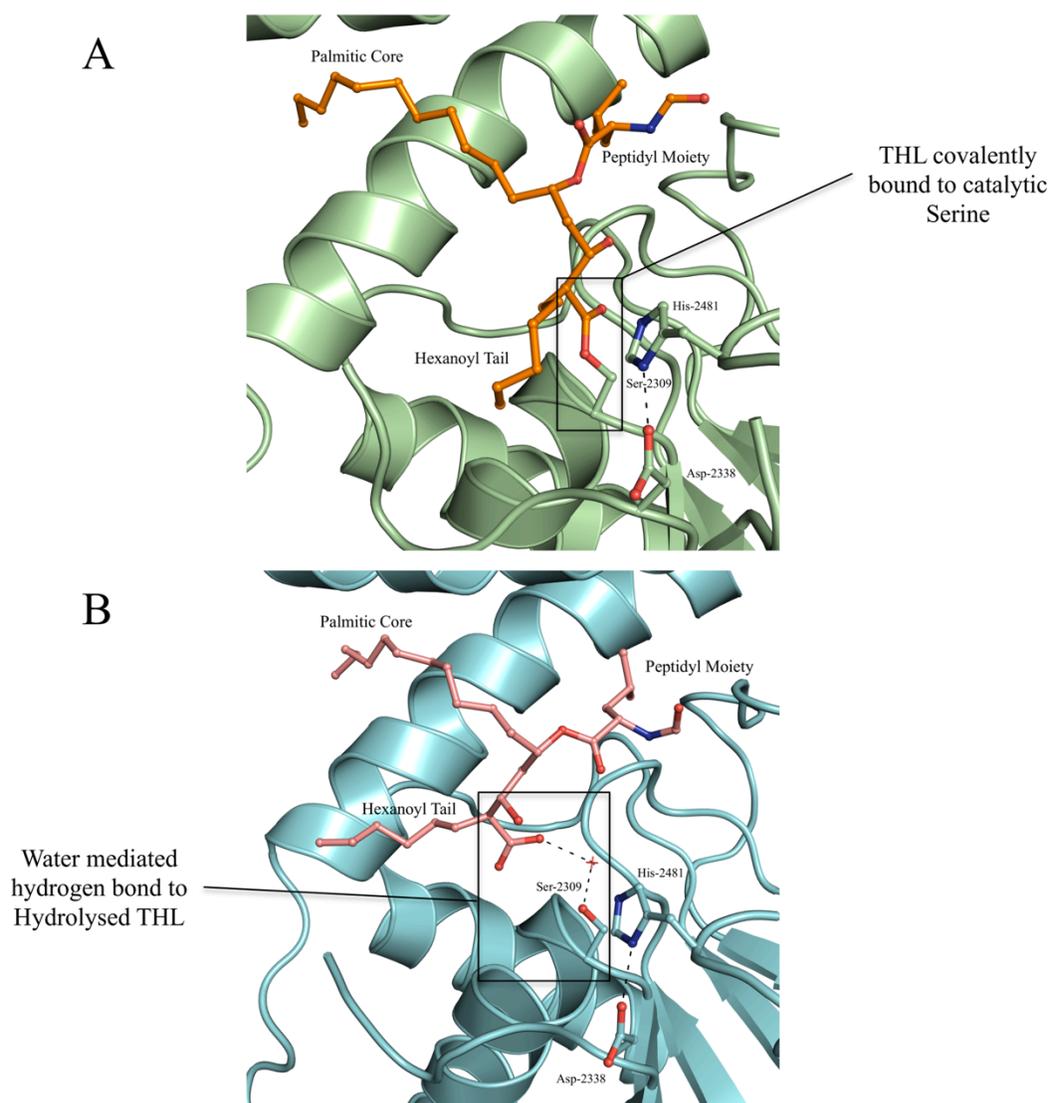


Figure 1.7: A structural basis for tetrahydrolipstatin (THL) inhibition against the human Fatty Acid Synthase II Thioesterase (FASIITE) domain (PDB ID: 2PX6)

- (A) THL is bound in the active site pocket of the FASIITE domain, where the free carboxylate ion is bound covalently to the catalytic serine. Structural components of the THL molecule, the palmitic core, peptidyl moiety and hexanoyl tail, are indicated.
- (B) The covalent bond between THL and FASIITE undergoes a water mediated hydrolytic attack, hydrolysing the THL molecule. The carboxylate ion is now bound to the catalytic serine via a network of water mediated hydrogen bonds.

Recently, a first generation inhibitor library of seventeen synthetic THL analogues was generated, which will further aid in drug design against Rv3802c (West, *et al.*, 2011). Out of the seventeen synthetic THL analogues generated, ten of those inhibited Rv3802c *in vitro* at an affinity much higher than native THL, with a number of analogues exhibiting nanomolar levels of inhibition (West, *et al.*, 2011). In addition to observing a greater level of inhibition, covalent irreversible inhibition was also reported for one of the compounds at concentrations between 1 mM – 5 mM *in vitro*, which exhibited an IC₅₀ value of 0.2 μM, compared to an IC₅₀ of 3.8 μM for native THL (West, *et al.*, 2011). The differences in these compounds lay in chemically altering the *N*-formyl-L-leucine peptidyl moiety side chain, with chemical changes such as the inclusion of lipophilic, aromatic, heteroaromatic and non-aromatic side chains. It was found that compounds bearing lipophilic and large palmityl side chains exhibited a drop in inhibition potency compared to native THL. Interestingly, the compounds containing nitrogen based aromatic side chains exhibited the highest rate of inhibition, with the most potent inhibitors containing *N*-acyl ester side chains. (West, *et al.*, 2011) The structure of Rv3802c in complex with THL would provide insight into how THL inhibits this enzyme. It would form a platform for rational design of more potent inhibitors of Rv3802c.

1.5: *M. tuberculosis* survival within host macrophages

M. tuberculosis pathogenicity relies heavily on the exploitation of the host-cell signalling pathways to allow for the persistence of the pathogen, enhancing its ability for intracellular survival within its host. Its unique cell wall offers pathogenic mycobacteria a physical advantage within the host alveolar lung macrophages; its primary reservoir of infection. In addition, pathogenic mycobacterial species have developed a multitude of mechanisms to parasitise the inhospitable environment of the macrophage, which is central to mycobacterial infection, latency and disease activation (Russell *et al.*, 2001; 2011; Deretic *et al.*, 2006).

Macrophages play a central role in both the innate and adaptive immune systems in the control and eradication of invading pathogenic organisms; fulfilling a range of activities moderated by both cytokine and tissue environment to which the cells are exposed. Macrophages phagocytose invading pathogenic organisms and dead cells and sort them into phagosomes; a complex intracellular compartment that facilitates ingested material eradication.

The hostile and acidic mature phagolysosomal environment facilitates the microbicidal and lytic activity against ingested material via the activity of hydrolytic enzymes, such as cathepsin D (Rossman, *et al.*, 1990), and superoxide bursts driven by the NADPH oxidase complex. The NADPH oxidase complex is an integral membrane complex that assembles on the surface of phagosomes during the process of phagocytosis (El-Benna, *et al.*, 2008; Sumimoto, *et al.*, 2008). Prior to lysosomal fusion, the maturing phagosome interacts with a variety of endosomal compartments that influence the processing and degradation of ingested material (Vieira, *et al.*, 2002; Desjardins, *et al.*, 2005; 2009; Stuart, *et al.*, 2007) (Figure 1.8). The process of endosomal fusion to the phagosome does not occur in a set sequence of events, but belongs to a highly plastic system modulated by extrinsic and intrinsic factors

of the innate and adaptive immune system (Koul *et al.*, 2004; Blander, *et al.*, 2004; Yates, *et al.*, 2007; Trost, *et al.*, 2009). Throughout the process of phagosomal maturation, the internal environment undergoes phagolysosomal acidification, reducing the pH of the phagosomal lumen from a near neutral pH to an acidic pH of 5.0, which facilitates the successful destruction of invading pathogens via the activation of hydrolytic enzymes. Phagosomal acidification is predominantly driven by the fusion of lysosomes, which exhibit an acidic pH of 4.5, as well as the actions of V-ATPase complex embedded in the endosomal membrane. The V-ATPase complex pumps protons into the lumen of the phagosome via the hydrolysis of ATP, rapidly acidifies the internal phagosomal environment (Hackam, *et al.*, 1997).

Upon eradication of the internalized material, degraded material is either ejected from the macrophage or utilized in the presentation of peptide antigens via cell surface antigen presenting Major Histocompatibility Complex (MHC) class II molecules, or lipid antigens via CD1 molecules; a monomorphic family of MHC-I like molecules (Brigl and Brenner, 2004), which in turn act to activate helper T cells for the activation of the adaptive immune system response. The complex membrane fusion events required for normal macrophage activity is highly regulated, and is essential for successful processing of invading pathogens.

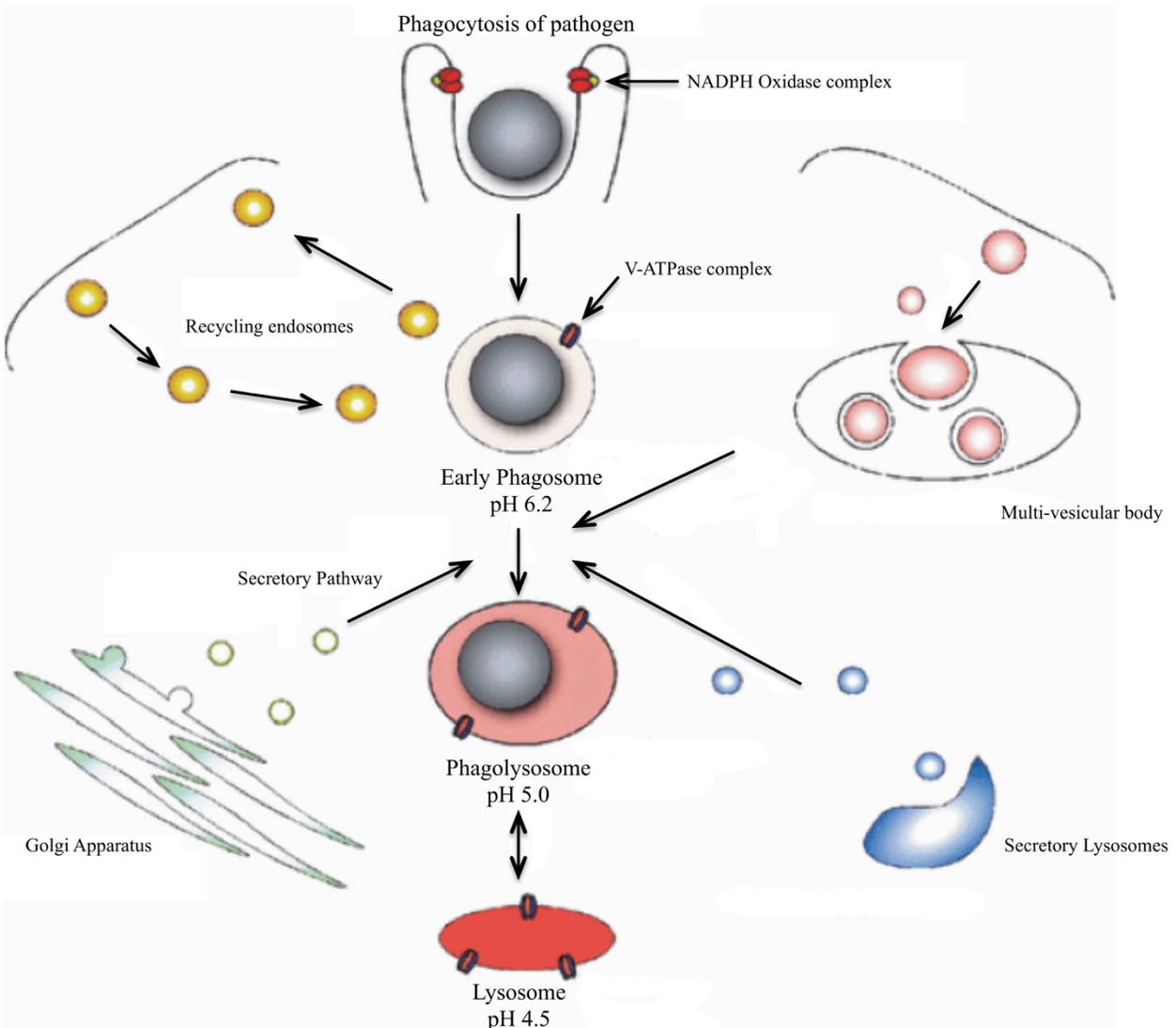


Figure 1.8: Overview of macrophage phagolysosome maturation.

The macrophage engulfs invading pathogens via phagocytosis and sorts them into early phagosomes. Via the modulation of a network of macrophage endosomes and lysosomes, the phagosome matures into the phagolysosome, facilitating the eradication of ingested pathogens. As the phagolysosome matures, the luminal pH drops from a near neutral pH to an acidic pH of 5.0, which facilitates the destruction of ingested pathogens. The membrane embedded V-ATPase complex indicated on the early phagosome also facilitates this acidification. The NADPH oxidase complex is responsible for the generation of reactive oxygen species, which provides a superoxide burst to enable successful macrophage killing. (Adapted from Russel, *et al.*, 2011)

1.5.1 Mechanisms of *M. tuberculosis* survival within host macrophages

Pathogenic *M. tuberculosis* exploits the host-cell signalling pathways that modulate the macrophage internal membrane fusion events to facilitate its survival (Figure 1.9). Disruption of normal macrophage activity enhances the intracellular survival of the pathogen via the disruption of phagolysosome maturation and the suppression of host immune responses. Upon phagocytosis, pathogenic mycobacteria are able to parasitise the internal environment of the phagosome, preventing normal macrophage activity.

Ingested pathogens are first exposed to superoxides, in particular hydrogen peroxide and hypervalent iron atoms, generated by the activity of the NADPH oxidase complex. Mycobacteria possess several enzymes and cell wall components for the avoidance of superoxide activity, which generally have toxic bactericidal activity. The superoxide dismutases SodA and SodC (Piddington, *et al.*, 2001), as well as the catalase peroxidase KatG (Rouse, *et al.*, 1996), the activating enzyme of the INH prodrug, act to convert reactive oxygen and nitrogen species to their molecular intermediates useable by the mycobacterium. Cell wall lipidoglycans, including LAMs (Chan, *et al.*, 1989; 1991) play a role in scavenging reactive oxygen species before entry into the mycobacterium, thus nullifying their bactericidal activity.

The success of the anti-pathogenic activity of the macrophage is the acidification of the phagosomal lumen during maturation, primarily driven by the activity of cell membrane embedded V-ATPase complexes and the fusion of late endosomes and lysosomes. Critical to the survival of pathogenic mycobacteria is its ability to arrest the pathways involved in phagosomal acidification, maintaining a near-neutral pH of 6.2-6.4 within its environment (Sturgill-Koszycki, *et al.*, 1994). Pathogenic mycobacteria achieve this by blocking the proton pump mechanisms of the V-ATPase complex and by manipulating the global phagosomal maturation process. Protein tyrosine phosphatase (PtpA) is

secreted into the phagosomal lumen and has been shown to inhibit phagosomal acidification by blocking V-ATPase complex trafficking and phagosomal acidification, thus enabling the survival of the mycobacterium within the phagosome (Wong, *et al.*, 2011).

Pathogenic mycobacteria modulate the global membrane fusion events involved in phagolysosomal maturation, blocking phagolysosome acidification and facilitating the intracellular survival of the pathogen. Critical to the membrane trafficking events involved in phagolysosomal maturation is the activity of the Rab protein family. Rabs are small GTP binding proteins that function via a molecular on-off switch mechanism that controls the maturation of the phagosome (Clemens, *et al.*, 2000a; 2000b). Through wholesale conversion of Rab5 in early endocytosis to Rab7 in the late endosome, the phagosome matures to allow for correct formation of the phagolysosome. (Kelley and Schorey, 2003; Rink *et al.*, 2005; Pereira-Leal and Seabra, 2001; Pfeffer, 2005; Fratti *et al.*, 2001). This conversion is modulated by the activity of the Rab5 modulator Early Endosomal Antigen 1 (EEA1), which is required to drive fusion of the phagosome with late endosomal and lysosomal compartments, and associates with phosphatidylinositol 3-phosphate (PI3P) on organelle membranes (Sturgill-Koszycki *et al.*, 1994; Vieira, *et al.*, 2004; Fratti *et al.*, 2003). The interaction between EEA1 and PI3P allows for the proper conversion between Rab5 and Rab7 to facilitate correct membrane trafficking within the endosomal system. The Rab7 marker allows for the fusion of late endosomes containing hydrolytic enzymes and lysosome associated membrane proteins (LAMP1), which act as lysosomal markers (Clemens and Horwitz, 1995) for the fusion of lysosomes (Figure 1.9).

Once sorted into the phagosome, pathogenic mycobacteria are able to block the conversion of Rab5 to Rab7, in a process known as Rab conversion block (Clemens, *et al.*, 2000a), utilizing key mechanisms unique to the mycobacteria. The phagosomal maturation pathway is blocked by the activity of the PI3P phosphatase SapM (Saleh and Belisle, 2000; Vergne *et al.*, 2005), and by the cell wall embedded LAM

(Sturgill-Koszycki *et al.*, 1994; Fratti *et al.*, 2003; Hmama *et al.*, 2004; Kang *et al.*, 2005). SapM is a secreted into the phagosomal lumen and upon gaining access to phagosomal PI3P, acts to hydrolyse PI3P, effectively blocking the phagosomal maturation process. SapM acts on the presence of excess PI3P that isn't initially blocked by the presence of mycobacterial LAM's. LAM acts to prevent the generation of PI3P by incorporating into the phagosomal membrane and inhibit the activity of the class III phosphoinositide 3-kinase vacuolar protein sorting 34 (Vsp34), limiting accumulation of PI3P in the phagosomal membrane (Vergne, *et al.*, 2004; 2005), thus blocking early phagosomal maturation.

The recruitment of EEA1 is modulated by the activation of two Ca^{2+} -dependant effector proteins: the sensory Ca^{2+} binding protein Calmodium (CaM) and multifunctional serine/threonine calmodium protein kinase II (CaMKII) (Malik, *et al.*, 2001). During phagosomal maturation, levels of intracellular Ca^{2+} increases to induces a conformational change in CaM which leads to the activation of CaMKII via autophosphorylation, which in acts to recruit EEA1 (Peters and Mayer, 1998). A pathogenic mycobacterium has the ability to interfere with Ca^{2+} signalling to impair the activity of CaMKII as a means of limiting EEA1 recruitment (Malik, *et al.*, 2000). The lipid effector molecule Man-LAM, which acts to inhibit the signalling cascades required to increase intracellular Ca^{2+} , prevents the activation of CaMKII by CaM and ultimately recruitment of EEA1 arresting phagosomal maturation (Rojas, *et al.*, 2000; Fratti, *et al.*, 2003; Vergne, *et al.*, 2003; 2005)

Despite the extensive research conducted on characterising the pathogen's ability to arrest phagosomal maturation, the understanding of the mechanisms of interaction between pathogenic mycobacteria and their host macrophages, with particular emphasis on its host cell machinery, is still incomplete. The identification and characterization of novel targets involved in the internal survival of pathogenic mycobacteria within its host macrophage will lead to a greater understanding of its pathogenesis, paving the way for the development of novel treatments.

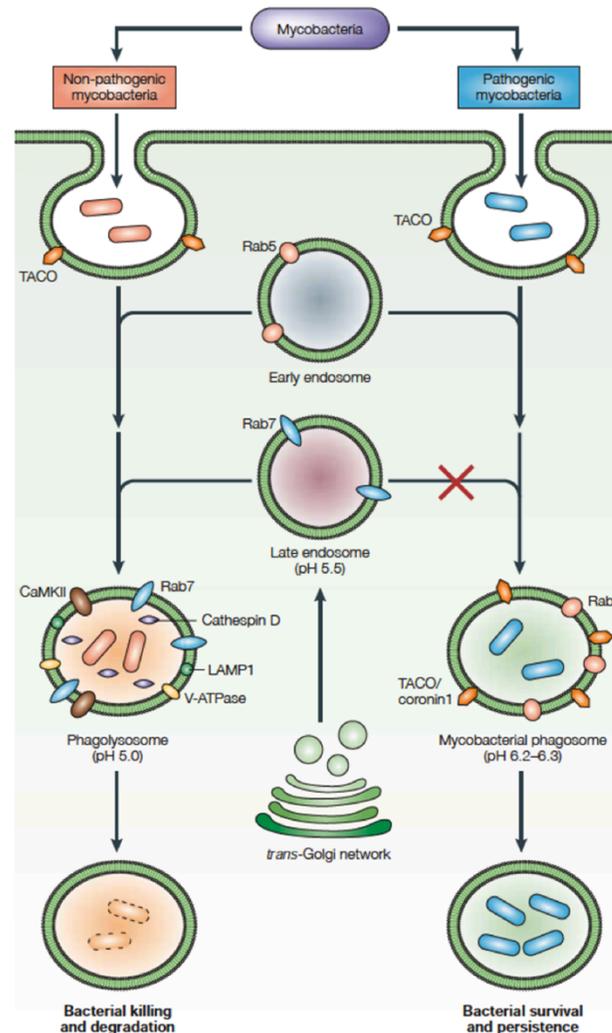


Figure 1.9: Phagosomal maturation in the presence of pathogenic and non-pathogenic mycobacteria.

Pathogenic mycobacteria employ a number of strategies to arrest phagosomal maturation and prolong its survival and persistence. In the presence of non-pathogenic mycobacteria, the wholesale conversion of Rab5 on early endosomes to Rab7 on late endosomes, modulated in part by Ca^{2+} -dependant CaMKII, facilitates the maturation of the phagolysosome for effective bacterial killing and degradation. Acidification of the phagolysosomal lumen to a pH of 5.0 activates hydrolytic enzymes, such as cathepsin D, for the eradication of enclosed material. Pathogenic mycobacteria have the ability to block this Rab5 to Rab7 conversion (red cross), thus preventing the correct maturation of the phagosome. The mycobacterial phagosomes retain a near neutral pH of 6.2-6.3 through this maturation blocking and by disabling the function of V-ATPase complex proton pump. (Koul *et al.*, 2004)

1.6: MSMEG_5817 – a novel drug target implemented in macrophage survival

A recently identified gene in *M. smegmatis*, a saprophytic fast growing mycobacterial species, has been found to be vital for mycobacterial survival within the macrophage. Despite its non-pathogenic nature, *M. smegmatis* has been adopted as a suitable organism for the evaluation of mycobacterial pathogenesis, due to its limited capabilities to persist and multiply within the macrophage environment, delaying phagosomal acidification (Kuehnel, *et al.*, 2001; Anes, *et al.*, 2006). A recent *Tn611* transposon mutagenesis library has been generated for *M. smegmatis*, allowing for the screening of atypical phenotypes in mutants for the identification of novel mycobacterial ORFs (Billman-Jacob, *et al.*, 1999; Patterson, *et al.*, 2000; Kovacevic, *et al.*, 2006). *Tn611*, a modified mobile genetic element transposon exhibiting kanamycin resistance, allows for the antibiotic selection of successfully created mutants (Pérez, *et al.*, 1998). The randomly inserted transposition element is delivered via a thermosensitive-shuttle vector allows for accurate and reproducible mutant generation for genotypic and phenotypic analysis of the *M. smegmatis* genome.

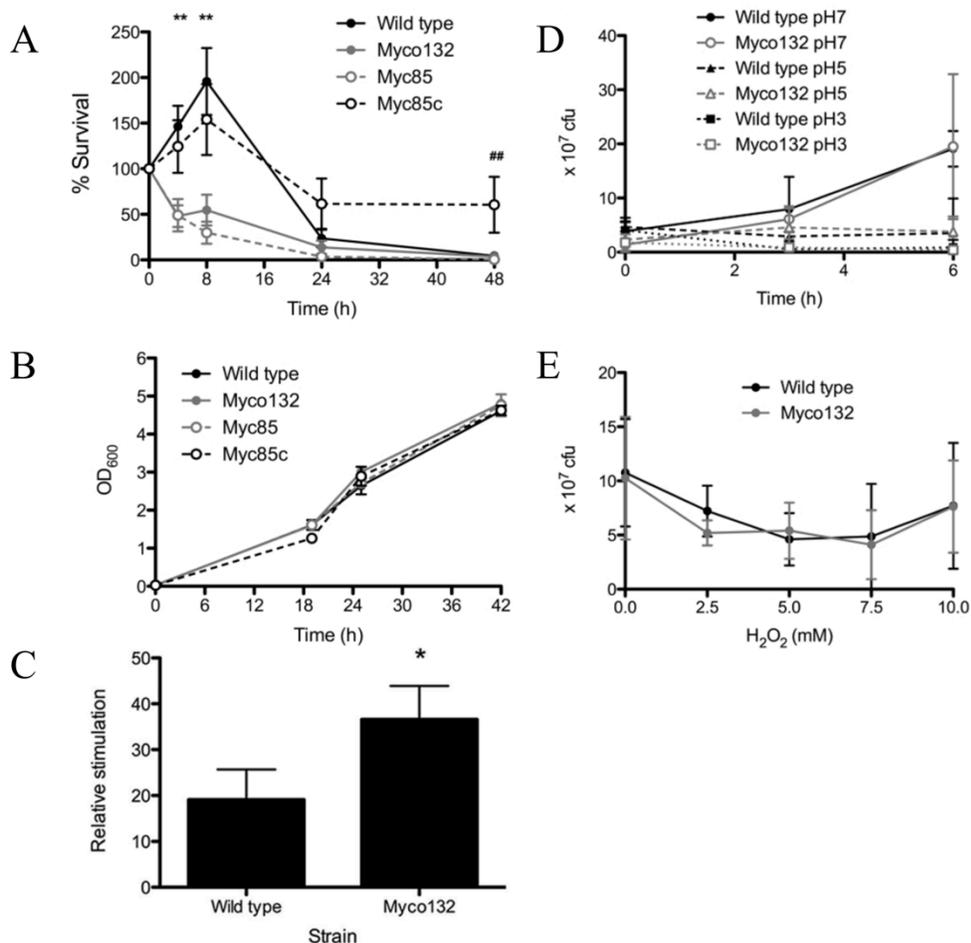


Figure 1.10: *in vivo* growth curves of the *M. smegmatis* transposon mutant Myco132.

- (A) Macrophage survivability of the transposon mutant Myco132 and targeted mutant Myc85 of MSMEGM_5817. Mycobacterial survival within macrophages is drastically reduced of the mutant compared to wild type *M. smegmatis*. Growth restored to near wild type levels upon complementation of the mutant.
- (B) Growth in liquid media shows no difference between MSMEG_5817 mutants and wild type, indicating the mutant is non-essential for normal mycobacterial growth
- (C) Increased NF- κ B activation in mouse model of the Myco132 mutant compared to wild type. * P < 0.05
- (D) Loss of Myco132 mutant survivability in macrophage is responsible due to loss of function, by normal macrophage killing. Growth in liquid medium replicating the acidic environment of the macrophage and (E) in the presence of the reactive oxidative species hydrogen peroxide (H₂O₂) indicate no change in growth between mutant and wild type. (Adapted from Pelosi, *et al.*, 2012)

One such mutant strain generated via this system, denoted as Myco132, exhibited altered colony morphology with the capacity for the uptake of dyes from liquid mycobacterial growth media (Pelosi, *et al.*, 2012). Upon further phenotypic characterization, it was observed that the mutants Myco132 and Myc85, a specific mutant targeting the gene of interest, exhibited no changes in normal growth within liquid media of *M. smegmatis*, indicating non-essentiality for mycobacterial growth. However, upon transfection of the mutant strains into J774A.1 macrophages, a drastic reduction of *M. smegmatis* survival was observed, with an approximate 72% reduction in survival compared to wild type *M. smegmatis*, within the first 8 hours of transfection (Figure 1.10a,b) (Pelosi, *et al.*, 2012). Survival to near wild type levels was observed upon transfection with a complement strain, denoted as Myc85c. To determine whether this decreased survivability of the mutant was as a result of more efficient anti-mycobacterial mechanisms of macrophages, growth sensitivity was assessed comparing wild type *M. smegmatis* to the Myco132 mutant within an acidic pH medium, as well as in the presence of H₂O₂ reactive oxygen intermediates. No differences in growth was observed between wild type *M. smegmatis* and the Myco132 mutants were observed in either low pH media or in the presence of H₂O₂, indicating that the decrease in macrophage survivability was a result of a loss of function (Figure 1.10d,e) (Pelosi, *et al.*, 2012). Its essentiality in the role of mycobacterial pathogenesis was reinforced by the significant increase in NF-κB activation in a mouse model compared to wild type *M. smegmatis* (Figure 1.10c) (Pelosi, *et al.*, 2012). As NF-κB activation plays a significant role in regulating the immune response to infectious agents via cytokine production stimulation, the increased activation of NF-κB within the Myco132 mutant may indicate the gene of interest may play a role in immune response suppression.

The gene disrupted by the transposon element was identified as *MSMEG_5817*. The identified gene shares significant sequence identity with a variety of pathogenic and non-pathogenic mycobacterial species, including in *M. tuberculosis* (Rv0807) and *M. leprae* (ML_2207), sharing a 66% and 62% sequence identity respectively. The ORF *MSMEG_5817* encodes a 128 amino acid protein of an unknown function. The closest sequence homologue to *MSMEG_5817*, a serine:pyruvate aminotransferase (SPT) from *Brevibacterium linens*, was identified via a BLASTp similar search, exhibiting a sequence identity of 46%. Despite sharing a moderately similar amino acid sequence, *MSMEG_5817* shares no domains characteristics of SPTs, and based on a pyruvate reduction assay conducted on the Myco132 mutant, is not functionally related to the SPTs (Pelosi, *et al.*, 2012). In a previous study, it was found that the *Rv0807* gene, the *M. tuberculosis* orthologue, was up-regulated in a deletion mutant of the *senX3-regX3* two component regulatory system (Parish, *et al.*, 2003). The *senX3-regX3* system regulates inorganic phosphate gene expression involved in phosphate acquisition by the mycobacterium essential for virulence, with deletion mutant studies revealing a decrease in mycobacterial survival within host macrophages (Parish, *et al.*, 2003; Glover, *et al.*, 2007). As it stands, the exact biological function of *MSMEG_5817* is unknown, but its importance for the mycobacterial survival within macrophages has been demonstrated, with a potential involvement in the *senX3-regX3* two component regulatory system.

1.7 – Aims of this study

With the prevalence of infection and disease by drug resistant strains of *M. tuberculosis* becoming increasingly wide spread, there is now an urgent need for the development of new anti-mycobacterial therapeutics to tackle this growing problem. The partial success of the pathogen is attributed to its unique lipid cell wall, which contributes to the persistence of the mycobacterium within host cell macrophages. Mycobacterial proteins involved in cell wall biosynthetic pathways and host-macrophage survivability make for attractive drug targets due to their essentiality for mycobacterial pathogenesis. Due to the inhibitory success of the first line drugs isoniazid and ethambutol against enzymes involved in biosynthesis of major cell wall components, alternate mycobacterial enzymes involved in these pathways may provide a variety of novel drug targets. Additionally, in further characterizing the mechanisms of mycobacterial persistence within host macrophages via parasitizing the phagosomal environment will not only provide insight into the mycobacterium's mechanisms of pathogenesis, but would also pave the way for further drug target discovery.

My study's central hypothesis is that the planned structural and functional analysis of novel mycobacterial proteins involved in mycolic acid biosynthesis and host macrophage survivability will provide fundamental insight into their modes of action, as well as provide a basis for rational drug design.

X-ray crystallography has played a pivotal role in the structural characterization of novel mycobacterial proteins in the past, with applications for planned rational drug design. Crystal structures of up to 10% of the overall ORFs present within the *M. tuberculosis* genome have been determined, which have provided a fundamental insight into their functionality and drug inhibition

mechanisms. The technique has previously been utilized to characterise the isoniazid drug target InhA; detailing the structural mechanisms of isoniazid inhibition and isoniazid resistance (Dias, *et al.*, 2007). In addition, the crystal structure of InhA has been solved in complex with a number of new inhibitors, which provide a structural model for the development of new therapeutics targeting the protein. X-ray crystallography has also provided a structural basis of essential mycobacterial functionality. The crystal structure of the AT domain of Pks13, the essential enzyme involved in the late stages of mycolic acid biosynthesis, has previously been determined in the presence of bound natural ligands, providing insight into the mechanisms of action of the Pks13 domain (Bergeret, *et al.*, 2012). These crystal structures of essential mycobacterial enzymes are only a select representation of the potential for structural determination via x-ray crystallography in understanding the pathogenesis of TB.

My study focuses on three specific mycobacterial proteins that have each previously been demonstrated to be essential for survival and pathogenesis of the mycobacterium. The essentiality of these targets has previously been determined through transposon size hybridization via microarray (Sasseti, *et al.*, 2003) and phenotypic analysis (Billman-Jacob, *et al.*, 1999; Patterson, *et al.*, 2000; Kovacevic, *et al.*, 2006) of the mycobacterial genome. The targets are the isoniazid drug target aldo-keto reductase AKR5H1, encoded by the gene Rv2971, Rv3802c; the essential thioesterase involved in the late stages of mycolic acid biosynthesis, and Rv0807; a protein of unknown function essential for mycobacterial survival within host macrophages first characterized in *M. smegmatis*.

X-ray crystallography was utilized to determine the crystal structure of each of the targets, allowing for a structural characterization of cofactor and drug binding pockets, gaining insight into inhibitor binding

mechanisms and providing a basis for rational drug design, as well as gaining insight into protein functionality.

Each chapter presented outlines the cloning, expression, purification, crystallization and structural determination of each of the targets. As permitted under Monash University regulations, the results chapters 2 to 6 are predominantly comprised of peer-reviewed journal articles that were published throughout the course of this study. Each of the studies were published as first-author papers.

Chapter 2 presents an overview of the crystal structure of the isoniazid drug target aldo-keto reductase ARK5H1, with focus on a structural characterization of the NADPH binding pocket architecture and isoniazid binding site. The work described was published in *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* in 2014, and includes the methods of cloning, expression and purification of recombinant AKR5H1 in *Escherichia coli*, with extensive details of its crystallization, data collection and structural determination outlined.

Chapter 3 focuses on the crystal structure of the essential lipase Rv3802c, in complex with the anti-TB drug THL. The crystal structure presented in Chapter 3 provides a structural basis for the targeted rational drug design against Rv3802c, using THL as a lead inhibitor. The work described includes methods of cloning, expression and purification of Rv3802c, with a focus on structural determination and characterization of Rv3802c in complex with THL.

Chapters 4, 5 and 6 focus on the structural and functional characterization of MSMEG_5817: The *M. smegmatis* orthologue of Rv0807. Chapter 4 comprises of a paper published in *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* in 2013, and focuses on the cloning, expression and purification of recombinant MSMEG_5817 from *Escherichia coli*, with a

focus on the crystallisation and x-ray data collection of native and selenomethionine derivatised MSMEG_5817. The work described in chapter 4 describes preliminary work that is expanded on in chapters 5 and 6.

Chapter 5 continues on from the work described in Chapter 4, and presents the crystal structure of MSMEG_5817 determined to 2.40 Å by 3-wavelength multiple anomalous dispersion (MAD) experimental phasing. The structure described presents a fundamental insight into the function of MSMEG_5817, further characterized by biochemical assays and alanine-scanning mutagenesis experiments. The work described in chapter 4 was published in the *Acta Crystallographica Section D: Biological Crystallography in 2014*, with a focus on the applications of x-ray crystallography for the determination of biological function of newly discovered proteins.

Chapter 6 focuses on the application of native source organism protein production for the purposes of x-ray crystallography, and presents the advantages of *M. smegmatis* as an expression system for structural analysis of mycobacterial proteins. The work described was prepared as submitted manuscript for the intention of publishing in *Protein Science* in 2015, and focuses on the cloning, expression and purification of recombinant MSMEG_5817 from *M. smegmatis*, describing a higher quality crystal structure of MSMEG_5817 determined to 1.70 Å.

The work described presents a range of applications for x-ray crystallography in the structural analysis of essential and novel mycobacterial proteins. With the predominant focus of rational drug design, these results demonstrate the applications of x-ray crystallography in characterizing the architecture of novel substrate and inhibitor binding pockets, as well as applications in gaining fundamental insight into the biological function of recently identified mycobacterial proteins.

Chapter 2: Structural investigation into the INH drug target, Rv2971

2.1 Introduction

Isoniazid, the first line drug against TB, inhibits 17 known targets within *M. tuberculosis* (Figure 2.1) (Argyrou *et al.*, 2006). Despite the mechanisms of INH activation being characterized, as well as the functionality of a number of the INH drug targets previously analysed, the mechanisms of INH inhibition is still widely unknown. As a means of further characterizing INH inhibition against its targets, x-ray crystallography has been utilized to gain structural insight into the mechanisms of inhibition.

This chapter focuses on the essential aldo-keto reductase, Rv2971, previously determined to be a drug target of INH (Argyrou *et al.*, 2006). The work described here contains a paper, published in the course of the research for this thesis in *Acta Crystallographica Section F - Structural Biology and Crystallization Communications* in 2014, and focuses on the crystallization and structural determination of Rv2971.

The crystal structure of its *M. smegmatis* orthologue, MSMEG_2407, has recently been determined to 1.60 Å, both in its APO form and in the presence of its NADPH cofactor (Scobles, *et al.*, 2010). The structure revealed an triosephosphate isomerase (α/β)₈ TIM barrel fold, comprised of an 8 bladed β -barrel structure stabilized by a series of α -helices, featuring an N-terminal β -turn lid domain. The crystal structure of the holoenzyme, with NADPH bound, allowed for an in depth structural analysis of the cofactor binding site, and a preliminary analysis of the INH-NADPH adduct binding site.

Subtle differences in the NADPH binding pocket between MSMEG_2407 and Rv2971 have a drastic effect on the substrate specificities and INH inhibition rates of each enzyme, as described in section 1.2. While the crystal structure of MSMEG_2407 does provide valuable insight into the cofactor and INH binding pocket, the differences in binding pocket architecture make this unsuitable for the analysis of INH inhibition of *M. tuberculosis*.

To counteract this, the key aim of this study was to solve the crystal structure of Rv2971, both in its APO and holoenzyme form, to further characterise the mechanisms of INH-NADPH adduct inhibition against the *M. tuberculosis* target. The crystal structure of Rv2971 in its APO form was successfully determined to 1.60 Å. Rv2971 exhibits the same TIM barrel fold as MSMEG_2407, with differences in amino acid sequence having a drastic effect on the NADPH binding pocket architecture. While no crystal structure of the holoenzyme was successfully determined, a number of water molecules and malonate ions originating from the crystallisation condition were found bound to the NADPH binding pocket, with a high degree of hydrogen bond conservation to the MSMEG_2407 holoenzyme crystal structure.

The successful determination of the crystal structure of Rv2971 to 1.60 Å allows for an in-depth analysis of the INH-NADPH adduct binding pocket, with differences in amino acid sequences compared to MSMEG_2407 explaining the differences in INH inhibition.

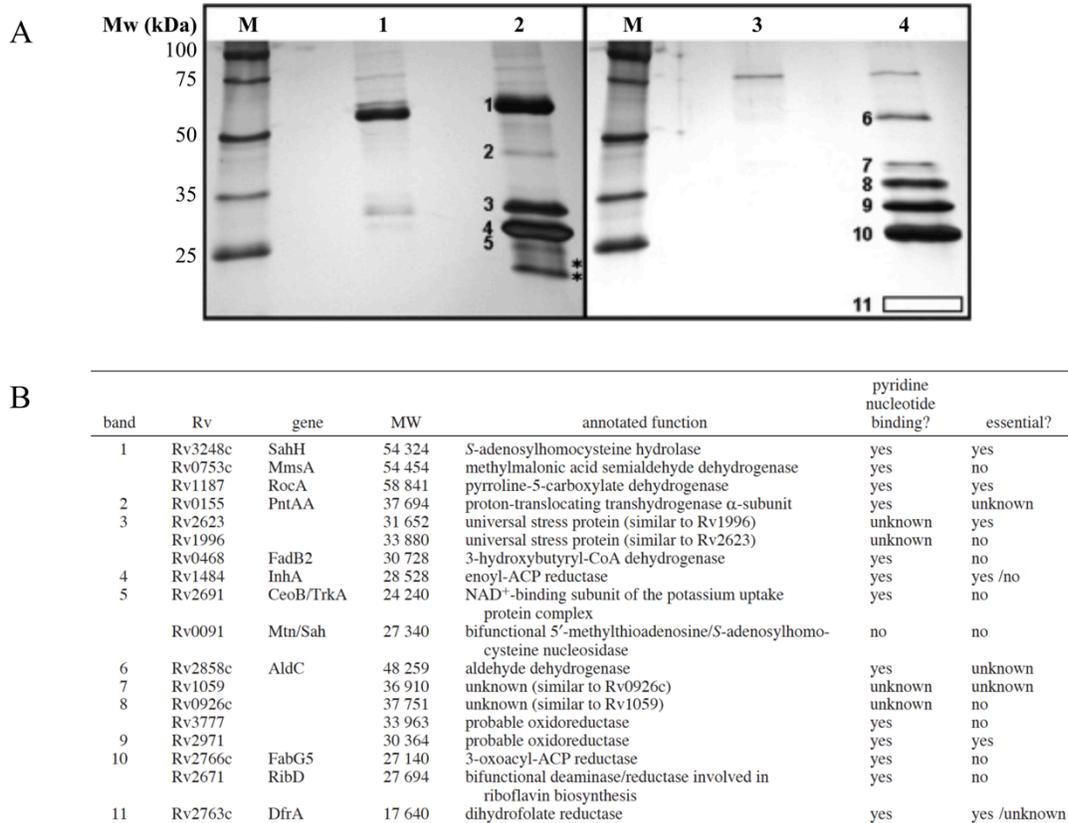


Figure 2.1: Determination of INH drug targets in *M. tuberculosis* by INH-NAD(P) affinity chromatography

(A) SDS-PAGE of INH-NAD(P) affinity chromatography eluates. Lanes 1 and 2 represent proteins eluted from NAD⁺-sepharose and INH-NAD-sepharose respectively, and lanes 3 and 4 represent proteins eluted from NADP⁺-sepharose and INH-NADP-sepharose respectively. Numbered bands represent isolated INH drug targets.

(B) Table targets summary of isolated INH-NAD(P) drug targets.

(Adapted from Argyrou *et al.*, 2006)

Declaration for Thesis Chapter 2

Declaration by candidate

In the case of Chapter 2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Expression, purification and x-ray data collection of Rv2971. Data processing and interpretation, manuscript and figure preparation.	80%

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Anggia Prasetyoputri	Purification and Crystallisation of Rv2971	N/A
Jamie Rossjohn	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Travis Beddoe	Conception, supervision, direction, data interpretation and manuscript preparation	N/A

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's Signature  Date 21/10/15

Main Supervisor's Signature  Date 21.10.15

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

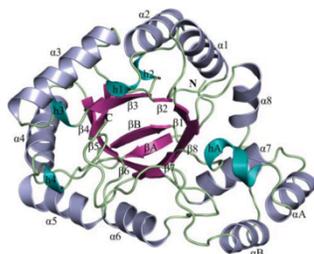
Adam Shahine,^a Anggia Prasetyoputri,^{a,b} Jamie Rossjohn^{a,c} and Travis Beddoe^{a,d*}

^aDepartment of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia, ^bResearch Centre for Biotechnology, Indonesian Institute of Sciences (LIPI), Indonesia, ^cInstitute of Infection and Immunity, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, Wales, and ^dDepartment of Agriculture Sciences, La Trobe University, Bundoora, Victoria, Australia

Correspondence e-mail:
travis.beddoe@monash.edu

Received 25 February 2014
Accepted 31 March 2014

PDB reference: Rv2971, 4otk



© 2014 International Union of Crystallography
All rights reserved

A structural characterization of the isoniazid *Mycobacterium tuberculosis* drug target, Rv2971, in its unliganded form

Aldo-keto reductases (AKR) are a large superfamily of NADPH-dependent oxidoreductases and play a role in detoxification of toxic metabolites. Rv2971, an AKR in *Mycobacterium tuberculosis*, has recently been identified as a target of isoniazid, a key first-line drug against tuberculosis. Here, the cloning, expression, purification, crystallization and structural characterization of Rv2971 are described. To gain insight into its function, the crystal structure of Rv2971 was successfully determined to 1.60 Å resolution in its unliganded form. The structure exhibits a TIM-barrel fold typical of AKRs, revealing structural characteristics essential for function and substrate specificities, allowing a structural comparison between Rv2971 and other mycobacterial AKRs.

1. Introduction

Mycobacterium tuberculosis, the aetiological agent of tuberculosis (TB), remains a major health burden globally, resulting in approximately 1.4 million deaths annually (World Health Organization, 2013). The emergence of drug-resistant strains underlines the urgent need for the development of new anti-mycobacterial therapeutics. Multiple drug resistant strains of TB (MDR-TB), and more recently extensively drug-resistant strains of TB (XDR-TB), are classified as expressing resistance to at least two of the first-line anti-TB drugs and a number of second-line anti-TB drugs, resulting in a limited means of treating diseased patients.

The key first-line drug in TB treatment is isoniazid (INH), a hydrazide prodrug activated by the catalase-peroxidase KatG to produce INH-NAD(P) adducts that exhibit potent anti-mycobacterial activity (Rozwarski *et al.*, 1998; Timmins & Deretic, 2006). The primary target of the INH-NAD(P) adduct is InhA, an enoyl-acyl carrier protein reductase that plays an essential role in mycolic acid biosynthesis (Dessen *et al.*, 1995). In addition to InhA, the drug acts against 17 recently identified targets within *M. tuberculosis* (Argyrou *et al.*, 2006). These targets exhibit a variety of essential roles implemented in the fitness and survival of the mycobacterium, and provide a variety of potential candidates for the development of new anti-TB therapeutics.

One of the targets identified is the essential aldo-keto reductase (AKR) AKR5H1 encoded by the gene *rv2971*. *rv2971* has previously been characterized as an essential gene for growth and survival in *M. tuberculosis* (Sassetti *et al.*, 2003) with a potential role in the detoxification of toxic metabolites (Grimshaw, 1992; Grant *et al.*, 2003; Penning & Drury, 2007). Recently, the catalytic activity of Rv2971 and its *M. smegmatis* orthologue MSMEG_2407 (67% sequence identity) has been characterized, revealing differences in dicarbonyl substrate specificities and affinities (Scoble *et al.*, 2010). The potency of INH was characterized against both enzymes, revealing a higher affinity against MSMEG_2407 compared with Rv2971, with apparent K_i values of 6.0 ± 1.2 and $31.0 \pm 1.4 \mu\text{M}$ reported, respectively (Scoble *et al.*, 2010).

To gain functional insight into the role of AKR5H1, the crystal structure of MSMEG_2407 has been determined in its apo form and in its holoenzyme form in the presence of NADPH (Scoble *et al.*, 2010). The crystal structure reveals an $(\alpha/\beta)_8$ -barrel topology, or a TIM-barrel fold, typical of other AKRs as previously described

structural communications

Table 1

Data-collection and refinement statistics for Rv2971.

Values in parentheses are indicative of the highest resolution shell.

Data collection	
Diffraction source	MX2, Australian Synchrotron
Detector	ADSC Quantum 315r CCD
Space group	$P3_21$
Unit-cell parameters (Å, °)	$a = b = 86.27$, $c = 86.29$, $\alpha = 90.0$ $\beta = 90.0$ $\gamma = 120.0$
Wavelength (Å)	0.95370
Resolution range (Å)	38.58–1.60 (1.66–1.60)
Total No. of reflections	525473 (76789)
No. of unique reflections	49373 (4857)
Completeness (%)	100.0 (100.0)
Mean $I/\sigma(I)$ †	12.72 (2.35)
$R_{\text{merge}}^{\ddagger}$ (%)	11.4 (91.7)
$R_{\text{p.i.m.}}^{\S}$ (%)	3.7 (29.1)
Multiplicity	10.6 (10.8)
Structural refinement	
Resolution range (Å)	38.59–1.60 (1.63–1.60)
R_{work}^{\S} (%)	13.60 (22.67)
R_{free}^{\S} (%)	16.90 (27.59)
R.m.s.d.‡ bond lengths (Å)	0.006
R.m.s.d.‡ bond angles (°)	1.05
Ramachandran plot	
Favoured (%)	98.20
Allowed (%)	1.80
Outliers (%)	0.00
No. of modelled non-H atoms	
Macromolecules	2128
Ligands	15
Water	311
Protein residues	276
Wilson B factor (Å ²)	18.51
Average B factors (Å ²)	
Protein atoms	20.40
Ligand molecules	20.40
Water molecules	33.70
PDB code	4otk

† I is the integrated intensity and $\sigma(I)$ is the estimated standard deviation of that intensity. ‡ $R_{\text{merge}} = \sum_{hkl} \sum_n |I_n(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_n I_n(hkl)$ where $I_n(hkl)$ is the scaled intensity of the n th measurement and $\langle I(hkl) \rangle$ is the mean intensity for that reflection. $R_{\text{p.i.m.}} = R_{\text{merge}}$ divided by the multiplicity. § $R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ for all data excluding the 5% that comprise the R_{free} used for cross-validation. ¶ Root-mean-square deviation from ideal values (Engh & Huber, 1991).

(Banner *et al.*, 1975), and enabled residues essential for enzyme kinetic activity and cofactor binding to be ascertained. While the crystal structure of MSMEG_2407 provides invaluable information on its functionality, Rv2971 possesses a number of differences in these binding pockets.

Here, we determined the crystal structure of Rv2971 in its unliganded form to a resolution of 1.60 Å. The structure enables the characterization of the substrate and NADPH binding pockets, revealing the subtle architectural differences between Rv2971 and MSMEG_2407, and providing insight into the differences in substrate specificities and INH inhibition levels. In addition, the Rv2971 structure may provide an initial template for the development of new anti-mycobacterial therapeutics.

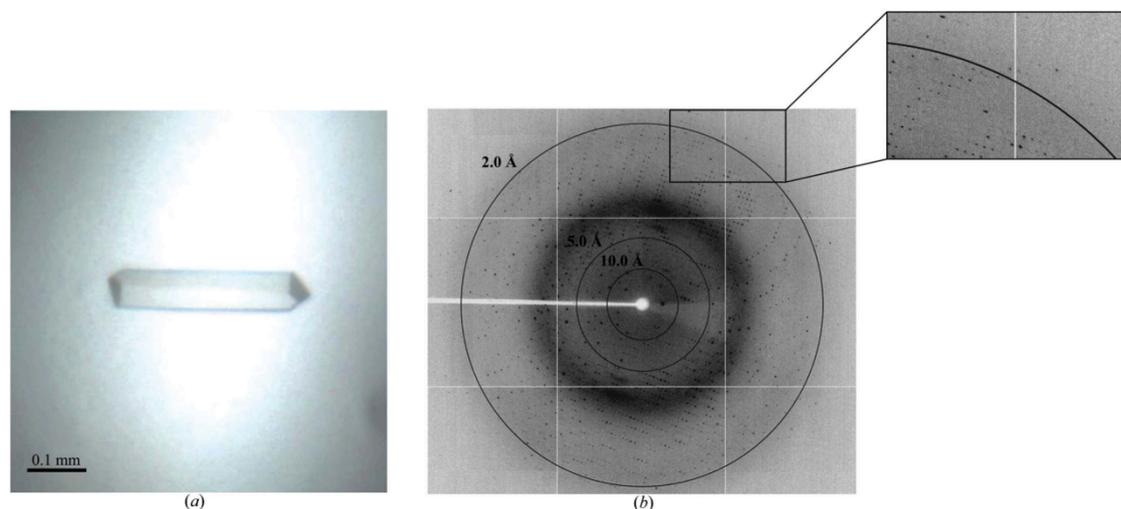
2. Materials and methods

2.1. Cloning, expression and purification of recombinant protein

The gene encoding Rv2971 was cloned, overexpressed and purified as soluble recombinant protein as previously described (Scoble *et al.*, 2010). For crystallization experiments, the purified recombinant protein was buffer-exchanged into 10 mM Tris-HCl pH 8.0, 200 mM NaCl and concentrated to 7 mg ml⁻¹. The concentration of purified Rv2971 was determined spectrophotometrically (NanoDrop 1000, Thermo Scientific) at 280 nm and was calculated using an extinction coefficient of 28 545 M⁻¹ cm⁻¹. The molecular weight, purity and identity of the protein were confirmed by SDS-PAGE and Western blotting with anti-hexahistidine antibody (R&D Systems).

2.2. Crystallization

The initial crystallization experiments involved screening 192 conditions from commercially available kits from Hampton Research (Crystal Screen HT and PEG/Ion HT) using a CrystalMation (Rigaku) integrated robotic workstation and the sitting-drop vapour-diffusion technique. Rod-shaped crystals were obtained initially at a

**Figure 1**

Crystallization and data collection of Rv2971. (a) Crystal of Rv2971 obtained from 7 mg ml⁻¹ protein solution (in 10 mM Tris-HCl pH 8.0, 200 mM NaCl) using 0.2 M sodium malonate pH 5.0, 19% (w/v) PEG 3350. (b) A typical 0.50° oscillation image from an Rv2971 crystal flash-cooled in 15% (v/v) glycerol.

structural communications

protein concentration of 7 mg ml^{-1} at 277 K after 24 h of equilibration against a crystallization condition comprised of 0.2 M sodium malonate pH 5.0, 20% (w/v) polyethylene glycol (PEG) 3350. Larger crystals of high diffraction quality were obtained by optimizing the crystal condition in 24-well Linbro plates (Hampton Research). The crystallization condition was optimized by fine-tuning the pH, protein concentration and precipitant concentration using a hanging drop consisting of 1 μl protein solution and 1 μl precipitant solution and a 500 μl reservoir volume (Fig. 1a). The best crystals appeared after 24 h of equilibration against a crystallization condition comprised of 0.2 M sodium malonate pH 5.0, 19% (w/v) PEG 3350 at a protein concentration of 7 mg ml^{-1} and grew to full size in 2 d.

2.3. X-ray data collection

For X-ray diffraction data collection, crystals of Rv2971 were transferred to a CryoLoop and soaked in a cryoprotectant consisting of 19% (v/v) PEG 3350, 0.2 M sodium malonate pH 5.0, 15% (w/v) glycerol before cooling to 100 K in a nitrogen-gas stream. A complete data set was collected from a single crystal on the MX2 beamline at the Australian Synchrotron using an ADSC Quantum 315rr CCD detector. A total of 360 frames of 0.5° were recorded with an exposure time of 1 s per frame (Fig. 1b). The data were processed using *iMosflm* (Battye *et al.*, 2011) and various programs from the CCP4 suite such as *SCALA* and *POINTLESS* (Winn *et al.*, 2011). The final statistics of data collection and processing are summarized in Table 1.

2.4. Structural determination

The crystal structure of Rv2971 was determined by the molecular-replacement method with *Phaser-MR* as part of the *PHENIX* program suite (Adams *et al.*, 2010). The coordinates of *M. smegmatis* AKR5H1 MSMEG_2407 (67% sequence identity; PDB entry 2wzm; Scoble *et al.*, 2010) were used as the search model. The resultant LLG score and TZF score obtained were 2010.15 and 43.8, respectively. For cross-validation, a random set of 5% of the total reflections were kept aside from the refinement and used for the calculation of R_{free} (Brünger, 1992). The initial model was refined as a rigid body followed by TLS refinement using TLS groups as suggested by the

TLSMD server (Painter & Merritt, 2006). The TLS refinement was performed according to four TLS groups, which are A8–A171, A172–A188, A189–A257 and A258–A283. The model obtained was manually adjusted using the interactive graphics program *Coot* (Emsley *et al.*, 2010) and refined using *phenix.refine* (Adams *et al.*, 2010) until no further improvement of the model could be made. Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession code 4okt. The final statistics of structural refinement are summarized in Table 1.

2.5. Structural analysis

Refinement validation was conducted using *MolProbity* and the *POLYGON* tools in the *PHENIX* program suite (Adams *et al.*, 2010; Chen *et al.*, 2010). Secondary structure was confirmed by the *STRIDE* plugin in *PyMOL* (Zhu, 2011). All structural superpositions were achieved using the SSM superpose feature of *Coot* (Emsley *et al.*, 2010). Electrostatic surface distribution calculations were performed using the *APBS* plugin for *PyMOL* (Baker *et al.*, 2001). The *M. smegmatis* AKRH5 MSMEG_2407 holoenzyme crystal structure (PDB entry 2wzm) and the apo crystal structure (PDB entry 2wzt; Scoble *et al.*, 2010) were used for structural alignments.

3. Results and discussion

3.1. Structural determination of Rv2971

To gain structural insight into the isoniazid drug target Rv2971, the crystal structure of the *M. tuberculosis* protein was determined. Recombinant *M. tuberculosis* Rv2971 was expressed, purified and crystallized using vapour-diffusion methods (Fig. 1a). Rv2971 was purified as a monomer as determined by size-exclusion chromatography, and routinely produced yields of 2 mg of pure protein per litre. Crystals belonged to the trigonal space group $P3_221$, with unit-cell parameters $a = b = c = 86.3 \text{ \AA}$, $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$. A complete data set was collected to a resolution of 1.60 \AA from a single crystal at 100 K (Fig. 1b), with data collection and processing summarized in Table 1. Based on Matthews coefficient calculations, one polypeptide chain (59.96% solvent content) could be accom-

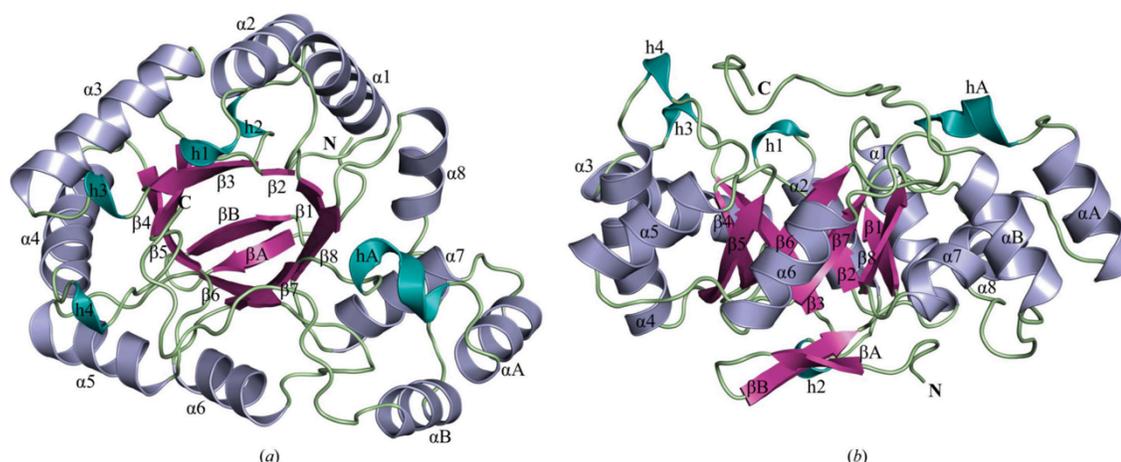


Figure 2

Overall crystal structure of Rv2971. Cartoon representation of the (a) top view and (b) side view rotated 90° of the overall structure. The C- and N-termini of the structure are labelled, with secondary-structure elements β -strands, α -helices and 3_{10} -helices, as calculated by *STRIDE* (Zhu, 2011), coloured magenta, blue and cyan, respectively. All figures were prepared in *PyMOL*.

structural communications

modated in the asymmetric unit, with a calculated V_M value of $3.07 \text{ \AA}^3 \text{ Da}^{-1}$ (Matthews, 1968).

The crystal structure of Rv2971 was determined to 1.60 Å resolution via molecular replacement using the *M. smegmatis* orthologue, MSMEG_2407, as a search model (PDB entry 2wzm), with final R_{work} and R_{free} statistics of 13.90 and 16.90%, respectively. Refinement statistics are summarized in Table 1. Of the construct utilized in crystallization, the large vector-derived N-terminal domain (MGSSH6SSGLVPRGSHMASMTGGQQMGRGSEF) was unstructured and residues 7–282 of Rv2971 were modelled into the electron density. The final model consisted of 276 amino-acid residues and 311 water molecules. The final model also included two malonate

ions and two chloride ions originating from the crystallization condition and crystallization buffer, respectively. A Ramachandran plot analysis of the final model by C^α geometry validation (Lovell *et al.*, 2003) showed that 98.2% of the residues are in the most favoured regions, while 1.8% of residues are within the allowed regions.

3.2. Overall crystal structure of Rv2971

The overall 1.60 Å resolution crystal structure of Rv2971 in its unliganded form (Figs. 2a and 2b) adopts an $(\alpha/\beta)_8$ -barrel topology, or a TIM-barrel fold, as previously described (Banner *et al.*, 1975). The structural fold exhibits typical characteristics of AKR substrates,

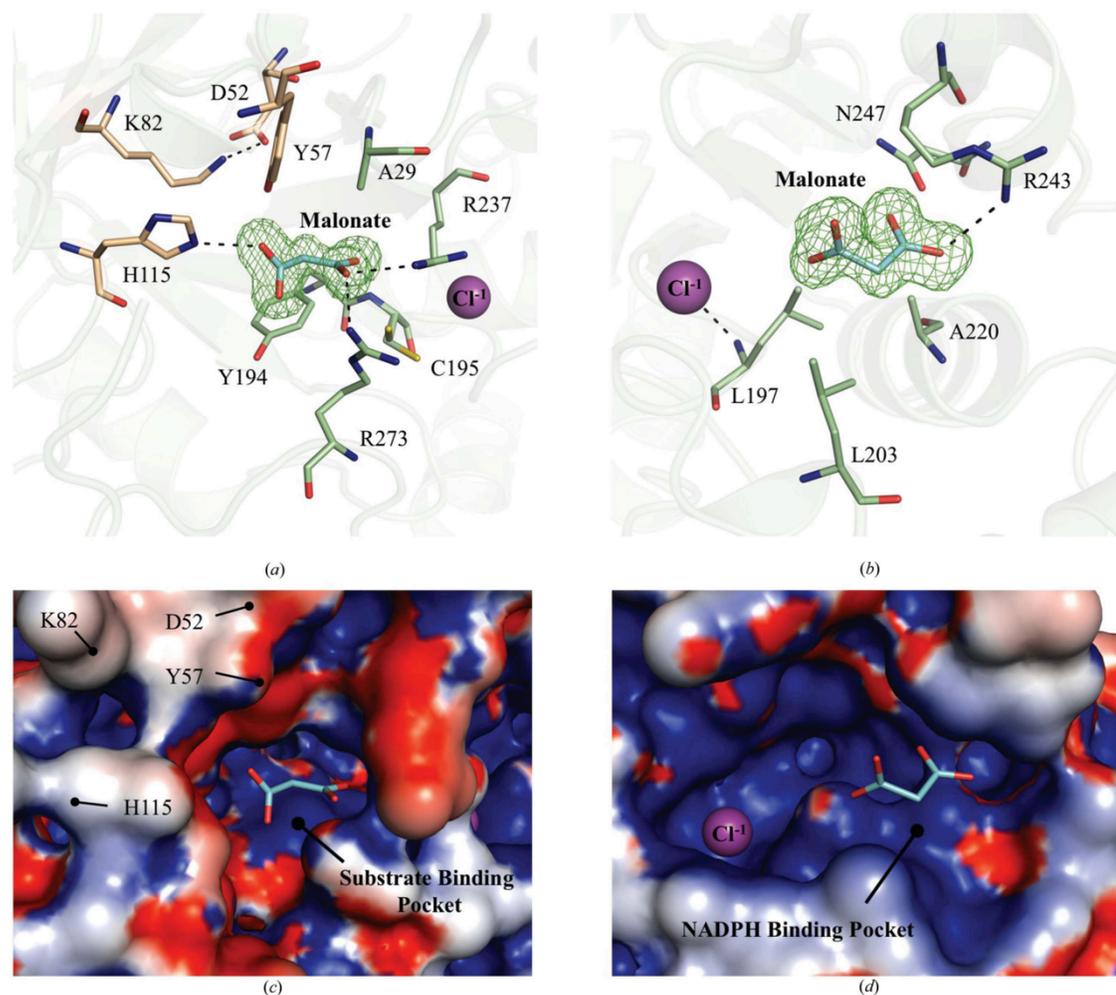


Figure 3 Rv2971 substrate and cofactor binding pocket architecture. Cartoon representation of (a) the substrate binding pocket and (b) the NADPH binding pocket. Two malonate ions and a single chloride ion bound during crystallization are represented as cyan sticks and magenta spheres, respectively. Electron density of maps is shown as $F_o - F_c$ simulated-annealing OMIT maps contoured at 3σ . Amino-acid residues involved in binding to the malonate and chloride ions are represented as green sticks. Hydrogen-bond formation between contact residues and ions is represented as black dashes, with bond distances ranging between 2.6 and 3.6 Å. The Asp52-Tyr57-Lys82-His15 catalytic tetrad is represented in wheat. The solvent-accessible surface representation coloured by electrostatic potential of the (c) substrate binding pocket and (d) NADPH binding pocket was calculated by APBS (Baker *et al.*, 2001). The potential contours are shown on a scale from $+5.0$ (blue) to $-5.0 k_B T e^{-1}$ (red); white indicates a value close to $0 k_B T e^{-1}$. Indicated are the positions of the Asp52-Tyr57-Lys82-His15 catalytic tetrad.

structural communications

where the TIM barrel is comprised of eight β -strands ($\beta 1$ – $\beta 8$), flanked by eight α -helices ($\alpha 1$ – $\alpha 8$) interwoven in an antiparallel manner between each β -strand. Situated at the N-terminus of the structure are two antiparallel β -stands (βA and βB) that cover the base of the protein structure (Fig. 2b). The TIM-barrel structure is punctuated by two α -helices (αA and αB) on the C-terminal side of the protein structure (Fig. 2a). Dispersed through the C-terminal side, or 'open rim' face, of the protein structure are a series of 3_{10} -helices (h1, h3, h4 and hA) that contribute to the overall α -helical content of the Rv2971 structure by forming the NADPH binding pocket.

3.3. Substrate and NADPH binding pockets of Rv2971

The crystal structure of Rv2971 was determined in the presence of two malonate ions bound within the substrate binding pocket and NADPH binding pocket. Attempts to soak NADPH into the crystals were unsuccessful because of the presence of the malonate ions blocking the binding pockets. Within the substrate binding pocket, the malonate ion is bound to Rv2971 *via* hydrogen bonding with Asp52 and His115 at a bond distance of 2.6 and 2.9 Å, respectively

(Fig. 3a). Residues Asp52 and His115 belong to the Asp-Tyr-Lys-His catalytic tetrad typical of AKRs. In the case of Rv2971, the catalytic tetrad is comprised of residues Asp52-Tyr57-Lys82-His115. A single chloride ion is bound between the substrate and NADPH binding pockets, and is bound to residue Leu197 *via* hydrogen bonding (Fig. 3b). A second malonate ion is bound within the NADPH pocket, binding residues involved in contacts with NADPH *via* hydrogen-bond formation, with bond distances ranging between 3.0 and 3.6 Å (Fig. 3b). Both the substrate and NADPH binding pockets are electrostatically positively charged (Figs. 3c and 3d) which, as well as being highly surface exposed, facilitates the substrate and NADPH binding capabilities of Rv2971.

3.4. Structural comparison with the *M. smegmatis* orthologue MSMEG_2407

With the successful determination of the crystal structure of Rv2971, a structural comparison was conducted with MSMEG_2407 to characterize the previously observed differences in enzyme kinetic activity (Scoble *et al.*, 2010). The crystal structures of Rv2971 and the

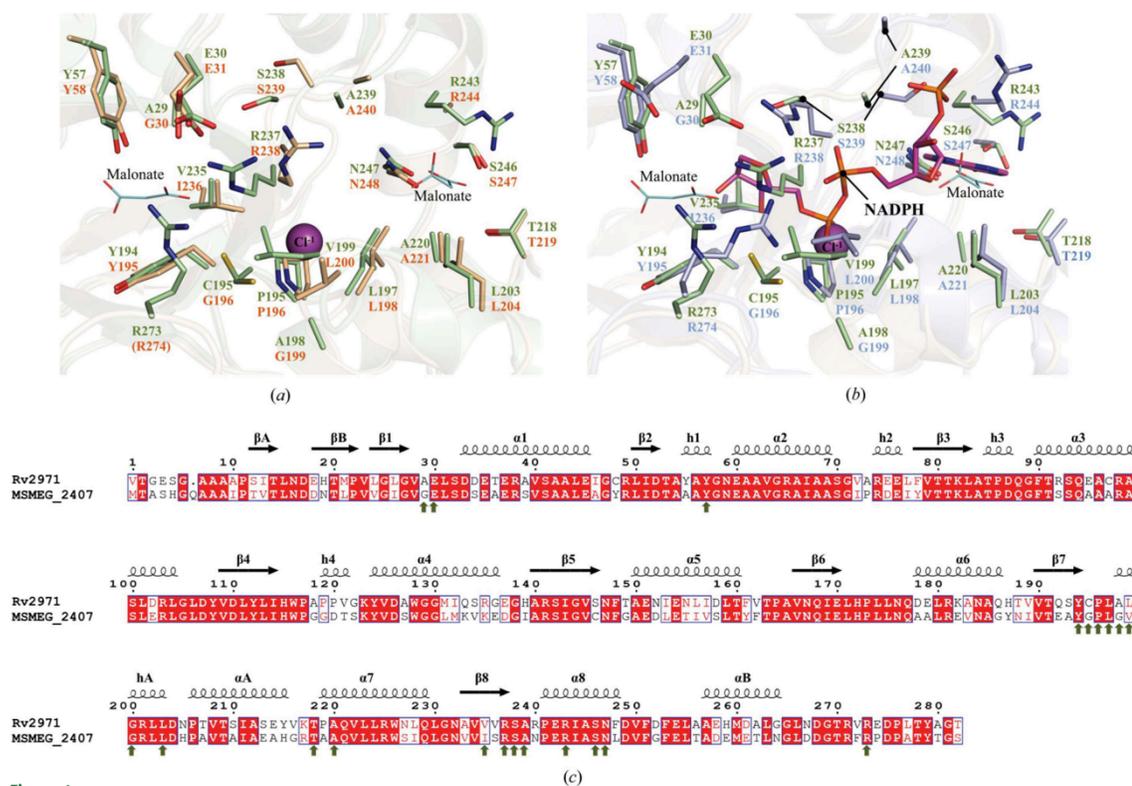


Figure 4

Structural comparison between Rv2971 and MSMEG_2407. Ball-and-stick representation of the NADPH binding pockets of Rv2971 and (a) MSMEG_2407 in its apo form (PDB entry 2wz1) and (b) in its holoenzyme form in the presence of NADPH (PDB entry 2wzm). Visualized is the structural comparison between amino-acid residues in contact with NADPH in the MSMEG_2407 holoenzyme structure with residues of identical position in Rv2971. Residues of Rv2971, the MSMEG_2407 apo form and the MSMEG_2407 holoenzyme are represented in green, wheat and blue, respectively. Bound malonate ions and a single chloride ion are represented as cyan lines and magenta spheres, respectively. The bound NADPH molecule in the MSMEG_2407 holoenzyme is represented in pink. Missing residues are indicated in parentheses. (c) Sequence alignment between Rv2971 and MSMEG_2407. The alignment was prepared with the program *ClustalW2* and visualized using *ESPrift* v.2.2. The secondary-structure elements correspond to the structure of Rv2971. Strict sequence-identical residues are denoted with a red background, while similar residues are visualized in red text with white background. Sequence similarities in groups are denoted by blue boxes. MSMEG_2407 residues interacting with NADPH by hydrogen bonding or van der Waals interactions are indicated by green arrows.

structural communications

apo form of MSMEG_2407 overlaid tightly, with an overall r.m.s.d. value of 0.85 Å observed. While Rv2971 and MSMEG_2407 share a sequence identity of 67%, a number of substitutions are present between amino-acid residues within the NADPH binding pocket that impact on the enzyme kinetic activity rates between the orthologues (Fig. 4). Of the 21 amino-acid residues from MSMEG_2407 interacting with NADPH, there is a difference of four residues in Rv2971. Residues Gly30, Gly196, Leu200 and Ile236 in MSMEG_2407 are substituted with residues Ala29, Cys195, Val199 and Val235 (Fig. 4). The most significant difference is the substitution of a glycine (Gly196) for cysteine (Cys195). Cys195 is present in two alternate conformations in the structure of Rv2971, both facing into the NADPH binding pocket. One of the conformations forms van der Waals interactions with the chloride ion present (Fig. 4a), and indicates the potential position of the residue in the presence of NADPH, owing to the position of a hydroxyl group of NADPH bound in MSMEG_2407 (Fig. 4b). The loss of glycine residues involved in NADPH contacts indicates a more constrained binding site, in particular the substitution of Gly30 for Ala29 at the nicotinamide binding site, and may provide a more complete holoenzyme crystal structure of AKR5H1, which appeared disordered in the MSMEG_2407 holoenzyme crystal structure (Scoble *et al.*, 2010).

4. Concluding remarks

While valuable structural insight into the architecture of Rv2971 has been obtained, the successful determination of further crystal structures of Rv2971, both in its holoenzyme form and in complex with the INH-NADP adduct, will aid in further characterization of its function, as well as in gaining a more in-depth characterization of INH resistance.

We thank the staff of the Australian Synchrotron and Monash Macromolecular Crystallization Facility for assistance with crystallization and X-ray data collection. This work was supported by the Australian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics and the National Health

and Medical Research Council of Australia. JR is an NHMRC Australia Fellow and TB is a Pfizer Australian Research Fellow.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Argyrou, A., Jin, L., Siconolfi-Baez, L., Angeletti, R. H. & Blanchard, J. S. (2006). *Biochemistry*, **45**, 13947–13953.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
- Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). *Nature (London)*, **255**, 609–614.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Dessen, A., Quémard, A., Blanchard, J. S., Jacobs, W. R. Jr & Sacchettini, J. C. (1995). *Science*, **267**, 1638–1641.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Grant, A. W., Steel, G., Waugh, H. & Ellis, E. M. (2003). *FEMS Microbiol. Lett.* **218**, 93–99.
- Grimshaw, C. E. (1992). *Biochemistry*, **31**, 10139–10145.
- Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* **D62**, 439–450.
- Penning, T. M. & Drury, J. E. (2007). *Arch. Biochem. Biophys.* **464**, 241–250.
- Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs, W. R. Jr & Sacchettini, J. C. (1998). *Science*, **279**, 98–102.
- Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
- Scoble, J., McAlister, A. D., Fulton, Z., Troy, S., Byres, E., Vivian, J. P., Brammananth, R., Wilce, M. C. J., Le Nours, J., Zaker-Tabrizi, L., Coppel, R. L., Crellin, P. K., Rossjohn, J. & Beddoe, T. (2010). *J. Mol. Biol.* **398**, 26–39.
- Timmins, G. S. & Deretic, V. (2006). *Mol. Microbiol.* **62**, 1220–1227.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- World Health Organization (2013). *Global Tuberculosis Report 2013*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/en/.
- Zhu, H. (2011). *DSSP and Stride Plugin for PyMOL*. Biotechnology Center (BIOTEC), TU Dresden, Germany. http://www.biotec.tu-dresden.de/~hongboz/dssp_pymol/dssp_pymol.html.

Chapter 3: Crystal structure of an essential lipase, Rv3802c, from *Mycobacterium tuberculosis*, in complex with the anti-TB drug

Tetrahydrolipstatin

3.1 Introduction

This next chapter corresponds to further work focused on the drug targets of *M. tuberculosis*. While chapter 2 focused on the structural characterization of a target of one of the key first line drugs against TB, this chapter focuses on a target inhibited by a new drug, tetrahydrolipstatin (THL), which has only been recently characterized to exhibit anti-TB activity. The chapter outlines the expression and purification of recombinant Rv3802c from *E. coli* and focuses on the crystallization, structural determination and analysis of Rv3802c in complex with the potential anti-TB drug THL.

Recently, the crystal structure of MSMEG_6394, the Rv3802c homologue from *M. smegmatis*, has been solved to a resolution of 2.9 Å (Crellin, *et al.*, 2010) (Appendix 1). Previous attempts to determine the crystal structure of Rv3802c have been unsuccessful, and it has been stipulated that this was due to a flexible N-terminal domain not visible in the MSMEG_6394 structure. The flexibility of this loop is hypothesised to have hindered effective protein crystal packing, preventing the generation of high diffracting quality crystals. The N terminal unstructured domain is comprised of 68 amino acids, with a majority being non-polar, aliphatic residues, and has been predicted to act as a transmembrane domain.

The N-terminal domain is conserved within a number of proteins annotated as putative cutinases within *M. tuberculosis*, including ML0099: the *M. leprae* Rv3802c homologue. The domain consists a

predicted N terminal transmembrane domain, as well as a signalling sequence and proline rich region, as characterized in *M. avium* orthologue (Miltner, *et al.*, 2005). It was speculated that the presence of this N-terminal sequence aid in playing a role in mycobacterial host cell invasion, as these characteristics are also observed in internalin A in *Listeria monocytogenes*, aiding in epithelial cell invasion (Miltner, *et al.*, 2005). While identifying the purpose of this flexible N-terminal domain would shed some light into the role of Rv3802c in mycolic acid biosynthesis, a main aim of the study was to provide structural insight into the mechanisms of THL inhibition.

The crystal structure of MSMEG_6394, while successfully solved to 2.9 Å, is not suitable model for the purposes of rational drug design against Rv3802c. Differences in enzyme substrate specificities and kinetic activity rates between Rv3802c and MSMEG_6394 have previously been reported, as well as differences in THL binding, with THL affinity for Rv3802c being three times greater than MSMEG_6394 (Crellin, *et al.*, 2010). This is due to subtle differences in the enzyme active site architecture attributed in part due to sequence identity, as MSMEG_6394 and Rv3802c share a sequence identity of 68%. Additionally, it is more suitable to design new therapeutics against targets originating from the target organism, as opposed to a model organism.

In order to produce higher quality diffracting crystals of Rv3802c, a construct was created lacking the flexible N-terminal domain as a means of aiding in better crystal contacts. While crystals of the Rv3802c lacking the N-terminal domain produced crystals, only crystals of Rv3802c in complex with THL produced crystals of high diffraction quality. The aim of the study was to solve the crystal structure of Rv3802c with THL to gain a structural insight into the inhibition mechanisms of the drug, as well as provide a basis for rational drug design.

The chapter describes the crystal structure of Rv3802c in complex with THL successfully determined to 2.9 Å. The Rv3802c construct lacking the N-terminal domain (Rv3802c^{Δ1-67}) was successfully cloned, expressed and purified, with minimal differences in enzyme kinetic rates or THL inhibition levels compared to the full Rv3802c protein. The crystal structure reveals the mechanisms of THL binding to the active site pocket of Rv3802c, with the free carboxylate ion of THL binding tightly to the catalytic serine 175 residues within the substrate-binding pocket of Rv3802c, which belongs to the serine – aspartic acid – histidine catalytic triad, typical of hydrolases. THL was bound in the hydrolyzed form, as previously described (Pemble, *et al.*, 2007). Initially, THL attaches to the catalytic serine residue in a covalent manner, where it undergoes a water-mediated nucleophilic attack: breaking the covalent attachment and binding it via hydrogen bond formation in its hydrolysed form. THL is bound to Rv3802c via hydrogen bond formation to the catalytic serine 175 residue, and is locked into the binding site via a hydrophobic locking mechanism facilitated by mobile α9-helix lid domain, typical of lipases.

The structure of Rv3802c in complex with THL reveals the structural mechanism of THL interactions with the substrate-binding pocket, revealing the structure-based mechanisms of THL inhibition. The successfully determined complex structure of Rv3802c with THL will provide a basis for the development of high affinity therapeutics by rational drug design.

3.2 Materials and Methods

3.2.1 Cloning of Rv3802c^{Δ1-67}

Rv3802c has been cloned as previously described (Crellin, *et al.*, 2010). The open reading frames of the Rv3802c^{Δ1-67} gene was PCR-amplified from *M. tuberculosis* genomic DNA using GoTaq Polymerase (Promega) with specific sense (5'-

GACGACGACAAGATGCAGGATGCGTCCTGCCCGACGTGCAGATGATC-3') and antisense (5'-GAGGAGAAGCCCGTTATCACCTATGTTTGGGGTGGGGCGCATTTTC-3') primers. The underlined sequences in the primers represent specific overhanging sequence utilized in the ligation independent cloning (LIC) (Aslanidis and de Jong, 1990) the LIC pET46 Ek/LIC vector kit following manufacturer's instructions (Novagen). The 810 bp PCR product was directly annealed to the pET46 Ek/LIC vector (Appendix 2), giving a protein with an enterokinase-cleavable N-terminal hexahistidine tag. The resultant N-terminal amino-acid sequence of the recombinant Rv3802c^{Δ1-67} is MAHHHHHHVDDDDK**QDASCPD**, where the highlighted sequence is the truncated Rv3802c^{Δ1-67} protein starting at residue 68. The clone obtained was confirmed by DNA sequencing.

3.2.2 Expression, Refolding and Purification of Rv3802c^{Δ1-67}

Recombinant Rv3802c and Rv3802c^{Δ1-67} were overexpressed as inclusion bodies in *E. coli* host strain BL21 (DE3) via isopropyl β-D-1-thiogalactopyranoside (IPTG) as previously described (Crellin, *et al.*, 2010). Recombinant protein was refolded and purified as previously described, with a few minor alterations. Harvested cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM EDTA, 1 % (v/v) Triton-X100, 1 % (w/v) sodium deoxycholate and 1 mM PMSF, then lysed by sonication. Inclusion bodies were harvested by centrifugation and washed twice by centrifugation in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM EDTA, 0.1 % (v/v) Triton-X100 and 0.1 % (w/v) sodium deoxycholate and once in 50 mM Tris-HCl pH 8.0, 100 mM NaCl. The washed inclusion bodies were solubilized in 20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 8 M Urea, 10 mM imidazole and applied to 5 mL Ni²⁺-chelating Sepharose (GE life sciences) affinity resin equilibrated in the same buffer. Unbound protein was removed with three column volumes of 20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 8 M urea and 20 mM imidazole, and bound protein was eluted with 5 column volumes of 20 mM Tris-HCl pH

8.0, 0.5 M NaCl, 8 M urea and 200 mM imidazole. Each fraction was collected and analysed by 15 % (w/v) SDS-PAGE.

Eluted Rv3802c and Rv3802c^{Δ1-67} was refolded via stepwise dialysis in three steps at 4 °C. Recombinant protein was dialysed against 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA and 7 mM 2-mercaptoethanol for 16 h, then dialysed against 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA for 24 h, then dialysed against 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM EDTA for 2 h. The refolded protein was applied to DEAE-cellulose (Sigma) equilibrated in the same buffer. Unbound protein was collected and bound protein was eluted in 5 column volumes in 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.5 mM EDTA. Each fraction was collected and analysed by 15 % SDS-PAGE. Eluted protein was concentrated and applied to size exclusion chromatography to isolate monomeric recombinant protein on a HiLoad Superdex 200 16/60 (Amersham-Pharmacia) equilibrated in 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM EDTA at a flow rate of 1 mL.min⁻¹. Protein containing fractions were visualized by 15 % SDS-PAGE and fractions containing monomeric protein were pooled and applied to a 5 mL Q-HiTrap FF column (Amersham-Pharmacia). The column was equilibrated in 10 mM Tris-HCl pH 8.0, 0.5 mM EDTA and protein was eluted with a linear gradient from 0 – 1 M NaCl in the same buffer at a flow rate of 0.3 mL.min⁻¹. Fractions containing purified recombinant protein were analysed by 15 % SDS-PAGE and pooled.

3.2.3 Circular Dichroism and Thermal Melting

Circular-dichroism spectra of Rv3802c and Rv3802c^{Δ1-67} were recorded on a Jasco J-815 circular-dichroism Spectropolarimeter at 20 °C with a data pitch of 0.1 nm and a scan speed of 100 nm.min⁻¹. Briefly, all samples were prepared at a concentration of 20 μM in 10 mM Tris-HCl pH 8.0, 100 mM NaCl buffer. Far-ultraviolet circular-dichroism spectra (260 –190 nm) were recorded with five data accumulations in a quartz cell with a path length of 0.1 cm. The resultant spectra were visualized using

Graphpad Prism software. Thermal Melting analysis of Rv3802c and Rv3802c^{Δ1-67} were recorded at a wavelength of 214 nm on a Jasco J-815 circular-dichroism Spectropolarimeter at a temperature range of 20 – 90 at of 1 °C.min⁻¹. All samples were prepared at a concentration of 20 μM in 10 mM Tris-HCl pH 8.0, 100 mM NaCl buffer. The resultant spectra were visualized using Graphpad Prism software.

3.2.4 Enzyme Kinetic and Inhibition Assays

The effects of removing the N terminal transmembrane domain on Rv3802c function was measured via *p*-nitrophenyl butyrate spectrophotometric kinetic assay, as previously described with a few minor alterations (West, *et al.*, 2009; Winkler and Stuckmann, 1979). *p*-nitrophenyl butyrate was prepared in 2-propanol at a range of concentrations between 0.3 – 4 mM and mixed 1:9 in assay solution containing 100 mM Tris pH 8.0, 0.4 % (v/v) Triton X100, 2.3 mg.ml⁻¹ sodium deoxycholate and 1 mg.ml⁻¹ gum arabic. 20 uL of enzyme at a concentration of 0.1 mg.ml⁻¹ was added to 240 uL of the above reaction mixture, yielding final *p*-nitrophenyl butyrate concentrations of 0.4 mM – 30 mM. The increase in production of *p*-nitrophenolate was continuously measured at 405 nm using a Fluostar OMEGA plate reader (BMG Labtech) for 10 minutes in 96 microwell plate at 37 °C. Concentrations were calculated by comparisons to a *p*-nitrophenol standard curve with an extinction coefficient of 18000 M⁻¹.cm⁻¹ used for initial velocity calculations. Results were visualized by the *Graphpad Prism* software, with error bars representing the mean initial velocity ± SEM of three independent experiments. All Michaelis-Menten kinetic analysis was calculated using *Graphpad Prism*.

Inhibition of Rv3802c^{Δ1-67} by THL was determined using a lipolytic Tween-20 cleavage assay previously described (West, *et al.*, 2009; West, *et al.*, 2011). Inhibitors were dissolved in 100 % (v/v)

DMSO and serial diluted across a 96-microwell plate, with a concentration range between 0 – 100 μ M. A reaction mixture was prepared with 66 mM CaCl₂ and 0.66 % (v/v) Tween-20 in 50 mM Tris pH 8.0, with 75 μ L aliquots added to the 96-microwell plate. 750 ng of enzyme was added to each reaction mixture for a final volume of 150 μ L. The turbidity was continuously measured at 405 nm using a Fluostar OMEGA plate reader (BMG Labtech) for 15 minutes at 37 °C. IC₅₀ values were calculated using *Graphpad Prism*, with error bars representing mean percentage relative activity \pm SEM of three independent experiments.

3.2.5 Cocomplexation and Crystallization of Rv3802c^{Δ1-67}-THL

Rv3802c^{Δ1-67} at 0.1 mg.ml⁻¹ was preincubated with THL (Orlistat, Sigma) at a 1:10 Protein:Inhibitor molar ratio for 16 h at 4 °C. Excess THL was removed via size exclusion chromatography using a Superdex 200 10/300 equilibrated in 10 mM Tris-HCl pH 8.0, 200 mM NaCl. For crystallization trials, Rv3802c^{Δ1-67}-THL was concentrated to 10 mg.ml⁻¹ and crystallized via the vapour diffusion hanging drop method consisting of equal volumes of 1 μ L protein and 1 μ L mother liquor. Crystals of diffraction quality were grown in 21 % (w/v) PEG 2000 MME, 0.1 M Bis-Tris pH 5.7 at 20 °C, with crystals growing to full size after 3 d.

3.2.6 X-ray Data Collection and Structural Determination of Rv3802c^{Δ1-67}-THL

Crystals of Rv3802c^{Δ1-67}-THL were flash frozen in liquid nitrogen with in a cryoprotectant of mother liquor + 10 % (v/v) PEG 400. Diffraction data was collected at 100 K at the Australia Synchrotron MX1 beamline using the ADSC-Quantum 210r CCD detector (Cowieson, et al., 2015). A total of 720 frames of 0.25 ° oscillation were recorded with an exposure time of 1 s. Data processing was performed with *iMOSFLM* (Leslie and Powel, 2007) and scaled using *SCALA* (Evans, 2006) as part of the *ccp4i suite* (Winn, et al., 2011). The structure was determined by molecular replacement using Phaser-MR (McCoy, et al., 2007) in the *ccp4i suite*. The search model used was the crystal structure of

MSMEG_6394 (PDB ID: 3AJA) with a sequence identity of 68 %. The crystal structure was refined using Phenix-refine (Adams, *et al.*, 2010) The molecule was manually adjusted in COOT (Emsley, *et al.*, 2010) and validated by Molprobit (Chen, *et al.*, 2010)

3.3 Results

3.3.1 Expression and purification of Rv3802c^{Δ1-67}

The coding region of Rv3802c was successfully cloned by ligation-independent cloning (LIC) and expressed in *E. coli* strain BL21 (DE3) resulting in protein with a hexahistidine purification tag at the N-terminus (Appendix 2). Previous crystallisation trials have yielded no diffraction quality crystals of Rv3802c (Crellin, *et al.*, 2010). Bioinformatic analysis revealed an N-terminal predicted transmembrane domain (TMHMM Server v. 2.0), which may impair crystallographic analysis, and is not present in the crystal structure of MSMEG_6394. To overcome this, a construct lacking the predicted N-terminal transmembrane domain (Rv3802c^{Δ1-67}) was created using the LIC method (Aslandis and de Jong, 1990). To confirm whether or not the predicted N-terminal transmembrane domain played a role in *in vitro* solubility or oligomeric state, both Rv3802c and Rv3802c^{Δ1-67} were expressed and purified. In both cases, Rv3802c was expressed as insoluble protein, indicating that the predicted transmembrane region does not contribute to the overall insolubility of the protein *in vitro*. Both Rv3802c and Rv3802c^{Δ1-67} were purified by Ni-NTA (Figure 3.1a-b), selecting for the N-terminal hexahistidine tag, as insoluble inclusion bodies before refolding by stepwise dialysis to allow correct formation of the two internal disulphide bonds.

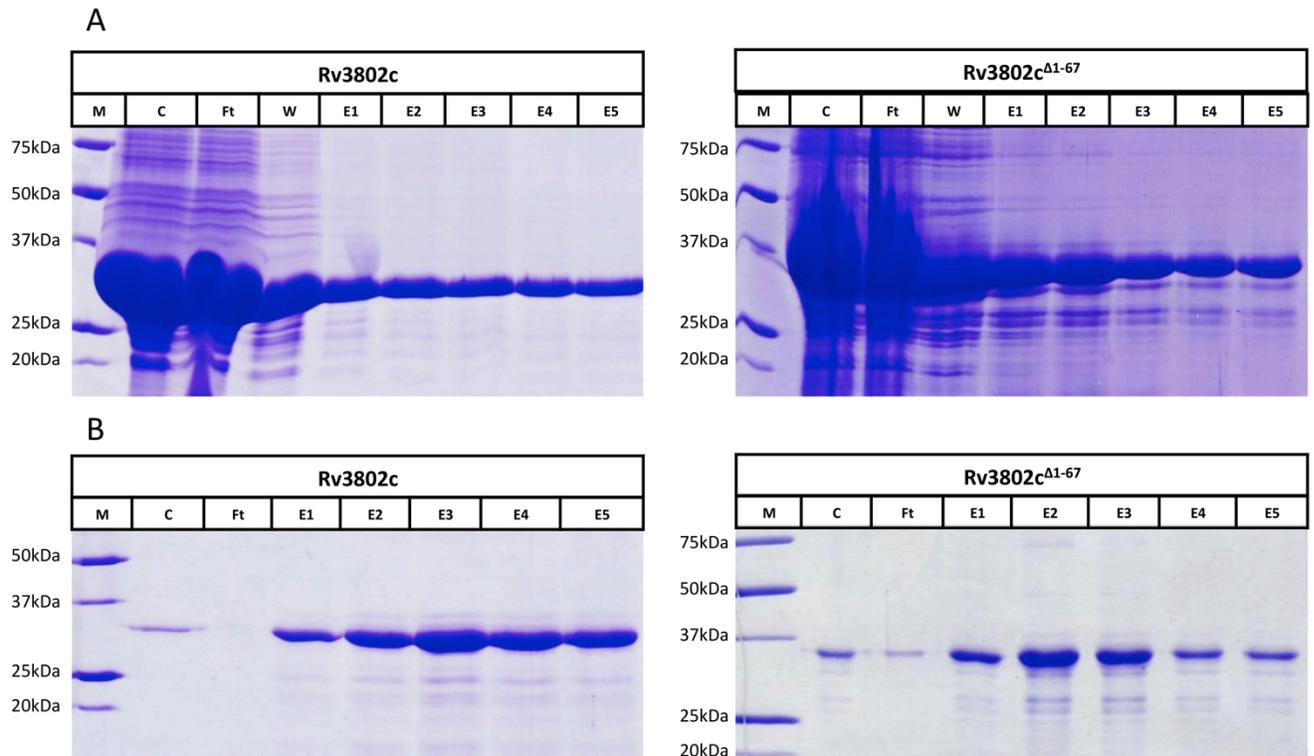


Figure 3.1: Expression of Rv3802c and Rv3802c^{Δ1-67}

- (A) Expressed protein was captured via Ni-Sepharose Affinity chromatography, with correlating 15% SDS-PAGE of Rv3802c and Rv3802c^{Δ1-67} shown. Crude sample (C) is applied to 5mL resin. Proteins with intact hexahistidine tags are bound to the resin, while non-binding or excess protein passes through as flowthrough (Ft). Resin is washed (W) in 20mM imidazole, and protein is eluted in 5x5mL fractions (E1-E5) in 200mM Imidazole. Flowthrough is re-purified until all protein is captured.
- (B) After capture, the protein is refolded and purified via DEAE-cellulose resin, with correlating 15% SDS-PAGE of Rv3802c and Rv3802c^{Δ1-67} shown. Refolded protein (C) is applied to DEAE-cellulose, with non-bound protein passing through as flowthrough (Ft). Protein is eluted in 5x5mL fractions (E1-E5) in 200mM NaCl. Molecular weight markers (M) are indicated in kDa.

Refolded protein was concentrated by DEAE-cellulose anion exchange (Figure 3.1b) before being purified via additional size-exclusion chromatography (Figure 3.2a-b) and anion exchange chromatography stages (Figure 3.2c-d), with recombinant protein appearing at an approximate size of 35 kDa for Rv3802c and 30 kDa for Rv3802c^{Δ1-67}, corresponding to the expected monomeric protein sizes as previously reported (Parker, *et al.*, 2007; Parker, *et al.*, 2009; West, *et al.*, 2009; Crellin, *et al.*, 2010). In the case of the native Rv3802c an additional dimeric peak was present on the size-exclusion chromatogram (Figure 3.2a), which was absent for Rv3802c^{Δ1-67}. This indicated that the presence of the N-terminal domain may have resulted in an artificial oligomeric state *in vitro*, as a result of improper reorganisation of the tertiary structure during the refolding process. Despite this, both Rv3802c and Rv3802c^{Δ1-67} were successfully purified as monomeric protein, and yielded ~2-3 mg of pure protein per litre of refold respectively.

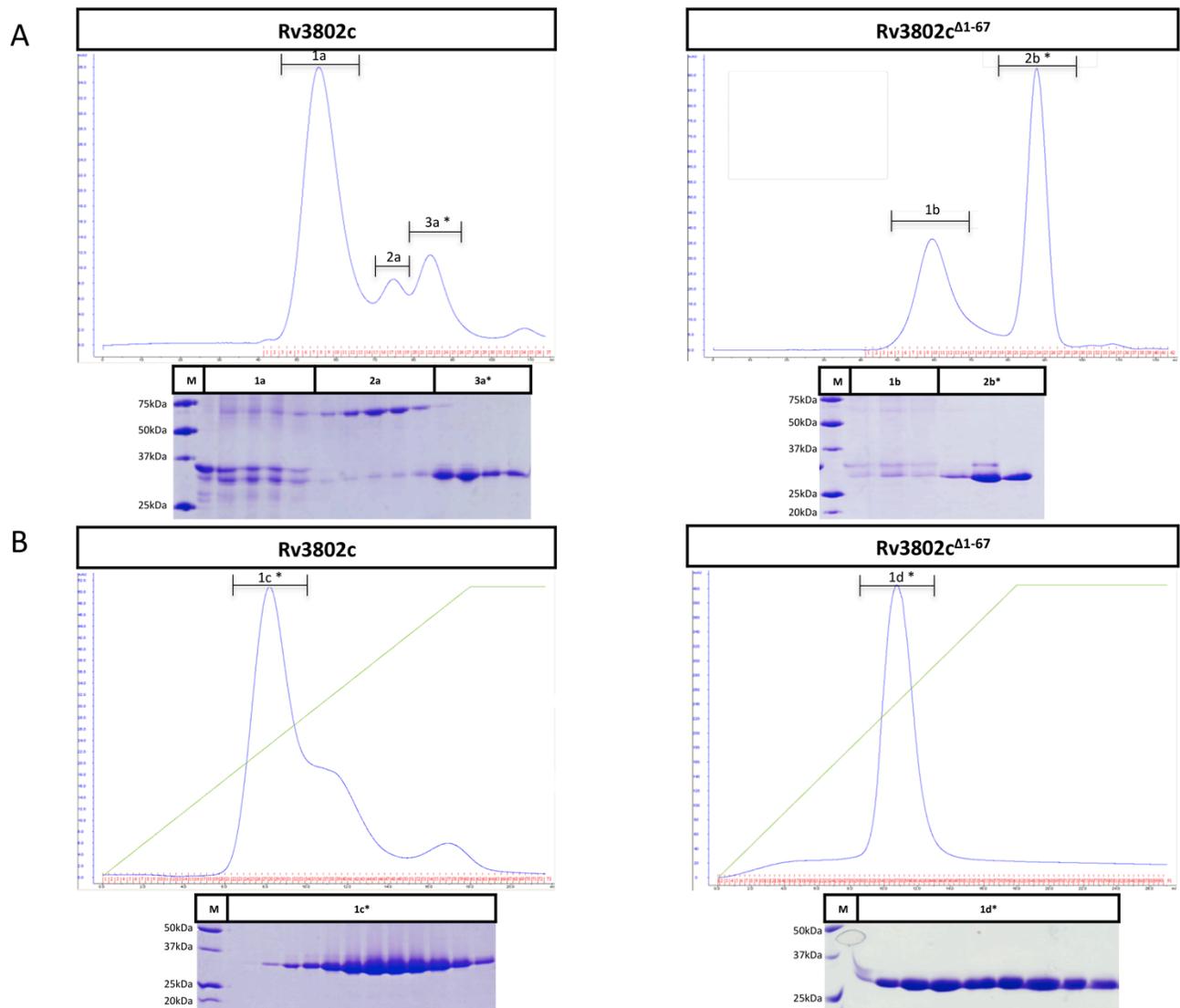


Figure 3.2: Purification of Rv3802c and Rv3802c^{Δ1-67}

- (A) Size exclusion chromatograms and correlating 15% SDS-PAGE of Rv3802c and Rv3802c^{Δ1-67}. Purification conducted via Superdex S200 16/60 chromatography column (Amersham-Pharmacia). Protein purified via size exclusion are indicated by an *. Molecular weight markers are indicated in kDa.
- (B) Purified protein is pooled and applied to Anion Exchange Chromatography using HiTrap Q FF 5mL column (Amersham-Pharmacia), with chromatograms and correlating 15% SDS-PAGE of (c) Rv3802c and (d) Rv3802c^{Δ1-67} visualized. Peaks from the anion exchange chromatography stage correlating to purified protein are indicated by an *. Molecular weight markers are indicated in kDa.

3.3.2 Circular Dichroism and Thermal Melt Analysis of Rv3802c

Circular dichroism analysis was conducted in order to confirm successful refolding of both Rv3802c and Rv3802c^{Δ1-67}, and to determine the effect of the predicted N-terminal transmembrane domain on overall protein fold and stability. The resulting CD Spectra conducted at 20 °C revealed secondary structure correlated to a typical α/β hydrolase fold protein for both Rv3802c and Rv3802c^{Δ1-67} (Figure 3.3a), confirming that both recombinant proteins appear to have formed the correct disulphide bond arrangement during the refolding step. The CD spectrum of Rv3802c are more characteristic of an α -helical trace than Rv3802c^{Δ1-67}, with a more prominent dip at 214 nm observed for Rv3802c than Rv3802c^{Δ1-67}, which correlates to the loss of the N-terminal predicted transmembrane domain.

To determine the effects of the loss of the N-terminal domain on the overall stability of Rv3802c^{Δ1-67}, thermal stability assays were conducted. Both proteins were gradually misfolded across a temperature gradient, with data collection occurring at wavelength of 214 nm. At the end of each thermal melting experiment, a CD spectrum was collected as previously described at the melting temperature, revealing both samples had unfolded during data collection (data not shown). Thermal melt (T_m) values were calculated in Graphpad Prism, using the Boltzmann sigmoidal equation (Niesen, *et al.*, 2007). The resultant T_m values of Rv3802c and Rv3802c^{Δ1-67} were 49.32 ± 0.16 °C and 45.78 ± 0.09 °C respectively (Figure 3.3b). A difference of ~ 3.5 °C indicate that the deletion of the N-terminal domain had only a minor effect on the overall stability of Rv3802c^{Δ1-67} compared to Rv3802c, confirming that both constructs are comparable in both protein fold and stability. The removal of the N-terminal domain as resulted in a significant change to the degree of cooperativity, as unfolding of Rv3802c is more gradual compared to Rv3802c^{Δ1-67}. The N-terminal domain may be acting to stabilise the protein *in vitro* via hydrophobic interactions through the predicted transmembrane region of the N-terminal domain.

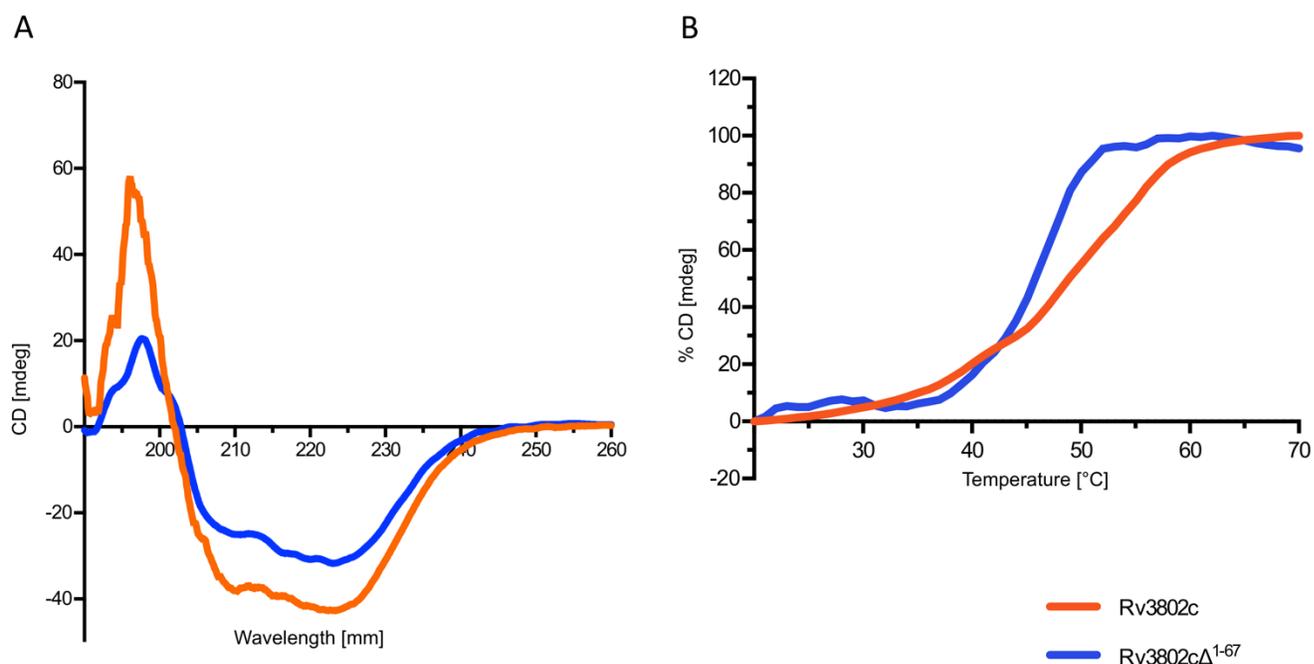


Figure 3.3: Circular Dichroism analysis of Rv3802c and Rv3802c^{Δ1-67}

- (A) Circular Dichroism spectra of Rv3802c (orange) and Rv3802c^{Δ1-67} (blue) were collected to ascertain folding of recombinant protein in solution. CD Spectra was conducted along 260nm – 190nm at a rate of 100 nm.min⁻¹ at 20°C, with a protein concentration of 0.5 mg.mL⁻¹. CD Spectra were collected in triplicate experiments.
- (B) Thermal melting analysis of Rv3802c (orange) and Rv3802c^{Δ1-67} (blue) was conducted to determine and compare thermal stabilities of the two proteins in solution. Temperature was gradually increased from 20°C-70°C at a rate of 1°C.min⁻¹. Results were normalized against the Rv3802c^{Δ1-67} spectra, resulting in % CD[mdeg] values on the Y-axis. All graphs were generated using *Graphpad Prism 6.0*.

3.3.3 Enzyme kinetic and inhibition assays of Rv3802c^{Δ1-67} against THL

With the successful production of recombinant Rv3802c^{Δ1-67}, effects of the removal of the N-terminal domain were characterized with regards to enzymatic function. The enzymatic activity of Rv3802c has previously been described (Crellin, *et al.*, 2010) using a synthetic colorimetric *p*-nitrophenyl butyrate lipolytic assay. Rv3802c cleaves the ester bond of *p*-nitrophenyl butyrate, releasing free *p*-nitrophenolate, which absorbs light at a wavelength of 405 nm that can be rapidly measured spectrophotometrically. The assay was conducted in a high throughput 96-well plate format in triplicate independent experiments. The enzymatic activity of Rv3802c^{Δ1-67} against *p*-nitrophenol butyrate was measured, and was directly compared to previously observed activity rates of Rv3802c.

The Michaelis-Menten kinetics of Rv3802c^{Δ1-67} were calculated, resulting in K_m and V_{max} values of $3.75 \text{ mM} \pm 0.61$ and $624 \text{ nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1} \pm 31.5$ respectively, which were compared to the Rv3802c K_m and V_{max} values of $4.52 \text{ mM} \pm 0.38$ and $241 \text{ nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1} \pm 7.8$ respectively (Figure 3.4a). The enzyme kinetic assays revealed no statistical significance between the K_m values of Rv3802c and Rv3802c^{Δ1-67}, indicating the removal of the N terminal domain does not affect substrate affinity of the enzyme. The V_{max} value of Rv3802c^{Δ1-67} was significantly higher than Rv3802c, and is most likely due to the substrate interacting with the exposed hydrophobic residues of the predicted transmembrane domain, reducing accessibility of the substrate to the substrate-binding site of the enzyme.

Additional to the kinetic activity, the inhibition of THL against Rv3802c^{Δ1-67} was also determined. The assay utilized was previously described for the analysis of THL and THL analogue inhibition rates against Rv3802c (West, *et al.*, 2009; West, *et al.*, 2011). The turbidimetric lipolytic assay utilized the cleavage of Tween-20 in the presence of calcium chloride, where precipitation occurs upon the formation of calcium salts between calcium chloride and the free fatty acid upon lipolytic cleavage of

the ester bond between the fatty acid and polysorbate headgroup. The turbidity is measured spectrophotometrically at a wavelength of 405 nm in 96-well plate format in triplicate independent experiments. A different assay was utilised to characterise inhibition due to increased sensitivity the turbidimetric lipolytic assay, as well as the significant decrease in background absorbance of the colorimetric *p*-nitrophenyl butyrate lipolytic assay. Inhibition of Rv3802c^{Δ1-67} by THL was determined across a 0 – 100 μM range and directly compared to previously observed inhibition levels against Rv3802c. Resultant IC₅₀ values of THL against Rv3802c^{Δ1-67} were calculated at 0.57 μM across a 0.37 – 1.02 μM 95 % confidence interval (Figure 3.4b). The values obtained for THL against Rv3802c were 1.82 μM across a 95 % confidence interval of 1.25 – 1.68 μM. The inhibition of Rv3802c^{Δ1-67} by THL are significantly tighter compared to Rv3802c. This could be due to either an increased affinity of THL against Rv3802c^{Δ1-67} compared to Rv3802c, or the loss of the predicted transmembrane domain has increased the amount of THL accessible to the enzyme active site within the assay. As a similar trend was observed in the *p*-nitrophenol butyrate enzyme assay, the latter is most likely. Despite this, Rv3802c^{Δ1-67} is enzymatically active at a comparable rate to Rv3802c, and can be inhibited by THL. The tighter affinity for THL against Rv3802c^{Δ1-67} compared to Rv3802c will aid in cocomplexation for crystallographic studies.

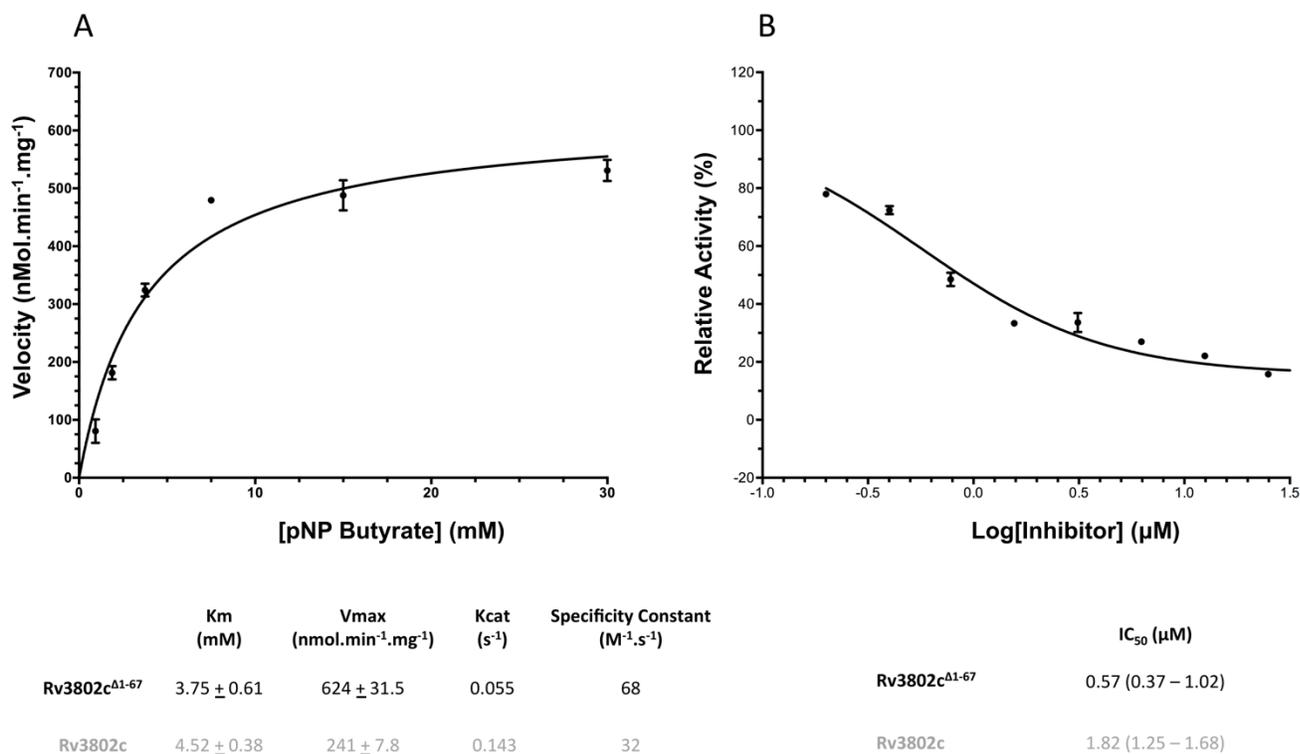


Figure 3.4 Enzyme kinetic activity and inhibition analysis of Rv3802c and Rv3802c^{Δ1-67}

- (A) The effects of removing the N-terminal transmembrane domain on Rv3802c function were measured via *p*-nitrophenyl butyrate spectrophotometric kinetic assay. No differences in substrate affinities were observed, with calculated K_m values for Rv3802c^{Δ1-67} 3.75 ± 0.61 mM being observed, compared to previously published values of 4.52 ± 0.38 mM (Crellin, *et al.*, 2010). An increase in V_{max} for Rv3802c^{Δ1-67} compared to Rv3802c indicated loss of the transmembrane domain allowed for higher accessibility to substrate *in vitro*.
- (B) Inhibition of Rv3802c^{Δ1-67} by THL was determined using a lipolytic Tween-20 cleavage assay. THL inhibited Rv3802c^{Δ1-67} tighter compared to Rv3802c, with IC_{50} values for Rv3802c^{Δ1-67} of 0.57 μM (0.37-1.02 μM) obtained, compared to previously published inhibition rates for Rv3802c of 1.82 μM (1.25-1.68 μM). The loss of the transmembrane domain resulted in higher accessibility of THL to Rv3802c *in vitro*. Duplicate readings were collected, with all experiments conducted in triplicate. Error bars represent Mean ± SEM. *Graphpad Prism* was used for all Michaelis-Menten kinetic analysis, as well as generating figures.

3.3.4 Crystallization and Structural Determination of Rv3802c^{Δ1-67}-THL

Crystallographic studies for both Apo-Rv3802c^{Δ1-67} and Rv3802c^{Δ1-67}-THL complex were conducted to characterize the structural mechanisms of THL binding, as well as to compare changes in the enzyme binding pocket in the presence of THL. Rv3802c^{Δ1-67} was complexed with THL at a 10-fold Molar excess for 16 h. Prior to crystallization, the Rv3802c^{Δ1-67}-THL complex was isolated and excess THL removed via size exclusion chromatography, followed by determining inhibition levels utilizing the lipolytic Tween-20 cleavage enzyme kinetic assay. The protein was successfully purified and was found to exhibit a ~90 % reduction in enzyme kinetic activity compared to Apo form, confirming the successful formation of the Rv3802c^{Δ1-67}-THL complex (data not shown).

Crystallisation trials of both Rv3802c^{Δ1-67}-THL and Apo-Rv3802c^{Δ1-67} were grown via the vapour diffusion method by incubating the protein and mother liquor at a 1:1 ratio. Crystals of Apo-Rv3802c^{Δ1-67} were grown in 24% (w/v) PEG 8000, 0.1 M Tris-HCl pH 7.6, 4 % (v/v) Acetonitrile after 24 h at 10 mg.mL⁻¹ with a rod shaped morphology (Figure 3.5a). Crystals of Rv3802c^{Δ1-67}-THL were grown in 21 % (w/v) PEG 2000 MME, 0.1 M Bis-Tris pH 5.7 after 3 d at 10 mg.mL⁻¹ with a hexagonal morphology (Figure 3.5b). Crystal morphologies and crystallisation conditions differed between Apo-Rv3802c^{Δ1-67} and Rv3802c^{Δ1-67}-THL samples, with Apo-Rv3802c^{Δ1-67} samples unable to crystallize in the same condition as the Rv3802c^{Δ1-67}-THL, and visa-versa. This indicated that the presence of THL within Rv3802c^{Δ1-67} has resulted in structural changes that had an effect on crystal contacts. To confirm that the crystals grown were Rv3802c^{Δ1-67}, crystals were washed 5 times in mother liquor and analysed on a 15 % SDS-PAGE against the protein sample used for crystallisation studies, and was found that the crystals were indeed Rv3802c^{Δ1-67} (Figure 3.5c). Crystals of Apo-Rv3802c^{Δ1-67} and Rv3802c^{Δ1-67}-THL were transferred for into a cryoprotectant consisting of mother liquor + 10 % (v/v) PEG 400 to protect the crystals against flash freezing and radiation damage.

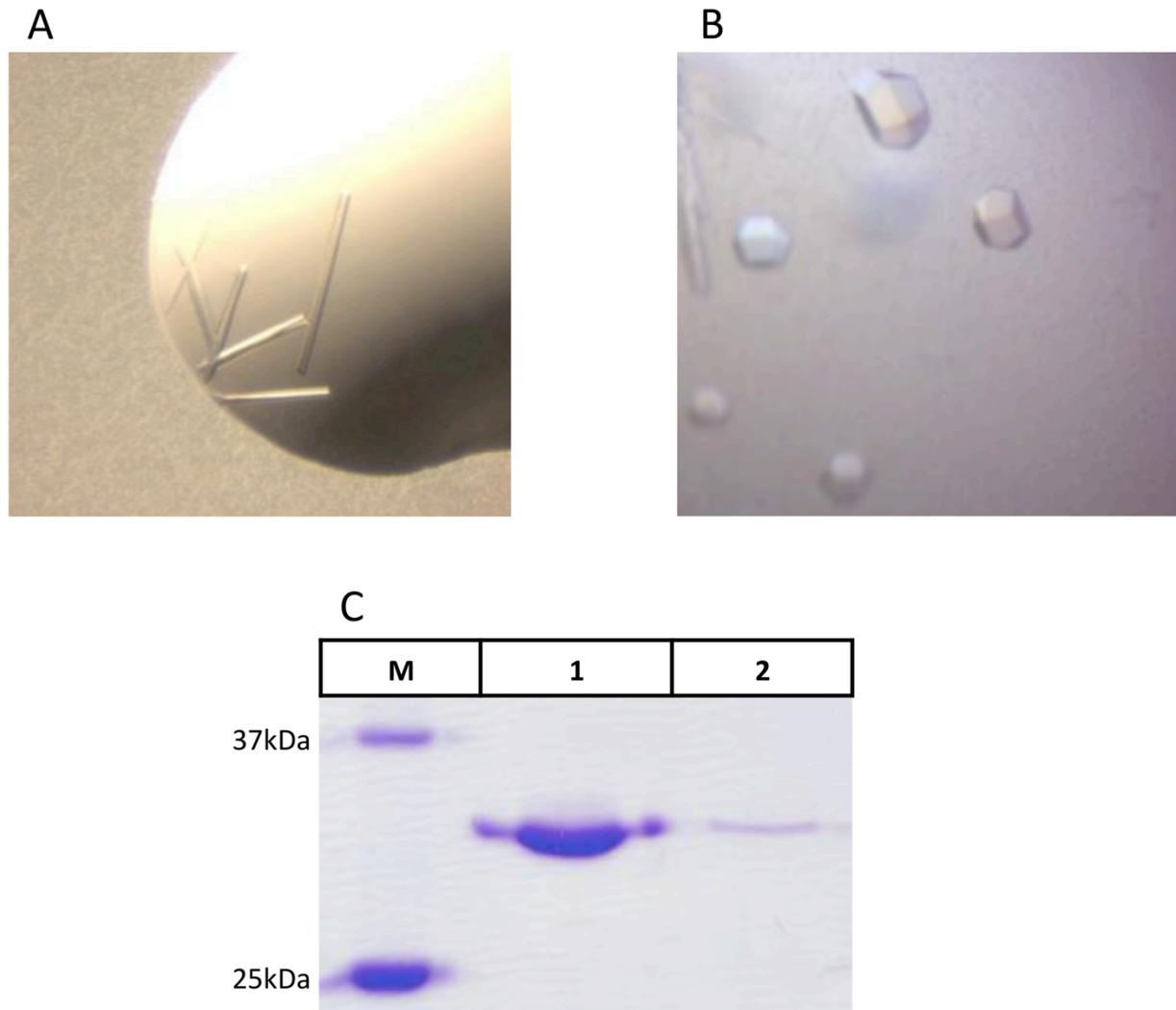


Figure 3.5 Crystals of Rv3802c^{Δ1-67} in its Apo form and in complex with THL

(A) Crystals of Rv3802c^{Δ1-67} exhibited a rod shaped morphology

(B) Crystals of Rv3802c^{Δ1-67}-THL exhibited hexagonal morphology

(C) 15% SDS-PAGE of Rv3802c^{Δ1-67}-THL protein sample used for crystallisation (1) and dissolved Rv3802c^{Δ1-67}-THL crystals (2). Molecular Weight markers are indicated in kDa.

X-ray diffraction data was collected on the MX1 beamline at the Australian Synchrotron, however a complete data set was collected only for Rv3802c^{Δ1-67}-THL. Despite extensive screening, crystals of Apo-Rv3802c^{Δ1-67} diffracted to an average maximum resolution of 7.0 - 8.0 Å, which is too low for extensive crystallographic structural analysis. A complete dataset was collected on a single crystal of Rv3802c^{Δ1-67}-THL, with crystals diffracting to a resolution of 2.9 Å (Table 3.1). The dataset was processed and scaled using *iMosflm* (Leslie and Powell, 2007) and *SCALA* (Evans, 2006) respectively, and the crystal was determined to belong to the primitive hexagonal P6₂22 space group. The unit cell dimensions for the structure was a= 126.15 Å, b= 126.15 Å, c = 234.01 Å, and $\alpha = 90.0^\circ$, $\beta = 90.0^\circ$, $\gamma = 120.0^\circ$. The crystal structure of Rv3802c^{Δ1-67}-THL was solved via molecular replacement by Phaser-MR (McCoy, *et al.*, 2007) using the crystal structure of the *M. smegmatis* homologue, MSMEG_6394, as the search model (PDB ID: 3AJA) sharing a sequence identity of 68 %, with a final TZF score of 14.4. Two copies of the molecule in the asymmetric unit were built, which corresponds to prediction determined by Matthews coefficient calculation. There was no evidence of higher order oligomerization within the crystal lattice, an observation that is consistent with the purification of Rv3802c^{Δ1-67} as a monomer. Wilson B factor was calculated to be 66.9 Å² by data quality assessment program xtriage (Adams, *et al.*, 2010). The crystal structure was manually adjusted using Coot (Emsley, *et al.*, 2010) and refined using phenix.refine as part of the phenix program suite (Adams, *et al.*, 2010).

Table 3.1: Data collection and refinement statistics for Rv3802c^{A1-67}-THL

Values in parentheses are for the outer resolution shell

	Rv3802c ^{A1-67} -THL
Space group	P6 ₂ 22
Temperature (K)	100
Resolution range (Å)	79.85 – 2.90 (3.06 – 2.90)
Wavelength (Å)	0.9357
Unit-cell parameters (Å, °)	a=126.15, b=126.15, c=234.01 α=90.0, β=90.0, γ=120.0
Total No. of reflections	409710 (58629)
No. of unique reflections	25218 (3591)
Multiplicity	16.2 (16.3)
Completeness (%)	100.0 (100.0)
CC (1/2)	0.998 (0.807)
R_{merge} (%)^a	14.0 (111.7)
R_{pim} (%)^b	3.6 (28.2)
Mean I/σ(I)^c	16.8 (2.5)
R_{work} (%)^d	20.5 (33.1)
R_{free} (%)^e	24.3 (39.8)
Number of non-hydrogen atoms	4048
Macromolecules	3895
Ligands	86
Water	67
Protein residues	529
R.M.S deviations from ideality	
Bond Lengths (Å)	0.008
Bond Angles (°)	1.32
Ramchandran Plot	
Favoured Region(%)	94.8
Allowed Region (%)	5.2
Outliers (%)	0.0
B-factors (Å²)	
Wilson B-factor	66.9
Average B-factors	69.2
Average Macromolecule	69.4
Average Ligand	77.0
Average Water	60.0

$$a \ R_{\text{merge}} = \frac{\sum hkl \sum_i |I_{hkl, i} - \langle I_{hkl} \rangle|}{\sum hkl \langle I_{hkl} \rangle}$$

$$b \ R_{\text{pim}} = \frac{\sum hkl [1/(N-1)]^{1/2} \sum_i |I_{hkl, i} - \langle I_{hkl} \rangle|}{\sum hkl \langle I_{hkl} \rangle}$$

c I is the integrated intensity and $\sigma(I)$ is the estimated standard deviation of that intensity

d $R_{\text{work}} = (\sum ||F_o| - |F_c||) / (\sum |F_o|)$ - for all data except as indicated in footnote e.

e $R_{\text{free}} = 5\%$ of data were used for the R_{free} calculation. ##

The hydrolysed form of THL was manually built into the crystal structure (Ligand ID: DH9) using Coot (Emsley, *et al.*, 2010) obtained from the complexation crystal structure of the Thioesterase Domain of Human Fatty Acid Synthase (FASTE) (PDB ID: 2PX6). At a resolution of 2.9 Å, the position of THL was unclear within the unbiased density, resulting in two possible orientations (Figure 3.6a-b). Modelling THL was based on the visibility of the electron density for palmitic core, as well as the proximity of the drug to the catalytic Ser175 residue. In both cases, the THL thioester is positioned within close proximity to Ser175, which may result in the formation of the acyl-enzyme intermediate as described in the FASTE-THL structure (see section 1.4.1) (Pemble, *et al.*, 2007), however orientation of the hexanoyl tail and the peptide moiety were reversed (Figure 3.6a-b). After refinement, the first orientation (Figure 3.6c) appeared to be most likely, with average B factors obtained of THL for each orientation being 67 Å² and 84 Å² respectively. Electron density is missing for the core of the molecule (Figure 3.6c), which is most likely due to flexibility of the region, as it is not involved in any direct contacts with Rv3802c.

The final R_{work} and R_{free} values were 20.5 % and 24.3 % respectively, with Ramachandran favoured, allowed and outlier regions of 94.8 %, 5.2 % and 0 % obtained respectively. The final refined structure of Rv3802c^{A1-67}-THL was given a MOLPROBITY score of 1.83 in the 100th percentile of all structures of comparable resolution.

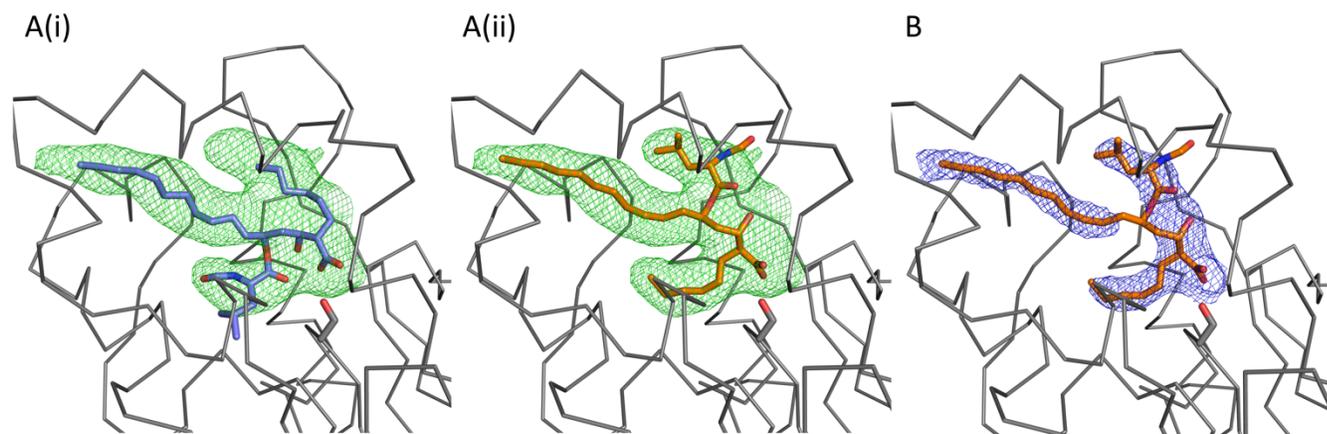


Figure 3.6: Modeling THL into unbiased electron density

- (A) Unbiased omit electron density in Chain B of the Rv3802c^{Δ1-67}-THL complex clearly revealed the presence of THL within the enzyme active site. Orientation of the peptidyl moiety and hexanoyl tail was unknown. Both possibilities were modeled based on hydrogen bonding between Ser175 of Rv3802c and the carboxylate ion of THL. After refinement, it was shown that the orientation with the peptidyl moiety facing the core of the protein (i) was incorrect, and the orientation with the hexanoyl tail facing the core of the protein (ii) was correct.
- (B) Refined density showed clear electron density for THL, with the exception of the core of the molecule. Fo-Fc unbiased omit electron density (green) and 2Fo-Fc simulated-annealing omit electron density (blue) were contoured to 3.0 σ and 0.9 σ simultaneously. THL is visualized in blue and orange respectively, and Rv3802c is visualized as grey ribbons. Ser175 is highlighted in stick representation.

3.3.5 Overview Structural Analysis of Rv3802c^{Δ1-67}-THL

The crystal structure of Rv3802c^{Δ1-67} confirmed that it is a member of the cutinase family of α/β -hydrolase fold with characteristic PE-PPE domains. The canonical α/β -hydrolase domain of Rv3802c includes a six-stranded parallel β -sheet (β 1- β 6) surrounded on both sides by four α -helices (α 1- α 3 and α 10) (Figure 3.7). Inserted between the β 5 and β 6 strands and between the β 6-strand and α 10-helix lies a second primarily helical domain (α 5- α 6, α 8- α 9) above the enzyme active site. This helical domain is representative of a “lid” domain typically found in esterases and lipases belonging to the α/β -hydrolase superfamily (Holmquist, *et al.*, 1995); facilitating substrate specificity and interface binding of this class of enzymes, aiding in the formation of the enzyme-substrate complex (Holmquist *et al.*, 1993; 1995; 2000). Two disulphide bonds, formed between Cys72-Cys164 and Cys264-Cys271, are required for the correct formation of the α/β hydrolase fold. (Figure 3.7a). The disulphide bond Cys72-Cys164 stabilises the overall α/β sandwich domain by linking the N-terminal region (Cys72) and C-terminal region (Cys164) of the α 2 helix. The disulphide bond Cys271-Cys271 contributes to the architecture of the enzyme active site by linking the β 6-strand and α 8-helix; acting as an intra-loop bridge that positions residues essential for enzymatic function within the enzyme active site (Figure 3.7a).

Critical to the lipolytic activity of α/β hydrolases is an active site consisting of a highly conserved catalytic triad, which in the case of Rv3802c is formed by Ser175-Asp268-His299 (Figure 3.7a). The nucleophilic Ser175 is located within the “nucleophilic elbow”; a feature conserved amongst α/β hydrolases comprised of a tight turn between the β 3-strand and the α 3-helix (Schrag and Cygler, 1997). At this position, the catalytic Ser175 O^γ is positioned at the mouth of the active site pocket for direct substrate interaction (Figure 3.7a). The catalytic Asp268 and His299 residues form a salt bridge, which act to stabilize both the enzyme active site, as well as the helical lid domain above the enzyme active site.

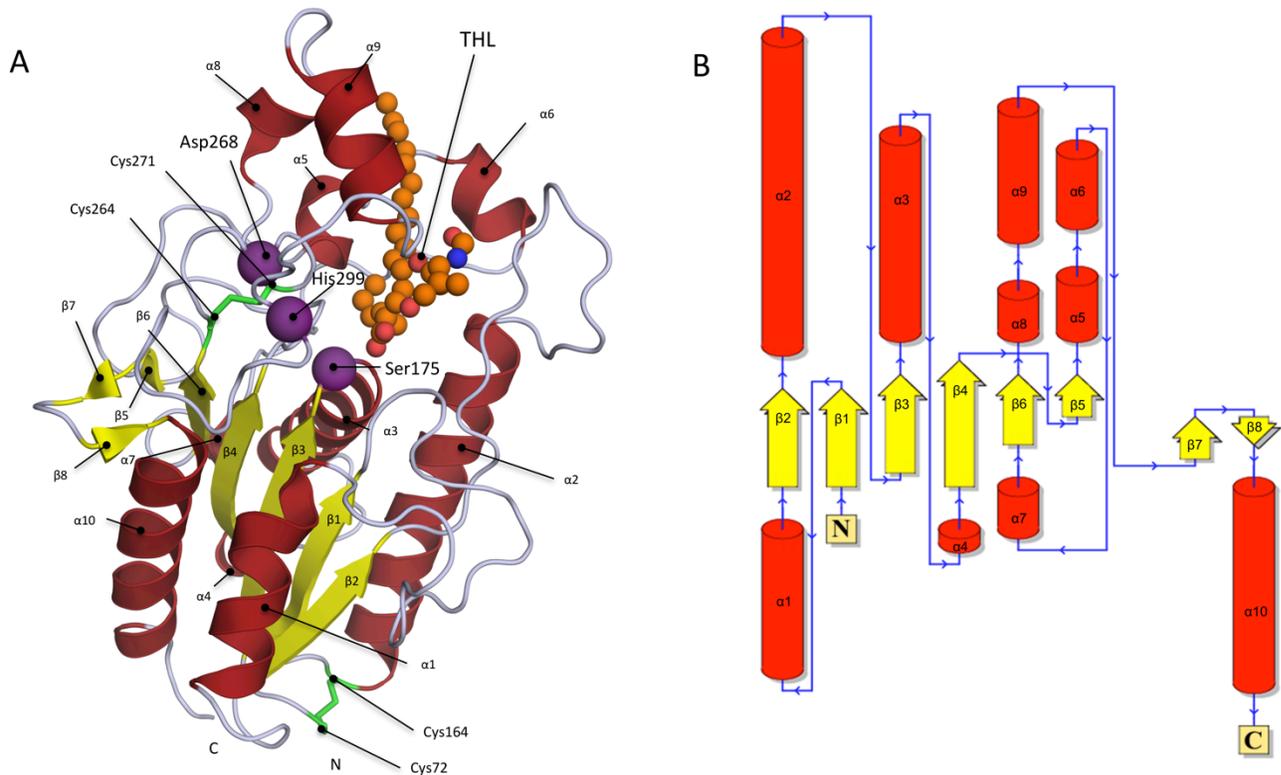


Figure 3.7: Overall structure of Rv3802c^{Δ1-67}-THL.

- (A) Rv3802c^{Δ1-67} adopts an α/β -hydrolase fold with six-stranded parallel β -sheet (β 1- β 6, yellow) surrounded by four α -helices (α 1-3, α 10, red). Intramolecular disulphide bonds (Cys72-Cys164 and Cys264-Cys271) are shown in green, and the catalytic triad typically found in serine esterases (Ser175, Asp268, His299) are shown in purple. THL is present within the enzyme active site is represented in orange.
- (B) Schematic of the secondary structure adapted from the Topology map generated by PDBsum³², with α -helices, β -strands and loops represented in red, yellow and blue respectively.

Previous studies have shown that single point mutations of either one of the three residues within the catalytic triad completely abolishes Rv3802c's ability to perform enzymatically (West, *et al.*, 2009).

3.3.6 Comparison of Rv3802c^{Δ1-67}-THL with Apo-MSMEG_6394

In the absence of a crystal structure of the apo-Rv3802c^{Δ1-67}, a direct comparison cannot be made on the structural changes of Rv3802c in the presence of THL. Despite this, a structural comparison can still be made between Rv3802c^{Δ1-67}-THL and apo-MSMEG_6394, the orthologue from *M. smegmatis* that shares a sequence identity of 68 %. The apo crystal structure of MSMEG_6394 overlays tightly with the Rv3802c^{Δ1-67}-THL crystal structure described here, with a core r.m.s.d value of 0.74 Å, as calculated by coot (Emsley, *et al.*, 2010) (Figure 3.8). However, as both crystal structures were solved at a resolution of 2.9 Å, only a comparison of the global structural changes between the two proteins in the presence of THL can be conducted.

In both MSMEG_6394 and Rv3802c, the helical lid domain sits on top of the enzyme active site, and in the absence of substrate, interacts directly with the enzyme cavity in a “closed” conformation to block solvent accessibility to the enzyme active site. In the presence of THL, the lid domain has shifted, exposing the active site cavity in an “open” conformation. A maximal shift of 11 Å is observed in the $\alpha 8$ and $\alpha 9$ lid domain helices in the presence of THL, with $\alpha 5$ and $\alpha 6$ remaining stationary (Figure 3.8). This shift facilitates substrate binding within the active site pocket, aided by a hinge loop at the direct N-terminal end of the $\alpha 9$ helix (Fig 3b). While sequence identity is not strictly conserved, the hinge loop is comprised of smaller, flexible amino acid residues. In the case of Rv3802c, the hinge loop spans residues Ala291-Gly292-Gly293-Ala24-Gly295. Due to its flexibility, electron density is lacking in chain A of the Rv3802c^{Δ1-67}-THL structure, and is not present in the crystal structure of MSMEG_6394.

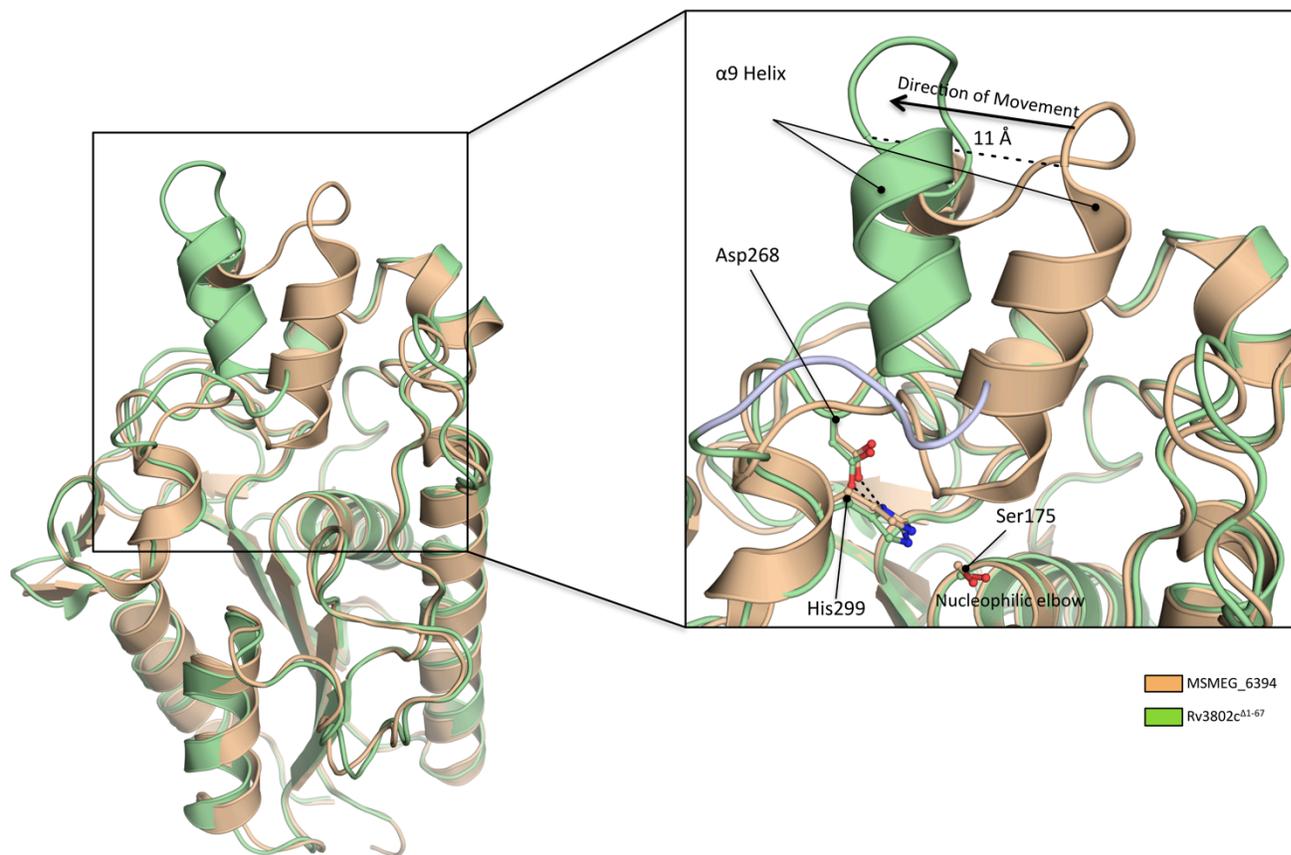


Figure 3.8 Rv3802c^{Δ1-67} active site cavity and structural overlay with MSMEG_6394

Structural overlay of Rv3802c^{Δ1-67} (green) and MSMEG_5817 (wheat). Shift in the $\alpha 9$ helix between the “closed” and “open” conformations of the lid region, with maximal calculated shift of 11 Å observed. The direction of movement of the $\alpha 9$ helix is indicated to facilitate substrate binding. The flexible hinge region facilitates the shift between “closed” and “open” conformations with a sequence of Ala291-Gly292-Gly293-Ala294-Gly295 (blue). Salt bridge formation between Asp268 and His299 of the catalytic triad aid in stabilising the conformational shift, with Ser175 located at the tip of the nucleophilic elbow, and faces into the enzyme active site.

In the closed confirmation, access to the catalytic serine is blocked, with the nucleophilic elbow existing in a small cavity with a volume of $\sim 240 \text{ \AA}^3$ (Figure 3.9a). Upon substrate binding, the open cavity extends to solvent expose the hydrophobic pocket and interface cavity, resulting in an increase in volume to $\sim 1530 \text{ \AA}^3$ (Figure 3.9b). Opening of the cavity is dependant on the presence of a lipid based substrate, where the lipid tail binds directly to the exposed hydrophobic core: both as a means of stabilizing the enzyme active site, as well as to lock and position the substrate for direct contact with the catalytic serine.

3.3.7 Structural Analysis of the THL Binding Site

The Rv3802c^{Δ1-67}-THL complex crystallised with two molecules (Chain A and B) present in the asymmetric unit. While electron density was visible for the hydrolysed THL product in both chain A and chain B, the drug could only be confidently built into the unbiased density of a single molecule (Chain B) of Rv3802c^{Δ1-67}. While previously reported that THL binds in covalently to the catalytic serine as an ester in an extended conformation (Pemble, *et al.*, 2007), only the hydrolysed product is present within the structure (Figure 3.10). THL is readily hydrolysed in aqueous solution, which is observed due to the 16 h co-crystallisation stage and 5 d crystallisation incubation period.

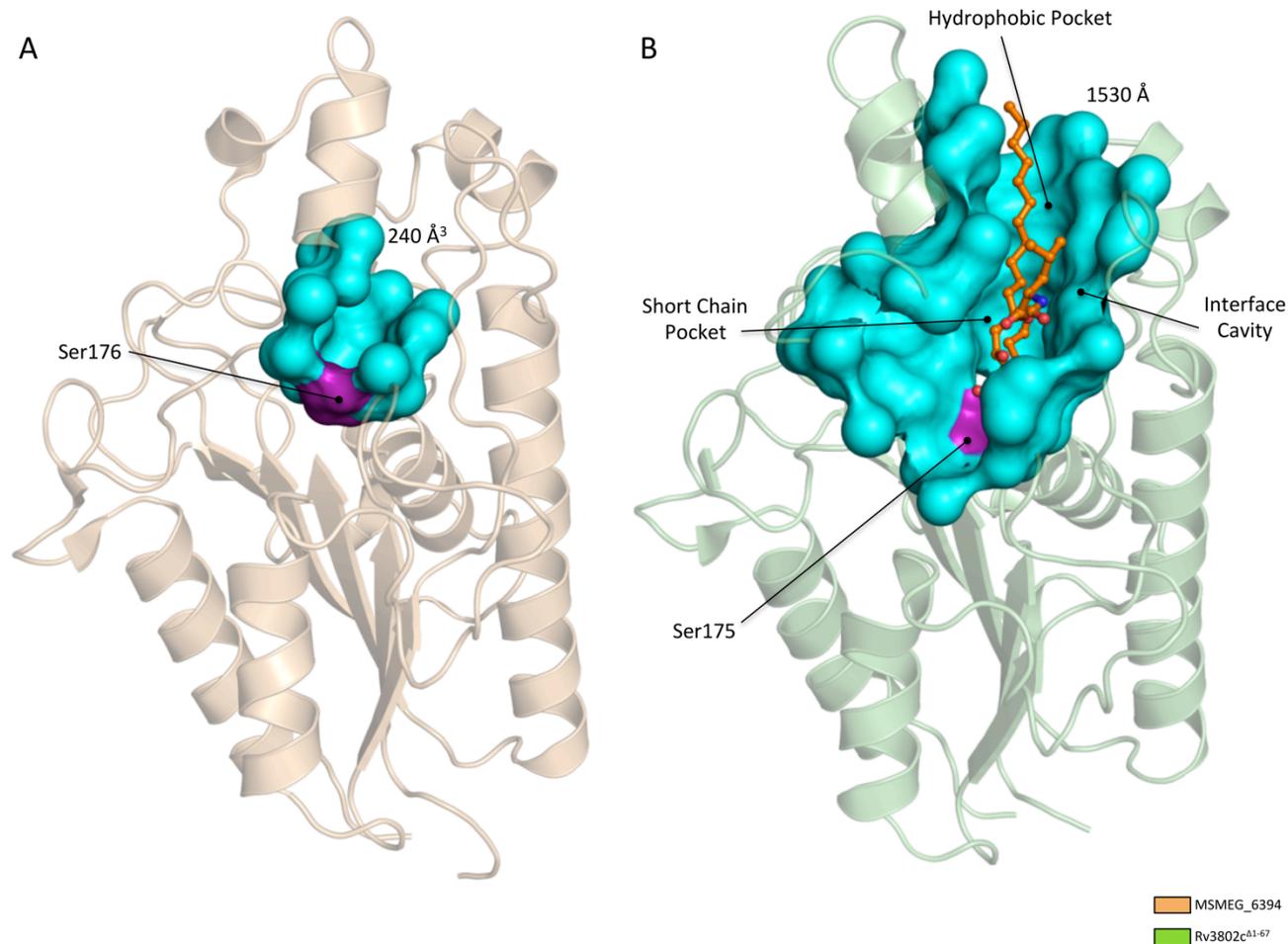


Figure 3.9 Solvent accessibility between “open” and “closed” conformations

Calculated solvent exposed surface of (a) MSMEG_6394 in the “closed” confirmation, and of (b) Rv3802c^{Δ1-67}-THL in the “open” confirmation. Solvent accessibility volumes were calculated at 240Å³ and 1530Å³ respectively. Solvent exposed surface is represented in cyan, with the catalytic Serine Residue represented in purple. Hydrophobic pocket, short chain pocket and the interface cavity which each interact with THL are indicate. Solvent accessible surfaces of the enzyme active site were calculated by CASTp (Dundas, *et al.*, 2006).

Previous models suggest that THL forms a stable acyl-enzyme intermediate, mediated by a covalent ester bond formation between the catalytic serine and the β -lactone form (Hadvary, *et al.*, 1991; Luthi-Peng, *et al.*, 1992; Pemble, *et al.*, 2007). Upon undergoing water mediated nucleophilic attack, the acyl-enzyme intermediate collapses to form a carboxylate at the C1 position. In the case of the Rv3802c^{Δ1-67}-THL complex, we observe this form in chain B and potentially chain A. The hydrolysed product of THL has previously been observed to interact with the catalytic serine via water mediated hydrogen bond formation (Pemble, *et al.*, 2007), however in the case of the Rv3802c^{Δ1-67}-THL complex, the catalytic Ser175 forms a direct hydrogen bond with the carboxylate ion at the C1 position (Figure 3.10).

In addition to the hydrogen bond interaction with the hydrolysed product, THL interacts with the surface of the active site pocket through a variety of chemical moieties (Figure 3.10b). The palmitic core, spanning between C1 – C16, binds in the hydrophobic pocket ~ 17 Å in length formed between lid domain $\alpha 6$ and $\alpha 9$ helices. In its APO form, the helical lid domain is held in a closed conformation stabilised by a network of Leucine residues on the $\alpha 5$, $\alpha 6$ and $\alpha 9$ helices. Upon THL binding, the lid domain clamps down on the palmitic core in the “open” conformation (Figure 3.10). Leu238 and Leu243 on the $\alpha 6$ helix, Leu287 and Leu290 on the $\alpha 9$ helix, along with Leu232 on the $\alpha 5$ helix, form direct hydrophobic bonds with the palmitic core of THL (Figure 3.10b).

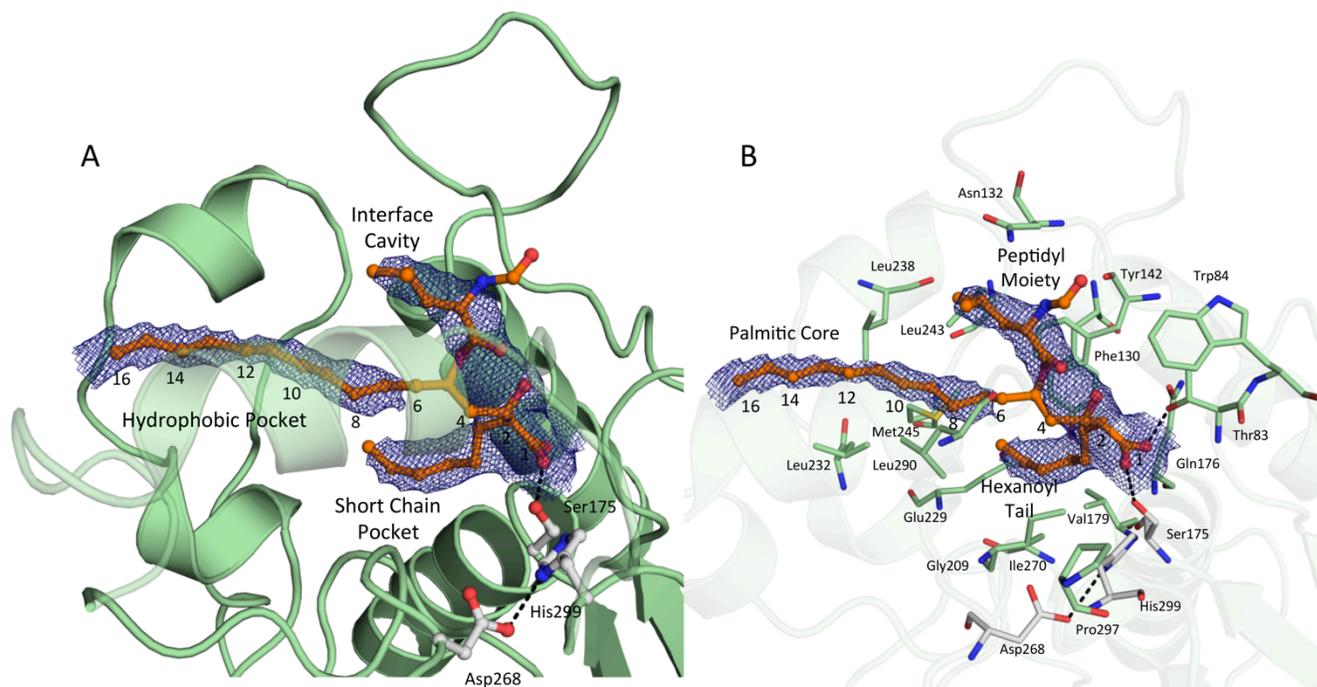


Figure 3.10 Structural characterizations of THL bound within the Rv3802c enzyme active site.

- (A) THL (orange) bound to the active site of Rv3802c Chain B, with carbon numbers indicated between 1-16. The THL scaffold is divided into three fragments: peptidyl moiety (*N*-formyl- L-leucine substituent extending off the C5 carbon atom), Hexanoyl tail (C2 substituent) and palmitic core. Direct hydrogen bond formation is observed between Ser175 and THL.
- (B) Detailed interactions between Rv3802c and THL. Each constituent interacts with the hydrophobic core (Palmitic Core), Interface cavity (Peptidyl Moiety) and Short Chain Pocket (Hexanoyl Tail). Contact residues are represented in stick form (green). The catalytic triad Ser175-Asp268-His299 are highlighted as white sticks. Black dashes represent salt bridge formation between Asp268 and His299, as well as hydrogen bond formation between THL and Ser175 and Thr83 respectively. 2Fo-Fc simulated-annealing omit electron density (blue) contoured at 0.9 σ generated by PyMOL.

The hexanoyl tail binds into the short chain pocket buried at the core of the active site pocket (Figure 3.10 & 3.11). The binding is predominantly mediated through hydrophobic interaction with Ile170: binding both the hexanoyl tail and the C2 carbon of the palmitic core (Figure 3.10). In addition, van der Waals interactions with Tyr142, Gln176 and Glu229 act to stabilise the hexanoyl tail within the short chain pocket (Figure 3.10b). The peptidyl moiety (or the *N*-formyl-L-leucine moiety) binds at the predominantly surface exposed interface cavity directly above the Ser175 nucleophilic elbow (Figure 3.10b). The *N*-formylamide group interacts predominantly with Asn132, located at the rim of the cavity, through van der Waals interactions (Figure 3.10b). Trp84 and Gly292 act to lock the L-Leucine component in place through hydrophobic bond formation (Figure 3.10b).

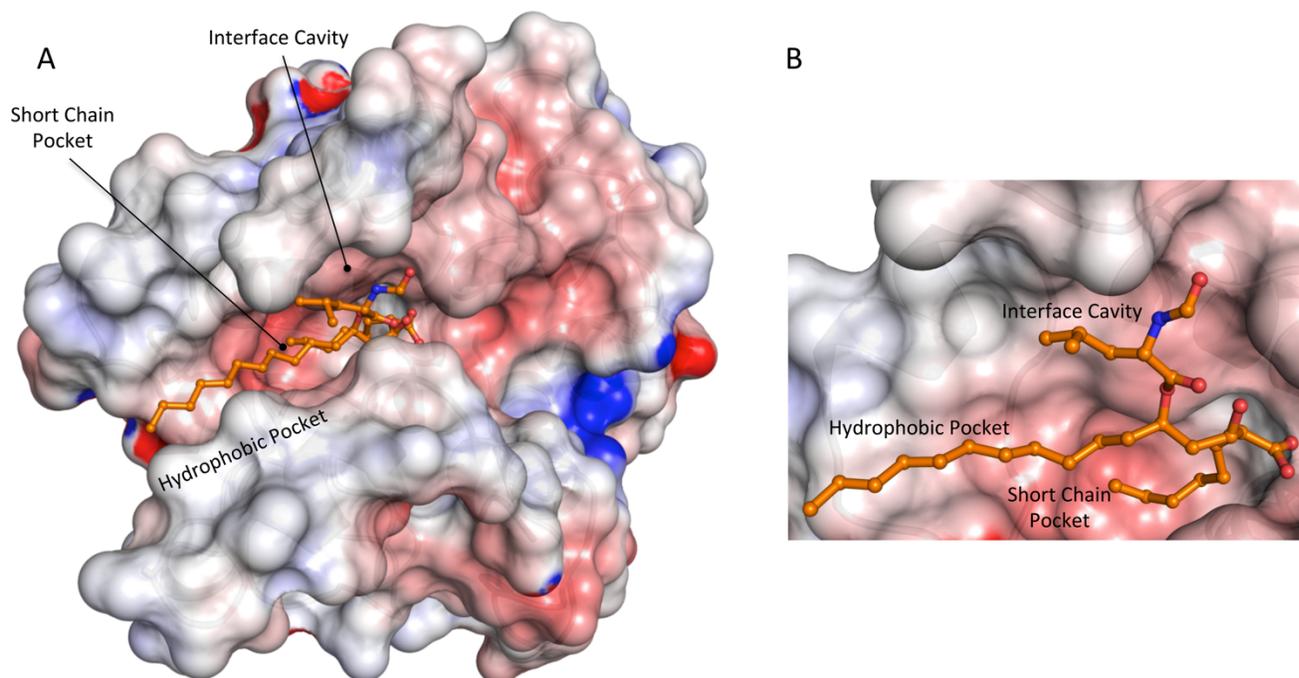


Figure 3.11: Surface electrostatic characterisation of Rv3802c

- (A) Top-view of the Rv3802c^{Δ1-67}-THL complex, with the hydrophobic pocket, short chain pocket and interface cavity highlighted.
- (B) Image rotated 90° and zoomed to visualize the electrostatic characterization of THL binding within the enzyme active site. Electrostatic surface calculations were conducted using the APBS plugin (Baker, *et al.*, 2001) in PyMOL. The potential contours are shown on a scale from +6.0 (Blue) to -6.0 $k_B T e^{-1}$ (red), with white indicating a value close to 0 $k_B T e^{-1}$. Both short chain pockets and interface cavity are slightly electrostatically negative, with the hydrophobic pocket-exhibiting close to zero electrostatic charge.

3.4 Discussion

Central to the process of rational drug design is a high-resolution experimental model, as obtained by x-ray crystallography or nuclear magnetic resonance (NMR), of the target protein and the initial drug compound. The model provides a view of the molecular mechanisms of drug inhibition, by visualization of contacts between protein residues and the initial drug. This supplies the initial blueprint for the rational redesigning of the initial drug compound to form additional and tighter interactions with the target protein, resulting in a higher affinity inhibition. The Rv3802c^{Δ1-67}-THL crystal structure provides the starting point for the rational design of higher affinity compounds against Rv3802c.

While THL is a US Food and Drug Administration (FDA)-approved anti-obesity drug marketed under the name Orlistat, it has been shown to have potent anti-TB effects, inhibiting and disrupting cell wall formation in several mycobacterial species (Kremer, *et al.*, 2005). Rv3802c is the most characterised mycobacterial target of THL, exhibiting targeted *in vitro* and *in vivo* inhibition. In addition, a number of additional cutinase-like proteins with characteristic α/β hydrolase folds from *M. tuberculosis*, Cfp21 (Rv1984) and Cut4 (Rv3452), have been recently identified as THL inhibition targets (Dedieu, *et al.*, 2013). Cfp21 and Cut4, along with Rv3802c, are three of the seven identified cutinase-like proteins (CLP) within the *M. tuberculosis* H37Rv genome (Dedieu, *et al.*, 2013), and despite displaying varying immunological and enzymatic effects (West, *et al.*, 2008), are all inhibited tightly by THL. Characterisation the THL inhibition rates against the remaining four CLPs would provide further insight into the mechanisms of THL against *M. tuberculosis*. Due to its potent anti-TB effects, and its inhibition of multiple essential mycobacterial targets, THL is an attractive initial drug for the development of new anti-TB compounds through rational design.

In the absence of structural data at the time, a THL lead compound library was created (West, *et al.*, 2011); comprising of a number of higher affinity compounds, with the highest affinity THL compounds increasing *in vitro* inhibition against Rv3802c and *in vivo* inhibition against *M. tuberculosis* 10-fold, into the nanomolar range. The initial THL compound (Figure 3.12a) was modified with changes to the peptidyl moiety, substituting the *N*-formyl-L-leucine moiety for alternative side chains to probe the enzyme active site for higher affinity interactions. These included lipophilic side chains to probe hydrophobicity, aromatic and non-aromatic rings for additional hydrogen bond formation and to test rigidity of the enzyme active site pocket, and finally flexible non-aromatic heterocyclic side chains as an alternative to rigid aromatic side chains (West, *et al.*, 2011).

It was found that the compounds containing the additional ester side chains resulted in a dramatic loss of potency, both *in vitro* and *in vivo* (West, *et al.*, 2011) (Figure 3.12b). Within the Rv3802c^{Δ1-67}-THL crystal structure, the peptidyl moiety docks within the polar interface cavity, which has been shown to exhibit a highly electrostatically negative charge (Figure 3.11b). Coupled with size restrictions, incorporation of a lipophilic side chain of varying lengths results in a loss of inhibition rates against Rv3802c. As the reduction of *in vivo* inhibition rates against *M. tuberculosis* was observed, it is a possibility that many targets of THL exhibit architectural similarities of the enzyme active site interface cavity.

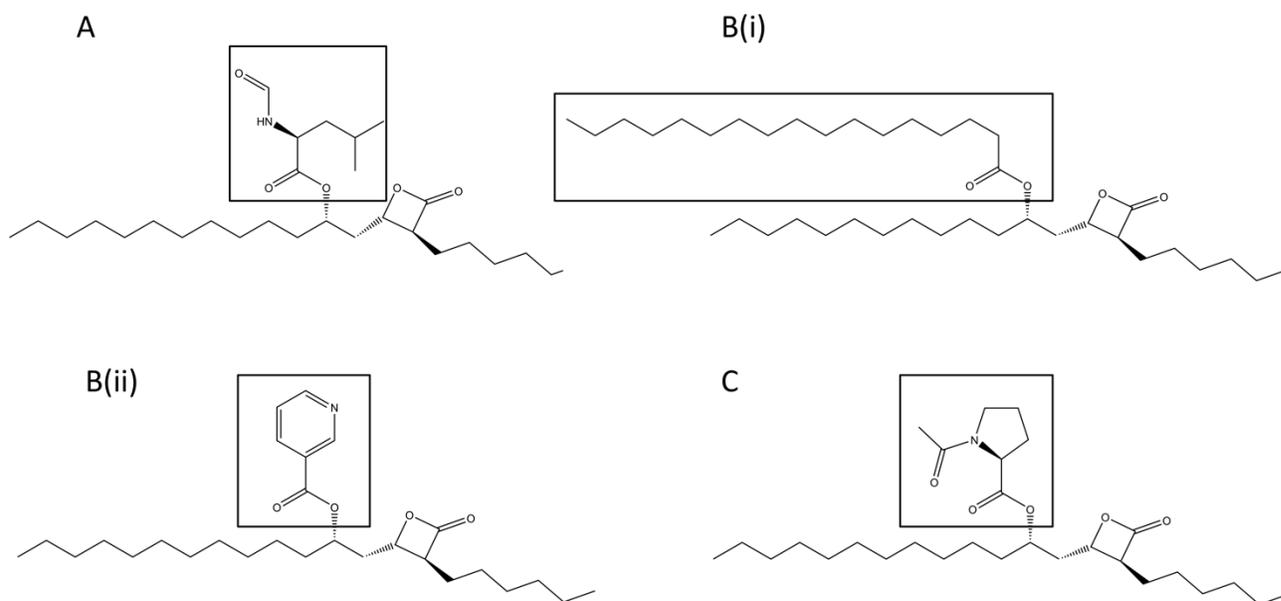


Figure 3.11 Chemical structure of THL and modified THL compound

- (A) The original THL structure was modified at the site of the peptidyl moiety (highlighted) as a means of obtaining higher affinity inhibition against Rv3802c.
- (B) Compounds exhibiting (i) hydrophobic tails and (ii) a pyrrole group resulted in a decrease in inhibition rates, both against Rv3802c *in vitro* and *M. tuberculosis in vivo*.
- (C) Altering the peptidyl moiety to include non-aromatic heterocyclic side chains with prolyl ester side chains resulted in a 10-fold increase in inhibition rates against Rv3802c *in vivo*.

Improvements in inhibition was observed with polar side chains and, with the exception of a compound with a pyrrole side chain (Figure 3.12c), all resulted in significantly tighter inhibition compared to THL (West, *et al.*, 2011). Non-aromatic heterocyclic side chains resulted in nanomolar range inhibition, with the inclusion of L- and D-prolyl ester side chains providing the best results (Figure 3.12d). In the case of the Rv3802c^{Δ1-67}-THL structure, there are no close interactions observed between the interface cavity and the peptidyl moiety. With the inclusion of larger side chains comprised of mixed aromatic and aliphatic groups, access to additional contact residues to Rv3802c may have been provided, allowing for tighter inhibition. It would be possible to dock these improved THL compounds into the Rv3802c^{Δ1-67} crystal structure to gain insight into their observed binding affinities, however due to the limited resolution of the crystal structure, as well as a lack of electron density of the peptidyl moiety, only an inadequate comparison can be drawn. As a means of validation, a crystal structure of Rv3802c in complex with these higher affinity compounds would provide the best insight into the mechanisms of tighter inhibition levels compared to THL; thus ultimately providing an additional model for downstream rational drug design.

In conclusion, the crystal structure of the essential lipase, Rv3802c, has been solved lacking the N-terminal predicted transmembrane domain in complex with the anti-TB drug THL. The crystal structure revealed that Rv3802c possesses a characteristic α/β hydrolase fold typical of serine esterases, with conservation in the positions of the catalytic Ser175 – Asp268 – His299 triad. In the presence of THL, the lid domain that enables substrate binding is present in an “open” conformation, with conformational changes facilitated by a flexible loop domain. The lid domain acts as a hydrophobic clamp that locks the palmitic core moiety of the THL molecule. THL is present in the crystal structure in its hydrolysed form, in which the β -lactone component undergoes a water mediated nucleophilic attack facilitated by Ser175 to convert it to a carboxylate ion. Hydrogen bond formation between the carboxylate and Ser175 are observed. While no close interactions are observed between Rv3802c and

the peptidyl moiety, several lead compounds created based on the original THL drug involve changes to the chemical structure of the peptidyl moiety. The next stage is to further define the binding mechanisms of the higher affinity lead compounds to THL by obtaining their complexation crystal structures in the hopes of developing higher affinity compounds targeted specifically to Rv3802c. This method can be expanded further for additional lead inhibitor compounds, or for additional drug targets of THL, which is described in more detail in chapter 7.

Chapter 4: Crystallization and X-ray Data Collection of MSMEG_5817

4.1 Introduction

Chapter 4 is focused on preliminary results obtained for the *M. smegmatis* protein, MSMEG_5817, of unknown function; with follow up work described in chapters 5 and 6. The previous two chapters have focused on utilizing x-ray crystallography as a means of characterizing the NADPH and INH binding sites of the aldo-keto reductase Rv2971, and for understanding the structural mechanisms of THL inhibition of the essential lipase Rv3802c, both for the purposes of rational drug design. In the case of MSMEG_5817, x-ray crystallography is employed as a means of gaining insight into a protein's function based on its x-ray crystal structure. Chapter 4 contains a paper published in April 2013 in *Acta Crystallographica Section F - Structural Biology and Crystallization Communications*, and describes the preliminary cloning, overexpression and purification of recombinant protein, with a focus on the crystallization and x-ray data collection of MSMEG_5817 for structural determination.

MSMEG_5817 has recently been identified as a novel protein within *M. smegmatis* essential for survival within host macrophages, via a presently unknown mechanism (Pelosi *et al.*, 2012). Previous attempts to characterise the function of the MSMEG_5817 family have been unsuccessful (Pelosi *et al.*, 2012). As a means of gaining insight into its biological function, the aim of my project was to solve the crystal structure of MSMEG_5817. The crystal structure will provide valuable insight into its biological role by identifying structural orthologues that have been functional characterised. Further insight into the function of MSMEG_5817 will be performed by variety of biochemical assays based on the function of its identified structural orthologues.

The paper describes the methods utilized in recombinant protein production, crystallization and x-ray data collection. The paper describes the crystallization and x-ray data collection of crystals generated by two derivatives of MSMEG_5817. The unique nature of the MSMEG_5817 protein family means that no structural homologues exist within the PDB exhibiting significant sequence identity. For that reason, x-ray data collection was conducted on crystals generated from both the native MSMEG_5817 and a selenomethionine derivative, to be utilized in structural determination by experimental phasing. The paper describes the x-ray data collection strategies for crystals generated by both derivatives to be utilized in structural determination by experimental phasing. Initially, soluble recombinant protein of MSMEG_5817 and its *M. tuberculosis* orthologue, Rv0807, was produced, however only crystals of high diffraction quality were generated for MSMEG_5817. In addition, MSMEG_5817 also contains two methionine residues in its amino acid sequence, while Rv0807 only contains one, making MSMEG_5817 more suitable for the selenomethionine phasing experiments. Structural characterization of MSMEG_5817 will be further elaborated upon in Chapter 5.

Declaration for Thesis Chapter 4

Declaration by candidate

In the case of Chapter 4, the nature and extent of my contribution to the work was the following:

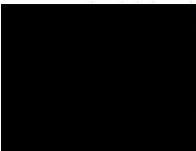
Nature of contribution	Extent of contribution (%)
Expression, purification, and crystallisation of native and Se-Met derivative MSMEG_5817. X-ray data collection of native MSMEG_5817. Data analysis, manuscript and figure preparation.	70%

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Phooi Chan	Cloning and expression trials	15%
Dene Littler	X-ray data collection of Se-Met MSMEG_5817 crystal	N/A
Julian Vivian	X-ray data collection of Se-Met MSMEG_5817 crystal	N/A
Rajini Brammananath	Supervision of Phooi Chan, data interpretation and manuscript preparation	N/A
Paul Crellin	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Ross Coppel	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Jamie Rossjohn	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Travis Beddoe	Conception, supervision, direction, data interpretation and manuscript preparation	N/A

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's Signature  Date 21/10/15

Main Supervisor's Signature  Date 21.10.15

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the



Acta Crystallographica Section F

Structural Biology
and Crystallization
Communications

ISSN 1744-3091

Adam E. Shahine,^{a,b} Phooi Y.
Chan,^{b,c} Dene Littler,^{a,b}
Julian Vivian,^{a,b} Rajini
Brammananth,^{b,c} Paul K.
Crellin,^{b,c} Ross L. Coppel,^{b,c}
Jamie Rossjohn^{a,b} and Travis
Beddoe^{a,b*}

^aDepartment of Biochemistry and Molecular
Biology, Monash University, Clayton, Victoria,
Australia, ^bAustralian Research Council (ARC)
Centre of Excellence in Structural and Functional
Microbial Genomics, Monash University,
Clayton, Victoria, Australia, and ^cDepartment of
Microbiology, Monash University, Clayton,
Victoria, Australia

Correspondence e-mail:
travis.beddoe@monash.edu

Received 22 March 2013
Accepted 16 April 2013



© 2013 International Union of Crystallography
All rights reserved

Cloning, expression, purification and preliminary X-ray diffraction studies of a mycobacterial protein implicated in bacterial survival in macrophages

Mycobacterium species have developed numerous strategies to avoid the antimycobacterial actions of macrophages, enabling them to survive within the generally inhospitable environment of the cell. The recently identified MSMEG_5817 protein from *M. smegmatis* is highly conserved in *Mycobacterium* spp. and is required for bacterial survival in macrophages. Here, the cloning, expression, purification and crystallization of MSMEG_5817 is reported. Crystals of MSMEG_5817 were grown in 1.42 M Li₂SO₄, 0.1 M Tris-HCl pH 7.7, 0.1 M sodium citrate tribasic dihydrate. Native and multiple-wavelength anomalous dispersion (MAD) data sets have been collected and structure determination is in progress.

1. Introduction

Macrophages are key components of host defence pathways that play a role in the control and eradication of foreign microbes, including pathogenic and nonpathogenic mycobacteria. Pathogenic mycobacterial species have developed numerous strategies to avoid the antimycobacterial actions of macrophages and to survive within the environment of the cell, frequently resulting in disease (Chandra *et al.*, 2011). Upon infection, the invading mycobacteria are internalized by host macrophages and taken into phagosomes, where they are able to survive for long periods of time. Mycobacteria create a hospitable environment within phagosomes through the alteration of host signalling pathways that mediate correct vesicular membrane trafficking and the formation of the phagolysosome. The resultant phagosomes fail to mature and do not fuse with late endosomes, and the lysosomes fail to recruit lysosomal hydrolases that are required for the degradation of internalized microbes (Armstrong & Hart, 1971; Fratti *et al.*, 2003). Identification of the mechanisms utilized by invading mycobacteria to disrupt host-macrophage antibacterial activities is paramount for the understanding of mycobacterial virulence and disease. While *Mycobacterium smegmatis* is generally considered to be nonpathogenic, it does have a limited capacity to survive and multiply within macrophages and to delay phagosomal acidification, making it a suitable model system to study intracellular mycobacterial survival (Anes *et al.*, 2006). In previous studies, the members of a random Tn611 transposon mutant library of *M. smegmatis* were screened for a number of atypical phenotypes indicative of cell-wall defects (Billman-Jacobe *et al.*, 1999; Kovacevic *et al.*, 2006; Patterson *et al.*, 2000). One mutant, designated Myco132, exhibited altered colony morphology and an increased ability to absorb dyes from growth media (Pelosi *et al.*, 2012). The gene disrupted in Myco132 was identified by sequencing as MSMEG_5817 (Pelosi *et al.*, 2012). MSMEG_5817 encodes a conserved hypothetical protein of 128 amino-acid residues that has orthologues in a number of pathogenic mycobacteria. Inactivation of the MSMEG_5817 gene led to accelerated cell death of *M. smegmatis* in J774A.1 macrophages within the first 8 h (Pelosi *et al.*, 2012). Complementation of the mutant with an active copy of the gene restored survival to near-wild-type levels. In addition to decreased survival in macrophages, there was also a significant increase in NF- κ B activation in macrophages infected with the MSMEG_5817 mutant compared with those

crystallization communications

infected by wild-type (WT) *M. smegmatis* (Pelosi *et al.*, 2012). Since NF- κ B activation plays key roles in regulating host immune responses through cell activation and the induction of cytokine production, MSMEG_5817 may be involved in suppressing NF- κ B activation to promote survival. Although MSMEG_5817 is implicated in bacterial survival within macrophages, it is not important for growth, with cultures grown *in vitro* in rich medium showing no differences in growth rates between Myco132 and WT *M. smegmatis* (Pelosi *et al.*, 2012). In this paper, we report the expression, purification and crystallization of this protein. The structural determination of MSMEG_5817 will provide valuable insight into the mechanistic pathways that mediate the function of this protein.

2. Materials and methods

2.1. Cloning

The MSMEG_5817 gene was PCR-amplified from *M. smegmatis* genomic DNA using HotStar HiFidelity polymerase (Qiagen) and specific sense (5'-GGAATTGC**ATATGGCCAGCCGCGTAGTG**-C-3') and antisense (5'-**GGAATTCCTACGGGTCGATACGCACCACGGG**-3') primers. The highlighted sequences in the primers represent *NdeI* and *EcoRI* restriction sites, respectively. The 567 bp PCR-amplified product was purified using the MO BIO kit (MO BIO Laboratories) and the fragment was digested with *NdeI* and *EcoRI* (NEB) and cloned into pET-28c(+) vector (Novagen), giving a protein with a thrombin-cleavable N-terminal hexahistidine tag. The resultant N-terminal amino-acid sequence of the recombinant MSMEG_5817 is MGSSHHHHHSSGLVPRGSH**MASRR**, where the sequence in bold is the MSMEG_5817 protein starting at residue 1. The clone obtained was confirmed by DNA sequencing.

2.2. Overexpression of recombinant protein

Recombinant protein was overexpressed in *Escherichia coli* host strain BL21 (DE3). The clone was transformed into competent BL21 (DE3) cells and transformants were selected on LB agar plates containing 34 $\mu\text{g ml}^{-1}$ kanamycin. Pre-inoculum cultures were grown overnight in LB-kanamycin medium from a single colony. The starter culture was used at a dilution of 1:100 to inoculate 800 ml fresh LB medium containing 34 $\mu\text{g ml}^{-1}$ kanamycin and was grown at 310 K until the OD₆₀₀ reached 0.6. Expression of the protein was induced by the addition of 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) and the cells were allowed to grow for an additional 4 h at 310 K. The cells were harvested by centrifugation and stored overnight at 193 K. Expression of the selenomethionine derivative of MSMEG_5817 was performed in *E. coli* host strain B834 (DE3). Pre-inoculum cultures grown overnight in LB-kanamycin medium from a single colony were then diluted in 6 l fresh LB medium containing 34 $\mu\text{g mg}^{-1}$ kanamycin and grown at 310 K until the OD₆₀₀ reached 0.6. The cells were harvested by centrifugation and washed twice with 400 ml M9 minimal salt medium consisting of 47 mM NH₄Cl, 55 mM KH₂PO₄, 120 mM Na₂HPO₄, 43 mM NaCl. The washed cells were resuspended in 400 ml M9 minimal medium containing trace metals, 0.5% glucose, 5.1 μM biotin, 4.2 μM thiamine, 2.5 mM MgCl₂, 0.75 mM CaCl₂, 34 $\mu\text{g ml}^{-1}$ kanamycin and 255 μM selenomethionine. The cultures were recovered at 310 K for 1 h before induction with 0.5 mM IPTG. The cells were allowed to grow for an additional 16 h. The cells were harvested by centrifugation and stored overnight at 193 K.

2.3. Purification of recombinant protein

The native protein and selenomethionine-derivatized protein were purified using the same methods. Cells were thawed and resuspended in 100 ml cold lysis buffer consisting of 50 mM sodium phosphate pH 8.0, 300 mM NaCl. Cells were lysed using a French Press on ice in the presence of 0.2 μM phenylmethanesulfonyl fluoride (PMSF). Cell debris was removed from the lysate by centrifugation. All subsequent purification steps were performed at 293 K. The lysate was added to a column packed with TALON metal-affinity resin (Roche) pre-equilibrated with lysis buffer. The protein was allowed to bind before being washed with one column volume of 50 mM sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole. The protein was eluted using five column volumes of 50 mM sodium phosphate pH 6.0, 300 mM NaCl, 250 mM imidazole. The eluted protein was concentrated to 500 μl by low-speed centrifugation using a Vivaspin 2 3000 Da molecular-weight cutoff concentrator (Sartorius Stedim Biotech). The protein was further purified by size-exclusion chromatography using an ÄKTAbasic fast protein liquid-chromatography (FPLC) system with a Sepharose 75 10/30 column (GE Healthcare Life Sciences). Chromatography was carried out in 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.0 buffer containing 100 mM NaCl. Purified protein was concentrated to 20 mg ml⁻¹ and stored at 277 K for use in crystallization trials. The concentration of the MSMEG_5817 protein was determined spectrophotometrically (Nanodrop 1000, Thermo Scientific) at 280 nm and was calculated using an extinction coefficient of 16 500 M⁻¹ cm⁻¹. The molecular weight, purity and identity of the protein were confirmed by SDS-PAGE (Fig. 1) and Western blotting with anti-hexahistidine antibody (R&D Systems).

2.4. Crystallization

The initial crystallization experiments involved screening 480 conditions from commercially available kits from Hampton Research (Crystal Screen, Crystal Screen 2, Index, PEG/Ion and SaltRx) and Emerald BioSystems (Wizard I and II) using a CrystalMation integrated robotic workstation (Rigaku) and the sitting-drop vapour-diffusion technique. Initial small spheroid crystals were obtained in 1.50 M Li₂SO₄, 0.1 M Tris-HCl pH 8.5 (SaltRx screen, Hampton

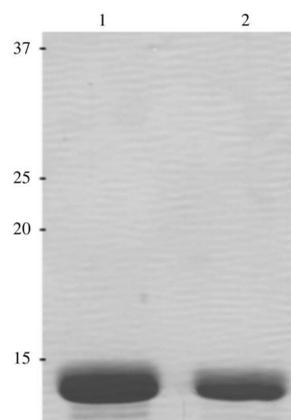


Figure 1
18% SDS-PAGE of purified native MSMEG_5817 (lane 1) and purified selenomethionine-derivatized MSMEG_5817 (lane 2). Molecular weights are indicated in kDa on the left.

crystallization communications

Table 1

Crystal and data-collection statistics.

Values in parentheses are for the outer shell.

	Selenomethionine-derivatized MSMEG_5817			
	Native MSMEG_5817	Inflection (λ_1)	Peak (λ_2)	High-energy remote (λ_3)
Diffraction source	MX2, Australian Synchrotron	MX2, Australian Synchrotron	MX2, Australian Synchrotron	MX2, Australian Synchrotron
Detector	ADSC Quantum 315r CCD			
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Temperature (K)	100	100	100	100
Unit-cell parameters (\AA , $^\circ$)	$a = 50.5, b = 50.5, c = 205.5,$ $\alpha = \beta = \gamma = 90.0$	$a = 49.9, b = 49.9, c = 205.0,$ $\alpha = \beta = \gamma = 90.0$	$a = 49.9, b = 49.9, c = 205.0,$ $\alpha = \beta = \gamma = 90.0$	$a = 49.9, b = 49.9, c = 205.0,$ $\alpha = \beta = \gamma = 90.0$
Resolution range (\AA)	50.05–2.40 (2.67–2.40)	40.3–2.40 (2.53–2.40)	41.06–2.47 (2.60–2.47)	41.4–2.50 (2.64–2.50)
Wavelength (\AA)	0.95369	0.97927	0.97949	0.95369
Total No. of reflections	20386	42924	38677	37473
No. of unique reflections	1516	1549	1429	1377
R_{merge}^\dagger (%)	9.9 (56.4)	9.4 (56.1)	10.5 (68.9)	10.3 (51.1)
Mean $I/\sigma(I)^\ddagger$	15.9 (4.6)	28.7 (7.5)	24.0 (6.0)	23.7 (7.7)
Completeness (%)	98.7 (98.2)	100 (100)	100 (100)	100 (100)
Anomalous redundancy		15.1 (15.1)	14.7 (14.8)	14.9 (14.9)

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the scaled intensity of the i th measurement of reflection hkl and $\langle I(hkl) \rangle$ is the mean intensity for that reflection. $^\ddagger I$ is the integrated intensity and $\sigma(I)$ is the estimated standard deviation of that intensity.

Research). Optimization of the native MSMEG_5817 crystallization conditions was performed by fine-tuning the protein and precipitant concentrations in 24-well Linbro plates (Hampton Research) using a hanging drop consisting of 1 μ l protein solution and 1 μ l precipitant solution and a 500 μ l reservoir volume. Additional optimization of the crystallization conditions was performed using Additive Screen (Hampton Research), in which drops consisting of 0.9 μ l protein solution, 0.9 μ l precipitant solution and 0.2 μ l additive solution were equilibrated against a 500 μ l reservoir volume consisting of 1.42 M Li_2SO_4 , 0.1 M Tris–HCl pH 7.7. The best crystals appeared after 3 d of equilibration against the crystallization condition and grew to full size in 5 d using 0.1 M sodium citrate tribasic dihydrate as an additive (Fig. 2a). Crystals of selenomethionine-derivatized MSMEG_5817

could only be obtained using the cross-seeding technique (Fig. 2b). Cross-seeding was performed using the streak-seeding method. Hanging-drop experiments were performed with drops consisting of 0.9 μ l protein solution, 0.9 μ l precipitant solution and 0.2 μ l 0.1 M sodium citrate tribasic dihydrate equilibrated against a 500 μ l reservoir volume. The drops were allowed to equilibrate for 5 d. A cat whisker was used to transfer seeds of the native MSMEG_5817 crystal into clear equilibrated drops. The best selenomethionine-derivatized MSMEG_5817 crystals grew after 24 h using 1.30 M Li_2SO_4 , 0.1 M Tris–HCl pH 7.7, 0.1 M sodium citrate tribasic dihydrate.

2.5. X-ray data collection

For X-ray diffraction data collection, crystals of the native and the selenomethionine-derivative protein were transferred to a CryoLoop and soaked in a cryoprotectant consisting of 1.5 M LiSO_4 , 0.1 M Tris–HCl pH 7.7, 15% glycerol before cooling to 100 K in a stream of nitrogen gas. A complete data set was collected from a single crystal on the MX2 beamline at the Australian Synchrotron using an ADSC Quantum 315r CCD detector. For the native crystal, a total of 360 frames of 0.5° oscillation were recorded with an exposure time of 1 s per frame. Using a single selenomethionine-derivative crystal, a fluorescence scan selecting the L_{III} edge of selenium was conducted to determine the inflection, peak and high-energy remote levels. Three complete data sets were collected from a single crystal at the required wavelengths; a total of 360 images at 0.5° oscillation were recorded with an exposure time of 1 s per frame (Fig. 3). The data were processed using *MOSFLM* (Leslie & Powell, 2007) and various programs from the *CCP4* suite (Winn, 2011). The final statistics of data collection and processing are summarized in Table 1.

3. Results and discussion

The *M. smegmatis* gene *MSMEG_5817* encoding a 128-residue protein was successfully cloned and the protein was expressed, purified and crystallized using vapour-diffusion methods. The MSMEG_5817 protein was purified to homogeneity as determined by SDS–PAGE (Fig. 1) and routinely produced yields of 5 mg l^{-1} . Crystals suitable for X-ray crystallographic studies were obtained using 1.42 M Li_2SO_4 , 0.1 M Tris–HCl pH 7.7 with 0.1 M sodium citrate tribasic dihydrate as an additive (Fig. 2a). A complete

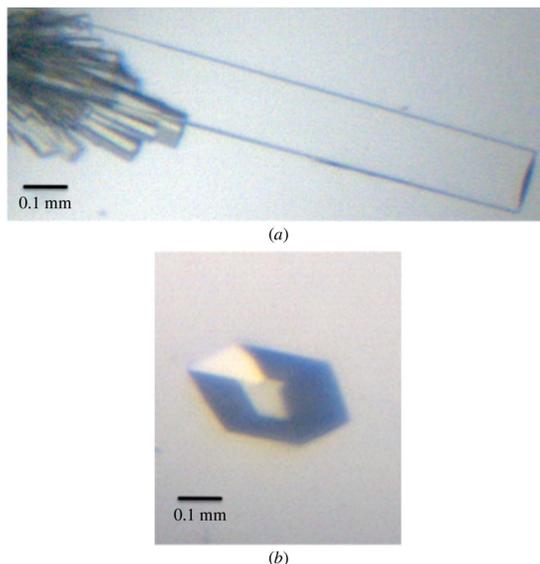


Figure 2

Crystals of (a) native MSMEG_5817 and (b) selenomethionine-derivatized MSMEG_5817 obtained from 20 mg ml^{-1} protein solution (in 20 mM MES pH 6.0, 100 mM NaCl) using 1.42 M lithium sulfate, 0.1 M Tris pH 7.7, 0.1 M sodium citrate tribasic dihydrate.

crystallization communications

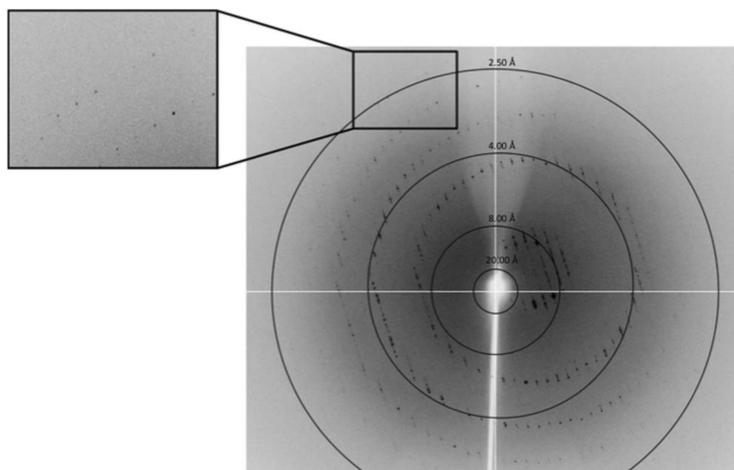


Figure 3
A typical 0.50° oscillation image from a selenomethionine-derivatized MSMEG_5817 crystal flash-cooled in 15% glycerol.

diffraction data set was collected to 2.40 Å resolution at 100 K. Preliminary crystallographic analysis indicated that the crystals belonged to space group $P4_32_12$, with unit-cell parameters $a = 50.5$, $b = 50.5$, $c = 205.5$ Å, $\alpha = \beta = \gamma = 90.0^\circ$. The data-collection and processing statistics are summarized in Table 1. Based on Matthews coefficient calculation, two molecules (42.68% solvent content) could be accommodated in the asymmetric unit, with an acceptable V_M of around 2.14 Å³ Da⁻¹ (Matthews, 1968). Molecular replacement was unsuccessful for MSMEG_5817 owing to the existing model (PDB entry 1iw7; *Thermus thermophilus* RNA polymerase holoenzyme; Vassylyev *et al.*, 2002) having relatively low sequence similarity. A selenomethionine derivative was expressed and the protein was purified using the same methods as used for the native protein. High-quality crystals were generated in the same crystallization condition by cross-seeding using crystals of the native protein (Fig. 2*b*). Three data sets were collected from a single crystal at 100 K based on experimentally determined inflection ($\lambda_1 = 0.97927$ Å), peak ($\lambda_2 = 0.97949$ Å) and high-energy remote ($\lambda_3 = 0.95369$ Å) wavelengths. A strong anomalous signal from selenium was detected to 2.5 Å resolution in each of the data sets (Fig. 3). Structural determination by multiple-wavelength anomalous dispersion is currently under way.

We thank the staff of the Australian Synchrotron and Monash Macromolecular Crystallization Facility for assistance with crystallization and X-ray data collection. This work was supported by the Australian Research Council (ARC) Centre of Excellence in Struc-

tural and Functional Microbial Genomics and the National Health and Medical Research Council of Australia. JR is an NHMRC Australia Fellow and TB is a Pfizer Australian Research Fellow.

References

- Anes, E., Peyron, P., Staali, L., Jordao, L., Gutierrez, M. G., Kress, H., Hagedorn, M., Maridonneau-Parini, I., Skinner, M. A., Wildeman, A. G., Kalamidas, S. A., Kuehnel, M. & Griffiths, G. (2006). *Cell. Microbiol.* **8**, 939–960.
- Armstrong, J. A. & Hart, P. D. (1971). *J. Exp. Med.* **134**, 713–740.
- Billman-Jacobe, H., McConville, M. J., Haites, R. E., Kovacevic, S. & Coppel, R. L. (1999). *Mol. Microbiol.* **33**, 1244–1253.
- Chandra, N., Kumar, D. & Rao, K. (2011). *Tuberculosis*, **91**, 487–496.
- Fratti, R. A., Chua, J., Vergne, I. & Deretic, V. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 5437–5442.
- Kovacevic, S., Anderson, D., Morita, Y. S., Patterson, J., Haites, R., McMillan, B. N., Coppel, R., McConville, M. J. & Billman-Jacobe, H. (2006). *J. Biol. Chem.* **281**, 9011–9017.
- Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, edited by R. J. Read & J. L. Sussman, pp. 41–51. Dordrecht: Springer.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Patterson, J. H., McConville, M. J., Haites, R. E., Coppel, R. L. & Billman-Jacobe, H. (2000). *J. Biol. Chem.* **275**, 24900–24906.
- Pelosi, A., Smith, D., Brammananth, R., Topolska, A., Billman-Jacobe, H., Nagley, P., Crellin, P. K. & Coppel, R. L. (2012). *PLoS One*, **7**, e31788.
- Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. & Yokoyama, S. (2002). *Nature (London)*, **417**, 712–719.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.

Chapter 5: Structural and Functional characterization of MSMEG_5817

5.1 Introduction

Chapter 5 continues from the crystallization and data collection of MSMEG_5817 crystals, described in chapter 4. Chapter 5 contains a paper published in *Acta Crystallographica Section D – Biological Crystallography*, and describes the crystal structure of MSMEG_5817 solved to 2.40 Å. Previous attempts to characterise the exact role of MSMEG_5817 within host survival have been unsuccessful. In obtaining the crystal structure of MSMEG_5817, we have gained valuable insight into its biological function.

The crystal structure of MSMEG_5817 has provided unique insights into its structural architecture. MSMEG_5817 exhibits a two-layer α/β sandwich fold, with a high content of hydrophobic residues forming the architecture of a ligand binding cavity. MSMEG_5817 was found to exhibit structural homology to the sterol carrier protein (SCP) family, whose primary role is in the binding and mediation of biologically relevant sterols and lipids. Based on these findings, we hypothesised that the MSMEG_5817 family is a new class of SCP-like proteins.

This hypothesis was supported by a series of biochemical assays, in which binding to a number of phospholipids was observed, exhibiting a differing substrate specificity compared to the human SCP (hSCP). MSMEG_5817 was found to bind specifically to phospholipids with highly electronegative headgroups, whereas hSCP bound phospholipids non-specifically. In addition, no lipid binding was observed by MSMEG_5817 against sterols and fatty acids tested compared to hSCP, which exhibits high affinity for these lipids tested.

The paper also describes mutant analysis as a means of probing the potential binding pocket of MSMEG_5817. Generated by alanine-scanning mutagenesis, mutants were selected based on sequence identity and structural homology to hSCP. The generated mutants were analysed for their essentiality not only in ligand binding, but for maintaining structural viability of the overall protein fold.

The crystal structure of MSMEG_5817 was successfully solved to 2.40 Å, revealing structural characteristics similar to the SCP class of proteins; indicated MSMEG_5817 may be functionally related to the SCP's. This was confirmed biochemically, with phospholipid binding detected by MSMEG_5817. Mutant analysis revealed structurally important residues, with a potentially different ligand binding mechanism to the SCP's. These results have provided valuable insight into the function of this class of proteins, paving the way for future characterisation of the MSMEG_5817 protein family with the *Corynebacterae* suborder of bacteria.

Declaration for Thesis Chapter 5

Declaration by candidate

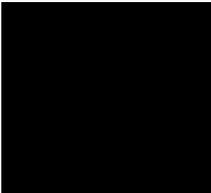
In the case of Chapter 5, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Structural determination of MSMEG_5817. Cloning, expression, purification of hSCP. PIP-Strip and ELISA analysis of MSMEG_5817 compared to hSCP. Alanine-scanning mutagenesis generation of nine mutants, expression and purification of six mutants. Circular dichroism and thermal melting of MSMEG_5817 and 6 mutants. Data analysis, figure and manuscript preparation.	80%

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Dene Littler	Assistance in structural determination	N/A
Rajini Brammananth	Supervision of Phooi Chan, data interpretation and manuscript preparation	N/A
Phooi Chan	Initial cloning of MSMEG_5817	5%
Paul Crellin	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Ross Coppel	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Jamie Rossjohn	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Travis Beddoe	Conception, supervision, direction, data interpretation and manuscript preparation	N/A

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's Signature		Date 21/10/15
Main Supervisor's Signature		Date 21/10/15

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

A structural and functional investigation of a novel protein from *Mycobacterium smegmatis* implicated in mycobacterial macrophage survivability

Adam Shahine,^{a,b} Dene Littler,^{a,b}
Rajini Brammananath,^{b,c}
Phooi Y. Chan,^{b,c} Paul K.
Crellin,^{b,c} Ross L. Coppel,^{b,c}
Jamie Rossjohn^{a,b,d,*} and
Travis Beddoe^{a,b,e,*}

^aDepartment of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia, ^bAustralian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia, ^cDepartment of Microbiology, Monash University, Clayton, Victoria 3800, Australia, ^dInstitute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, Wales, and ^eDepartment of Agriculture Sciences, La Trobe University, Bundoora, Victoria, Australia

Correspondence e-mail:
jamie.rossjohn@monash.edu,
t.beddoe@latrobe.edu.au

The success of pathogenic mycobacterial species is owing in part to their ability to parasitize the generally inhospitable phagosomal environment of host macrophages, utilizing a variety of strategies to avoid their antimycobacterial capabilities and thereby enabling their survival. A recently identified gene target in *Mycobacterium smegmatis*, highly conserved within *Mycobacterium* spp. and denoted MSMEG_5817, has been found to be important for bacterial survival within host macrophages. To gain insight into its function, the crystal structure of MSMEG_5817 has been solved to 2.40 Å resolution. The structure reveals a high level of structural homology to the sterol carrier protein (SCP) family, suggesting a potential role of MSMEG_5817 in the binding and transportation of biologically relevant lipids required for bacterial survival. The lipid-binding capacity of MSMEG_5817 was confirmed by ELISA, revealing binding to a number of phospholipids with varying binding specificities compared with *Homo sapiens* SCP. A potential lipid-binding site was probed by alanine-scanning mutagenesis, revealing structurally relevant residues and a binding mechanism potentially differing from that of the SCPs.

Received 11 December 2013
Accepted 13 May 2014

PDB reference:
MSMEG_5817, 4nss

1. Introduction

The genus *Mycobacterium* includes a number of deadly human pathogens, most notably *M. tuberculosis*, the causative agent of tuberculosis (TB). Despite extensive usage of the BCG vaccine and a series of drugs against TB, approximately one-third of the world's population is still infected, resulting in approximately 1.4 million deaths in 2011 (World Health Organization, 2012). The emergence of multiple drug-resistant (MDR) and extensively drug-resistant (XDR) strains of *M. tuberculosis* emphasizes the urgent need for the development of new antimycobacterials with novel modes of activity.

Pathogenic mycobacterial species have developed a multitude of mechanisms to parasitize the inhospitable environment of the macrophage, which is central to mycobacterial infection, latency and disease activation (Russell, 2001, 2011; Deretic *et al.*, 2006). Macrophages act as the main reservoir of infection for invading mycobacterial species, which contrasts with the accepted stance that these cells successfully act as our initial line of defence against bacterial infection. Invading mycobacteria are internalized by host macrophages through phagocytosis and are sorted into phagosomes, in which they are retained. At this stage, invading mycobacteria are able to arrest phagolysosome maturation, prolonging their survival (Armstrong & Hart, 1971; Vergne *et al.*, 2004). Invading mycobacteria are retained within the phagosome until the infected macrophage dies through necrosis or apoptosis.

Invading mycobacteria manipulate the host-signalling pathways required for correct vesicular trafficking and

research papers

therefore the infected phagosomes fail to mature and do not fuse with late endosomes for subsequent destruction by the lysosomes (Armstrong & Hart, 1971; Sturgill-Koszycki *et al.*, 1994; Deretic *et al.*, 2006). Additionally, invading mycobacteria have the ability to modulate the internal pH of the phagosomal lumen, retaining its neutral vacuolar pH of 6.2–6.4 and thus creating a relatively hospitable environment for survival (Via *et al.*, 1997). Elucidation of the mechanisms utilized by invading mycobacteria to corrupt host macrophage activities is vital for a greater insight into mycobacterial virulence and disease.

Invading mycobacteria utilize a wide variety of proteins and effector molecules to parasitize the macrophage phagosome, resulting in the absence of mature lysosomal hydrolases and incomplete luminal acidification (Via *et al.*, 1997). Invading mycobacteria have the ability to block phagolysosome biogenesis by modulating the function of macrophage small GTP-binding proteins (Rabs; Clemens *et al.*, 2000a). The Rab protein family, whose role is to control the identity of intracellular organelles and direct membrane trafficking *via* a molecular on–off switch mechanism (Clemens *et al.*, 2000a,b), plays a crucial role in macrophage phagolysosome formation. Through the wholesale conversion of Rab5 (early endocytosis) to Rab7 (late endosome), the phagosome matures to allow the correct generation of the phagolysosome (Fratti *et al.*, 2001; Pereira-Leal & Seabra, 2001; Kelley & Schorey, 2003; Pfeffer, 2005; Rink *et al.*, 2005). Normal Rab5 to Rab7 conversion is modulated by the Rab5 modulator EEA1 (early endosomal autoantigen 1) that associates with phosphatidylinositol 3-phosphate (PI3P) on organelle membranes, an interaction that is essential for proper membrane trafficking within the endosomal system (Sturgill-Koszycki *et al.*, 1994; Fratti *et al.*, 2003). Invading mycobacteria have the ability to block the conversion of Rab5 to Rab7 in a process known as the Rab conversion block (Clemens *et al.*, 2000a), utilizing two key mechanisms unique to mycobacteria. The lipoglycan liparabinomannan (LAM; Sturgill-Koszycki *et al.*, 1994; Fratti *et al.*, 2003; Hmama *et al.*, 2004; Kang *et al.*, 2005) and the enzyme SapM (a PI3P phosphatase; Saleh & Belisle, 2000; Vergne *et al.*, 2005) act to prevent the generation of PI3P within the macrophage and to break down excess PI3P, respectively, ensuring blocking of phagosome maturation (Deretic *et al.*, 2006). While we have a basic knowledge of mycobacterial phagosome maturation blocking, it is likely that additional mechanisms remain to be found. The identification and characterization of the mechanisms of macrophage survivability by pathogenic mycobacterial species will aid in further understanding mycobacterial pathogenesis, paving the way for the development of novel treatments.

The saprophytic *M. smegmatis* has been exploited as a model organism to investigate *M. tuberculosis* pathogenesis. As such, *M. smegmatis* has been used to investigate host–pathogen interactions in macrophages, as *M. smegmatis* has been shown to be able to mimic the growth and survival of *M. tuberculosis* within macrophages (Anes *et al.*, 2006). To investigate key factors in mycobacterial survival in macrophages, a random transposon mutant library of *M. smegmatis*

was screened for reduced growth in macrophages (Billman-Jacobe *et al.*, 1999; Patterson *et al.*, 2000; Kovacevic *et al.*, 2006; Pelosi *et al.*, 2012). One such mutant, denoted Myco132, resulted in accelerated cell death of *M. smegmatis* in J774A.1 macrophages within the first 8 h (Pelosi *et al.*, 2012). The defective gene in Myco132 was identified as *MSMEG_5817*, which encodes a conserved hypothetical protein of 128 amino-acid residues with orthologues in a large variety of pathogenic and nonpathogenic mycobacteria (Pelosi *et al.*, 2012). However, *MSMEG_5817* is most likely not required for normal growth, as Rv0807, the *M. tuberculosis* homologue of *MSMEG_5817*, has been shown to be non-essential (Sasseti *et al.*, 2003; Griffin *et al.*, 2011). Additionally, there was an up-regulation of NF- κ B activation in macrophages infected with the *MSMEG_5817* mutant compared with those infected with wild-type *M. smegmatis* (Pelosi *et al.*, 2012). NF- κ B activation plays key roles in driving host immune responses through the production of cytokines, suggesting that *MSMEG_5817* may be involved in suppression of the NF- κ B activation pathway to promote survivability. In addition, other studies have shown that Rv0807 is upregulated upon phosphate starvation owing to the loss of the two-component *senX3–regX3* regulatory system, suggesting that Rv0807 may play a role in phosphate metabolism (Parish *et al.*, 2003).

Upon conducting a *BLASTp* similarity search utilizing the nonredundant protein-sequence database, it was found that *MSMEG_5817* shared high sequence identity with a vast number of orthologues within the Corynebacterineae sub-order, each of unknown function. Significant homology was also detected to a serine:pyruvate aminotransferase (SPT) from *Brevibacterium linens* BL2 (accession No. ZP_00381195). Upon comparing the SPT activities of wild-type *M. smegmatis* and the Myco132 mutant, it was found that the relative rate of pyruvate reduction was highly similar between the samples tested, indicating that *MSMEG_5817* does not encode a functional SPT (Pelosi *et al.*, 2012).

We have previously cloned and expressed *MSMEG_5817* as a recombinant protein from *Escherichia coli*, including the production of a selenomethionine derivative, and have performed purification, crystallization and preliminary X-ray diffraction studies to 2.40 Å resolution (Shahine *et al.*, 2013).

As a means of ascertaining its function, we set out to solve the crystal structure of *MSMEG_5817*. This structure would provide valuable insight into the structural characteristics of the protein, as well as to gauge a potential functional role through structural homology searches. Here, we describe the structural features of *MSMEG_5817* and present preliminary evidence for its functional role.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant *MSMEG_5817* protein

2.1.1. *MSMEG_5817* and alanine-scanning mutagenesis mutants for *E. coli* expression. The gene encoding *MSMEG_5817* was cloned and overexpressed in *E. coli*, and

research papers

both native and selenomethionine-derivative recombinant proteins were purified as described previously (Shahine *et al.*, 2013). *MSMEG_5817* mutants were generated from the parental pET-28c-*MSMEG_5817* plasmid DNA through site-directed mutagenesis. The mutants were PCR-amplified using *Pfu* DNA polymerase (Promega) in 10× *Pfu* buffer with specific sense and antisense primers. The PCR-amplified product was digested with *DpnI* (BioLabs) for 2 h at 37°C before being transformed into chemically competent *E. coli* XL1 Blue cells. Successfully cloned transformants were confirmed by DNA sequencing. The expression and purification of recombinant *MSMEG_5817* and its variants were conducted as described previously (Shahine *et al.*, 2013).

2.1.2. Homo sapiens sterol carrier protein (SCP). The expression vector pQTEV-SPC2 cloned into *E. coli* DH5α cells was purchased from Addgene (ID 31336). The sterol carrier protein from *H. sapiens* was expressed with an N-terminal hexahistidine tag with a TEV protease cleavage site. The pQTEV-SCP2 vector was transformed into *E. coli* BL21 (DE3) cells for recombinant protein expression, with successful transformants selected on LB-agar in the presence of 100 µg ml⁻¹ ampicillin. Recombinant protein was expressed and purified in the same manner as described previously (Shahine *et al.*, 2013).

2.2. Crystallization, data collection and data processing of MSMEG_5817

Crystals of both native and selenomethionine-derivative *MSMEG_5817* were obtained as described previously, with data collection and data processing conducted as described previously (Shahine *et al.*, 2013).

2.3. Structural determination of MSMEG_5817

The structure of *MSMEG_5817* was determined as follows. Using anomalous differences, two selenium positions were identified and refined using *AutoSol* as part of the *PHENIX* program package using experimentally determined f' and f'' values (Terwilliger *et al.*, 2009) that were obtained previously (Shahine *et al.*, 2013) with default settings. The top solution resulted in a FOM value of 0.59 and an estimated CC × 100 value of 66.7 ± 12.9. Upon density modification the final R factor was 28.32%. The resultant density-modified Fourier map, based on the three selenomethionine-derivative *MSMEG_5817* data sets, was used for automated model building with the *AutoBuild* program of the *PHENIX* program suite (Terwilliger *et al.*, 2008) with default settings. 210 of 256 residues were successfully built, with two molecules in the unit cell with an R_{work} of 24.68% and an R_{free} of 29.11%. The model built by *AutoBuild* was used as a molecular-replacement model for a 2.40 Å resolution native data set using *Phaser-MR* in the *PHENIX* program suite (Adams *et al.*, 2010), with the top solution solved with a TFZ value of 10.0. For cross-validation, a random set of 5% of the total reflections were kept aside from the refinement and used for the calculation of R_{free} (Brünger, 1992). The model was manually adjusted using the interactive graphics program *Coot* (Emsley

Table 1

Structural refinement and validation of *MSMEG_5817*.

Resolution range (Å)	40.42–2.40
Completeness (%)	98.00
$R_{\text{work}}^{\dagger}$ (%)	18.92
$R_{\text{free}}^{\dagger}$ (%)	22.52
R.m.s.d.s	
Bond lengths (Å)	0.010
Bond angles (°)	1.04
Ramachandran plot	
Favoured (%)	98.20
Allowed (%)	0.90
Outliers (%)	0.90
No. of modelled residues	230
No. of water molecules	71
Average B factors (Å ²)	
Protein atoms	61.05
Water molecules	60.93
PDB code	4nss

$\dagger R_{\text{work}} = \sum_{\text{hkl}} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{\text{hkl}} |F_{\text{obs}}|$ for all data excluding the 5% that comprise the R_{free} set used for cross-validation.

et al., 2010) and refined using *BUSTER* (Bricogne *et al.*, 2011) until no further improvement could be made, resulting in a final R_{work} of 18.92% and R_{free} of 22.52% (Table 1). The final round of *BUSTER* refinement was conducted with the -TLSBasic and -autoncs refinement settings. Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession code 4nss.

2.4. Structural analysis of MSMEG_5817

Refinement validation was conducted using *MolProbity* and the *POLYGON* tool in the *PHENIX* program suite (Adams *et al.*, 2010). All structural superpositions were achieved using the *SSM Superpose* feature of *Coot* (Emsley *et al.*, 2010). Secondary structure was confirmed using the *STRIDE* plugin in *PyMOL* (Zhu, 2011). Surface-area calculations were performed using the *APBS* plugin for *PyMOL* (Baker *et al.*, 2001). *H. sapiens* sterol carrier protein-2 (PDB entry 1ikt; Haapalainen *et al.*, 2001), *Aedes aegypti* sterol carrier protein-2 (PDB entry 1pz4; Dyer *et al.*, 2003) and *Oryctolagus cuniculus* sterol carrier protein-2 domain (PDB entry 1c44; Choinowski *et al.*, 2000) were used for all structural alignments.

2.5. Circular-dichroism spectra

Circular-dichroism spectra of *MSMEG_5817* and variants were recorded on a Jasco J-815 circular-dichroism spectropolarimeter at 20°C with a data pitch of 0.1 nm and a scan speed of 100 nm min⁻¹. Briefly, all *MSMEG_5817* samples were prepared at a concentration of 30 µM, with the exception of the M109A and L121A mutants, which were prepared at a concentration of 10 µM, in 50 mM sodium phosphate pH 6.0, 100 mM NaCl buffer. Far-ultraviolet circular-dichroism spectra (260–200 nm) were recorded with five data accumulations in a quartz cell with a path length of 0.1 cm. The resultant spectra were visualized using *Graphpad Prism*.

2.6. Thermal stability analysis via RT-PCR Thermofluor

Thermal stability assays of *MSMEG_5817* and variants were performed using a real-time detection instrument

research papers

(Corbett Rotor-Gene 300). Briefly, all MSMEG_5817 samples were prepared at a concentration of 30 μM in 50 mM sodium phosphate pH 6.0, 300 mM NaCl buffer. The fluorescent dye SYPRO Orange (LifeTechnologies) was added to each sample at a 10 \times concentration to enable monitoring of the protein-unfolding process. Samples were heated from 30 to 90°C at a rate of 1°C min⁻¹ and the changes in fluorescent intensity were recorded accordingly with an excitation wavelength of 530 nm and an emission wavelength of 660 nm.

2.7. Phospholipid binding of MSMEG_5817 by chemiluminescent enzyme-linked immunoassay (ELISA)

The phospholipid-binding capabilities of MSMEG_5817 and variants were determined by immunoassay. Selected phospholipids (Avanti), fatty acid and cholesterol (Sigma) stocks were created by dissolving 1 mg in chloroform and

freeze-drying under a stream of nitrogen gas before resuspension in 1:1 chloroform:methanol at a final concentration of 100 μM . Briefly, for direct plating assays lipids were diluted 1:10 in 10 mM phosphate-buffered saline (PBS) to yield a final concentration of 10 μM in 100 μl and were plated on Nunc-Immuno MicroWell Maxisorp 96-well solid plates overnight at 293 K. Excess phospholipids were washed in PBS + 1% (v/v) Tween 20 before blocking in PBS + 1% (v/v) Tween 20 + 3% (w/v) bovine serum albumin (BSA) for 2 h and washing again. MSMEG_5817 and variants were titrated between 10 and 100 μM and allowed to bind for 1 h before washing. The presence of bound MSMEG_5817 protein was detected by HRP anti-hexahistidine antibody (R&D Systems) at a 1:2000 dilution. SureBlue TMB Microwell Peroxidase Substrate (KPL) was used to detect the presence of antibodies, with the reaction stopped by 0.1 M HCl. Absorbance was detected by FLUOstar Omega (BMG Labtech) at a wavelength of 450 nm.

All experiments were conducted in triplicate, with data visualized and statistical *t*-tests conducted using *GraphPad Prism*.

3. Results and discussion

3.1. Structural determination of MSMEG_5817

To gain insight into the function of MSMEG_5817 and its orthologues, the crystal structure of the MSMEG_5817 protein was determined. As no significant structural homologue was available, the crystal structure of MSMEG_5817 was determined *via* multiple-wavelength anomalous dispersion (MAD) using data collected at three experimentally determined wavelengths. Initial experimental phases were obtained with *AutoSol* as part of the *PHENIX* program package (Terwilliger *et al.*, 2009), using experimentally determined f' and f'' values. The highly symmetrical primitive tetragonal $P4_32_12$ space group, with unit-cell parameters $a = 50.5$, $b = 50.5$, $c = 205.57$ Å, could accommodate two polypeptide molecules in the asymmetric unit (molecules *A* and *B*), corresponding to a Matthews coefficient of 2.2 Å³ Da⁻¹ at 15 kDa per molecule (Matthews, 1968).

The structure of MSMEG_5817 was determined to a resolution of

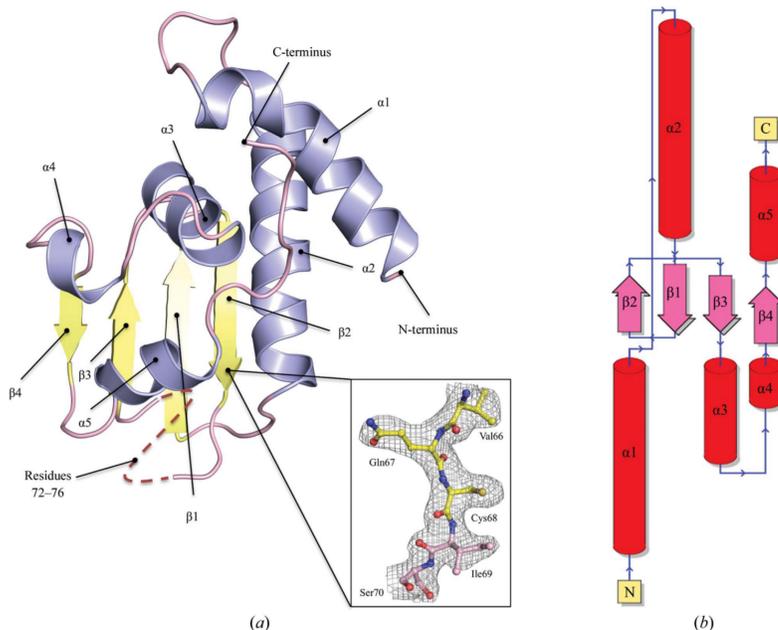


Figure 1

Overall crystal structure of MSMEG_5817. (a) Cartoon representation of the 2.4 Å resolution MSMEG_5817 crystal structure. The structure is colour-coded by secondary-structure element, with α -helices, β -strands and loops coloured blue, yellow and pink, respectively. The structure comprises a single antiparallel β -sheet consisting of four β -strands, $\beta 1$ (residues 55–60), $\beta 2$ (residues 63–68), $\beta 3$ (residues 81–84) and $\beta 4$ (residues 108–110), with a classical β -bulge occurring between the $\beta 2$ strand and the $\beta 3$ strand. Two β -hairpin structures are present, with the first spanning between the $\beta 3$ strand and the $\beta 4$ strand and the second occurring through the $\beta 1$ strand sandwiched between the $\beta 3$ strand and the $\beta 4$ strand. The latter contributes to a ψ -loop spanning residues 64–85. Residues 72–76 missing from the crystal structure contribute to the formation of the ψ -loop. The β -sheet is stacked by a layer of five α -helices: $\alpha 1$ (residues 11–24), $\alpha 2$ (residues 33–50), $\alpha 3$ (residues 87–95), $\alpha 4$ (residues 87–95) and $\alpha 5$ (residues 100–104). A highly hydrophobic cavity is formed between the β -sheet and the $\alpha 3$, $\alpha 4$ and $\alpha 5$ helices. The quality of the final density is visualized (inset) around residues 66–70, centred on Ile69, which are visualized in ball-and-stick representation and colour-coded based on their secondary structure. The σ_A -weighted $(2m|F_o| - D|F_c|)$ electron-density maps are contoured at 1.0σ . (b) Topology diagram of the MSMEG_5817 structure colour-coded by secondary-structure element, with α -helices, β -strands and loops coloured red, pink and blue, respectively.

research papers

2.40 Å, with final R_{work} and R_{free} statistics of 18.92 and 22.52%, respectively. The refinement statistics are summarized in Table 1. The construct utilized in crystallization included residues 1–128 of MSMEG_5817 with ten vector-derived residues and an N-terminal hexahistidine purification tag. Of this construct, residues 9–128 in molecule *A* and residues 8–128 in molecule *B* were modelled from the electron density. Electron density is missing for one segment, consisting of residues 72–76 of molecule *A* and 72–77 of molecule *B*, which could not be traced in the electron-density map and therefore was not included in the final model. Molecules *A* and *B* can be superimposed with a core root-mean-square deviation (r.m.s.d.) value of 0.40 Å. There was no evidence of higher order oligomerization within the crystal lattice, which coincides with the purification of MSMEG_5817 as a monomer by size-exclusion chromatography (Shahine *et al.*, 2013). The final model of MSMEG_5817 consisted of 230 amino-acid residues and 71 water molecules. A Ramachandran plot analysis of the final model using *MolProbity* (Chen *et al.*, 2010) showed that 98.20% of the residues are in the most favoured regions, while 0.90% of the residues are in the allowed region and 0.90% of the residues are in the outlier region. The outlier residues are Ile69 in both molecules *A* and *B*.

3.2. Crystal structure of MSMEG_5817

The 2.40 Å resolution MSMEG_5817 crystal structure (Fig. 1*a*) reveals a single β -sheet comprised of four antiparallel β -strands surrounded by five α -helices, adopting an $\alpha 1$ – $\alpha 2$ –

$\beta 1$ – $\beta 2$ – $\beta 3$ – $\alpha 3$ – $\alpha 4$ – $\beta 4$ – $\alpha 5$ arrangement (Fig. 1*b*). The overall tertiary structure of MSMEG_5817 exhibits a Rossmann-like fold α/β two-layer sandwich (Rao & Rossmann, 1973; García *et al.*, 2000; Sillitoe *et al.*, 2013). A tightly packed cavity exists at the interface between the α -helical layer and the β -sheet, which may provide an interface for ligand binding. The four antiparallel β -strands form the first layer of the protein structure, arranged in a $-1\ 2\ 1$ topology. A single antiparallel classic β -bulge occurs between Val60 on the $\beta 3$ strand and Phe63 and Val64 on the $\beta 2$ strand across a β -turn. This type IV hydrogen-bonded β -turn is comprised of residues Val60-Pro61-Pro62-Phe63-Val64, with an i to $i + 3$ C^α distance of 5.8 Å. A large β -hairpin structure is formed between $\beta 3$ and $\beta 4$ with a hairpin class of 19:21, which includes $\alpha 3$ and $\alpha 4$ that form part of the structural architecture over the β -sheet. A secondary β -hairpin structure is present in the form of a ψ -loop, in which the $\beta 1$ strand is sandwiched in between the $\beta 2$ and $\beta 3$ strands, forming hydrogen bonds between the two strands. The ψ -loop, which is generally considered to be rare in protein structures (Hutchinson & Thornton, 1996), results in a 12-residue loop including an extended conserved residue patch spanning residues 74–83. The architecture of the ψ -loop is stabilized by hydrogen-bond formation between the C^α atom of Ile69 with Ser55 on the $\beta 1$ strand and Ala51 at the C-terminal end of $\alpha 2$, with bond distances of 2.7 and 2.9 Å, respectively. Owing to the rarity of ψ -loops within protein structures, the presence of Ile69 as a Ramachandran outlier is explained. The Cys68 residue is situated at the C-terminal end of $\beta 2$ before the ψ -loop and faces into the core of the protein.

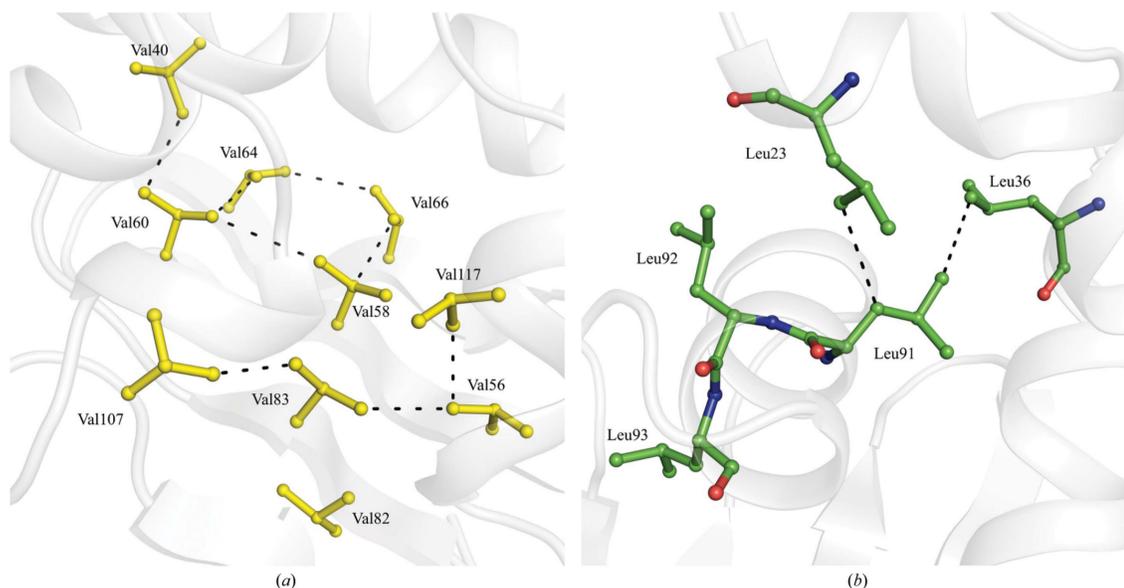


Figure 2

Hydrophobic residue content. (*a*) Stick representation of valine residue side chains comprising the hydrophobic core of MSMEG_5817 (yellow). Black dashed lines represent cross-valine hydrophobic interactions that stabilize the hydrophobic cavity of the protein, with bond distances ranging between 3.7 and 4.3 Å. (*b*) Coordination of the conserved Leu91 residue positioned on the $\alpha 3$ helix. Hydrophobic interactions between Leu91 and Leu23 on the $\alpha 1$ helix and Leu36 on the $\alpha 2$ helix are denoted by black dashed lines, with bond distances of 3.7 and 3.6 Å, respectively.

research papers

MSMEG_5817 is comprised of a high percentage of hydrophobic residues, predominantly alanine, valine and leucine, which make up 39.1% of its amino-acid sequence. The valine residues, which contribute 11.7% of the amino-acid sequence, are predominantly positioned on the $\beta 1$ – $\beta 3$ strands, with side chains facing the interior of the structure (Fig. 2*a*). The valine residues act to contribute to the overall hydrophobicity of the protein core. Interestingly enough, Val82, located on the $\beta 3$ strand, is the only valine residue with a solvent-exposed side chain within the structure (Fig. 2*a*). The 11 leucine residues, comprising 8.6% of the total amino-acid sequence, contribute to stabilizing the α -helical architecture

surrounding the β -sheet (Fig. 2*b*). These helix–helix interactions are primarily driven by a stretch of three leucine residues, Leu89, Leu90 and Leu91, positioned in the $\alpha 3$ helix. Leu91 acts to stabilize helix–helix interactions between the $\alpha 3$ helix and the $\alpha 1$ and $\alpha 2$ helices through hydrophobic interactions with neighbouring leucine residues (Fig. 2*b*). Additional hydrophobic interactions are made between Leu91 and Leu23 in the $\alpha 1$ helix and Leu36 in the $\alpha 2$ helix, with bond lengths of 3.7 and 3.6 Å, respectively (Fig. 2*b*).

Despite the highly hydrophobic nature of the MSMEG_5817 structure, the recombinant protein was successfully purified as a soluble protein. This is aided by the

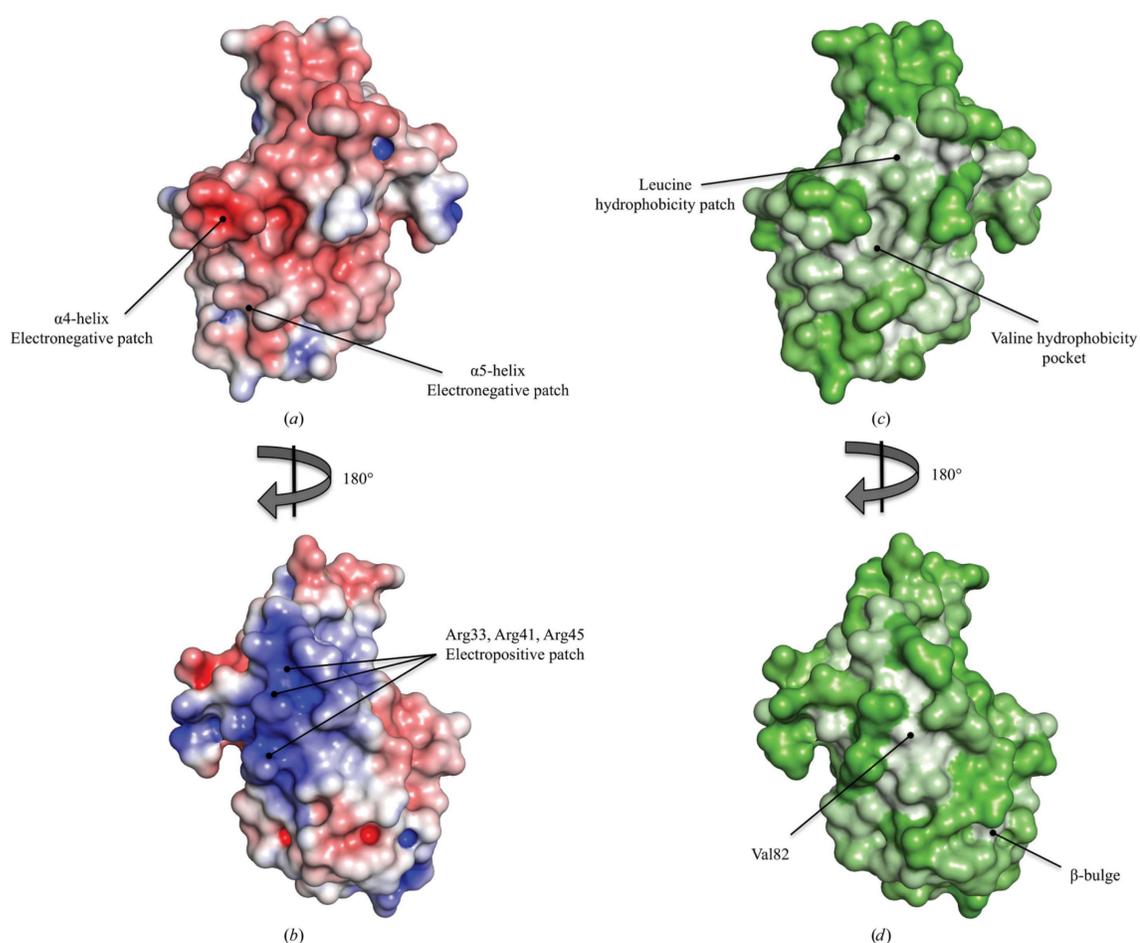


Figure 3 Surface electrostatic and hydrophobic properties of the MSMEG_5817 crystal structure. (a) Front and (b) rear views of the solvent-accessible surface representation coloured by electrostatic potential as calculated by *APBS* (Baker *et al.*, 2001). The potential contours are shown on a scale from $+5.0k_B T e^{-1}$ (blue) to $-5.0k_B T e^{-1}$ (red); white indicates no change. (a) The positions of two electronegative patches involving residues Asp99 and Asp102 on the $\alpha 4$ helix, as well as residues Glu116 and His119 on the $\alpha 5$ helix. (b) The position of an extended electropositive patch involving residues Arg33, Arg41 and Arg45. (c) Front and (d) rear views of a surface representation coloured by hydrophobicity according to the normalized consensus hydrophobicity scale (Eisenberg *et al.*, 1984). Colour representation with a colour gradient of most hydrophobic to least hydrophobic ranging from white to green, respectively. Indicated are the positions of (c) the leucine hydrophobic patch and valine hydrophobic pocket and (d) the solvent-exposed Val82 residue and the classical β -bulge.

research papers

high percentage of solvent-exposed arginine residues, which make up 9.4% of the amino-acid sequence and are positioned throughout the surface of the protein. The localization of three arginine residues, Arg33, Arg41 and Arg45, situated in the $\alpha 2$ helix contributes to a highly electrostatically positive region facing the rear of the structure (Fig. 3b).

In addition, there is an extended negative electrostatic patch spanning between the α -helical layer across the β -sheet, comprised of $\alpha 3$, $\alpha 4$ and $\alpha 5$. This primarily involves Asp99 and Asp102 in the $\alpha 4$ helix and Glu116 and His119 in the $\alpha 5$ helix (Fig. 3a). In each case, the residues are poorly conserved throughout the orthologues analysed, save for Glu116, which is substituted by an aspartic acid in *M. leprae*. In each case, the side chains of these negatively charged residues face away from the core of the structure and, if this is indeed a ligand-binding cavity, would not be involved in direct ligand contacts.

A differing trend was observed upon comparing electrostatic potential with hydrophobic content. The highest localization of hydrophobicity is situated within the cavity sandwiched between the α -helical layer and the β -sheet (Fig. 3c). This region of hydrophobicity primarily involves the

valine residues spanning the β -sheet, combined with the leucine residues linking the $\alpha 3$ helix to the $\alpha 1$ and $\alpha 2$ helices, as described previously. The rear of the protein structure reveals a small hydrophobic patch indicative of the anti-parallel β -bulge spanning between the $\beta 2$ and $\beta 3$ strands (Fig. 3d). This localization of hydrophobicity at the cavity interface, coupled with the lack of electrostatically charged residues facing inwards, may suggest that if this is indeed a binding cavity then MSMEG_5817 may be involved in facilitating the binding of nonpolar ligands.

3.3. Comparison of MSMEG_5817 to structural homologues

Owing to the lack of conserved domains, and the absence of characterized proteins with amino-acid sequence identity, the function of MSMEG_5817 is presently unknown. In order to gain functional insight based on the crystal structure of MSMEG_5817, a DALI search of the Protein Data Bank was conducted (Holm & Rosenström, 2010). The closest structural homologues identified from the search results were a number of sterol carrier proteins (SCPs), the function of which is to

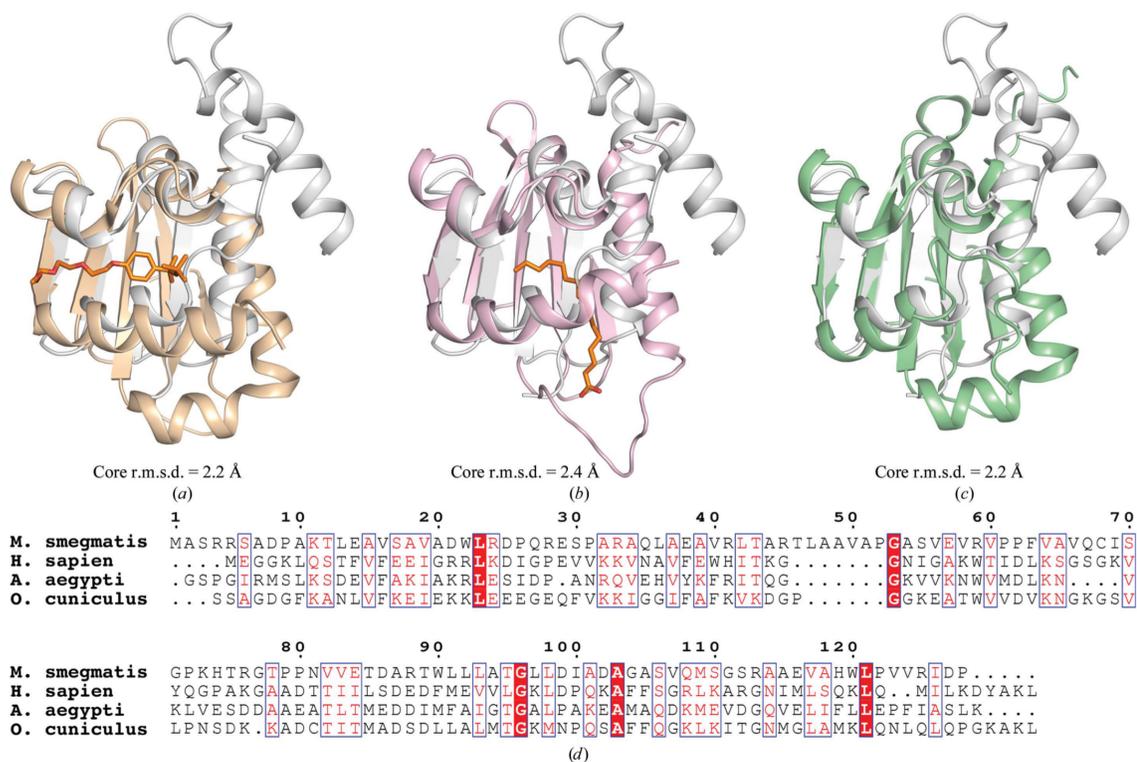


Figure 4

Cartoon structure superposition of MSMEG_5817 and sterol carrier proteins (SCPs). MSMEG_5817 (grey) is overlaid with (a) *H. sapiens* SCP (hSCP; wheat), (b) *A. aegypti* SCP (AaSCP; magenta) and (c) *O. cuniculus* SCP (OeSCP; green), with core r.m.s.d. values of 2.2, 2.4 and 2.2 Å, respectively. The presence of bound Triton X-100 (a) and palmitic acid (b) in the SCP structures is shown in orange stick representation. (d) Sequence alignment between MSMEG_5817 and identified SCPs. The alignment was prepared with ChustaW2 and visualized using ESPript v.2.2. Strictly sequence-identical residues are denoted with a red background, while similar residues are visualized in red text with a white background. Sequence similarities in groups are denoted by blue boxes. Gaps are represented by dots.

bind and transport sterols (Avdulov *et al.*, 1999) and lipids (Schroeder *et al.*, 1995) across cellular membranes (Kernstock & Girotti, 2007). Three SCPs were identified from *H. sapiens* (PDB entry 1ikt; Haapalainen *et al.*, 2001; Szyperki *et al.*, 1993), *A. aegypti* (PDB entry 1pz4; Dyer *et al.*, 2003) and *O. cuniculus* (PDB entry 1c44; Choinowski *et al.*, 2000), with core r.m.s.d. values of 2.2, 2.4 and 2.2 Å, respectively (Figs. 4a, 4b and 4c). The SCPs exhibit a Rossmann-like fold α/β two-layer sandwich, structurally similar to MSMEG_5817. The crystal structures of the *H. sapiens* SCP (hSCP) and *A. aegypti* SCP (AaSCP) were solved in the presence of a lipid bound within the binding groove: Triton X-100 (Haapalainen *et al.*, 2001) and palmitic acid (Dyer *et al.*, 2003), respectively. The sequence identities between MSMEG_5817 and the SCPs are very low, with AaSCP having the highest percentage identity at 15.7% (Fig. 4d). Despite this, the secondary-structure elements overlay tightly between MSMEG_5817 and the SCPs, with the overall protein architecture being predominantly retained.

While the similarities between the tertiary structures of MSMEG_5817 and the SCPs are apparent, there are some subtle differences in their secondary-structure compositions. The SCPs exhibit a larger binding cleft arranged in a $-1 -1 3X -1$ topology comprised of a five-stranded antiparallel β -sheet with a slight curvature. In comparison, the

MSMEG_5817 β -sheet is arranged in a $-1 2 1$ topology comprised of a flat four-stranded antiparallel β -sheet (Figs. 4a, 4b and 4c). The larger β -sheets present in the SCPs accommodate a larger lipid-binding interface. In addition, the MSMEG_5817 structure contains a N-terminal helix that is not conserved in other SCPs (Figs. 4a, 4b and 4c) that results in r.m.s.d.s that are larger than normal between conserved proteins. However, the N-terminal helix is conserved across other Corynebacterineae orthologues (Fig. 5), suggesting a conserved function within this family.

The previously determined crystal structures of hSCP and AaSCP both contain a bound lipid within their respective binding cavities. Upon comparison, the residues involved in contacts between the SCPs and the respective lipid differ between the two proteins. In the case of the AaSCP, the basal α -helix unravels to accommodate binding against the polar carboxylic acid head group. The residues binding the lipid tail are predominantly conserved with respect to hSCP, principally comprising large aliphatic residues situated on the α -helices surrounding the β -sheet. This supports the function of the SCPs as nonspecific lipid-binding proteins exhibiting a large binding cavity accommodating a broad range of lipids both in functionality and size.

In comparison, the positions of the lipid-binding residues are poorly conserved between MSMEG_5817 and the SCPs.

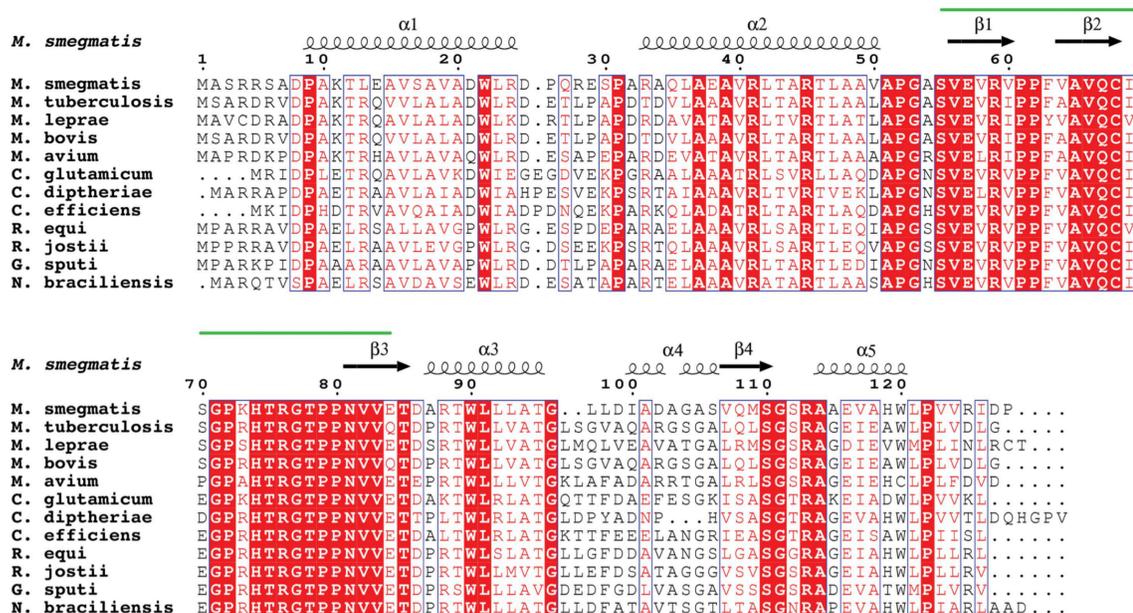


Figure 5

Amino-acid sequence alignment of members of the MSMEG_5817 protein family. This list includes the following: *Mycobacterium smegmatis* (YP_890045.1), *M. tuberculosis* (NP_215322.1), *M. leprae* (NP_302448.1), *M. bovis* (NP_854488.1), *M. avium* (WP_003875797.1), *Corynebacterium glutamicum* (NP_601783.1), *C. diphtheriae* (YP_005138847.1), *C. efficiens* (WP_006769165), *Rhodococcus equi* (YP_004005639.1), *R. jostii* (YP_704781.1), *Gordonia sputi* (WP_005204186.1) and *Nocardia brasiliensis* (YP_006805579.1). The alignment was prepared with *ClustalW2* and was visualized using *ESPrpt* v2.2. The secondary-structure elements correspond to the structure of MSMEG_5817. Strict sequence-identical residues are denoted with a red background, while similar residues are visualized in red text with a white background. Sequence similarities in groups are denoted by blue boxes. The green line indicates the position of the ψ -loop.

research papers

The only conserved residue present within the binding groove of MSMEG_5817, assuming a similar binding mechanism to the SCPs, is Leu121. The Leu121 side chain faces the binding groove situated on the α 5 helix and may be involved in direct contact with a bound ligand.

3.4. Sequence similarity of MSMEG_5817 to other *Corynebacterineae* orthologues

In order to gain insight into the functional role of MSMEG_5817, we compared the sequence of MSMEG_5817 with those of its orthologues from the closely related myco-

bacterial species *M. tuberculosis* and *M. leprae*, as well as those from other representatives of the suborder *Corynebacterineae* (Fig. 5). The sequences are highly conserved, with a minimum of 53% pairwise sequence conservation. A *BLASTp* search was unable to detect any putative conserved domains. The longest stretch of conserved residues is located from residues 74 to 83. A high degree of variation is observed in residues 97–105 that represent the α 4 helix. The highly conserved valine residues Val56, Val66, Val82 and Val83 in MSMEG_5817 maintain the potential hydrophobic binding pocket. The conservation of these valine residues across all orthologues, as well as the presence of additional hydrophobic residues replacing the other valine positions in MSMEG_5817, suggests that all orthologues may bind nonpolar ligands.

3.5. Phospholipid-binding capabilities of MSMEG_5817 by ELISA

Owing to the conservation of the hydrophobic binding pocket and the structural similarities of MSMEG_5817 to the SCP class of proteins, we hypothesized that MSMEG_5817 and its orthologues may be involved in the binding of biologically relevant lipids. SCPs have previously been characterized to bind to cholesterol, a range of phospholipids and saturated and unsaturated fatty acids of varying lengths (van Amerongen *et al.*, 1989; Wirtz & Gadella, 1990; Schroeder *et al.*, 1995; Avdulov *et al.*, 1999; Kernstock & Girotti, 2007).

The lipid-binding capabilities of MSMEG_5817 were probed by immobilized lipid ELISA (enzyme-linked immunoassay). hSCP has previously been observed to bind tightly to cholesterol, with a reported affinity of 4.0 nM (Stolowich *et al.*, 1999). The lipid-binding capabilities of MSMEG_5817 was initially probed against cholesterol, as well as a number of highly prevalent biological fatty acids, including palmitic acid, which was present in the crystal structure of AaSCP (Dyer *et al.*, 2003). Recombinant hSCP was found to bind oleic acid, palmitic acid and cholesterol, as observed previously (Fig. 6a; Dyer *et al.*, 2003; Kernstock & Girotti, 2007). In comparison, MSMEG_5817 was found to not bind any of the initially tested lipids. In addition, no growth defects were observed upon growing MSMEG_5817 mutant on minimal medium where oleic acid, palmitic acid and cholesterol were the only carbon sources (data not shown). This observation is in keeping with a recent study looking at genes that are essential for cholesterol catabolism in *M. tuberculosis*, with Rv0807, the *M. tuberculosis* homologue of MSMEG_5817, being shown not to be required for growth that is dependent on cholesterol as sole carbon source (Griffin *et al.*, 2011).

Further experiments exploring the potential lipid-binding capabilities of MSMEG_5817 were performed using commercially available lipid strips (Echelon). It was shown that MSMEG_5817 bound to a range of biologically relevant phospholipids; however, no binding was observed to cholesterol (data not shown). It has previously been reported that hSCP is capable of binding phosphatidylcholine (PC) as well as a range of phosphatidylinositols with varying specificities (Gadella & Wirtz, 1991; Avdulov *et al.*, 1999; Zhou *et al.*, 2004).

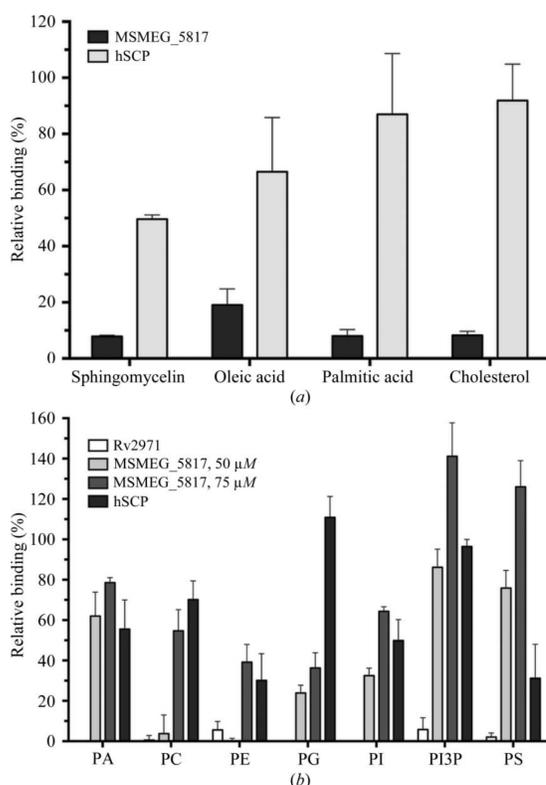


Figure 6 Comparative lipid-binding capabilities of MSMEG_5817 and *H. sapiens* SCP (hSCP) as determined by ELISA. (a) The lipid-binding capabilities of MSMEG_5817 were assessed against lipids previously determined to be bound by hSCP with high affinity (Schroeder *et al.*, 2007). Data were normalized against the results of hSCP binding to cholesterol as a positive control. Limited to no binding was observed for MSMEG_5817, indicating a difference in lipid specificity compared with hSCP. (b) A range of biologically relevant phospholipids were selected to probe the lipid-binding capabilities of MSMEG_5817. Data were normalized against hSCP as a positive control and the *M. tuberculosis* Aldo-keto reductase Rv2971 as a negative control. The results indicate binding to a number of phospholipids in a dose-dependent response manner. The ELISA experiment was conducted as described in §2.7. Data are represented as relative percentage binding against the controls. Results were visualized using *GraphPad Prism*, with error bars representing the mean percentage \pm SEM of triplicate experiments.

research papers

Verification of phospholipid binding by MSMEG_5817 was also performed using immobilized lipid ELISA. MSMEG_5817 was found to bind a number of phospholipids

with a dose-dependent response, with the strongest binding observed to phosphatidic acid (PA), phosphatidylserine (PS) and phosphatidylinositol 3-phosphate (PI3P) (Fig. 6b). Weak

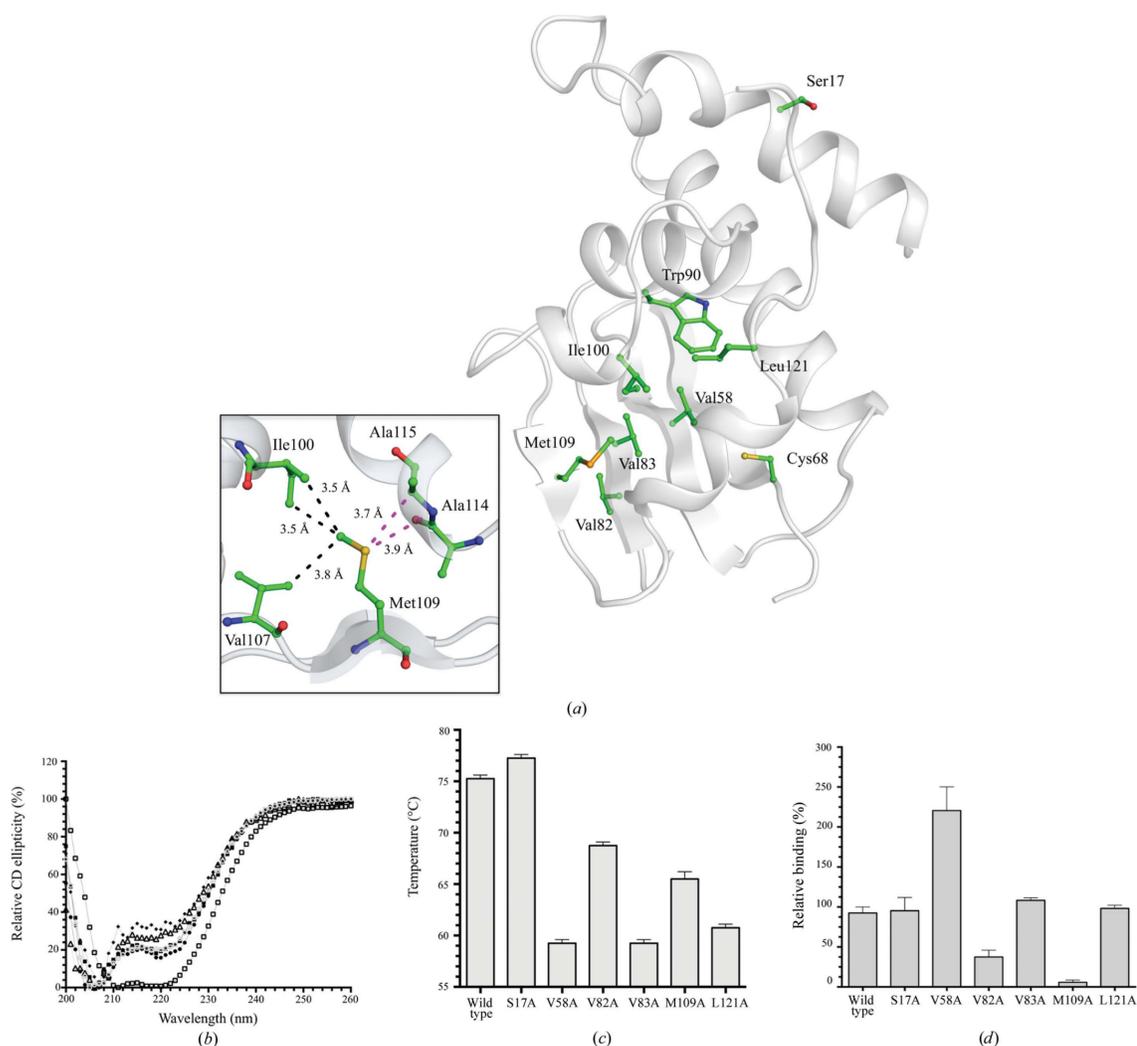


Figure 7

Mutant analysis of MSMEG_5817 by alanine-scanning mutagenesis. (a) Stick representation of the positions of the residues mutated to alanine. Residues for mutagenesis studies were selected based on sequence identity between MSMEG_5817 orthologues and conservation with SCP ligand-binding residues. Inset: stick representation of amino-acid residue contacts with Met109 positioned in the β_4 strand. The C^δ atom of Met109 forms two hydrophobic interactions with Ile100 with bond distances of 3.5 Å, as well as a single hydrophobic interaction with Val107 with a bond distance of 3.8 Å, represented by the black dashed line. In addition, the S^δ atom of Met109 forms van der Waals interactions with the main chains of Ala114 and Ala115 in the α_5 helix, with bond distances of 3.7 and 3.9 Å, represented by magenta dashed lines. (b) Circular-dichroism spectra of MSMEG_5817 and mutants. Far-UV CD spectra were recorded of wild-type MSMEG_5817 (filled circles) and the S17A (filled squares), V58A (triangles), V82A (filled diamonds), V83A (circles), M109A (squares) and L121A (diamonds) mutants between wavelengths of 200 and 260 nm. Spectra were visualized using *GraphPad Prism*. (c) Thermal stability assay of MSMEG_5817 mutants compared with the wild type, as measured by ThermoFluor. The greatest reduction in thermal stability was observed for the V58A and V63A mutants. Results are represented as the mean temperature \pm SEM in triplicate experiments. (d) Lipid-binding capabilities of MSMEG_5817 mutants by ELISA. Binding capabilities to PI3P of MSMEG_5817 mutants compared with wild-type MSMEG_5817 as measured by ELISA as described in §2.7. The greatest reduction in binding capabilities was observed for the V82A and M109A mutants. Results were normalized against the wild type (100%) and a PBS blank (0%). Results are represented as mean percentages \pm SEM in triplicate experiments. All graphs were visualized using *GraphPad Prism*.

research papers

binding was observed with phosphatidylglycerol (PG) and phosphatidylinositol (PI) and minimal or no binding was observed with phosphatidylcholine (PC) and phosphatidylethanolamine (PE). Stronger binding was observed with PI3P compared with PI, indicating substrate specificity for the additional phosphate on the PI3P head group. No binding was observed for the saturated and unsaturated fatty acids tested, which contribute to phospholipid fatty-acid tails, indicating substrate specificity for the larger phospholipid head groups of PI3P and PS. In contrast, recombinant hSCP was found to bind to a range of the phospholipids, with the highest binding to PG, while the Aldo-keto reductase (Rv2971) from *M. tuberculosis* showed no binding to phospholipids (Fig. 6b). These results indicate a difference in substrate specificity between MSMEG_5817 and hSCP, with MSMEG_5817 binding preferentially to phospholipids with predominantly electro-negative head groups.

3.6. Probing the putative MSMEG_5817 binding pocket

In the absence of a crystal structure of MSMEG_5817 with a bound ligand, the potential ligand-binding residues were probed by utilizing alanine-scanning mutagenesis. The lipid ELISA indicated differences in binding mechanisms and substrate specificities between MSMEG_5817 and hSCP. The mutated residues were selected based on sequence identity within the MSMEG_5817 orthologue family and the SCP family, along with positional conservation of residues involved in ligand contacts within the SCPs (Fig. 7a).

The mutants were expressed and purified as soluble protein, with the exception of W90A and C68A, which were expressed as insoluble protein, indicating essentiality in maintaining the structural integrity of MSMEG_5817. No expression of I100A was detected by Western blotting. Each of the mutant proteins was purified by size-exclusion chromatography, revealing a strong peak relating to monomeric protein. In the case of M109A and L121A, a high degree of soluble aggregate was present and was successfully removed by size-exclusion chromatography.

In order to ascertain the effects of the mutants on the overall secondary-structure composition of MSMEG_5817, circular dichroism (CD) in the far-UV range was conducted. The CD spectra measured between wavelengths of 260 and 200 nm indicated conservation of α -helical content between wild-type MSMEG_5817 and each of the purified mutants, with the exception of M109A (Fig. 7b). The changes within the M109A mutant may indicate destabilization of the β 4 sheet, resulting in a stronger α -helical representative trace, with a prominent dip at 220 nm observed in the CD spectra. Met109 forms hydrophobic interactions with Ile100 present on the α 4 helix, as well as van der Waals interactions between the S atom and the C α atom of residue Ala115 in the α 5 helix (inset in Fig. 7a), which is conserved amongst the orthologues (Fig. 5). This could be an indication as to the lack of expression of I100A, as well as the high degree of soluble aggregate observed for M109A. However, Met109 is poorly conserved between orthologues and is only present in *M. leprae* (Fig. 5).

Within mycobacterial species a leucine residue is predominantly present at position 109, with smaller residues present amongst the broader Corynebacteria orthologues.

The structural integrity of each of the mutants was analysed by RT-PCR Thermofluor using the fluorescent dye SYPRO Orange to determine thermal stability (T_m) values (Ericsson *et al.*, 2006). A significant drop in the T_m value was observed for each of the mutants, with the exception of S17A, which acted as the control mutant. The greatest decrease in the T_m value was observed for the internal valine mutants V58A and V83A, with a decrease of $\sim 15^\circ\text{C}$ from $\sim 70^\circ\text{C}$ for wild-type MSMEG_5817 to $\sim 54^\circ\text{C}$ for the mutants (Fig. 7c). These results indicate that the internal valine residues tested are vital for maintaining the hydrophobic integrity of the protein core, stabilizing the overall protein fold (Fig. 4a).

To gain insight into the effects of the mutant binding capabilities, an ELISA was conducted using PI3P as the strongest binding lipid. The results of the ELISA indicated no statistical differences in binding capabilities between the wild-type and the internal V83A ($p < 0.19$) and L121A ($p < 0.57$) mutants (Fig. 7d). The V58A ($p < 0.05$) mutation resulted in a significant increase in lipid binding, suggesting that this mutation affects the environment in the potential binding pocket, thus increasing the affinity for the phospholipid PI3P.

The binding of PI3P by wild-type MSMEG_5817 and the control S17A mutant was the same. There was a complete loss of PI3P binding by the M109A mutant ($p < 0.001$), which may be owing to alteration of the overall secondary-structure composition of MSMEG_5817 or to direct residue binding. Interestingly enough, there was a 60% reduction in binding of PI3P by the solvent-exposed V82A mutant ($p < 0.05$), which may indicate differences from the binding mechanism shown by hSCP, which has a preference for cholesterol over phospholipids.

4. Conclusions

Vital to the success of mycobacterial pathogenesis is the ability of mycobacteria to parasitize the host macrophage phagosome, creating a hospitable environment for survival and replication. The key to controlling mycobacterial infection is to identify the novel biochemical pathways utilized for this process for the purposes of rational drug design. The recently identified gene product of MSMEG_5817 from *M. smegmatis* has been found to be vital for mycobacterial survival within host macrophages (Pelosi *et al.*, 2012). The MSMEG_5817 gene encodes a protein of 128 amino-acid residues of unknown function that is highly conserved within the Corynebacterineae suborder. In the present study, we have determined the crystal structure of MSMEG_5817. The crystal structure revealed a Rossmann-like fold $\alpha\beta$ two-layer sandwich forming a highly hydrophobic interface cavity and with high structural homology to the SCP family. Coupled together, this supports the hypothesis that MSMEG_5817 may be involved in the interaction of apolar ligands through its hydrophobic cavity.

research papers

Investigation of the lipid-binding capabilities of MSMEG_5817 revealed binding to a number of phospholipids, with the highest specificities for those with predominantly negatively charged head groups, namely PA, PS and PI3P. This specificity was in direct contrast to the lipid-binding capabilities of hSCP. Alanine-scanning mutagenesis of the hydrophobic cavity of MSMEG_5817 demonstrated that the conserved Val82 residue plays an important role in ligand binding, with the ligand-binding interface potentially located at the base of the structure, centred on Val82 and the ψ -loop, in a similar position to the AaSCP binding site. The result for the M109A mutant was surprising owing to changes in secondary structure and the complete loss of ligand-binding capability despite its poor conservation within the MSMEG_5817 family. This suggests that the potential interface has delicate architecture.

While binding to a number of phospholipids was observed, further studies are required to identify the natural ligand and identify the true function of the MSMEG_5817 family. Owing to the binding of MSMEG_5817 to PI3P, we speculate that it may have a potential role in processing host-cell PI3P, thereby contributing to the block in phagolysosomal maturation; however, confirmatory studies are required. Nevertheless, the structural and functional characterization presented in this report has paved the way for the further characterization of this family of proteins and their functional role within the Corynebacterineae suborder of bacteria.

We thank the staff of the Australian Synchrotron and Monash Macromolecular Crystallization Facility for assistance with crystallization and X-ray data collection. This work was supported by the Australian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics and the National Health and Medical Research Council of Australia. JR is an NHMRC Australia Fellow and TB is a Pfizer Australian Research Fellow.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst. D* **66**, 213–221.
- Amerongen, A. van, Demel, R. A., Westerman, J. & Wirtz, K. W. A. (1989). *Biochim. Biophys. Acta*, **1004**, 36–43.
- Anes, E., Peyron, P., Staali, L., Jordao, L., Gutierrez, M. G., Kress, H., Hagedorn, M., Maridonneau-Parini, I., Skinner, M. A., Wildeman, A. G., Kalamidas, S. A., Kuehnel, M. & Griffiths, G. (2006). *Cell. Microbiol.* **8**, 939–960.
- Armstrong, J. A. & Hart, P. D. (1971). *J. Exp. Med.* **134**, 713–740.
- Avdulov, N. A., Chochina, S. V., Igbavboa, U., Warden, C. S., Schroeder, F. & Wood, W. G. (1999). *Biochim. Biophys. Acta*, **1437**, 37–45.
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
- Billman-Jacobe, H., McConville, M. J., Haites, R. E., Kovacevic, S. & Coppel, R. L. (1999). *Mol. Microbiol.* **33**, 1244–1253.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P. S. A., Smart, O. S., Vornrhein, C. & Womack, T. O. (2011). *BUSTER*. Cambridge: Global Phasing Ltd.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D* **66**, 12–21.
- Choinowski, T., Hauser, H. & Piontek, K. (2000). *Biochemistry*, **39**, 1897–1902.
- Clemens, D. L., Lee, B.-Y. & Horwitz, M. A. (2000a). *Infect. Immun.* **68**, 2671–2684.
- Clemens, D. L., Lee, B.-Y. & Horwitz, M. A. (2000b). *Infect. Immun.* **68**, 5154–5166.
- Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., Davis, A., de Haro, S., Naylor, J., Lee, H.-H. & Vergne, I. (2006). *Cell. Microbiol.* **8**, 719–727.
- Dyer, D. H., Lovell, S., Thoden, J. B., Holden, H. M., Rayment, I. & Lan, Q. (2003). *J. Biol. Chem.* **278**, 39085–39091.
- Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 140–144.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
- Fratti, R. A., Backer, J. M., Gruenberg, J., Corvera, S. & Deretic, V. (2001). *J. Cell Biol.* **154**, 631–644.
- Fratti, R. A., Chua, J., Vergne, I. & Deretic, V. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 5437–5442.
- Gadella, T. W. J. Jr & Wirtz, K. W. A. (1991). *Biochim. Biophys. Acta*, **1070**, 237–245.
- García, F. L., Szyperski, T., Dyer, J. H., Choinowski, T., Seedorf, U., Hauser, H. & Wüthrich, K. (2000). *J. Mol. Biol.* **295**, 595–603.
- Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J. & Sasseti, C. M. (2011). *PLoS Pathog.* **7**, e1002251.
- Haapalainen, A. M., Van Aalten, D. M. F., Meriläinen, G., Jalonen, J. E., Piriälä, P., Wierenga, R. K., Hiltunen, J. K. & Glumoff, T. (2001). *J. Mol. Biol.* **313**, 1127–1138.
- Hmama, Z., Sendide, K., Talal, A., Garcia, R., Dobos, K. & Reiner, N. E. (2004). *J. Cell Sci.* **117**, 2131–2140.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
- Kang, P. B., Azad, A. K., Torrelles, J. B., Kaufman, T. M., Beharka, A., Tibesar, E., DesJardin, L. E. & Schlesinger, L. S. (2005). *J. Exp. Med.* **202**, 987–999.
- Kelley, V. A. & Schorey, J. S. (2003). *Mol. Biol. Cell.* **14**, 3366–3377.
- Kernstock, R. M. & Girotti, A. W. (2007). *Anal. Biochem.* **365**, 111–121.
- Kovacevic, S., Anderson, D., Morita, Y. S., Patterson, J., Haites, R., McMillan, B. N. I., Coppel, R., McConville, M. J. & Billman-Jacobe, H. (2006). *J. Biol. Chem.* **281**, 9011–9017.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Parish, T., Smith, D. A., Roberts, G., Betts, J. & Stoker, N. G. (2003). *Microbiology*, **149**, 1423–1435.
- Patterson, J. H., McConville, M. J., Haites, R. E., Coppel, R. L. & Billman-Jacobe, H. (2000). *J. Biol. Chem.* **275**, 24900–24906.
- Pelosi, A., Smith, D., Brammananth, R., Topolska, A., Billman-Jacobe, H., Nagley, P., Crellin, P. K. & Coppel, R. L. (2012). *PLoS One*, **7**, e31788.
- Pereira-Leal, J. B. & Seabra, M. C. (2001). *J. Mol. Biol.* **313**, 889–901.
- Pfeffer, S. R. (2005). *J. Biol. Chem.* **280**, 15485–15488.
- Rao, S. T. & Rossmann, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.
- Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. (2005). *Cell*, **122**, 735–749.
- Russell, D. G. (2001). *Nature Rev. Mol. Cell Biol.* **2**, 569–577.
- Russell, D. G. (2011). *Immunol. Rev.* **240**, 252–268.
- Saleh, M. T. & Belisle, J. T. (2000). *J. Bacteriol.* **182**, 6850–6853.
- Sasseti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
- Schroeder, F., Atshaves, B. P., McIntosh, A. L., Gallegos, A. M., Storey, S. M., Parr, R. D., Jefferson, J. R., Ball, J. M. & Kier, A. B. (2007). *Biochim. Biophys. Acta*, **1771**, 700–718.
- Schroeder, F., Myers-Payne, S. C., Billheimer, J. T. & Wood, W. G. (1995). *Biochemistry*, **34**, 11919–11927.

research papers

- Shahine, A. E., Chan, P. Y., Littler, D., Vivian, J., Brammananth, R., Crellin, P. K., Coppel, R. L., Rossjohn, J. & Beddoe, T. (2013). *Acta Cryst. F* **69**, 566–569.
- Sillitoe, I., Cuff, A. L., Dessailly, B. H., Dawson, N. L., Furnham, N., Lee, D., Lees, J. G., Lewis, T. E., Studer, R. A., Rentsch, R., Yeats, C., Thornton, J. M. & Orengo, C. A. (2013). *Nucleic Acids Res.* **41**, D490–D498.
- Stolowich, N., Frolov, A., Petrescu, A. D., Scott, A. I., Billheimer, J. T. & Schroeder, F. (1999). *J. Biol. Chem.* **274**, 35425–35433.
- Sturgill-Koszycki, S., Schlesinger, P. H., Chakraborty, P., Haddix, P. L., Collins, H. L., Fok, A. K., Allen, R. D., Gluck, S. L., Heuser, J. & Russell, D. G. (1994). *Science*, **263**, 678–681.
- Szyperski, T., Scheek, S., Johansson, J., Assmann, G., Seedorf, U. & Wüthrich, K. (1993). *FEBS Lett.* **335**, 18–26.
- Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst. D* **65**, 582–601.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst. D* **64**, 61–69.
- Vergne, I., Chua, J., Lee, H.-H., Lucas, M., Belisle, J. & Deretic, V. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 4033–4038.
- Vergne, I., Chua, J., Singh, S. B. & Deretic, V. (2004). *Annu. Rev. Cell Dev. Biol.* **20**, 367–394.
- Via, L. E., Deretic, D., Ulmer, R. J., Hibler, N. S., Huber, L. A. & Deretic, V. (1997). *J. Biol. Chem.* **272**, 13326–13331.
- Wirtz, K. W. A. & Gadella, T. W. J. Jr (1990). *Experientia*, **46**, 592–599.
- World Health Organization (2012). *Global Tuberculosis Report*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/en/.
- Zhou, M., Parr, R. D., Petrescu, A. D., Payne, H. R., Atshaves, B. P., Kier, A. B., Ball, J. M. & Schroeder, F. (2004). *Biochemistry*, **43**, 7288–7306.
- Zhu, H. (2011). *DSSP & Stride Plugin for PyMOL*. Biotechnology Center (BIOTEC), TU Dresden, Germany.

Chapter 6: Crystal Structure of MSMEG_5817 produced in *M.*

smegmatis

6.1 Introduction

In the previous two chapters, structural determination of MSMEG_5817 produced in *E. coli* was described, with insight into its biological function characterized based on its 2.40 Å crystal structure. As a means of retaining native protein fold and functionality, and to obtain a higher quality crystal structure, the crystal structure of MSMEG_5817 was determined utilizing recombinant protein produced within its native host *M. smegmatis*.

The advantage of employing recombinant protein produced from its source organism for x-ray crystallographic studies is the potential for the co-expression and purification in the presence of naturally bound ligands. Previously published studies have utilized this approach for the characterization of essential mycobacterial proteins in the presence of native ligands, as described in section 1.7. While the crystal structure of MSMEG_5817 provided some insight into its biological function, its biochemical pathway is presently unknown, with its natural ligand presently unknown.

Chapter 6 comprises of a manuscript prepared for publication in *Protein Science*, describing the recombinant protein expression in *M. smegmatis*, as well as the crystallization, x-ray data collection and structural determination to 1.70 Å. While no native ligands bound to the protein were present, a higher quality crystal structure of MSMEG_5817 was obtained. Additional structural elements absent in the 2.40 Å crystal structure were present, namely the missing residues belonging to the highly conserved Ψ-loop. The higher quality crystal structure allows a more in depth structural analysis of

MSMEG_5817, providing a more representative crystal structure of the MSMEG_5817 protein family within the *Corynebacterae* suborder.

The successful determination of the crystal structure of MSMEG_5871 has revealed no structural differences between structures determined utilizing both *E. coli* and *M. smegmatis* as an expression system. These findings have broader implications in the structural genomic characterization of *M. tuberculosis* proteins in the utilization of *M. smegmatis* as an expression system. *M. smegmatis* provides an alternate expression system in cases where soluble recombinant TB proteins cannot be produced in *E. coli*, which represents a major bottleneck in TB structural genomics. Despite protein production in *M. smegmatis* proving to be highly successful, this method is highly underutilized, with only 8 unique crystal of mycobacterial proteins determined to date (RCSB PDB: <http://www.rcsb.org/pdb/home/home.do>) As a means of retaining native fold, as well as the potential for targeting natural ligands, *M. smegmatis* utilized as an expression system represents a significant technique for the structural characterization of TB targets.

Declaration for Thesis Chapter 6

Declaration by candidate

In the case of Chapter 6, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Expression, purification, and crystallisation of native and Se-Met derivative MSMEG_5817. X-ray data collection of native MSMEG_5817. Data analysis, manuscript and figure preparation.	80%

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

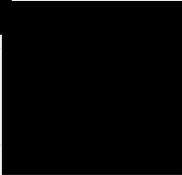
Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Rajini Brammananath	Supervision of Phooi Chan, assistance in <i>M. smegmatis</i> expression system, data interpretation and manuscript preparation	N/A
Phooi Chan	Cloning into <i>M. smegmatis</i>	10%
Paul Crellin	Direction, data interpretation and manuscript preparation	N/A
Ross Coppel	Direction, data interpretation and manuscript preparation	N/A
Jamie Rossjohn	Conception, supervision, direction, data interpretation and manuscript preparation	N/A
Travis Beddoe	Conception, supervision, direction, data interpretation and manuscript preparation	N/A

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's
Signature

 **Date**
21/10/15

Main
Supervisor's
Signature

 **Date**
21/10/15

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

**The high-resolution structure of a novel uncharacterized
Mycobacterium smegmatis gene product MSMEG_5817 expressed
within homologous expression system.**

Adam Shahine¹, Rajini Brammananth², Phooi Chan², Paul Crellin², Ross L. Coppel², Jamie Rossjohn^{1,3*} and Travis Beddoe^{1,4*}

¹ Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia.

² Infection and Immunity Program and Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia.

³ Institute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, UK.

⁴ Department of Animal, Plant and Soil Science and Centre for AgriBioscience (AgriBio), La Trobe University, Bundoora, Victoria, Australia

* Joint senior and corresponding authors, email: t.beddoe@latrobe.edu.au or Jamie.Rossjohn@monash.edu

Running Title: 1.70 Å crystal structure of MSMEG_5817 from native host

Manuscript pages: 16

Tables and Figures:

Table 1: Crystallographic data summary table

Figure 1: Overall crystal structure of MSMEG_5817 solved to 1.70 Å

Figure 2: Overview of the Ψ-loop

Figure 3: Stabilization of the Ψ-loop

Supplementary material (Single figure): Inter-asymmetric unit interactions between two units, with the two molecules (Molecule A and Molecule B) highlighted in each unit.

Abstract

The genus *Mycobacterium* comprises a number of deadly human pathogens, including *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB). An urgent need has arisen for drug development against TB due to the increasing prevalence of drug resistance. X-ray crystallography has provided a pivotal platform for the structural characterization of mycobacterial drug targets, proteins involved in pathogenicity, and essential proteins of unknown function. Despite this, a major bottleneck exists in the production of soluble recombinant protein and diffraction-quality crystals from *M. tuberculosis*. *Mycobacterium smegmatis*, a fast-growing, saprophytic mycobacterial species, has provided a means for overcoming this bottleneck; both as an effective expression system for mycobacterial proteins themselves and as a source of orthologues with related functions. MSMEG_5817 is an *M. smegmatis* protein implicated in bacterial survival within host macrophages and we have determined its crystal structure to a resolution of 1.70 Å utilizing recombinant protein produced within a homologous expression system. The high-resolution structure reported here reveals previously unresolved structural details from a previously determined MSMEG_5817 structure (2.4Å), such as the conformation of the usually disordered ψ -loop that is highly conserved throughout the protein family. The new MSMEG_5817 crystal structure allows for a higher quality functional characterization of this poorly characterized protein family.

Key words: X-ray Crystallography, *Mycobacterium tuberculosis*, *Mycobacterium smegmatis*, Native Expression Host, MSMEG_5817, Ψ -loop.

Introduction

Bacteria of the genus *Mycobacterium* include a number of devastating pathogens, such as *Mycobacterium tuberculosis*, the causative agent of human tuberculosis (TB). Approximately one third of the world's population is currently infected with *M. tuberculosis*, resulting in 1.4 million deaths per year.¹ The emergence of drug resistant strains of *M. tuberculosis* in many countries has made treatment substantially more difficult, resulting in an urgent need for the development of novel drugs.

Structural biology has been valuable in the understanding of mycobacterial biology and drug development. Of the 3,999 open reading frames present within the *M. tuberculosis* H37Rv genome², approximately 400 unique crystal structures have been deposited into the protein data bank (PDB, <http://www.rcsb.org/pdb/home/home.do>). While this signifies considerable progress, primarily driven by the formation of the TB Structural Genomics Consortium in 2000³, there still exists substantial difficulty in expressing and crystallising recombinant TB proteins. As a means of overcoming this bottleneck, suitable orthologues from *Mycobacterium smegmatis* are often selected for structural analysis, with such orthologues making up 15% of deposited mycobacterial structures. *M. smegmatis* is a saprophytic mycobacterial species with limited capabilities for survival and growth within host macrophages, making it an suitable model organism for the study of mycobacterial pathogenesis.⁴

The fast growing, non-pathogenic properties of *M. smegmatis* make it an attractive recombinant mycobacterial protein expression system.⁵ In the absence of soluble recombinant protein expression within *Escherichia coli*, *M. smegmatis* has been a useful replacement expression system⁶, with 40 mycobacterial structures currently produced via this manner. Expression in mycobacteria hosts causes minimal disruption to the post-translational and folding pathways of TB proteins. Moreover it has previously allowed for the retention of native substrates to aid in functional characterization⁷.

The preservation of a protein's structural and functional integrity in this system makes it an attractive mechanism for the characterization of novel proteins of unknown function. One such example is a recently identified gene product from *M. smegmatis*, MSMEG_5817, implicated in bacterial survival within host macrophages.⁸ The *M. tuberculosis* protein Rv0807 did not crystallize readily in our hands, whereas we were able to obtain a structure of the *M. smegmatis* orthologue at 2.40 Å resolution, providing insight into the protein's function^{9,10}. It was found that MSMEG_5817 shared structural homology with the sterol carrier protein (SCP) family¹⁰, whose role is to mediate the binding and transport of sterols and lipids across cellular membranes¹¹⁻¹³.

The lipid binding capabilities of MSMEG_5817 were confirmed, revealing binding to a range of phospholipids, with a potentially different binding mechanism to that of the SCPs.¹⁰

Here we describe the 1.70 Å crystal structure of MSMEG_5817 produced within a homologous expression system, revealing additional structural elements not observed within its previously determined crystal structure. Our analysis of this higher resolution protein structure provides a more in depth characterisation of MSMEG_5817.

Results

The construct included residues 1– 128 of MSMEG_5817 with 7 vector-derived residues and an C-terminal hexahistidine purification tag was successfully produced in homologous *M. smegmatis* expression system. Protein purification routinely yielded 10 – 12 mg.L⁻¹ of pure MSMEG_5817, significantly higher than a similar construct produced from a heterologous *E. coli* system⁹. High diffraction quality crystals of MSMEG_5817 grew in 16 % Polyethylene glycol 6000, 0.1 M Tris-HCl pH 8.3, 0.2 M MgCl₂ at 20 °C. No crystals of the *M. smegmatis* produced MSMEG_5817 were obtained in the crystallization condition used for the *E. coli* produced MSMEG_5817.^{9,10} The crystal structure was solved by molecular replacement, using the *M. smegmatis* MSMEG_5817 produced in *E. coli* (PDB 4NSS). Crystallographic data collection and refinement statistics are summarised in Table 1.

Two nearly identical copies of MSMEG_5817 were present within the asymmetric unit, with a core R.M.S.D. of 0.23 Å upon structural overlay. Residues 1 – 128 of MSMEG_5817 were crystallized; including 7 N-terminal vector derived residues and a C-terminal hexahistidine tag. Of this construct, residues 8 – 126 in Molecule A [Fig 1(A)] and residues 1 – 126 in molecule B [Fig 1(B)] were modelled from the electron density, with the inclusion of the 7 N-terminal vector derived residues in molecule B, denoted as residues -6 to 0. Electron density is missing for one segment, between residues 73-77, in molecule B [Fig 1(B)].

Table 1: Crystallographic Data Summary Table

Values in parenthesis represent highest resolution shell

	MS-MSMEG_5817
Space group	P22 ₁ 2 ₁
Temperature (K)	100
Resolution range (Å)	30.22 – 1.70 (1.79 – 1.70)
Wavelength (Å)	0.9357
Unit-cell parameters (Å, °)	a=42.6, b=70.5, c=100.4 α=90.0, β=90.0, γ=90.0
Total No. of reflections	342332 (430235)
No. of unique reflections	33747 (4709)
Multiplicity	10.1 (9.1)
Completeness (%)	99.2 (97.2)
CC (1/2)	0.999 (0.901)
R_{merge} (%)^a	5.8 (33.7)
R_{pim} (%)^b	2.9 (18.6)
Mean I/σ(I)^c	20.9 (4.8)
R_{work} (%)^d	17.52
R_{free} (%)^e	20.95
Number of non-hydrogen atoms	2190
Macromolecules	1874
Ligands	40
Water	276
Protein residues	247
R.M.S deviations from ideality	
Bond Lengths (Å)	0.008
Bond Angles (°)	1.08
Ramchandran Plot	
Favoured Region (%)	98.0
Allowed Region (%)	1.2
Outliers (%)	0.8
B-factors (Å²)	
Wilson B-factor	15.71
Average B-factors	28.4
Average Macromolecule	26.50
Average Ligand	51.30
Average Water	38.50

$$a R_{\text{merge}} = \sum hkl \sum i |I_{hkl, i} - \langle I_{hkl} \rangle| / \sum hkl \langle I_{hkl} \rangle$$

$$b R_{\text{pim}} = \sum hkl [1/(N-1)]^{1/2} \sum i |I_{hkl, i} - \langle I_{hkl} \rangle| / \sum hkl \langle I_{hkl} \rangle$$

c *I* is the integrated intensity and $\sigma(I)$ is the estimated standard deviation of that intensity

d $R_{\text{work}} = (\sum | |F_o| - |F_c| |) / (\sum |F_o|)$ - for all data except as indicated in footnote e.

e $R_{\text{free}} = 5\%$ of data were used for the R_{free} calculation. ##

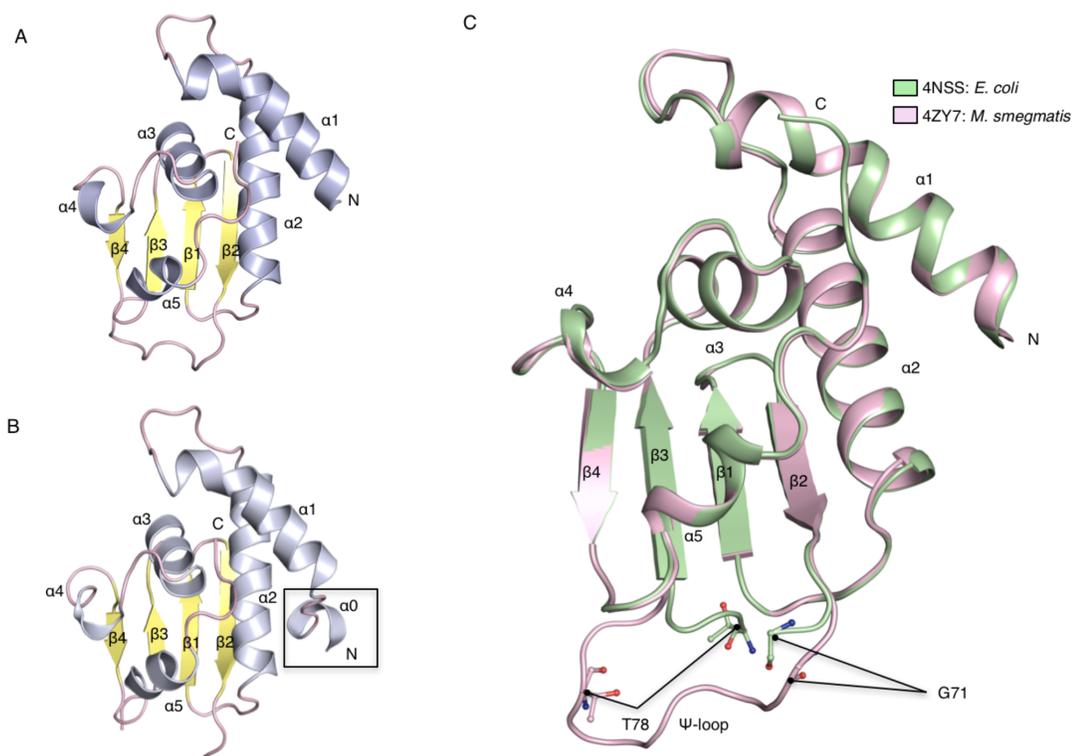


Figure 1: Overall crystal structure of MSMEG_5817 solved to 1.70 Å. Cartoon representation of the MSMEG_5817 crystal structure of (A) Chain A and (B) Chain B, highlighting secondary structure elements. The $\alpha 0$ helix is visible only in Chain B (outlined). (C) Structural overlay the MSMEG_5817 crystal structures as produced in *M. smegmatis* (pink) and *E. coli* (green). The two crystal structures overlay with a core R.M.S.D value of 0.26 Å. The N-terminal (G71) and C-terminal (T78) residues surrounding the gap in the ψ -loop presented in the *E. coli* produced crystal structure are also highlighted.

The structure of MSMEG_5817 consists of a four stranded antiparallel β -sheet stacked by a layer of five α -helices in a Rossman-like two-layer α/β sandwich fold [Fig. 1]. Crystals of MSMEG_5817 grew in the orthorhombic space group P22₁2₁, which contrasts with the primitive tetragonal space group P4₃2₁2 of the MSMEG_5817 used as the search model for molecular replacement⁹. The two structures overlay closely, with a core R.M.S.D value of 0.26 Å obtained, showing a virtually identical fold [Fig 1(C)]. The higher resolution of the *M. smegmatis*-derived structure allowed an additional 14 residues to be modelled, which corresponded to the previously described ψ -loop^{10,14,15}, that spans residues 55-85, including all secondary structure elements [Fig 2]. The ψ -loop in MSMEG_5817 adopts a type 1 topology^{14,15} [Fig 2(i)], which is common in a wide variety of protein families¹⁶⁻²². The ψ -loop sits on two flexible hinges: residues S70-G71 at the C-terminal end of the β 2 strand, and P79-P80 at the N terminal end of the β 3 strand. The residues in between these hinges were disordered in the previous structure whereas they are visible in the *M. smegmatis* ψ -loop and do not appear to be influenced by crystal packing interactions in either crystal form¹⁰. The hinge residues contribute to the flexibility of this loop region and the intervening residues are highly conserved amongst Mycobacterial orthologues; 23 of the 30 residues within the ψ -loop are strictly conserved. Compared to the *M. tuberculosis* orthologue, Rv0807, the only residue difference within the ψ -loop is K73, which is substituted for an arginine at this position¹⁰.

The main stabilizing features of the MSMEG_5817 ψ -loop are contacts made by residue I69, which is characterized as a Ramachandran outlier on both chains in the *M. smegmatis* [Table 1] and *E. coli*¹⁰ crystal structures. I69 is the C-terminal residue of the β 2 strand and sits at the interface between the core of the domain and the flexible ψ -loop, making contacts with both [Fig 3]. This includes van der Waals interactions with the side chains of E116, V117 and Y120 within the central core of the fold, found on the α 3 helix [Fig 3]. The β -branched character of I69 is conserved amongst all Mycobacterial orthologues of MSMEG_5817.

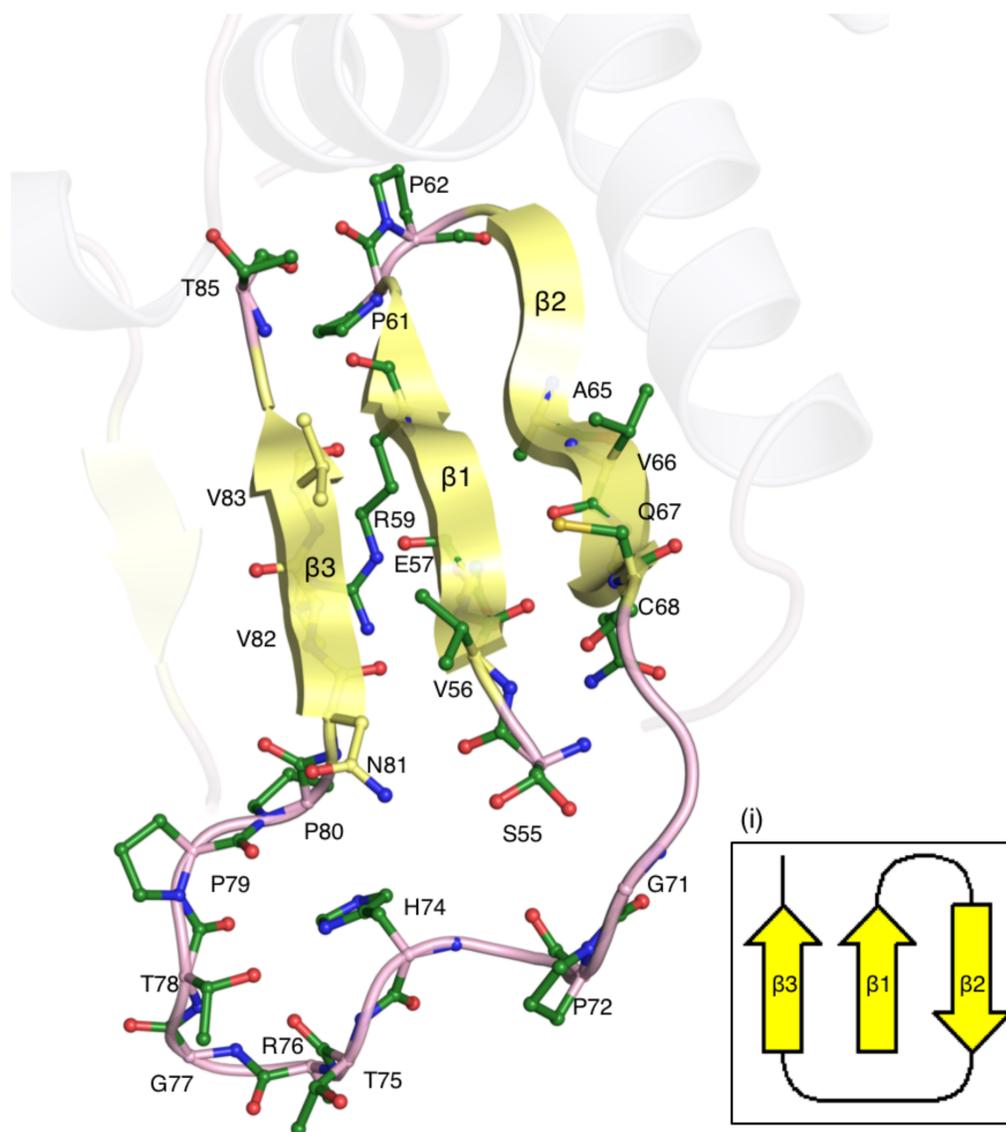


Figure 2: Overview of the Ψ -loop. Cartoon representation of the β 1- β 2- β 3 strand organization in MSMEG_5817 characteristic of a type 1 Ψ -loop (i) Topology map of MSMEG_5817 Ψ -loop strand organization. The 23 strictly conserved residues (green) within the Ψ -loop are highlighted in stick-and-ball representation. Residue S55 exists as two alternate conformations.

In addition to the newly resolved ψ -loop, the N-terminal residues 1 – 7 are visualized within molecule B. These residues are unobserved within the previously determined structure of MSMEG_5817 and reveal an additional α -helix (denoted as $\alpha 0$), spanning residues 1 – 5 [Fig 1(B)]. The architecture of the N-terminal region is influenced by the 7 residues derived from the pJAM2 vector, with the sequence PEVVFGS. The vector-derived serine is included within the $\alpha 0$ helix, indicating that the additional secondary structure element may not be biologically relevant. In the case of this structure, the vector-derived residues aid crystallization by acting as an inter-asymmetric unit locking mechanism, stabilized predominantly through hydrogen bond and salt bridge formation [Sup Fig 1].

Discussion

Identifying the phospholipid-binding site of MSMEG_5817 is central to characterizing the function of this protein family, but has remained elusive¹⁰. Understanding the role of the ψ -loop, which is highly conserved amongst MSMEG_5817 orthologues, will aid in characterizing this function. Not all SCPs contain ψ -loop equivalents, but for proteins in which it is present the equivalent residues are often seen to be essential for maintaining structural integrity and protein function¹⁷. It is unclear whether the hydrophobic core of MSMEG_5817 opens to allow phospholipids to enter in a manner similar to that of other proteins within the SCP family. A lack of sequence or motif similarities between the ψ -loops of different protein family members suggests substrate binding by MSMEG_5817 may have some unique features. In addition, the presence of the ψ -loop in MSMEG_5817 and not in the SCP family supports the role that MSMEG_5817 binds phospholipids in a differing manner to the SCPs¹⁰.

The improved diffraction resolution compared to the previously determined MSMEG_5817 structure appears to be fortuitously driven by the presence of additional crystal packing residues derived from the *M. smegmatis* expression vector, no additional post-translational modifications were observed. Thus, producing MSMEG_5817 utilizing the homologous expression system yields one major advantage: mainly a much-improved yield when compared to heterologous expression system such as an *E. coli* system. Previous alanine-scanning mutagenesis studies had shown that mutation of V82 within this ψ -loop reduces the protein's affinity for phospholipids¹⁰, leading to the hypothesis that this region of MSMEG_5817 is part of the phospholipid-binding site. As the ψ -loop was not ordered in the previous structure, the role of V82 could not be confirmed. In the present structure, vector derived residues appear to limit the movement of the

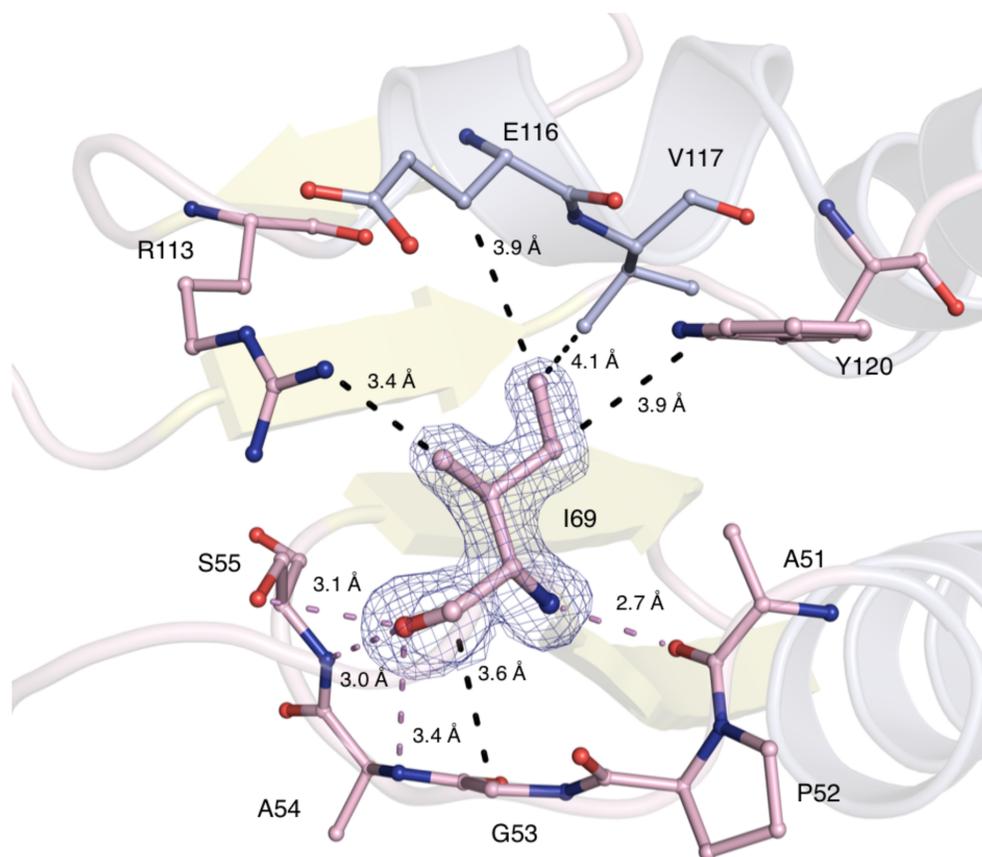


Figure 3: Stabilization of the Ψ -loop. Hydrogen bond formation between the backbone of I69 and neighbouring residues of the Ψ -loop are shown in pink, with bond distances ranging between 2.7 Å and 3.4 Å. Hydrophobic and Van Der Waals interactions between I69 and residues on the α 3 helix are shown in black, with bond distances ranging between 3.4 Å and 4.1 Å. Secondary structure elements are also highlighted. The σ_A -weighted ($2m|F_o| - D|F_c|$) electron-density map in blue for I69 are contoured at 2.0 σ .

β 1 and β 3 hinges locking the ψ -loop into a single conformation that facilitates its visualisation, confirming that V82 is orientated towards the potential lipid binding site.

In conclusion, the crystal structure of MSMEG_5817 derived from recombinant protein produced in its native host has produced a higher quality crystal structure, revealing previously unobserved structural elements. The complete visualization of the ψ -loop allows for a more in depth characterization of the protein structure, which exhibits differences with protein families that also contain this structural motif. Aided by crystal contacts with the vector-derived sequence, a more complete structural model was derived, allowing for visualization of the ψ -loop and an additional α 0 helix, which will aid in further biological characterisation of this class of protein.

Materials and Methods

MSMEG_5817 expression and purification

The *MSMEG_5817* gene was cloned into the pJAM2 vector, containing a C-terminal hexahistidine purification tag, and transformed into wild type mc²155 *M. smegmatis* cells for expression. Cell cultures for overexpression of recombinant protein were grown as previously described, with minor alterations.⁵ Pre-inoculum cultures were grown for 4 d at 37 °C in 7H9 Middlebrook medium containing 100 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin and supplemented with 0.025 % (v/v) Tween 80 and 10 % (v/v) ADS enrichment (11 mM glucose, 14.5 mM NaCl, 76 μM Bovine Serum Albumin). The starter culture was used at a dilution of 1:50 to inoculate 500 mL M63 medium (76 mM $(\text{NH}_4)_2\text{SO}_4$, 500 mM KH_2PO_4 , 5 μM FeSO_4 , pH 7.0) supplemented with 1 mM MgSO_4 , 0.5 % (v/v) Tween-80, 2 % (w/v) succinate, 2 % (w/v) acetamide, 0.008 % (w/v) glucose and 100 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin. Cultures were grown at 37 °C for 3 d. Cells were harvested by centrifugation and stored at -80 °C until use. Recombinant protein was purified to homogeneity as previously described^{9,10}, with yields of 10-12 $\text{mg}\cdot\text{L}^{-1}$ routinely obtained.

Crystallization

Initial crystallization conditions for MSMEG_5817 were found through commercially available screens (Hampton Research and Qiagen). Crystals were produced by the hanging drop method in 24-well Linbro plates (Hampton Research), with each crystallization drop consisting of 1 μL protein solution (15 – 20 $\text{mg}\cdot\text{mL}^{-1}$ in 20 mM MES pH 6.0, 100 mM NaCl) mixed with 1 μL mother liquor and a 500 μL reservoir volume. Thin plate crystals of poor diffraction quality appeared within 12 – 24 hours in 20 % (v/w) PEG 8000, 0.1 M Tris-HCl pH 8.5, 0.2 M MgCl_2 at 20 °C. Optimal crystals for diffraction data collection were obtained in 16 % (v/v) PEG 6000, 0.1 M Tris-HCl pH 8.3, 0.2 M MgCl_2 at 20 °C after 24 – 48 hours. For x-ray diffraction data collection, crystals were transferred into a CryoLoop and soaked in a cryoprotectant of 20 %

(v/w) PEG 6000, 0.1 M Tris-HCl pH 8.3, 0.2 M MgCl₂, 10 % (v/v) Ethylene Glycol before flash freezing at -173 °C in a stream of nitrogen gas.

Data collection, structure solution and refinement

X-ray diffraction data for MSMEG_5817 plate crystals was collected at the Australian Synchrotron MX1 beamline using the ADSC-Quantum 210r CCD at -173 °C and $\lambda = 0.9537 \text{ \AA}^{23}$. The data was processed in the orthorhombic space group P22₁2₁ to 1.70 Å using *iMOSFLM*²⁴ and processed using SCALA as part of the *CCP4* suite²⁵. The unit cell parameters were $a = 42.61 \text{ \AA}$, $b = 70.49 \text{ \AA}$, $c = 100.37 \text{ \AA}$, and $\alpha/\beta/\gamma = 90.0^\circ$. There were two molecules in the asymmetric unit, corresponding to a Matthews coefficient of $2.47 \text{ \AA}^3 \text{ D}^{-1}$.²⁶ The structure was determined using molecular replacement as implemented by Phaser-MR as part of the PHENIX protein suite²⁶. The search model used was MSMEG_5817 as produced in *Echerichia coli* (PDB entry: 4NSS), yielding the final ensemble model. Resultant TFZ and LLG values obtained were 9.7 and 27.9 respectively. Model building and refinement were performed using the programs Coot²⁸ and Phenix-refine as part of the Phenix program suite²⁷. Figures were prepared using PyMOL. Secondary structure was confirmed using the DSSP/STRIDE plugin for PyMOL²⁹. Model quality was evaluated by PROCHECK³⁰ and Molprobit³¹.

Atomic coordinates

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB code 4ZY7).

Acknowledgments

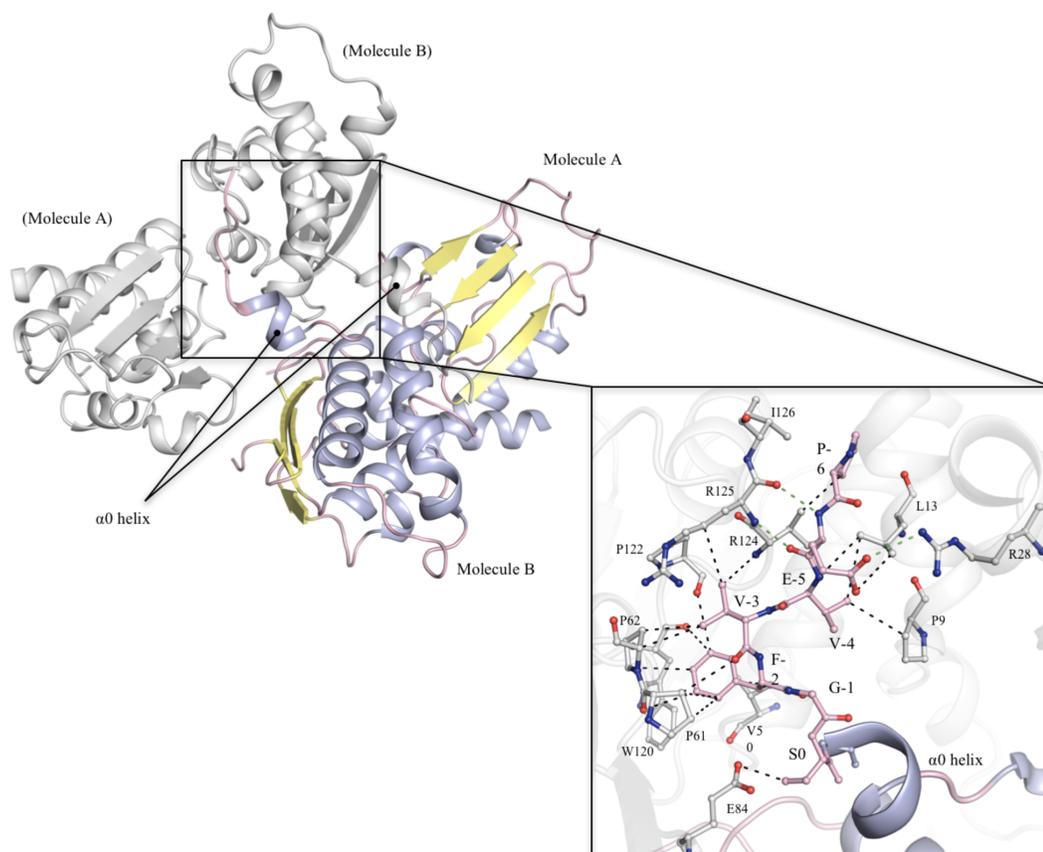
We thank the staff of Monash Macromolecular Crystallisation Facility for their assistance in crystallisation and Australian Synchrotron for the X-ray data collection and support provided. This research was supported by the Australian Research Council (ARC). JR was supported by an NHMRC Australia Fellowship.

References

1. WHO (2012) Global Tuberculosis Report 2012. World Health Organization.
2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence. *Nature* 393: 537-544.
3. Goulding CW, Apostol M, Anderson DH, Gill HS, Smith CV, et al. (2002) The TB structural genomics consortium: Providing a structural foundation for drug discovery. *Current Drug Targets - Infectious Disorders* 2: 121-141.
4. Anes E, Peyron P, Staali L, Jordao L, Gutierrez MG, et al. (2006) Dynamic life and death interactions between mycobacterium smegmatis and J774 macrophages. *Cellular Microbiology* 8: 939-960.
5. Triccas JA, Parish T, Britton WJ, Gicquel B (1998) An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*. *FEMS Microbiology Letters* 167: 151-156.
6. Harth G, Lee BY, Horwitz MA (1997) High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major *Mycobacterium tuberculosis* extracellular proteins considered to be leading vaccine candidates and drug targets. *Infection and Immunity* 65: 2321-2328.
7. Gokulan K, O'Leary SE, Russell WK, Russell DH, Lalgondar M, et al. (2013) Crystal structure of *Mycobacterium tuberculosis* polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *Journal of Biological Chemistry* 288: 16484-16494.
8. Pelosi A, Smith D, Brammananth R, Topolska A, Billman-Jacobe H, et al. (2012) Identification of a Novel Gene Product That Promotes Survival of *Mycobacterium smegmatis* in Macrophages. *PLoS ONE* 7: e31788.
9. Shahine A, Chan P, Littler D, Vivian J, Brammananth R, et al. (2013) Cloning, expression, purification and preliminary X-ray diffraction studies of a mycobacterial protein implicated in bacterial survival in macrophages. *Acta Crystallographica Section F: Crystallization Communications* F69: 566-569.
10. Shahine A, Littler D, Brammananth R, Chan PY, Crellin PK, et al. (2014) A structural and functional investigation of a novel protein from *Mycobacterium smegmatis* implicated in mycobacterial macrophage survivability. *Acta Crystallographica Section D: Biological Crystallography* D70: 2264-2276

11. Schroeder F, Myers-Payne SC, Billheimer JT, Wood WG (1995) Probing the ligand binding sites of fatty acid and sterol carrier proteins: Effects of ethanol. *Biochemistry* 34: 11919-11927.
12. Avdulov NA, Chochina SV, Igbavboa U, Warden CS, Schroeder F, et al. (1999) Lipid binding to sterol carrier protein-2 is inhibited by ethanol. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1437: 37-45.
13. Kernstock RM, Girotti AW (2007) Lipid transfer protein binding of unmodified natural lipids as assessed by surface plasmon resonance methodology. *Analytical Biochemistry* 365: 111-121.
14. Hutchinson EG & Thornton JM (1990) HERA—A program to draw schematic diagrams of protein secondary structures. *Proteins* 8: 203–212
15. Hutchinson EG & Thornton JM (1996) PROMOTIF—a program to identify and analyse structural motifs in proteins. *Protein Science* 5: 212-220.
16. Song H, Parsons MR, Roswell S, Leonard G, Phillips SE (1999) Crystal structure of intact elongation factor EF-Tu from *Escherichia coli* in GDP conformation at 2.05 Å resolution. *JMB* 285: 1245-1256
17. Turco MM & Sousa MC (2014) The structure and specificity of the Type III secretion system effector NleC suggest a DNA mimicry mechanism of substrate recognition. *Biochemistry* 53(31): 5131-5139
18. Castillo RM, Mizuguchi K, Dhanaraj V, Albert A, Blundell TL, et al, (1999) A six-stranded double-psi β barrel is shared by several protein superfamilies. *Structure* 7(2): 227-236
19. Holmes MA, Tronrud DE, Matthews BW (1983), Structural analysis of the inhibition of thermolysin by an active-site-directed irreversible inhibitor. *Biochemistry* 4(22): 236-240
20. Khan AR, Khazanovich-Bernstein N, Bergmann EM, James MNG (1999) Structural aspects of activation pathways of aspartic protease zymogens and viral 3C protease precursors. *PNAS* 96: 10968-10975
21. Bernstein NK, Cherney MM, Loetscher H, Ridley RG, James MNG (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nature Structure Biology* 6: 32-37
22. Kay J, Meijer HJG, ten Have A, van Kan JAL (2011) The aspartic proteinase family of three *Phytophthora* species. *BMC Genomics* 12: 254

23. Cowieson, N.P., Aragao, D., Clift, M., Ericsson, D.J., Gee, C., Harrop, S.J., Mudie, N., Panjikar, S., Price, J.R., Riboldi-Tunncliffe, A., Williamson, R., & Caradoc-Davies, T., (2015), MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.*, 22, 187-190
24. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* 67: 271-281.
25. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 67: 235-242.
26. Matthews BW (1968) Solvent content of protein crystals. *Journal of Molecular Biology* 33: 491-497.
27. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 66: 213-221.
28. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66: 486-501.
29. Zhu H (2011) DSSP & Stride plugin for PyMOL. Biotechnology Center (BIOTEC), TU Dresden.
30. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26: 283-291.
31. Chen VB, Arendall Iii WB, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* 66: 12-21.
32. Brunger AT (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-475.



Supplementary Figure 1: Inter-asymmetric unit interactions between two units, with the two molecules (Molecule A and Molecule B) highlighted in each unit. Indicated is the α_0 -helix observed on Molecule B. (Insert) Crystal contacts between the vector-derived residues (pink) and neighbouring asymmetric unit molecule residues (grey). Insert is rotated 45° to the right. Black dashes indicate Van Der Waals and hydrophobic interactions, whereas the green dashes indicate salt bridge formation. Distances range between 2.9 Å and 4.0 Å.

Chapter 7: Discussion

The emergence of drug resistant strains of *M. tuberculosis* has placed a global strain on mechanisms of treatment against TB. This has resulted in an urgent need in the development of new anti-TB therapeutics via novel mechanisms of action. To fully gauge the devastating effects of the pathogen, further characterisation of *M. tuberculosis* pathogenesis is required, with a focus on the mechanisms of survival within host macrophages and the essential biosynthetic pathways involved in the generation of its lipid rich cell wall. These unique characteristics of mycobacteria are largely attributed to the success of the pathogen, and result in significant difficulties in its eradication. While an extensive amount of research has been conducted on *M. tuberculosis*, there are still significant gaps in our knowledge on the mechanisms of mycobacterial pathogenesis and host survival, as well as mechanisms of drug resistance.

The work conducted in the course of this thesis highlights the characterization of three novel mycobacterial proteins essential for the pathogen's survival, with a predominant focus on their structural characterization through x-ray crystallography. The three crystal structures of mycobacterial protein targets highlights a range of applications for x-ray crystallography in the characterization of mycobacterial pathogenesis and mechanisms of drug inhibition, for the ultimate purpose of rational drug design.

7.1: Structural genomics and TB

Structural genomics is a relatively new field, beginning in the late 1990s, and focuses on the large-scale generation of protein structures of novel importance. Now, structural genomics is a mature field, and has primarily been driven by new technological advancements in high throughput soluble protein expression, purification and protein crystallisation, as well as the improvement of tuneable synchrotron radiation for the purposes of x-ray crystallography. Technological improvements such as the

establishment of high-throughput structural characterization on a collaborative level have increased impact of determined crystal structures and the advancement of the structural genomics field, allowing for the formation of the Protein Structure Initiative: Biology (PSI: Biology) (DePietro *et al.*, 2013). As the whole annotated genomes of a multitude of organisms have become increasingly more available, targeted structural genomic centres and consortiums have been developed. These structural genomic groups focus on significant human pathogens, with a key aim in the determination of protein structures for the purposes characterizing pathogenesis and the development of novel therapeutics.

With regards to *M. tuberculosis*, the TB Structural Genomics Consortium (TBSGC) that began in 2000 (<http://www.webtb.org/>), encompassing global organization of core structural biology facilities, with the primary goal of the 3-D structural determination of all mycobacterial proteins. These protein structures have the potential to advance understanding of mycobacterial pathogenesis, with implications in rational drug design (Terwilliger *et al.*, 2008). By 2011, approximately 250 unique crystal structures of *M. tuberculosis* proteins have been determined by members of the consortium, which accounts for over one third of total *M. tuberculosis* structures deposited in the PDB (Chim *et al.*, 2011).

Targets for structural determination are predominantly selected through manual and bioinformatic filters to focus on essential and novel protein targets. This is based on sequence identity with predetermined protein structures, as well as filtering out targets with biophysical characteristics unsuitable for structural analysis, such as protein insolubility, membrane association and protein size. On the genomic level, targets are selected based on their novelty and essentiality for mycobacterial growth and survival. With the annotation of the *M. tuberculosis* H37Rv genome in 1998, extensive research has been conducted on determining essential genes for survival through transposon site

hybridization analysis and phenotypic annotation (Cole *et al.*, 1998; Sasseti *et al.*, 2003; Griffin *et al.*, 2011), which has allowed for the effective selection of mycobacterial targets for structural characterization.

Each of the three mycobacterial protein targets described throughout the course of this thesis were selected based on previous genomic annotation, as well as significant experimental work recently conducted (Scoble *et al.*, 2010, Crellin *et al.*, 2010, Pelosi *et al.*, 2012), and represent three targets essential for *M. tuberculosis* growth and survival. The crystal structures of each of the targets have been successfully determined; each revealing novel structural features that aid in functional characterization, with the potential for future therapeutic development.

7.2 Drug discovery and structure based rational drug design against TB

One of the central aims of the project was the determination of the crystal structures of these essential mycobacterial targets for their use in structure based rational drug design. This process utilizes the crystal structure of a target protein for the design of a high affinity inhibitor. A crystal structure of the target protein in complex with the inhibitor of interest allows for the in depth analysis of the interaction mechanisms of the inhibitor. This allows for the redesigning of the inhibitor to create a compound of higher affinity for drug design.

With regards to the crystal structures described, the crystal structures of Rv2971 and Rv3802c were determined for the ultimate aim of rational drug design, with both INH and THL respectively. In both cases, the crystal structures of their *M. smegmatis* orthologues have previously been determined (Scoble *et al.*, 2010; Crellin *et al.*, 2010). Yet while these crystal structures reveal vitally important information in the structure and function on these two protein targets, they remain unsuitable for the

purposes of both the characterization of inhibition mechanisms, as well as for the purposes of rational drug design.

The orthologue approach in rational drug design is suitable in cases where structural data on the target protein is lacking. In these cases, structures of orthologues of high sequence identity in related species are determined and, while this approach is not ideal, can be utilized for the purposes of drug design. In the case of Rv3802c, the orthologue approach is unsuitable for this purpose, where THL inhibits Rv3802c 3 times tighter than its *M. smegmatis* orthologue, MSMEG_6394, which has previously had its crystal structure determined (Crellin *et al.*, 2010). Differences in sequence identity between the two orthologues play, in part, a role in THL inhibition variations. Despite the orthologues sharing a 66% sequence identity, residues apparently involved in contacts with THL are strictly conserved, save for Serine 292, which is substituted for Alanine 291 in MSMEG_6394. A higher resolution crystal structure of Rv3802c would be essential to further characterize the THL binding cavity, allowing for a more in depth comparison with MSMEG_6394. The most significant variations are present within the $\alpha 9$ helical lid domain and the flexible loop facilitating movement of the helix, which plays a vital role in substrate binding. Variations in plasticity between the helical lid domains of the two orthologues aids in explaining differences in both substrate and THL inhibition specificities. Despite a high sequence identity, subtle architectural differences in the substrate binding machinery between the two orthologues have a significant impact on THL specificities, reiterating the fact that structural characterization of the target itself, and not the orthologue, is more suitable for rational drug design.

To truly gauge the mechanisms of THL inhibition against *M. tuberculosis*, additional drug targets must be identified. Rv3802c belongs to the cutinase-like protein (Culp) family, where there are 7 Culps present within *M. tuberculosis*, which displays a range of substrate specificities (West *et al.*, 2009).

Coupled with Rv3802c belonging to the highly prevalent α/β -hydrolase fold superfamily, it is most likely that there are additional THL targets within the *M. tuberculosis* proteome. As a means of identifying additional drug targets of THL through experimental means, the next stage would be to conduct a THL affinity chromatographic assay, in a similar manner as previously described with the INH that identifying Rv2971 as an INH drug target (Argyrou *et al.*, 2006). This assay can be expanded with the utilization of the THL compound library, which contains THL analogous compounds with a higher affinity against Rv3802c (West *et al.*, 2011), allowing for an in-depth characterisation of the inhibition mechanisms of THL to aid in rational drug design.

This approach has been utilized in identifying a number of targets in *Trypanosoma brucei*, the parasite responsible for African sleeping sickness, against THL and THL analogues, via an *in situ* parasite-based proteome wide profiling method (Yang *et al.*, 2012). A total of 30 targets were identified via this manner, with targets predominantly functionally annotated in fatty acid synthesis, phospholipase or lipid transfer proteins, including targets belonging to the α/β hydrolase fold family (Yang *et al.*, 2012). This approach may yield similar success levels upon identifying additional THL targets within *M. tuberculosis* aside from Rv3802c: aiding the broad characterization of the inhibition mechanisms of THL.

Identifying targets against a broad range of drugs can further expand this type of experiment. For example, a collection of 400 anti-malarial inhibitors and chemotypes, termed the “malaria box” (Guiguemde *et al.*, 2012; Spangenberg *et al.*, 2013), may provide a starting point for identifying a range of both new drug targets and new anti-TB compounds. While developed to specifically target malaria causing *Plasmodium* spp., such as *Plasmodium falciparum*, application of the malaria box can be expanded to alternative pathogenic characterisation, such as *M. tuberculosis*. In identifying new lead

inhibitors, this allows for the determination of affinities between identified drug targets and lead inhibitors, as well as provide a basis for new structural genomic characterisation of TB drug targets.

A number of new lead inhibitors against TB have recently been identified following a similar principle. A recent study involving high-throughput screening in the identification of new lead inhibitor compounds against *M. tuberculosis* H37Rv has been conducted, involving the GlaxoSmithKline's (GSK) corporate compound collection (~ 2 million compounds) (Ballel *et al.*, 2013). The collection of compounds were selected based on the established "rule of 5" parameters indicating efficiency of lead compounds as effective drugs (Lipinski *et al.*, 2001). A total of 177 compounds were identified to be highly selective potent inhibitors of *M. tuberculosis* H37Rv, with *M. bovis* BCG implemented as a TB surrogate, and exhibit non-cytotoxic properties against human hepatocarcinoma (HepG2) cells (Figure 7.2a) (Ballel *et al.*, 2013), with the most potent compounds belonging to 7 family clusters of compounds (Figure 7.2b). Biochemical characterization of the targets of these new inhibitors points towards the novel mycobacterial mechanisms of cell wall biosynthesis and respiratory mechanisms (Ballel *et al.*, 2013); cementing the specificities of these new inhibitors against *M. tuberculosis*.

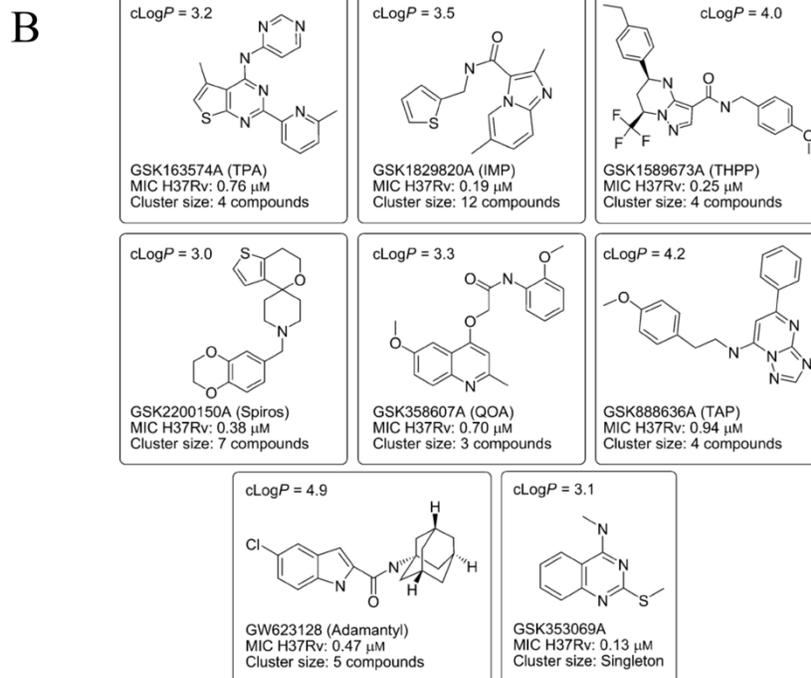
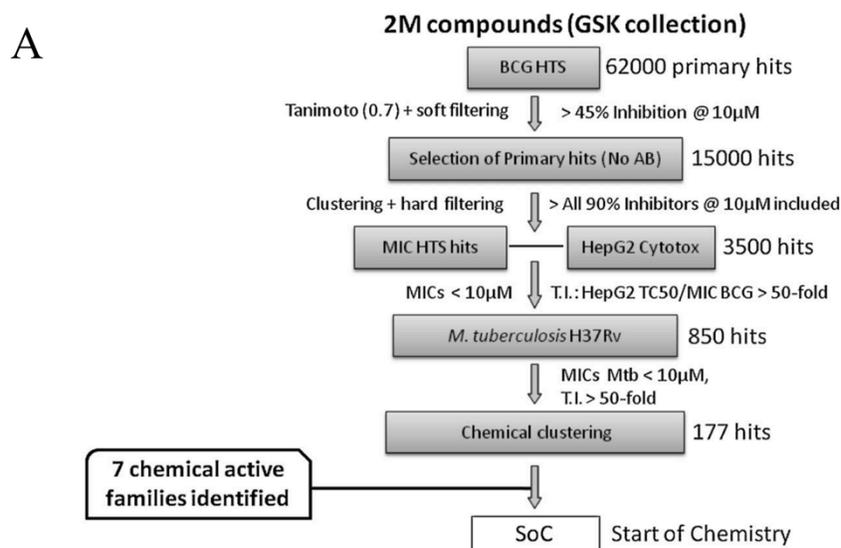


Figure 7.2: Lead inhibitor clusters identified via high-throughput screening

(A) Progressional flow diagram of the high-throughput screening used in identifying the 177 new lead drug compounds.

(B) Seven chemical families, as well as a single drug, with potent anti-TB capabilities identified via high-throughput screening using compounds from the GSK corporate compound collection. Indicated for each drug family visualized are the chemical identifiers, minimal inhibitory concentration (MIC) values against *M. tuberculosis* H37Rv and cluster sizes. For each compound family, an octanol-water partition coefficient (cLog P) value of less than 5.0 was obtained. (Adapted from Ballel *et al.*, 2013)

The application of broad high throughput screening as a means of identifying new lead inhibitors represents a significant advancement in the development of new anti-TB drugs. The more frequent utilization of high-throughput screening calls into question the efficiency of the single target approach: where structure based rational drug design is centred on identifying a high affinity inhibitor against a predetermined protein target. This approach has been utilized with regards to the TB targets presented, in particular the structural determination of Rv3802c in complex with THL. While the successful determination of its crystal structure may now provide a basis for rational drug design with THL, the mechanisms of inhibition, as well as the number of TB targets against the drug, are still widely unknown. Furthermore, in the case of isoniazid, while each of the drug targets have previously been identified, including Rv2971 presented earlier, it presently does not act as a new lead inhibitor against the continuously establishing drug resistant strains of *M. tuberculosis*. In both cases, the drugs THL and INH are “old” drugs, in the sense that they have previously been implemented for purposes other than the treatment of multiple drug resistant TB. For these drugs to be utilized as new lead inhibitors, they must undergo significant a significant drug design process, aided by x-ray crystallography for the high-resolution visualization of inhibition mechanisms.

The future of drug discovery and drug design against TB is highly dependant on the identification of new lead inhibitor compounds. While the implementation of the single target approach for drug discovery has become a widely utilized method, there is a high level of attrition in these projects reaching beyond the discovery phase, with an ever-decreasing number of drugs successfully passing the clinical trial phases (Payne *et al.*, 2006; Brown., 2007). Aside from time and cost restrictions, this is limited due to factors beyond structural or biochemical characterization of new drugs and drug targets, including drug specificities and inhibitor potencies, cytotoxicity, and the poor translation from the R&D stage to their clinical therapeutic impact. To overcome the many factors impairing successful

therapeutic development and distribution against TB, the identification of new lead inhibitory compounds are required that overcome these issues upon initial discovery. The high-throughput screening of the GSK compound collection against TB represents the most significant progress forward in discovery of new inhibitors (Ballel *et al.*, 2013). Further characterization of the protein targets inhibited by the newly identified targets will play an essential role in not only further understanding the biology of *M. tuberculosis*, but will aid significantly in the rational drug design of new therapeutics.

To advance the employment of these compounds as new anti-TB therapeutics, the next stage would be to identify and characterise the mechanisms of inhibition against the identified protein targets from *M. tuberculosis*. This is where x-ray crystallography plays crucial role as a means of characterizing the mechanisms of inhibition of the compounds against the identified targets; thus aiding in rational drug design. The mechanisms of identifying the inhibitor targets presents an opportunity to provide an effective alternative to the single target approach for structural based rational drug design described throughout this thesis. Drug targets can be identified via *in vitro* affinity chromatography and immunoprecipitation pulldown assays from whole cell lysate using each of the lead compounds identified, or at least representative members of the 7 compound family clusters. Examples of this assay were described previously in the identification of INH drug targets from *M. tuberculosis* (Argyrou *et al.*, 2006) and THL drug targets from *T. brucei* (Yang *et al.*, 2012). The limitation of this assay is the possible inability to select for membrane associated drug targets, which could not be isolated via this method due to insolubility issues of potential drug targets and the compounds themselves; heavily restricting the number of identified TB drug targets. Assays such as phage display for peptide drugs (Molek *et al.*, 2011), and with the implementation of nanodisks for non-peptide drugs (Pavlidou *et al.*, 2013), as well as the AVEXIS screen (Bartholdson *et al.*, 2013) may be implemented in the presentation of inhibitory compounds to membrane-associated targets in *M. tuberculosis* H37Rv

in vivo. Upon target identification, inhibitor affinities can be determined using techniques such as surface plasmon resonance and absorbance based enzyme kinetic assays as a means of characterising *in vitro* affinities. Ultimately, crystal structures of identified targets in complex with lead inhibitors would provide in depth analysis of the inhibition mechanisms and structural based rational drug design.

The utilization of high throughput inhibitor screens against *M. tuberculosis* provides the strongest basis for accelerating drug discovery for TB. Coupled with the biochemical assays described for the identification and characterization of drug targets, as well as x-ray crystallography for structural characterization of inhibitory mechanisms, significant progress can be made towards the development of new anti-TB therapeutics utilizing structural based rational drug design. This significantly aids in advancing the progress of TB structural genomics in further understanding mycobacterial biology and pathogenesis, yet predominantly in the development of new anti-TB therapeutics to combat multiple drug resistant strains of *M. tuberculosis*.

7.3 Hypothetical mycobacterial proteins

An advantage of conducting a broad drug binding screen using compound libraries such as the GSK compound collection of the malaria box is in the identification of novel drug targets, in particular targeting hypothetical mycobacterial proteins or mycobacterial proteins of unknown function. A recent analysis of the *M. tuberculosis* H37Rv genome has annotated 1096 ORFs (27% of overall genome) to code for hypothetical proteins. The study continued to attempt to predict the function of a majority of these hypothetical proteins through bioinformatic and manual data mining analysis, allowing for the functional prediction of 622 ORF's (15% of the overall genome), with the remaining 474 ORFs (12% of the overall genome) correlating to novel mycobacterial proteins (Figure 7.3) (Doerks *et al.*, 2012).

While the functions of over half the ORFs encoding hypothetical proteins were successfully predicted, further experimental analysis of these hypothetical proteins is required to confirm the predictions.

Characterizing the functionality of hypothetical proteins within *M. tuberculosis*, particularly those essential for the mycobacterium's survival, would greatly increase our understanding of its pathogenesis. Previous assessment of ORF essentiality of the *M. tuberculosis* H37Rv genome has revealed a number of genes essential for mycobacterial growth annotated as hypothetical proteins (Sasseti *et al.*, 2003). Recent genomic analysis of the *M. tuberculosis* H37Rv genome predicting 261 genes coding hypothetical proteins were found to be essential genes; correlating to 38.2% of total essential genes within the mycobacterium's genome (Xu *et al.*, 2013). Further genomic characterization of orthologous mycobacterial species, in particular *M. smegmatis*, has revealed additional mycobacterial genes essential for pathogenesis (Billman-Jacobe *et al.*, 1999; Patterson *et al.*, 2000; Kovacevic *et al.*, 2006; Pelosi *et al.*, 2012). Despite its saprophytic nature, *M. smegmatis* can be successfully utilized to characterize mycobacterial pathogenesis due to its fast growing nature and limited ability to replicate and survive within host macrophages, with less safety restrictions than working with *M. tuberculosis*. While *M. smegmatis* is non-pathogenic and exhibits a differing biology to *M. tuberculosis*, the high percentage of orthologous ORFs between genomes allows for both a functional and structural characterization of *M. tuberculosis* orthologues in *M. smegmatis*.

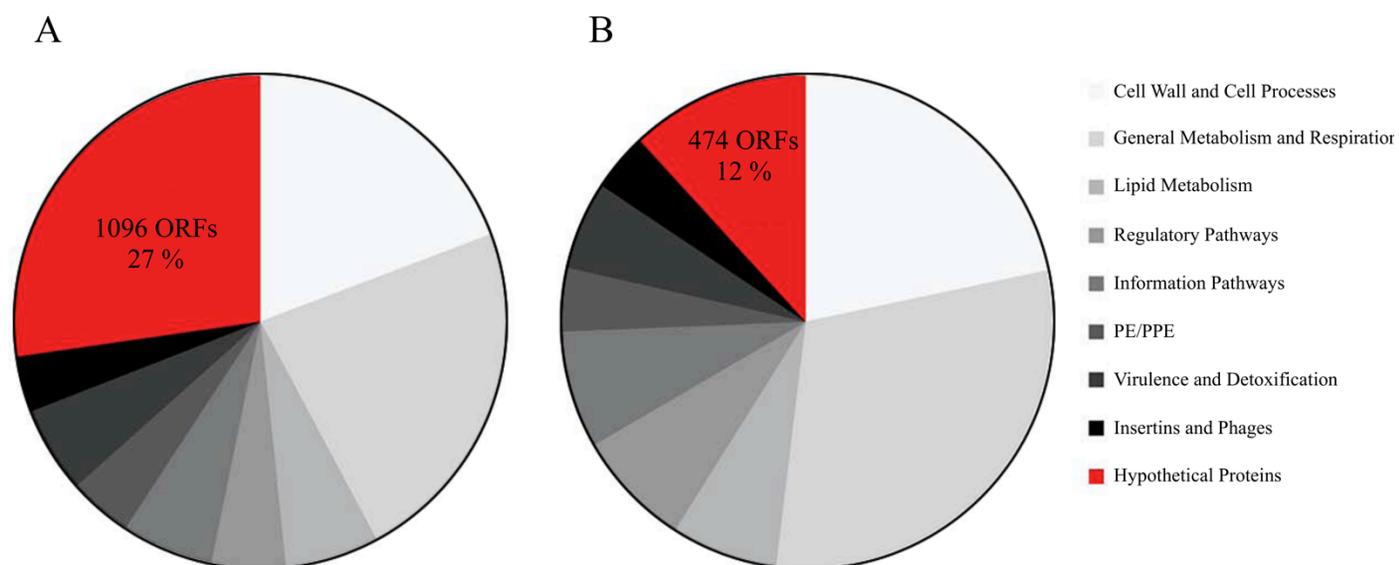


Figure 7.3: Graphical representation of annotated ORF function in the *M. tuberculosis* H37Rv genome

The graphs highlight the number and percentages against the complete genome of ORF's annotated as hypothetical proteins.

(A) Before computational functional prediction of hypothetical proteins in *M. tuberculosis*, 1096 ORF's, approximately 27% of the 4019 ORF *M. tuberculosis* H37Rv genome, encode for hypothetical proteins with no functional characterization available

(B) After computational function prediction of hypothetical proteins in *M. tuberculosis*, 622 ORF's had functional roles predicted, with a remaining 474 ORF's, approximately 12% of overall *M. tuberculosis* H37Rv genome, representing novel mycobacterial genes with no functional annotation.

(Adapted from Doerks *et al.*, 2012)

This is evident in the identification and characterization of the newly discovered gene in *M. smegmatis*, identified as *MSMEG_5817*, essential for the mycobacterium's survival within its host macrophages. The high degree of conservation indicates the essentiality of this protein family within *mycobacterial spp.*, with its orthologue in *M. tuberculosis*, Rv0807, sharing a sequence identity of 66%. *MSMEG_5817* and its orthologues are annotated as a hypothetical protein, and prior to the work described here, very little was known as to its biological function (Glover *et al.*, 2007; Pelosi *et al.*, 2012).

The successful determination of its crystal structure revealed that *MSMEG_5817* might play a functional role in the binding of biologically relevant lipids, which was confirmed experimentally in its ability to bind a range of phospholipids. In addition, with the crystal structure of *MSMEG_5817* determined, we were able to probe its potential lipid-binding site based on highly conserved residues between mycobacterial orthologues, as well as structural homology to the SCPs. Based on the functional characterization of *MSMEG_5817* based on its crystal structure, it was determined that *MSMEG_5817* binds a range of phospholipids in a potentially different binding mechanism to the SCPs. This successfully demonstrates applications of structural biology as a means of gaining functional insight into the role of hypothetical proteins in TB.

7.4 Overcoming bottlenecks in TB Structural biology

As with previously described work in relation to Rv2971 and Rv3802c, only the crystal structures of their *M. smegmatis* orthologues (MSMEG_2407 and MSMEG_6394 respectively) were determined, where crystals of diffraction quality were not obtained for both Rv2971 and Rv3802c (Crellin *et al.*, 2010; Scoble *et al.*, 2010). Only through in depth crystallization screening and by reengineering the protein constructs utilized in crystallization experiments could the structures of Rv2971 and Rv3802c be determined. The same trend was observed for MSMEG_5817 and its orthologue Rv0807. While both proteins were successfully expressed and purified as soluble recombinant protein, only MSMEG_5817 produced crystals of high diffraction quality, despite extensive crystallization screening. This highlights a prevalent bottleneck in TB structural biology, where targets selected for structural characterization through bioinformatic and manual filters were unable to crystallize for x-ray crystallographic analysis, which is generally overcome by orthologue selection based on sequence identity. However as has been previously described, while the crystal structures of MSMEG_2407 and MSMEG_6394 provided valuable insight into functionality, they were unsuitable for the purposes of rational drug design due to differences in their respective enzyme active architecture with their *M. tuberculosis* orthologues.

An additional major bottleneck in TB structural biology is the inability to produce soluble recombinant protein of a number of *M. tuberculosis* targets selected for structural characterization. A means of overcoming this problem is by utilizing *M. smegmatis* as an expression system for the production of recombinant mycobacterial targets for structural characterization. Due to its saprophytic, fast growing nature, large-scale expression in *M. smegmatis* can be conducted in only 3-5 days, with the same safety prerequisites as with overexpression in *E. coli*. In addition, the use of the native bacterial species for the production of recombinant mycobacterial proteins is the ability to retain native structural fold and

functionality (Triccas *et al.*, 1998), including retention of natural substrates and ligands (Bergeret *et al.*, 2012), as well as post-translational modifications lost in using *E. coli* (Dobos *et al.*, 1995). While this is a powerful technique in overcoming the solubility bottleneck for mycobacterial proteins, it is highly underutilized in TB structural characterization, with only 8 unique mycobacterial structures deposited in the PDB produced in this manner. A more prevalent use of *M. smegmatis* as an expression system for mycobacterial targets for structural characterization would ensure a greater success rate for TB structural genomics projects.

As described in chapter 6, recombinant MSMEG_5817 was successfully expressed and purified in its native organism *M. smegmatis*, allowing for the determination of a higher quality crystal structure. A significantly higher recombinant protein expression yield was observed when compared to expression in *E. coli*, demonstrating a more efficient time and cost benefit in utilizing *M. smegmatis* as an expression system. MSMEG_5817 is a special case in the fact that soluble recombinant protein was produced in both *E. coli* and *M. smegmatis* expression systems, with recombinant protein from both expression systems purified via the same protocol, predominantly due to the highly polar surface of its protein fold. Yet despite the similarities, the crystal structure obtained using recombinant protein from *M. smegmatis* revealed additional structural features not visible in the originally determined structure, allowing for a more in depth structural representation of MSMEG_5817 and its orthologues.

7.5 Conclusions and Future Directions

Much of the past work performed in structural annotation of the *M. tuberculosis* proteome has focused on previously characterized protein targets, with a high degree structural data available on drug targets. While extensive progress has been made in characterizing the *M. tuberculosis* proteins through x-ray crystallography in the last 20 years, there is limited to no structural data available on up to 90% of the *M. tuberculosis* proteome. Through applications of x-ray crystallography in the characterization of *M. tuberculosis* targets, a more in depth understanding of the mycobacterium's pathogenesis and survival will be obtained, paving the way for anti-TB therapeutic development.

The work described throughout the course of this thesis has focused on three novel mycobacterial protein targets, with successfully determined crystal structures of each targets described. These cases represent a range of applications of x-ray crystallography in the structural characterization of novel drug targets, providing invaluable structural and functional insight. Bottlenecks representing major obstacles in TB structural genomics have hindered a more rapid pace in the structural characterization of the *M. tuberculosis* drug targets. For instance, while structural determination of orthologous proteins in alternate mycobacterial species provides valuable functional insight, they remain unsuitable for structural based drug inhibition characterization in *M. tuberculosis* targets, due to amino acid sequence and architectural variation. This has been overcome in the case of the *M. tuberculosis* targets Rv2971 and Rv3802c, whose crystal structures may now provide a basis for rational drug design.

The structural characterization of Rv2971, the essential aldo-keto reductase believed to be involved in removal of toxic metabolites, provides additional insight into the mechanisms of INH drug inhibition; the key first line anti-TB drug. INH has been found to inhibit a broad range of targets with varying

functionality in *M. tuberculosis*, and through the structural determination of these targets can we gauge the true complex mechanisms of INH inhibition.

The mechanisms used to identify INH drug targets can be applied to identifying additional drug targets of newly characterized anti-TB drug THL. As it stands, the most highly characterized drug target of THL is the essential lipase, Rv3802c, believed to play an essential role in the late stages of mycolic acid biosynthesis. We now have structural insight into the binding mechanism of THL against Rv3802c, and while further experimentation is required to confirm this, is believed to inhibit Rv3802c in a similar manner to the previously described structure of the human FAS II Thioesterase domain in complex with THL. Once additional drug targets of THL have been identified via techniques such as a THL affinity chromatography, we will gain a deeper understanding of the mechanism of THL inhibition against *M. tuberculosis*. Despite this, the successfully determined crystal structure of Rv3802c and THL may now provide a basis for future rational drug design.

One of the more powerful, yet underutilized, applications of x-ray crystallography is the ability to apply structural information on hypothetical proteins to gain insight into its biological function. In regards to the *M. tuberculosis* genome, approximately 27% of ORFs are annotated as hypothetical proteins. An in depth characterization its hypothetical proteins will allow us to truly gauge the novel mechanisms of *M. tuberculosis* pathogenesis.

One such example is Rv0807, whose *M. smegmatis* orthologue, MSMEG_5817, has recently been identified and found to be essential in macrophage host survivability via a presently unknown mechanism. The successfully determined crystal structure of MSMEG_5817 has revealed a potential role in the binding of biologically relevant lipids in facilitating mycobacterial survival. The hypothesis

was successfully tested experimentally, revealing binding to a range of phospholipids. The crystal structure of MSMEG_5817 also allowed for the successful determination of amino acid residues essential for structural and functional viability via additional experimentation. While further work is required to identify and characterise binding to its natural ligand, the successfully determined crystal structure of MSMEG_5817 is representative of this family of proteins within mycobacteria. If it were to be utilized as a drug target, once function has been characterized and if interaction with a suitable inhibitory compound is described, the crystal structure of Rv0807 would be required for drug development against *M. tuberculosis* to be effective.

The work described throughout the course of this thesis signifies significant progress in the structural characterization of novel *M. tuberculosis* proteins involved in essential pathways in mycobacterial growth and survival, utilizing a variety of applications of x-ray crystallography. While additional work is required to further characterise the drug binding mechanisms and functionality of these targets, the crystal structures described now allows for future drug binding characterization and rational drug design. In addition, significant obstacles in structural biology have been overcome utilizing a number of techniques, such as the use of *M. smegmatis* as an expression system for the production of mycobacterial proteins, which has applications for furthering the progress of TB structural genomics. Future experimentation expanding on the work described here will further our global understanding of *M. tuberculosis* pathogenesis: allowing for the development of novel therapeutics to combat drug resistance in this devastating human pathogen.

Chapter 8: References

- Alcaide, F., G. E. Pfyffer and A. Telenti (1997). "Role of embB in natural and acquired resistance to ethambutol in mycobacteria." Antimicrobial Agents and Chemotherapy **41**(10): 2270-2273.
- Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: A comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D: Biological Crystallography **66**(2): 213-221.
- Amin, A. G., R. Goude, L. Shi, J. Zhang, D. Chatterjee and T. Parish (2008). "EmbA is an essential arabinosyltransferase in *Mycobacterium tuberculosis*." Microbiology **154**(1): 240-248.
- Anes, E., P. Peyron, L. Staali, L. Jordao, M. G. Gutierrez, H. Kress, M. Hagedorn, I. Maridonneau-Parini, M. A. Skinner, A. G. Wildeman, S. A. Kalamidas, M. Kuehnel and G. Griffiths (2006). "Dynamic life and death interactions between *Mycobacterium smegmatis* and J774 macrophages." Cellular Microbiology **8**(6): 939-960.
- Argyrou, A., L. Jin, L. Siconilfi-Baez, R. H. Angeletti and J. S. Blanchard (2006). "Proteome-wide profiling of isoniazid targets in *Mycobacterium tuberculosis*." Biochemistry **45**(47): 13947-13953.
- Argyrou, A., M. W. Vetting, B. Aladegbami and J. S. Blanchard (2006). "*Mycobacterium tuberculosis* dihydrofolate reductase is a target for isoniazid." Nature Structural and Molecular Biology **13**(5): 408-413.
- Aslanidis, C., & de Jong, P. J. (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." Nucleic Acids Research **18**(20): 6069-6074.
- Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: Application to microtubules and the ribosome." Proceedings of the National Academy of Sciences of the United States of America **98**(18): 10037-10041.
- Bass Jr, J. B., L. S. Farer, P. C. Hopewell, R. O'Brien, R. F. Jacobs, F. Ruben, D. E. Snider Jr and G. Thornton (1994). "Treatment of tuberculosis and tuberculosis infection in adults and children." American Journal of Respiratory and Critical Care Medicine **149**(5): 1359-1374.
- Bergeret, F., S. Gavalda, C. Chalut, W. Malaga, A. Quémard, J. D. Pedelacq, M. Daffé, C. Guilhot, L. Mourey and C. Bon (2012). "Biochemical and structural study of the atypical acyltransferase domain from the mycobacterial polyketide synthase Pks13." Journal of Biological Chemistry **287**(40): 33675-33690.
- Besra, G. S., T. Sievert, R. E. Lee, R. A. Slayden, P. J. Brennan and K. Takayama (1994). "Identification of the apparent carrier in mycolic acid synthesis." Proceedings of the National Academy of Sciences of the United States of America **91**(26): 12735-12739.
- Bhatt, A., A. K. Brown, A. Singh, D. E. Minnikin and G. S. Besra (2008). "Loss of a Mycobacterial Gene Encoding a Reductase Leads to an Altered Cell Wall Containing β -oxo- Mycolic Acid Analogs

and Accumulation of Ketones." Chemistry and Biology **15**(9): 930-939.

Billman-Jacobe, H., M. J. McConville, R. E. Haites, S. Kovacevic and R. L. Coppel (1999). "Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of *Mycobacterium smegmatis*." Molecular Microbiology **33**(6): 1244-1253.

Blander, J. M. and R. Medzhitov (2004). "Regulation of Phagosome Maturation by Signals from Toll-Like Receptors." Science **304**(5673): 1014-1018.

Boshoff, H. I., V. Mizrahi and C. E. Barry Iii (2002). "Effects of pyrazinamide on fatty acid synthesis by whole mycobacterial cells and purified fatty acid synthase I." Journal of Bacteriology **184**(8): 2167-2172.

Brennan, P. J. (2003). "Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*." Tuberculosis **83**(1-3): 91-97.

Brennan, P. J. and D. C. Crick (2007). "The cell-wall core of *Mycobacterium tuberculosis* in the context of drug discovery." Current Topics in Medicinal Chemistry **7**(5): 475-488.

Brigl, M. and M. B. Brenner (2004). CD1: Antigen presentation and T cell function. Annual Review of Immunology. **22**: 817-890.

Brindley, D. N., S. Matsumura and K. Bloch (1969). "*Mycobacterium phlei* fatty acid synthetase - A bacterial multienzyme complex." Nature **224**(5220): 666-669.

Chan, J., X. Fan, S. W. Hunter, P. J. Brennan and B. R. Bloom (1991). "Lipoarabinomannan, a possible virulence factor involved in persistence of *Mycobacterium tuberculosis* within macrophages." Infection and Immunity **59**(5): 1755-1761.

Chan, J., T. Fujiwara, P. Brennan, M. McNeil, S. J. Turco, J. C. Sibille, M. Snapper, P. Aisen and B. R. Bloom (1989). "Microbial glycolipids: possible virulence factors that scavenge oxygen radicals." Proceedings of the National Academy of Sciences of the United States of America **86**(7): 2453-2457.

Chatterjee, D. and K. H. Khoo (1998). "Mycobacterial lipoarabinomannan: An extraordinary lipoheteroglycan with profound physiological effects." Glycobiology **8**(2): 113-120.

Chen, V. B., W. B. Arendall Iii, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: All-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D: Biological Crystallography **66**(1): 12-21.

Chim, N., J. E. Habel, J. M. Johnston, I. Krieger, L. Miallau, R. Sankaranarayanan, R. P. Morse, J. Bruning, S. Swanson, H. Kim, C. Y. Kim, H. Li, E. M. Bulloch, R. J. Payne, A. Manos-Turvey, L. W. Hung, E. N. Baker, J. S. Lott, M. N. G. James, T. C. Terwilliger, D. S. Eisenberg, J. C. Sacchettini and C. W. Goulding (2011). "The TB Structural Genomics Consortium: A decade of progress." Tuberculosis **91**(2): 155-172.

Choi, K. H., L. Kremer, G. S. Besra and C. O. Rock (2000). "Identification and substrate specificity of β -Ketoacyl (Acyl carrier protein) synthase III (mtFabH) from *Mycobacterium tuberculosis*." Journal of Biological Chemistry **275**(36): 28201-28207.

- Clemens, D. L. and M. A. Horwitz (1995). "Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited." Journal of Experimental Medicine **181**(1): 257-270.
- Clemens, D. L., B. Y. Lee and M. A. Horwitz (2000). "Deviant expression of Rab5 on phagosomes containing the intracellular pathogens *Mycobacterium tuberculosis* and *Legionella pneumophila* is associated with altered phagosomal fate." Infection and Immunity **68**(5): 2671-2684.
- Clemens, D. L., B. Y. Lee and M. A. Horwitz (2000). "*Mycobacterium tuberculosis* and *Legionella pneumophila* phagosomes exhibit arrested maturation despite acquisition of Rab7." Infection and Immunity **68**(9): 5154-5166.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry Iii, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell (1998). "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence." Nature **393**(6685): 537-544.
- Corbett, E. L., C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione and C. Dye (2003). "The growing burden of tuberculosis: Global trends and interactions with the HIV epidemic." Archives of Internal Medicine **163**(9): 1009-1021.
- Crellin, P. K., J. P. Vivian, J. Scoble, F. M. Chow, N. P. West, R. Brammananth, N. I. Proellocks, A. Shahine, J. Le Nours, M. C. J. Wilce, W. J. Britton, R. L. Coppel, J. Rossjohn and T. Beddoe (2010). "Tetrahydrolipstatin inhibition, functional analyses, and three-dimensional structure of a lipase essential for mycobacterial viability." Journal of Biological Chemistry **285**(39): 30050-30060.
- Crick, D. C., S. Mahapatra and P. J. Brennan (2001). "Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*." Glycobiology **11**(9): 107R-118R.
- Dedieu, L., C. Serveau-Avesque and S. Canaan (2013). "Identification of Residues Involved in Substrate Specificity and Cytotoxicity of Two Closely Related Cutinases from *Mycobacterium tuberculosis*." PLoS ONE **8**(7).
- Depietro, P. J., E. S. Julfayev and W. A. McLaughlin (2013). "Quantification of the impact of PSI: Biology according to the annotations of the determined structures." BMC Structural Biology **13**(1).
- Deretic, V., S. Singh, S. Master, J. Harris, E. Roberts, G. Kyei, A. Davis, S. de Haro, J. Naylor, H. H. Lee and I. Vergne (2006). "*Mycobacterium tuberculosis* inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism." Cellular Microbiology **8**(5): 719-727.
- Desjardins, M. (2009). "The Good Fat: A Link between Lipid Bodies and Antigen Cross-presentation." Immunity **31**(2): 176-178.
- Desjardins, M., M. Houde and E. Gagnon (2005). "Phagocytosis: The convoluted way from nutrition to

- adaptive immunity." Immunological Reviews **207**: 158-165.
- Dessen, A., A. Quemard, J. S. Blanchard, W. R. Jacobs Jr and J. C. Sacchettin (1995). "Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*." Science **267**(5204): 1638-1641.
- Devedjiev, Y., Z. Dauter, S. R. Kuznetsov, T. L. Z. Jones and Z. S. Derewenda (2000). "Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å." Structure **8**(11): 1137-1146.
- Dias, M. V. B., I. B. Vasconcelos, A. M. X. Prado, V. Fadel, L. A. Basso, W. F. de Azevedo Jr and D. S. Santos (2007). "Crystallographic studies on the binding of isonicotinyl-NAD adduct to wild-type and isoniazid resistant 2-trans-enoyl-ACP (CoA) reductase from *Mycobacterium tuberculosis*." Journal of Structural Biology **159**(3): 369-380.
- Dobos, K. M., K. Swiderek, K. H. Khoo, P. J. Brennan and J. T. Belisle (1995). "Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*." Infection and Immunity **63**(8): 2846-2853.
- Doerks, T., V. van Noort, P. Minguéz and P. Bork (2012). "Annotation of the m. tuberculosis hypothetical orfeome: Adding functional information to more than half of the uncharacterized proteins." PLoS ONE **7**(4).
- Dundas, J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang (2006). "CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues." Nucleic Acids Research **34**(WEB. SERV. ISS.): W116-W118.
- El-Benna, J., P. M. C. Dang and M. A. Gougerot-Pocidaló (2008). "Priming of the neutrophil NADPH oxidase activation: Role of p47phox phosphorylation and NOX2 mobilization to the plasma membrane." Seminars in Immunopathology **30**(3): 279-289.
- Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallographica Section D: Biological Crystallography **66**(4): 486-501.
- Escuyer, V. E., M. A. Lety, J. B. Torrelles, K. H. Khoo, J. B. Tang, C. D. Rithner, C. Frehel, M. R. McNeil, P. Brennan, J. and Chatterjee (2001). "The Role of the embA and embB Gene Products in the Biosynthesis of the Terminal Hexaarabinofuranosyl Motif of *Mycobacterium smegmatis* Arabinogalactan." Journal of Biological Chemistry **276**(52): 48854-48862.
- Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallographica Section D: Biological Crystallography **62**(1): 72-82.
- Fratti, R. A., J. M. Backer, J. Gruenberg, S. Corvera and V. Deretic (2001). "Role of phosphatidylinositol 3-kinase and Rab5 effectors in phagosomal biogenesis and mycobacterial phagosome maturation arrest." Journal of Cell Biology **154**(3): 631-644.
- Fratti, R. A., J. Chua, I. Vergne and V. Deretic (2003). "*Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest." Proceedings of the National Academy of Sciences of the United States of America **100**(9): 5437-5442.
- Gagneux, S. (2009). "Fitness cost of drug resistance in *Mycobacterium tuberculosis*." Clinical

Microbiology and Infection **15**(SUPPL. 1): 66-68.

Goude, A. G. A., D. Chatterjee, T. Parish (2009). "The EmbC arabinosyltransferase is inhibited by ethambutol in *Mycobacterium tuberculosis*." Unpublished.

Grant, A. W., G. Steel, H. Waugh and E. M. Ellis (2003). "A novel aldo-keto reductase from *Escherichia coli* can increase resistance to methylglyoxal toxicity." FEMS Microbiology Letters **218**(1): 93-99.

Griffin, J. E., J. D. Gawronski, M. A. DeJesus, T. R. Ioerger, B. J. Akerley and C. M. Sassetti (2011). "High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism." PLoS Pathogens **7**(9).

Grimshaw, C. E. (1992). "Aldose reductase: Model for a new paradigm of enzymic perfection in detoxification catalysts." Biochemistry **31**(42): 10139-10145.

Grosset, J. (2003). "*Mycobacterium tuberculosis* in the extracellular compartment: An underestimated adversary." Antimicrobial Agents and Chemotherapy **47**(3): 833-836.

Guiguemde, W. A., A. A. Shelat, J. F. Garcia-Bustos, T. T. Diagana, F. J. Gamo and R. K. Guy (2012). "Global phenotypic screening for antimalarials." Chemistry and Biology **19**(1): 116-129.

Hackam, D. J., O. D. Rotstein, W. J. Zhang, N. Demaurex, M. Woodside, O. Tsai and S. Grinstein (1997). "Regulation of phagosomal acidification. Differential targeting of Na⁺/H⁺ exchangers, Na⁺/K⁺-ATPases, and vacuolar-type H⁺-ATPases." Journal of Biological Chemistry **272**(47): 29810-29820.

Herrmann, J. L. and P. H. Lagrange (2005). "Dendritic cells and *Mycobacterium tuberculosis*: Which is the Trojan horse?" Pathologie Biologie **53**(1): 35-40.

Hershkovitz, I., H. D. Donoghue, D. E. Minnikin, G. S. Besra, O. Y. C. Lee, A. M. Gernaey, E. Galili, V. Eshed, C. L. Greenblatt, E. Lemma, G. K. Bar-Gal and M. Spigelman (2008). "Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the Eastern mediterranean." PLoS ONE **3**(10).

Hmama, Z., K. Sendide, A. Talal, R. Garcia, K. Dobos and N. E. Reiner (2004). "Quantitative analysis of phagolysosome fusion in intact cells: Inhibition by mycobacterial lipoarabinomannan and rescue by an α , 25-dihydroxyvitamin D3-phosphoinositide 3-kinase pathway." Journal of Cell Science **117**(10): 2131-2139.

Holmquist, M., M. Martinelle, P. Berglund, I. G. Clausen, S. Patkar, A. Svendsen and K. Hult (1993). "Lipases from *Rhizomucor miehei* and *Humicola lanuginosa*: Modification of the lid covering the active site alters enantioselectivity." Journal of Protein Chemistry **12**(6): 749-757.

Holmquist, M., I. G. Clausen, S. Patkar, A. Svendsen and K. Hult (1995). "Probing a functional role of Glu87 and Trp89 in the lid of *Humicola lanuginosa* lipase through transesterification reactions in organic solvent." Journal of Protein Chemistry **14**(4): 217-224.

Holmquist, M. (2000). "Alpha/Beta-Hydrolase Fold Enzymes: Structures, Functions and Mechanisms." Current Protein and Peptide Science **1**(2): 209-235.

- Hotelier, T., L. Renault, X. Cousin, V. Negre, P. Marchot and A. Chatonnet (2004). "ESTHER, the database of the α/β -hydrolase fold superfamily of proteins." Nucleic Acids Research **32**(DATABASE ISS.):D145-D147.
- Houben, E. N. G., L. Nguyen and J. Pieters (2006). "Interaction of pathogenic mycobacteria with the host immune system." Current Opinion in Microbiology **9**(1): 76-85.
- Jackson, M., C. Raynaud, M. A. Lan elle, C. Guillhot, C. Laurent-Winter, D. Ensergueix, B. Gicquel and M. Daff  (1999). "Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope." Molecular Microbiology **31**(5): 1573-1587.
- Kang, P. B., A. K. Azad, J. B. Torrelles, T. M. Kaufman, A. Beharka, E. Tibesar, L. E. DesJardin and L. S. Schlesinger (2005). "The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis." Journal of Experimental Medicine **202**(7): 987-999.
- Keiser, K. J., Rubin, E. J (2014). "How sisters grow apart: mycobacterial growth and division." Nature Reviews Microbiology **12**: 550-562.
- Kaur, D., M. E. Guerin, H. Skovierov , P. J. Brennan and M. Jackson (2009). Chapter 2: Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*. **69**: 23-78.
- Kelley, V. A. and J. S. Schorey (2003). "Mycobacterium's arrest of phagosome maturation in macrophages requires Rab5 activity and accessibility to iron." Molecular Biology of the Cell **14**(8): 3366-3377.
- Koglin, A., F. L hr, F. Bernhard, V. V. Rogov, D. P. Frueh, E. R. Strieter, M. R. Mofid, P. G ntert, G. Wagner, C. T. Walsh, M. A. Marahiel and V. D tsch (2008). "Structural basis for the selectivity of the external thioesterase of the surfactin synthetase." Nature **454**(7206): 907-911.
- Konno, K., F. M. Feldmann and W. McDermott (1967). "Pyrazinamide susceptibility and amidase activity of tubercle bacilli." American Review of Respiratory Disease **95**(3): 461-469.
- Koul, A., T. Herget, B. Klebl and A. Ullrich (2004). "Interplay between mycobacteria and host signalling pathways." Nature Reviews Microbiology **2**(3): 189-202.
- Kovacevic, S., D. Anderson, Y. S. Morita, J. Patterson, R. Haites, B. N. I. McMillan, R. Coppel, M. J. McConville and H. Billman-Jacobe (2006). "Identification of a novel protein with a role in lipoarabinomannan biosynthesis in mycobacteria." Journal of Biological Chemistry **281**(14): 9011-9017.
- Kremer, L., C. De Chastellier, G. Dobson, K. J. C. Gibson, P. Bifani, S. Balor, J. P. Gorvel, C. Locht, D. E. Minnikin and G. S. Besra (2005). "Identification and structural characterization of an unusual mycobacterial monomeromycolyl-diacylglycerol." Molecular Microbiology **57**(4): 1113-1126.
- Kremer, L., K. M. Nampoothiri, S. Lesjean, L. G. Dover, S. Graham, J. Betts, P. J. Brennan, D. E. Minnikin, C. Locht and G. S. Besra (2001). "Biochemical Characterization of Acyl Carrier Protein (AcpM) and Malonyl-CoA:AcpM Transacylase (mtFabD), Two Major Components of *Mycobacterium tuberculosis* Fatty Acid Synthase II." Journal of Biological Chemistry **276**(30): 27967-27974.
- Kridel, S. J., F. Axelrod, N. Rozenkrantz and J. W. Smith (2004). "Orlistat Is a Novel Inhibitor of Fatty

Acid Synthase with Antitumor Activity." Cancer Research **64**(6): 2070-2075.

Kuehnel, M. P., R. Goethe, A. Habermann, E. Mueller, M. Rohde, G. Griffiths and P. Valentin-Weigand (2001). "Characterization of the intracellular survival of *Mycobacterium avium ssp. paratuberculosis*: Phagosomal pH and fusogenicity in J774 macrophages compared with other mycobacteria." Cellular Microbiology **3**(8): 551-566.

Lea-Smith, D. J., J. S. Pyke, D. Tull, M. J. McConville, R. L. Coppel and P. K. Crellin (2007). "The reductase that catalyzes mycolic motif synthesis is required for efficient attachment of mycolic acids to arabinogalactan." Journal of Biological Chemistry **282**(15): 11000-11008.

Lee, R. E., K. Mikusova, P. J. Brennan and G. S. Besra (1995). "Synthesis of the mycobacterial arabinose donor β -D-arabinofuranosyl-1-monophosphoryldecaprenol, development of a basic arabinosyl-transferase assay, and identification of ethambutol as an arabinosyl transferase inhibitor." Journal of the American Chemical Society **117**(48): 11829-11832.

Leslie, A. W. and H. Powell (2007). Processing diffraction data with mosflm. Evolving Methods for Macromolecular Crystallography. R. Read and J. Sussman, Springer Netherlands. **245**: 41-51.

Linne, U., D. Schwarzer, G. N. Schroeder and M. A. Marahiel (2004). "Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis." European Journal of Biochemistry **271**(8): 1536-1545.

Lipinski, C. A., F. Lombardo, B. W. Dominy and P. J. Feeney (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Advanced Drug Delivery Reviews **46**(1-3): 3-26.

Longhi, S. and C. Cambillau (1999). "Structure-activity of cutinase, a small lipolytic enzyme." Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids **1441**(2-3): 185-196.

Lüthi-Peng, Q., H. P. Märki and P. Hadváry (1992). "Identification of the active-site serine in human pancreatic lipase by chemical modification with tetrahydrolipstatin." FEBS Letters **299**(1): 111-115.

Mahapatra, S., D. C. Crick, M. R. McNeil and P. J. Brennan (2008). "Unique structural features of the peptidoglycan of *Mycobacterium leprae*." Journal of Bacteriology **190**(2): 655-661.

Malik, Z. A., G. M. Denning and D. J. Kusner (2000). "Inhibition of Ca^{2+} signaling by *Mycobacterium tuberculosis* is associated with reduced phagosome-lysosome fusion and increased survival within human macrophages." Journal of Experimental Medicine **191**(2): 287-302.

Malik, Z. A., S. S. Iyer and D. J. Kusner (2001). "*Mycobacterium tuberculosis* phagosomes exhibit altered calmodulin-dependent signal transduction: Contribution to inhibition of phagosome-lysosome fusion and intracellular survival in human macrophages." Journal of Immunology **166**(5): 3392-3401.

Marrakchi, H., K. H. Choi and C. O. Rock (2002). "A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*." Journal of Biological Chemistry **277**(47): 44809-44816.

Marrakchi, H., S. Ducasse, G. Labesse, H. Montrozier, E. Margeat, L. Emorine, X. Charpentier, M. Daffé

- and A. Quémard (2002). "MabA (FabG1), a *Mycobacterium tuberculosis* protein involved in the long-chain fatty acid elongation system FAS-II." Microbiology **148**(4): 951-960.
- Martinez, C., A. Nicolas, H. Van Tilbeurgh, M. P. Egloff, C. Cudrey, R. Verger and C. Cambillau (1994). "Cutinase, a lipolytic enzyme with a preformed oxyanion hole." Biochemistry **33**(1): 83-89.
- Meniche, X., C. Labarre, C. De Sousa-D'Auria, E. Huc, F. Laval, M. Tropis, N. Bayan, D. Portevin, C. Guilhot, M. Daffé and C. Houssin (2009). "Identification of a stress-induced factor of *Corynebacterineae* that is involved in the regulation of the outer membrane lipid composition." Journal of Bacteriology **191**(23):7323-7332.
- McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**(4): 658-674.
- Mikuová, K., T. Yagi, R. Stern, M. R. McNeil, G. S. Besra, D. C. Crick and P. J. Brennan (2000). "Biosynthesis of the galactan component of the mycobacterial cell wall." Journal of Biological Chemistry **275**(43): 33890-33897.
- Miltner, E., K. Daroogheh, P. K. Mehta, S. L. G. Cirillo, J. D. Cirillo and L. E. Bermudez (2005). "Identification of *Mycobacterium avium* genes that affect invasion of the intestinal epithelium." Infection and Immunity **73**(7): 4214-4221.
- Mohan, S., T. M. Kelly, S. S. Eveland, C. R. H. Raetz and M. S. Anderson (1994). "An *Escherichia coli* Gene (FabZ) Encoding (3R)-Hydroxymyristoyl Acyl Carrier Protein Dehydrase: Relation to fabA and Suppression of Mutations in Lipid a Biosynthesis." Journal of Biological Chemistry **269**(52): 32896-32903.
- Molek, P., B. Strukelj and T. Bratkovic (2011). "Peptide phage display as a tool for drug discovery: Targeting membrane receptors." Molecules **16**(1): 857-887.
- Niesen, F., H. Berglund, H. Vedadi, M (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nat Protoc. **2**: 2212–2221.
- Parish, T., D. A. Smith, G. Roberts, J. Betts and N. G. Stoker (2003). "The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence." Microbiology **149**(6):1423-1435
- Parker, S. K., R. M. Barkley, J. G. Rino and M. L. Vasil (2009). "*Mycobacterium tuberculosis* Rv3802c encodes a phospholipase/thioesterase and is inhibited by the antimycobacterial agent tetrahydrolipstatin." PLoS ONE **4**(1).
- Parker, S. K., K. M. Curtin and M. L. Vasil (2007). "Purification and characterization of mycobacterial phospholipase A: An activity associated with mycobacterial cutinase." Journal of Bacteriology **189**(11): 4153-4160.
- Patterson, J. H., M. J. McConville, R. E. Haites, R. L. Coppel and H. Billman-Jacobe (2000). "Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis." Journal of Biological Chemistry **275**(32): 24900-24906.
- Pavlidou, M., K. Hänel, L. Möckel and D. Willbold (2013). "Nanodiscs Allow Phage Display Selection for

- Ligands to Non-Linear Epitopes on Membrane Proteins." PLoS ONE **8**(9).
- Payne, D. J., M. N. Gwynn, D. J. Holmes and D. L. Pompliano (2007). "Drugs for bad bugs: Confronting the challenges of antibacterial discovery." Nature Reviews Drug Discovery **6**(1): 29-40.
- Pelosi, A., D. Smith, R. Brammananth, A. Topolska, H. Billman-Jacobe, P. Nagley, P. K. Crellin and R. L. Coppel (2012). "Identification of a novel gene product that promotes survival of *Mycobacterium smegmatis* in macrophages." PLoS ONE **7**(2).
- Pemble Iv, C. W., L. C. Johnson, S. J. Kridel and W. T. Lowther (2007). "Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat." Nature Structural and Molecular Biology **14**(8): 704-709.
- Penning, T. M. and J. E. Drury (2007). "Human aldo-keto reductases: Function, gene regulation, and single nucleotide polymorphisms." Archives of Biochemistry and Biophysics **464**(2): 241-250.
- Pereira-Leal, J. B. and M. C. Seabra (2001). "Evolution of the Rab family of small GTP-binding proteins." Journal of Molecular Biology **313**(4): 889-901.
- Pérez, E., J. A. Gavigan, I. Otal, C. Guilhot, V. Pelicic, B. Giquel and C. Martín (1998). "Tn611 transposon mutagenesis in *Mycobacterium smegmatis* using a temperature-sensitive delivery system." Methods in molecular biology (Clifton, N.J.) **101**: 187-198.
- Peters, C. and A. Mayer (1998). "Ca²⁺/calmodulin signals the completion of docking and triggers a late step of vacuole fusion." Nature **396**(6711): 575-580.
- Pfeffer, S. R. (2005). "Structural clues to rab GTPase functional diversity." Journal of Biological Chemistry **280**(16): 15485-15488.
- Piddington, D. L., F. C. Fang, T. Laessig, A. M. Cooper, I. M. Orme and N. A. Buchmeier (2001). "Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst." Infection and Immunity **69**(8): 4980-4987.
- Portevin, D., C. De Sousa-D'Auria, C. Houssin, C. Grimaldi, M. Chami, M. Daffé and C. Guilhot (2004). "A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms." Proceedings of the National Academy of Sciences of the United States of America **101**(1):314-319.
- Quémard, A., J. C. Sacchettin, A. Dessen, C. Vilcheze, R. Bittman, W. R. Jacobs Jr and J. S. Blanchard (1995). "Enzymatic Characterization of the Target for Isoniazid in *Mycobacterium tuberculosis*." Biochemistry **34**(26): 8235-8241.
- Rallis, E., V. Paparizos, K. Kyriakis and A. Katsambas (2009). "Treatment of epidermodysplasia verruciformis in human immunodeficiency virus-positive patients." Journal of the European Academy of Dermatology and Venereology **23**(2): 195-196.
- Raviglione, M. C. and I. M. Smith (2007). "XDR tuberculosis - Implications for Global Public Health." New England Journal of Medicine **356**(7): 656-659.

- Rink, J., E. Ghigo, Y. Kalaidzidis and M. Zerial (2005). "Rab conversion as a mechanism of progression from early to late endosomes." Cell **122**(5): 735-749.
- Rojas, M., L. F. Garcia, J. Nigou, G. Puzo and M. Olivier (2000). "Mannosylated lipoarabinomannan antagonizes *Mycobacterium tuberculosis*- induced macrophage apoptosis by altering CA^{2+} -dependent cell signaling." Journal of Infectious Diseases **182**(1): 240-251.
- Rossmann, M. D., B. T. Maida and S. D. Douglas (1990). "Monocyte-derived macrophage and alveolar macrophage fibronectin production and cathepsin D activity." Cellular Immunology **126**(2): 268-277.
- Rouse, D. A., J. A. DeVito, Z. Li, H. Byer and S. L. Morris (1996). "Site-directed mutagenesis of the katC gene of *Mycobacterium tuberculosis*: Effects on catalase-peroxidase activities and isoniazid resistance." Molecular Microbiology **22**(3): 583-592.
- Rozwarski, D. A., G. A. Grant, D. H. R. Barton, W. R. Jacobs Jr and J. C. Sacchettini (1998). "Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*." Science **279**(5347): 98-102.
- Russell, D. G. (2001). "*Mycobacterium tuberculosis*: Here today, and here tomorrow." Nature Reviews Molecular Cell Biology **2**(8): 569-577.
- Russell, D. G. (2011). "*Mycobacterium tuberculosis* and the intimate discourse of a chronic infection." Immunological Reviews **240**(1): 252-268.
- Saleh, M. T. and J. T. Belisle (2000). "Secretion of an acid phosphatase (SapM) by *Mycobacterium tuberculosis* that is similar to eukaryotic acid phosphatases." Journal of Bacteriology **182**(23): 6850-6853.
- Sassetti, C. M., D. H. Boyd and E. J. Rubin (2003). "Genes required for mycobacterial growth defined by high density mutagenesis." Molecular Microbiology **48**(1): 77-84.
- Schaeffer, M. L., G. Agnihotri, C. Volker, H. Kallender, P. J. Brennan and J. T. Lonsdale (2001). "Purification and Biochemical Characterization of the *Mycobacterium tuberculosis* β -Ketoacyl-acyl Carrier Protein Synthases KasA and KasB." Journal of Biological Chemistry **276**(50): 47029-47037.
- Schaloske, R. H. and E. A. Dennis (2006). "The phospholipase A2 superfamily and its group numbering system." Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids **1761**(11): 1246-1259.
- Scoble, J., A. D. McAlister, Z. Fulton, S. Troy, E. Byres, J. P. Vivian, R. Brammananth, M. C. J. Wilce, J. Le Nours, L. Zaker-Tabrizi, R. L. Coppel, P. K. Crellin, J. Rossjohn and T. Beddoe (2010). "Crystal structure and comparative functional analyses of a *Mycobacterium* Aldo-Keto reductase." Journal of Molecular Biology **398**(1): 26-39.
- Scorpio, A. and Y. Zhang (1996). "Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus." Nature Medicine **2**(6):662-667.
- Seidel, M., L. J. Alderwick, H. L. Birch, H. Sahm, L. Eggeling and G. S. Besra (2007). "Identification of a novel arabinofuranosyltransferase AftB involved in a terminal step of cell wall arabinan biosynthesis in

- Corynebacterianae*, such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*." Journal of Biological Chemistry **282**(20): 14729-14740.
- Schrag, J. D. and M. Cygler (1997). Lipases and α/β hydrolase fold. Methods in Enzymology. **284**: 85-107.
- Smith, S., A. Witkowski and A. K. Joshi (2003). "Structural and functional organization of the animal fatty acid synthase." Progress in Lipid Research **42**(4): 289-317.
- Spangenberg, T., J. N. Burrows, P. Kowalczyk, S. McDonald, T. N. C. Wells and P. Willis (2013). "The Open Access Malaria Box: A Drug Discovery Catalyst for Neglected Diseases." PLoS ONE **8**(6).
- Stoeckle, M. Y., L. Guan, N. Riegler, I. Weitzman, B. Kreiswirth, J. Kornblum, F. Laraque and L. W. Riley (1993). "Catalase-peroxidase gene sequences in isoniazid-sensitive and -resistant strains of *Mycobacterium tuberculosis* from New York City." Journal of Infectious Diseases **168**(4): 1063-1065.
- Stuart, L. M., J. Boulais, G. M. Charriere, E. J. Hennessy, S. Brunet, I. Jutras, G. Goyette, C. Rondeau, S. Letarte, H. Huang, P. Ye, F. Morales, C. Kocks, J. S. Bader, M. Desjardins and R. A. B. Ezekowitz (2007). "A systems biology analysis of the *Drosophila* phagosome." Nature **445**(7123): 95-101.
- Sturgill-Koszycki, S., P. H. Schlesinger, P. Chakraborty, P. L. Haddix, H. L. Collins, A. K. Fok, R. D. Allen, S. L. Gluck, J. Heuser and D. G. Russell (1994). "Lack of acidification in *Mycobacterium* phagosomes produced by exclusion vesicular proton-ATPase." Science **263**(5147): 678-681.
- Sumimoto, H. (2008). "Structure, regulation and evolution of Nox-family NADPH oxidases that produce reactive oxygen species." FEBS Journal **275**(13): 3249-3277.
- Takayama, K., C. Wang and G. S. Besra (2005). "Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*." Clinical Microbiology Reviews **18**(1): 81-101.
- Terwilliger, T. C., D. Stuart and S. Yokoyama (2009). Lessons from structural genomics. Annual Review of Biophysics. **38**: 371-383.
- Timmins, G. S. and V. Deretic (2006). "Mechanisms of action of isoniazid." Molecular Microbiology **62**(5):1220-1227.
- Triccas, J. A., T. Parish, W. J. Britton and B. Gicquel (1998). "An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*." FEMS Microbiology Letters **167**(2): 151-156.
- Trost, M., L. English, S. Lemieux, M. Courcelles, M. Desjardins and P. Thibault (2009). "The Phagosomal Proteome in Interferon- γ -Activated Macrophages." Immunity **30**(1): 143-154.
- Vergne, I., J. Chua and V. Deretic (2003). "Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca^{2+} /calmodulin-PI3K hVPS34 cascade." Journal of Experimental Medicine **198**(4): 653-659.
- Vergne, I., J. Chua, H. H. Lee, M. Lucas, J. Belisle and V. Deretic (2005). "Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*." Proceedings of the National Academy of Sciences of the United States of America **102**(11): 4033-4038.

- Vergne, I., J. Chua, S. B. Singh and V. Deretic (2004). "Cell biology of *Mycobacterium tuberculosis* phagosome." Annual Review of Cell and Developmental Biology **20**: 367-394.
- Vieira, O. V., R. J. Botelho and S. Grinstein (2002). "Phagosome maturation: Aging gracefully." Biochemical Journal **366**(3): 689-704.
- Vissa, V. D. and P. J. Brennan (2001). "The genome of *Mycobacterium leprae*: A minimal mycobacterial gene set." Genome Biology **2**(8): 1023.1021-1023.1028.
- Wang, F., P. Jain, G. Gulten, Z. Liu, Y. Feng, K. Ganesula, A. S. Motiwala, T. R. Ioerger, D. Alland, C. Vilch ze, W. R. Jacobs Jr and J. C. Sacchettini (2010). "*Mycobacterium tuberculosis* dihydrofolate reductase is not a target relevant to the antitubercular activity of isoniazid." Antimicrobial Agents and Chemotherapy **54**(9): 3776-3782.
- West, N. P., F. M. Chow, E. J. Randall, J. Wu, J. Chen, J. M. Ribeiro and W. J. Britton (2009). "Cutinase-like proteins of *Mycobacterium tuberculosis*: characterization of their variable enzymatic functions and active site identification." The FASEB journal : official publication of the Federation of American Societies for Experimental Biology **23**(6): 1694-1704.
- West, N. P., T. M. Wozniak, J. Valenzuela, C. G. Feng, A. Sher, J. M. C. Ribeiro and W. J. Britton (2008). "Immunological diversity within a family of cutinase-like proteins of *Mycobacterium tuberculosis*." Vaccine **26**(31): 3853-3859.
- West, N. P. C., K. M., M. Xue, E. J. Randall, W. J. Britton and R. K. Payne (2011). "Inhibitors of an essential mycobacterial cell wall lipase (Rv3802c) as tuberculosis drug leads." Chemical Communications **47**:5166-5168
- WHO (2015). Global Tuberculosis Report 2014. Geneva, World Health Organization.
- Winkler, U. K. and M. Stuckmann (1979). "Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*." Journal of Bacteriology **138**(3): 663-670.
- Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallographica Section D: Biological Crystallography **67**(4): 235-242.
- Wong, D., H. Bach, J. Sun, Z. Hmama and Y. Av-Gay (2011). "*Mycobacterium tuberculosis* protein tyrosine phosphatase (PtpA) excludes host vacuolar-H⁺-ATPase to inhibit phagosome acidification." Proceedings of the National Academy of Sciences of the United States of America **108**(48): 19371-19376.
- Xu, G., B. Liu, F. Wang, C. Wei, Y. Zhang, J. Sheng, G. Wang and F. Li (2013). "High-throughput screen of essential gene modules in *Mycobacterium tuberculosis*: A bibliometric approach." BMC Infectious Diseases **13**(1).
- Yang, P. Y., M. Wang, K. Liu, M. H. Ngai, O. Sheriff, M. J. Lear, S. K. Sze, C. Y. He and S. Q. Yao (2012). "Parasite-based screening and proteome profiling reveal orlistat, an FDA-approved drug, as a

potential anti-*Trypanosoma brucei* agent." Chemistry - A European Journal **18**(27): 8403-8413.

Yates, R. M., A. Hermetter, G. A. Taylor and D. G. Russell (2007). "Macrophage activation downregulates the degradative capacity of the phagosome." Traffic **8**(3): 241-250.

Zimhony, O., J. S. Cox, J. T. Welch, C. Vilchèze and W. R. Jacobs Jr (2000). "Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of *Mycobacterium tuberculosis*." Nature Medicine **6**(9): 1043-1047.

Chapter 9: Appendix

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 39, PP. 30050–30060, SEPTEMBER 24, 2010
 © 2010 BY THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC. PRINTED IN THE U.S.A.

Tetrahydrolipstatin Inhibition, Functional Analyses, and Three-dimensional Structure of a Lipase Essential for Mycobacterial Viability*

Received for publication, May 31, 2010, and in revised form, July 21, 2010. Published, JBC Papers in Press, July 23, 2010, DOI 10.1074/jbc.M110.150094

Paul K. Crellin^{†§1,2}, Julian P. Vivian^{§1}, Judith Scoble^{†¶}, Frances M. Chow^{||}, Nicholas P. West^{||**}, Rajini Brammananth^{†§}, Nicholas I. Proellocks[‡], Adam Shahine[¶], Jerome Le Nours^{†¶}, Matthew C. J. Wilce[¶], Warwick J. Britton^{||**}, Ross L. Coppel^{†§}, Jamie Rossjohn^{†¶3,4}, and Travis Beddoe^{†¶4,5}

From the [†]Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics and the Departments of [§]Microbiology and [¶]Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, the ^{||}Mycobacterial Research Program, Centenary Institute, Locked Bag 6, Newtown, New South Wales 2042, and the ^{**}Sydney Medical School, University of Sydney, Sydney, New South Wales 2006, Australia

The highly complex and unique mycobacterial cell wall is critical to the survival of *Mycobacteria* in host cells. However, the biosynthetic pathways responsible for its synthesis are, in general, incompletely characterized. Rv3802c from *Mycobacterium tuberculosis* is a partially characterized phospholipase/thioesterase encoded within a genetic cluster dedicated to the synthesis of core structures of the mycobacterial cell wall, including mycolic acids and arabinogalactan. Enzymatic assays performed with purified recombinant proteins Rv3802c and its close homologs from *Mycobacterium smegmatis* (MSMEG_6394) and *Corynebacterium glutamicum* (NCgl2775) show that they all have significant lipase activities that are inhibited by tetrahydrolipstatin, an anti-obesity drug that coincidentally inhibits mycobacterial cell wall biosynthesis. The crystal structure of MSMEG_6394, solved to 2.9 Å resolution, revealed an α/β hydrolase fold and a catalytic triad typically present in esterases and lipases. Furthermore, we demonstrate direct evidence of gene essentiality in *M. smegmatis* and show the structural consequences of loss of MSMEG_6394 function on the cellular integrity of the organism. These findings, combined with the predicted essentiality of Rv3802c in *M. tuberculosis*, indicate that the Rv3802c family performs a fundamental and indispensable lipase-associated function in mycobacteria.

The genus *Mycobacterium* contains a number of medically significant species, most notably the devastating human pathogen *Mycobacterium tuberculosis* that causes around 2 million deaths each year, the most by any single infectious agent. Despite the availability of a vaccine, the number of infected individuals worldwide continues to increase, as does the prev-

alence of drug-resistant forms of *M. tuberculosis* (1). A key virulence factor is the unique mycobacterial cell wall that consists of a core structure as follows: peptidoglycan covalently linked to arabinogalactan esterified with mycolic acids to form the mycolyl-arabinogalactan-peptidoglycan or “mAGP” complex and a series of free glycolipids, including trehalose monomycolates, trehalose dimycolates, phosphatidylinositol mannosides, and lipoarabinomannans (2), that facilitate vital interactions with host cells to initiate and maintain an infection. The essentiality of the core for mycobacterial growth and survival leads to the biosynthetic enzymes involved and being considered as ideal targets for drug development (3).

A subset of genes required for the late steps of mycolic acid and arabinogalactan biosynthesis are located in proximity to the genomes of mycobacteria and corynebacteria. These genes include a well characterized cell wall biosynthesis cluster encoding enzymes required for the activation (AccD4 and FadD32) and condensation (Pks13) (4) of mycolic acid intermediates prior to the final reduction step (5), and transfer of mature mycolic acids (6). Also within the cluster are genes required for arabinogalactan biosynthesis (*atfB*, *glf*, *glfT*, and *Rv3806c*) (7–10) which, like the mycolic acid biosynthesis genes, are essential for growth of *M. tuberculosis* (10).

Despite extensive functional characterization of this cluster over the last decade, the *in vivo* function of the product of one gene, *Rv3802c*, remains unknown, although mycolyltransferase (11) or Pks13-associated thioesterase (12) functions have been suggested. Although its genomic location strongly suggests a role in cell wall biosynthesis, definitive proof of such a role is lacking. The putative product of *Rv3802c* has a predicted signal sequence that contains a possible transmembrane domain, and it has been expressed to assess immunological responses (13) and enzymatic activities. The enzyme is one of seven cutinase-like proteins in *M. tuberculosis* and is retained in the cell wall, following translocation across the cell membrane (14). Previous studies have shown it to have phospholipase A and thioesterase activities (12), consistent with a role in mycolic acid biosynthesis, and significant lipase activity completely dependent on its Ser-Asp-His catalytic triad (14). A very recent study has suggested a role for Rv3802c in regulation of outer lipid composition in response to stress because the induction of the

* This work was supported in part by the Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics and the National Health and Medical Research Council of Australia.

The atomic coordinates and structure factors (code 3aja) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ Both authors should be considered as equal first authors.

² To whom correspondence should be addressed. Tel.: 61-3-9902-9148; Fax: 61-3-9902-9222; E-mail: paul.crellin@monash.edu.

³ Supported by an Australian Research Council Federation Fellowship.

⁴ Both authors should be considered as equal senior authors.

⁵ Supported by a Pfizer Senior Research Fellowship.

Structure/Function Analysis of a Mycobacterial Lipase

Corynebacterium glutamicum ortholog triggered an increase in mycolic acid biosynthesis as part of an outer membrane remodeling response to heat stress (15).

Recently, Rv3802c was identified as a major target of tetrahydrolipstatin (THL)⁶ (12). THL is a well characterized and irreversible inhibitor of serine esterases (16), originally identified for its specificity for pancreatic lipases, and thus was developed as an anti-obesity drug. THL has been reported to bind covalently to the catalytic serine residue of pancreatic lipase (17) and was found to have similar affinity for human fatty-acid synthase (18). In addition to its actions in humans, THL inhibits and disrupts cell wall formation in several mycobacterial species, with the exception of the nonpathogenic model species *Mycobacterium smegmatis* (19). Rv3802c was strongly inhibited by THL, whereas the nonorthologous but cutinase motif-bearing *M. smegmatis* lipase MSMEG_1403 was not inhibited at up to a 500:1 inhibitor to enzyme molar ratio (12).

To better understand the enzymology of Rv3802c, we expressed and purified the *M. tuberculosis* enzyme, its homolog in the nonpathogenic model species *M. smegmatis*, and the ortholog from a related species, *C. glutamicum*, and we assessed their inhibition by THL. We report here that the closest *M. smegmatis* homolog to Rv3802c, MSMEG_6394, is inhibited by THL. The crystal structure of MSMEG_6394, along with direct evidence of the essentiality of MSMEG_6394 in *M. smegmatis*, indicates a fundamental role of this lipase in *Mycobacteria*.

MATERIALS AND METHODS

Growth and Manipulation of *Escherichia coli* and *M. smegmatis*—*E. coli* DH5 α was used for plasmid preparations during cloning experiments, although BL21-DE3 was used for protein expression. Bacteria were routinely cultured at 30, 37, or 42 °C in solid and liquid Luria Burtani (LB) medium supplemented with kanamycin (Kn, 20 μ g/ml), streptomycin (Sm, 20 μ g/ml), ampicillin (100 μ g/ml), and sucrose (10% w/v), as appropriate. Tween 80 was added to 0.05% (v/v) to reduce clumping in mycobacterial liquid cultures. Competent *M. smegmatis* mc²155 cells were prepared as described previously (20) and electroporated using a Bio-Rad Gene Pulser with the following settings: 2.5 kV, 1000 ohms, 25 microfarads.

DNA Manipulations—PCRs were performed using ProofStart DNA polymerase (Qiagen) according to the manufacturer's instructions. Reactions consisted of a hot start (95 °C, 5 min) followed by 35 cycles of denaturation (95 °C, 1 min), annealing (55 °C, 1 min), and extension (72 °C, 2 min). Restriction enzymes and T4 DNA polymerase were from Roche Applied Science or New England Biolabs. Genomic DNA was prepared from mycobacteria as described previously (21). Southern blots (22) were performed using digoxigenin-labeled probes (Roche Applied Science) according to the manufacturer's instructions. *M. tuberculosis* H37Rv genomic DNA was obtained from Colorado State University (Fort Collins, CO).

Protein Expression and Refolding—Rv3802c was produced as described previously (14). Briefly, Rv3802c was amplified from

H37Rv genomic DNA and cloned into an *E. coli* expression vector (pET19b; Merck). Cytoplasmic, N-terminally His-tagged recombinant protein was expressed in *E. coli* BL21-DE3 and accumulated in cytoplasmic inclusion bodies following induction with isopropyl 1-thio- β -D-galactopyranoside (0.5 mM). Recombinant protein was solubilized in urea and purified by immobilized metal ion affinity chromatography before refolding by dialysis into Tris (50 mM, pH 8.0).

The MSMEG_6394 gene was amplified by PCR from *M. smegmatis* genomic DNA using primers A (5'-GGAATTGCA-TATGCGCCGTCGGACACCC) and B (5'-CCCCAAGCT-TCAACCGTGTTCGGATGGG) and cloned into pET28b (Merck) using NdeI and HindIII (underlined). Recombinant MSMEG_6394 was expressed in B834 *E. coli*, and inclusion bodies were prepared as in Kjer-Nielsen *et al.* (23). The recombinant protein was solubilized in buffer A (20 mM Tris-HCl, 0.5 M NaCl, 8 M urea, pH 8.0) and purified by immobilized metal ion affinity chromatography. The bound protein was eluted with buffer A + 0.2 M imidazole. The eluted protein was diluted 8-fold with buffer A and refolded by dialysis against 16 liters of buffer B (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 7 mM 2-mercaptoethanol) for 16 h. The refolded protein was further dialyzed against 16 liters of buffer C (10 mM Tris-HCl, pH 8.0, 20 mM NaCl, 2 mM EDTA). The refolded protein was concentrated using DEAE-cellulose (Sigma) and further purified by size-exclusion and ion-exchange chromatography.

The NCgl2775 gene was PCR-amplified from *C. glutamicum* ATCC 13032 genomic DNA using primers C (GGAATTGCA-TATGTCGGATGACTCAGATTCATTG) and D (GCCCAAGCTTATCCGTTGTCGATGAGGTTG), digested with NdeI and HindIII (underlined), and cloned into NdeI/HindIII-digested pET28b (Merck). Primer C was designed to bind downstream of the putative signal sequence such that the first codon after the ATG start codon encoded Ser²⁸ of NCgl2775. A sequenced clone was transformed into *E. coli* BL21 (DE3), and protein expression was induced at 30 °C with 1 mM isopropyl 1-thio- β -D-galactopyranoside. The recombinant NCgl2775 was found to be soluble and was purified by immobilized metal ion affinity chromatography using Talon metal affinity resin (Clontech).

Activity Assays—The esterase and lipase potentials of MSMEG_6394, Rv3802c, and NCgl2775 were determined as described previously (14, 24) with minor adaptations. The serine esterase substrate *p*-nitrophenyl butyrate was obtained from Sigma and was prepared in isopropyl alcohol at a range of concentrations between 200 and 1.875 mM. The substrate solutions were mixed 1:9 with a solution containing 50 mM sodium phosphate, pH 8.0, 2.3 mg/ml sodium deoxycholate, and 1 mg/ml gum arabic. To 20 μ l of enzyme solution (100 μ g/ml), 240 μ l of the above reaction mixture was added in a 96-well microwell plate, mixed, and incubated at 37 °C for 30 min. To quantify inhibition by THL, the assay was carried out with a final *p*-nitrophenyl butyrate concentration of 5 mM, close to the measured K_m value for each enzyme. THL was solubilized in DMSO and diluted in water to concentrations ranging between 100 μ M and 400 nM. Each enzyme was preincubated with inhibitor at a 1:1 ratio for 30 min at room temperature before the addition of the substrate mixture. The accumulation of *p*-nitro-

⁶ The abbreviations used are: THL, tetrahydrolipstatin; Kn, kanamycin; Sm, streptomycin; PDB, Protein Data Bank; r.m.s.d., root mean square deviation; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; TEM, transmission electron microscopy.

Structure/Function Analysis of a Mycobacterial Lipase

TABLE 1
Details of structural data collection and refinement

Dataset	Inflection	Peak	Remote
Data collection			
Wavelength	0.97941 Å	0.97930 Å	0.96411 Å
Symmetry	<i>I</i> ₄ ,22	<i>I</i> ₄ ,22	<i>I</i> ₄ ,22
Cell dimensions	<i>a</i> = <i>b</i> = 130.4, <i>c</i> = 209.5	<i>a</i> = <i>b</i> = 130.4, <i>c</i> = 209.5	<i>a</i> = <i>b</i> = 130.4, <i>c</i> = 209.5
Resolution	50 to 2.9 Å	50 to 3.1 Å	50 to 3.1 Å
Unique observations	20,363 (2909)	14,375 (2100)	16,769 (2412)
<i>R</i> _{pim} ^a	0.025 (0.420)	0.034 (0.305)	0.017 (0.336)
Mean(<i>I</i>)/S.D.(<i>I</i>)	11.3 (1.8)	9.9 (2.2)	15.2 (1.8)
Completeness	99.8% (100%)	87.0% (88.6%)	99.7% (100%)
Redundancy	6.4 (6.3)	4.1 (4.1)	6.4 (6.5)
Refinement			
<i>R</i> _{work} ^b	0.213		
<i>R</i> _{free}	0.250		
Residues			
Chain A	71–293, 298–334		
Chain B	71–334		
Total protein atoms	3924		
Total waters	32		
Bond lengths	0.016 Å		
Bond angles	1.508°		
Ramachandran analysis			
Favored	91.89%		
Allowed	7.34%		
Outlier	0.77%		
Average total <i>B</i> factor			
Protein atoms	102.5 Å ²		
Average isotropic <i>B</i> factor			
Non-protein atoms	45.9 Å ²		

^a $R_{pim} = \sqrt{(1/(n-1)) \sum (I_i - I_{mean})^2 / \sum I_i}$.^b $R_{work} = \sum_{hkl} |F_o| - |F_c| / \sum_{hkl} |F_o|$ for all data excluding the 5% that comprised the *R*_{free} used for cross-validation.

phenol was measured spectrophotometrically at 405 nm, and concentrations were calculated by comparisons to a *p*-nitrophenol standard curve. All assays were performed in triplicate. *K_m*, *V_{max}*, *K_p*, and IC₅₀ values for each enzyme were calculated using GraphPad Prism (GraphPad Software, version 4.03).

Crystallization and Data Collection—Crystals of selenomethionyl MSMEG_6394 were produced by the hanging-drop vapor diffusion method. Three μl of protein solution comprising MSMEG_6394 at 30 mg/ml, 10 mM Tris, pH 8.0, and 0.2 M NaCl were mixed with 1 μl of reservoir solution containing 2.1 M sodium formate and 0.1 M Bistris propane, pH 6.8. Crystallization trials were incubated at 21 °C. Crystals appeared after 5 days and typically grew to dimensions of 0.4 × 0.3 × 0.2 mm.

Prior to data collection, the crystals were soaked in reservoir solution with an additional 20% glycerol and flash-cooled to –173 °C in a stream of liquid nitrogen. X-ray diffraction data were collected from a single crystal mounted 300 mm from a Quantum-210 CCD detector at the BioCARS 14-BMD beamline of the Advanced Photon Source, Chicago. X-ray diffraction data were collected at wavelengths corresponding to the peak and inflection of the selenium absorption edge and at a high energy remote wavelength. The data were integrated with MOSFLM (25, 26) and scaled with SCALA (25, 27). Details of the data collection are summarized in Table 1.

Structure Determination and Refinement—The structure was determined using three wavelength multiple anomalous dispersion phasing. The positions of the 18 selenium sites, corresponding to two copies of MSMEG_6394 in the asymmetric unit, were calculated using SOLVE (28), and subsequent electron density modification was performed with RESOLVE (27, 28). Into the resultant electron density map an initial peptide-backbone trace was constructed with TEXTAL (29, 30). The structure was built with iterative cycles of manual building in

COOT (31) and maximum likelihood-based refinement with TLS using REFMAC (25, 32). Strict noncrystallographic restraints were maintained during refinement. The structure was validated using MOLPROBITY (33). Details of the refinement are summarized in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank under accession code 3AJA.

Construction and Analysis of a Conditional Knock-out of MSMEG_6394—MSMEG_6394 and flanking DNA were PCR-amplified from *M. smegmatis* mc²155 genomic DNA as a 2-kb fragment using primers E (5'-GATCAAGCTTACATGTCCG-GTGAGCTGG-3') and F (5'-GATCGGATCCGCGCACCTT-GGCCAGCG-3'), digested at the underlined restriction sites for HindIII and BamHI, and cloned into HindIII/BamHI-digested pUC18 (34). A nonpolar kanamycin resistance cassette carrying the *aphA3* gene was then inserted at a unique SphI site within MSMEG_6394 after T4 polymerase treatment to form blunt ends. The 2.8-kb HindIII-BamHI fragment containing MSMEG_6394::aphA3 was then transferred to BamHI-digested pPR27, following T4 polymerase treatment of both insert and vector. To generate single crossovers, this plasmid was introduced into *M. smegmatis* mc²155 by electroporation and selecting kanamycin-resistant clones at 30 °C. A 10-ml LB broth containing kanamycin was inoculated with a single colony and grown for 5 days at 30 °C to saturation. Serial dilutions were plated onto LB + Kn plates at 42 °C and incubated for 4 days to select for potential single crossovers. Colonies were screened for incorporation of the plasmid into the chromosome by growing 10-ml LB + Kn cultures to saturation at 42 °C, extracting genomic DNA, digesting with XbaI/BamHI, and performing a Southern hybridization with a probe specific for MSMEG_6394. Out of nine colonies tested, one single cross-

Structure/Function Analysis of a Mycobacterial Lipase

over, designated Myc46, was found and subjected to further manipulation. To derive a double crossover (a conditional knock-out), a complementing plasmid, containing *MSMEG_6394* carried on a 2.0-kb BamHI fragment cloned into the temperature-sensitive plasmid pCG76, was introduced into Myc46 by electroporation. Transformants were selected on LB/Kn/Sm plates at 30 °C, and a single colony was grown to saturation in LB/Kn/Sm broth at 30 °C. Serial dilutions were plated on LB/Kn/Sm plates containing sucrose and incubated at 30 °C. Potential conditional knock-out clones were grown in 10 ml of LB/Kan/Sm to saturation followed by genomic DNA extraction, digestion with XbaI/BamHI, and Southern blotting using an *MSMEG_6394*-specific probe. A confirmed conditional knock-out strain was designated 6394CKO. To derive growth curves, strains were cultured in 10 ml of LB/kanamycin/streptomycin at 30 °C for 3 days and then 5 ml was added to 200 ml of LB/Kan that had been pre-warmed to 30 or 42 °C. The cultures were sampled daily and serial dilutions plated onto LB/Kan at 30 °C. Following 5 days of incubation, colonies were counted to determine colony-forming units per ml.

Electron Microscopy—Bacteria were grown at 30 or 42 °C for 5 days on solid media containing appropriate antibiotics and fixed for 2 min with 2.5% glutaraldehyde in phosphate-buffered saline (PBS). Cells were scraped gently, transferred to centrifuge tubes, left in glutaraldehyde/PBS for 30 min, and centrifuged for 1 min at 5000 rpm. The solution was replaced with 2.5% glutaraldehyde and 0.05% ruthenium red in PBS and fixed overnight in the dark at 4 °C. Cells were rinsed in PBS three times for 5 min, post-fixed in 1% osmium tetroxide for 2 h, and then rinsed in water three times for 5 min. The cells were dehydrated in 10, 30, and 50% ethanol for 30 min in each concentration and then held in 70% ethanol for 4 days. Dehydration was completed in 90% ethanol for 30 min and 100% dry ethanol three times for 1 h.

Samples for TEM were then placed in propylene oxide for 1 h and then infiltrated with 25% firm grade Spurr's resin in propylene oxide for 3 days, followed by two times for 2 h in 50% Spurr's resin in propylene oxide, then 3 days in 75% Spurr's resin in propylene oxide, and 2 days in 100% Spurr's resin. Polymerization was completed at 60 °C overnight. Cells were then sectioned at 90 nm using a Reichert Ultracut S ultramicrotome and picked up onto 300 mesh copper grids. Staining was then performed with saturated uranyl acetate in 50% methanol for 10 min followed by saturated lead citrate in carbonate-free distilled water for 10 min. Sections were viewed with a Jeol 200CX TEM at 100 kV.

Samples for scanning electron microscopy were kept in 100% dry ethanol for 7 days and then rinsed in hexamethyldisilazane three times for 10 min. Drops of hexamethyldisilazane-containing cells were placed on a plastic film (unexposed, developed Ektachrome photographic emulsion) and allowed to air dry. The film was then mounted onto a double-sided carbon tape on aluminum stubs. Sputter was coated with gold using Balzers SCD 005 sputter-coating unit for 3 min at 25 mA. Cells were viewed at 10 kV and a working distance of 3 mm using a Hitachi S570 scanning electron microscope.

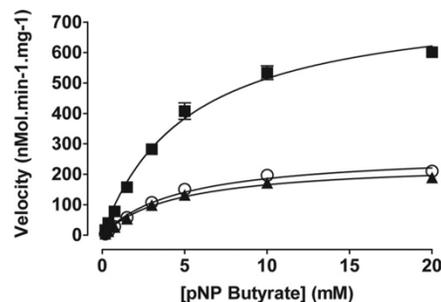


FIGURE 1. Kinetics of *p*-nitrophenyl butyrate hydrolysis. Activity of MSMEG_6394 (■), Rv3802c (▲), and NCgl2775 (○) at pH 8.0 is shown. *p*NP butyrate, *p*-nitrophenyl butyrate.

TABLE 2

Enzyme activity of MSMEG_6394, Rv3802c, and NCgl2775 in *p*-nitrophenyl butyrate hydrolysis

Data shown are the mean of three experiments. Errors shown are means \pm S.E.

	V_{max} $nmol \cdot min^{-1} \cdot mg^{-1}$	k_{cat} s^{-1}	K_m mM	Specificity constant $M^{-1} \cdot s^{-1}$
MSMEG_6394	783 ± 26.7	0.448	5.22 ± 0.45	86
Rv3802c	241 ± 7.8	0.143	4.52 ± 0.38	32
NCgl2775	273 ± 11.8	0.146	4.71 ± 0.52	31

RESULTS

Purification of Rv3802c, MSMEG_6394, and NCgl2775—Rv3802c is very well conserved among the mycobacteria with homologs present in all genomes sequenced to date, including the minimal genome of *Mycobacterium leprae* (14). The *M. smegmatis* homolog, MSMEG_6394, shares 69% sequence identity with Rv3802c, whereas the ortholog in *C. glutamicum*, NCgl2775, shares 60% sequence identity with Rv3802c. The genes encoding Rv3802c, MSMEG_6394, and NCgl2775 were PCR-amplified from their respective genomes without their putative signal sequences and cloned into inducible expression vectors for production in *E. coli*. The construct for Rv3802c expression had been used in an earlier study and was known to yield insoluble protein that could be refolded to an active conformation (14). Expression of MSMEG_6394 in this study also gave rise to insoluble material that was successfully purified and refolded for enzymatic and structural studies. Surprisingly, NCgl2775 was found to be soluble when overexpressed in *E. coli*, and purified material was used directly for enzyme assays.

Enzyme Kinetics and THL Inhibition Studies—The kinetic parameters of Rv3802c and its orthologs from *M. smegmatis* and *C. glutamicum* were measured in functional assays based on the hydrolysis of *p*-nitrophenyl butyrate at a range of substrates. All three enzymes displayed activity under the conditions tested (Fig. 1), with MSMEG_6394 demonstrating the highest activity, with a maximum enzyme velocity of $783 \text{ nmol} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$ (± 26.7) or more than three times that of Rv3802c. The observed specificity constants for all three enzymes were similar (Table 2).

Activity of all three enzymes was inhibited by THL with K_i of $0.8 \mu M$ (with 95% confidence interval of 0.59–1.27) for Rv3802c (Fig. 2 and Table 3). Similar levels of inhibition were recorded whether or not samples were preincubated with inhibitors, indicating rapid and potentially irreversible inhibition. For each

Structure/Function Analysis of a Mycobacterial Lipase

enzyme, 90% inhibition of activity was observed with THL concentrations less than 50 μM . This inhibition was not observed in either of the control enzymes tested, a lipase from *Candida rugosa* and a liver esterase from *Sus domesticus* (data not shown), indicating a level of specificity.

Three-dimensional Structure of MSMEG_6394—To gain functional insight into the role of Rv3802c and its orthologs, we

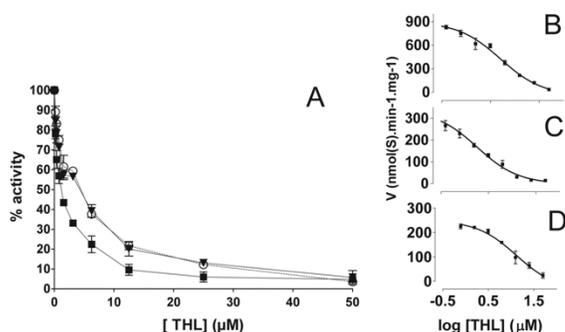


FIGURE 2. Inhibition of enzyme activity for *p*-nitrophenyl butyrate by THL. Enzymes were preincubated with inhibitor for 30 min before the addition of 5 mM substrate. *A*, residual activity of MSMEG_6394 (○, dashed line), Rv3802c (■, solid line), and NCgl2775 (▼, solid line). *B–D*, curve fitting of enzyme activity for MSMEG_6394 (*B*), Rv3802c (*C*), and NCgl2775 (*D*) was used to calculate K_i and IC_{50} values.

TABLE 3

Inhibition of *p*-nitrophenyl butyrate hydrolysis activity of MSMEG_6394, Rv3802c, and NCgl2775 by THL

Data shown are the mean of three experiments. Values in parentheses are 95% confidence intervals.

	IC_{50}	K_i
	μM	μM
MSMEG_6394	5.08 (3.42–7.55)	2.59 (1.74–3.86)
Rv3802c	1.82 (1.25–2.68)	0.87 (0.59–1.27)
NCgl2775	13.06 (6.69–25.53)	6.34 (3.24–12.38)

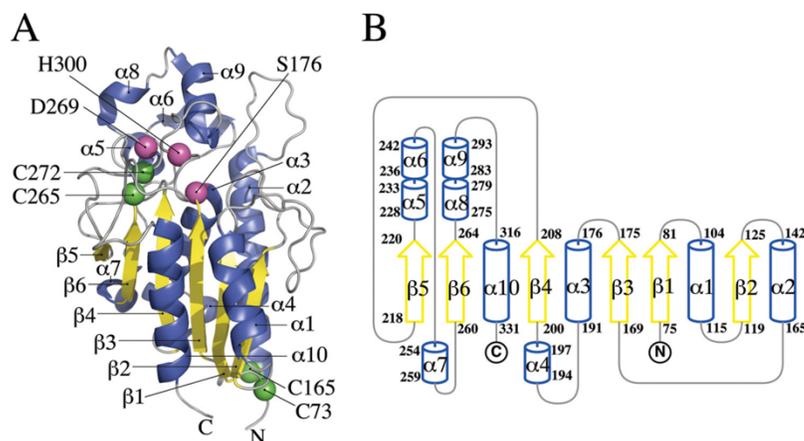


FIGURE 3. Crystal structure of MSMEG_6394. *A*, schematic representation of the MSMEG_6394 structure. The structure includes α/β hydrolase fold with six-stranded, parallel β -sheets ($\beta 1$ – $\beta 6$, yellow) bounded on both sides by four α -helices ($\alpha 1$ – $\alpha 3$ and $\alpha 10$, blue). A primarily helical “lid-like” domain ($\alpha 5$, $\alpha 6$, $\alpha 8$, and $\alpha 9$) sits atop the active site. The two disulfides in the structure are represented by green spheres. The catalytic triad residues Asp²⁶⁹, His³⁰⁰, and Ser¹⁷⁶ are shown as magenta spheres. *B*, topology diagram of MSMEG_6394 colored similarly to *A*.

determined the crystal structure of the *M. smegmatis* enzyme. Initial crystallization trials of all three proteins resulted in crystals for Rv3802c and MSMEG_6394; however, only diffraction quality crystals were obtained for MSMEG_6394. The structure of MSMEG_6394 was determined to a resolution of 2.9 Å and contained two copies of the molecule in the asymmetric unit (chains A and B). The construct used in crystallization included residues 36–337 of MSMEG_6394 as well as 20 vector-derived residues, including the His₆ purification tag. Of this construct, residues 71–334 were modeled from the electron density. There was no evidence of higher order oligomerization within the crystal lattice, an observation that is consistent with the purification of MSMEG_6394 as a monomer. The protein structure confirmed that MSMEG_6394 is a member of the cutinase family of α/β hydrolases (Fig. 3*A*). Indeed, the closest structural homologs found with a DALI search of the Protein Data Bank were the acetyl xylan esterase (PDB code 1BS9 (35), Z-score = 19.2, r.m.s.d. = 2.6 Å over 189 C α positions), cutinase-like protein (PDB code 2CZQ (36), Z-score = 18.8, r.m.s.d. = 2.4 Å over 185 C α positions), lysin B (PDB code 3HC7 (37), Z-score = 18.5, r.m.s.d. = 2.9 Å over 195 C α positions) and cutinase (PDB code 1CUS (38), Z-score = 16.1, r.m.s.d. = 2.8 Å over 173 C α positions). All of these structural homologs shared ~20% sequence identity with MSMEG_6394.

The canonical α/β hydrolase domain of MSMEG_6394 included a six-stranded, parallel β -sheet ($\beta 1$ – $\beta 6$) bounded on both sides by four α -helices ($\alpha 1$ – $\alpha 3$ and $\alpha 10$) (Fig. 3*B*). Inserted into this domain between the $\beta 5$ - and $\beta 6$ -strands and between the $\beta 6$ -strand and $\alpha 10$ -helix lie a second, primarily helical domain ($\alpha 5$, $\alpha 6$, $\alpha 8$, and $\alpha 9$) atop the proposed active site. This second domain is reminiscent of the “lid” domains found in other esterase and lipase members of the α/β hydrolase superfamily and is most often associated with substrate binding specificity and interfacial activation (39).

The four cysteine residues in the protein formed two disulfide bonds, Cys⁷³–Cys¹⁶⁵ and Cys²⁶⁵–Cys²⁷² (Fig. 3*A*). Both of these disulfides were structurally conserved with the acetyl xylan esterase and cutinase proteins (PDB codes 1BS9 and 1CUS), yet neither was found in lysin B (PDB code 3HC7). The first disulfide stabilizes the α/β -sandwich domain, linking the N terminus of the protein with the C-terminal region of the $\alpha 2$ -helix. The second disulfide formed an intra-loop bridge in the β -turns spanning the $\beta 6$ -strand and $\alpha 8$ -helix and is likely to be important in maintaining the conformation of the residue Asp²⁶⁹ from the catalytic triad.

Serine esterases typically have an active site comprising a catalytic triad of serine aspartate and histidine. The position of the catalytic triad of residues is strictly conserved

Structure/Function Analysis of a Mycobacterial Lipase

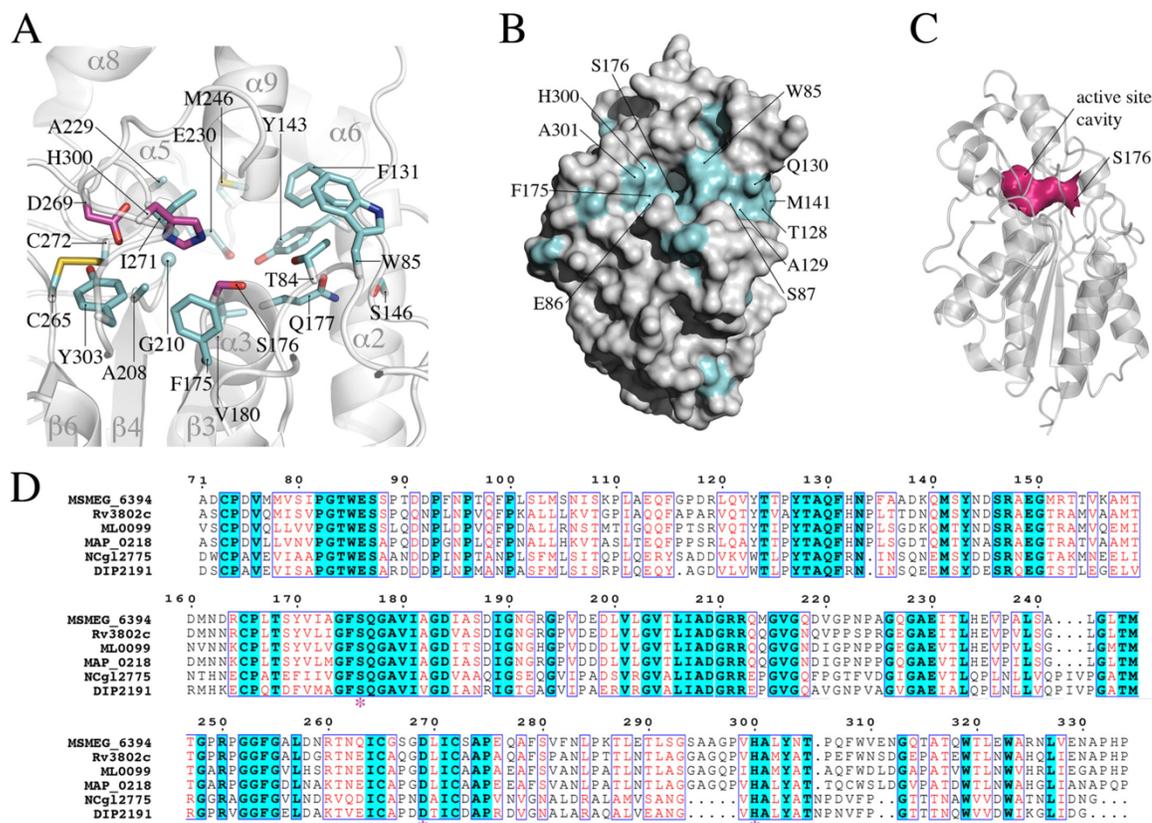


FIGURE 4. Analysis of the active site of MSMEG_6394. *A*, view of the active site of MSMEG_6394 shown in the orientation depicted in Fig. 3*A*. The secondary structure elements are shown in schematic representation, colored gray, and labeled. Residues that are strictly conserved in the sequence alignment shown in *D* are represented in *stick format*. The catalytic triad is colored magenta, and the other conserved residues are colored cyan. *B*, surface representation of MSMEG_6394 orientated similarly to Fig. 3*A*. Residues that are strictly conserved in the sequence alignment shown in *D* are colored cyan and labeled. The patch of sequence conservation leading to the active site cavity may be involved in substrate recognition. *C*, schematic representation of MSMEG_6394 rotated 90° from the orientation shown in Fig. 3*A*. The solvent-accessible surface of the active site cavity is shown in magenta (analyzed with CASTP (43)). The catalytic residue Ser¹⁷⁶ is labeled. *D*, sequence alignment of MSMEG_6394 homologs from related species. MSMEG_6394 is from *M. smegmatis*; Rv3802 is from *M. tuberculosis*; ML0099 is from *M. leprae*; MAP_0218 is from *Mycobacterium paratuberculosis*; NCgl2775 is from *C. glutamicum*, and DIP2191 is from *Corynebacterium diphtheriae*. The strictly conserved residues are colored cyan, and conserved residues (44) are colored red. Residues of the catalytic triad are indicated by an asterisk. Sequences were aligned with ClustalW (45) and annotated with ESPRIT (46).

between all the members of the cutinase family and coincides with the positions of Ser¹⁷⁶, Asp²⁶⁹, and His³⁰⁰ in MSMEG_6394 (Fig. 4*A*). The nucleophilic Ser¹⁷⁶ is located at the bend of the tight turn between the β 3-strand and α 3-helix, a feature known as the “nucleophilic elbow” that is conserved among all members of the α/β hydrolase superfamily (40). This nucleophilic elbow has the sequence Gly-Phe-Ser-Gln-Gly in MSMEG_6394, conforming to the consensus of Gly-(Phe/Tyr)-Ser-Gln-Gly with the other members of the cutinase family solved to date. Asp²⁶⁹ and His³⁰⁰ are positioned adjacent to each other, with Asp²⁶⁹ residing on the loop connecting the β 6-strand and α 7-helix and His³⁰⁰ within the linker between the β 6-strand and α 10-helix within the cutinase fold (Fig. 4*A*). Taken together, the structural features of MSMEG_6394 indicate that it has a catalytic mechanism akin to that of other serine esterases.

The catalytic Ser¹⁷⁶ O γ is positioned at the mouth of an 11-Å deep cavity that encloses the active site. This active site cavity is composed of residues from both the cutinase and lid domains of

the protein and has a solvent-accessible surface area of 80 Å² (Fig. 4, *B* and *C*). Of the residues that line the cavity, Thr⁸⁴, Phe¹³¹, Tyr¹⁴³, Ser¹⁷⁶, Gln¹⁷⁷, Val¹⁸⁰, Ala²⁰⁸, Gly²¹⁰, Ala²²⁹, Glu²³⁰, Met²⁴⁶, Ile²⁷¹, and His³⁰⁰ are all strictly conserved across MSMEG_6394 homologs in other mycobacterial and related species (Fig. 4, *A* and *D*). The exceptions are residues 291–293 that form the C terminus of the α 9-helix and line the upper part of the cavity (Fig. 4*D*). Similarly, there is strict conservation among mycobacterial species of the residues that form surface-exposed patches about the mouth of the active site cavity (Glu⁸⁶, Phe¹⁷⁵, His³⁰⁰, and Ala³⁰¹) and within a patch leading to the active site (Trp⁸⁵, Ser⁸⁷, Thr¹²⁸, Ala¹²⁹, Gln¹³⁰, and Met¹⁴¹) (Fig. 4*B*). The direct modeling of THL binding to MSMEG_6394 is complicated by the lack of structural homology of the lid domain between related proteins. However, the proximity of these conserved patches to the active site suggests that they, together with the residues lining the active site, may determine substrate recognition.

Structure/Function Analysis of a Mycobacterial Lipase

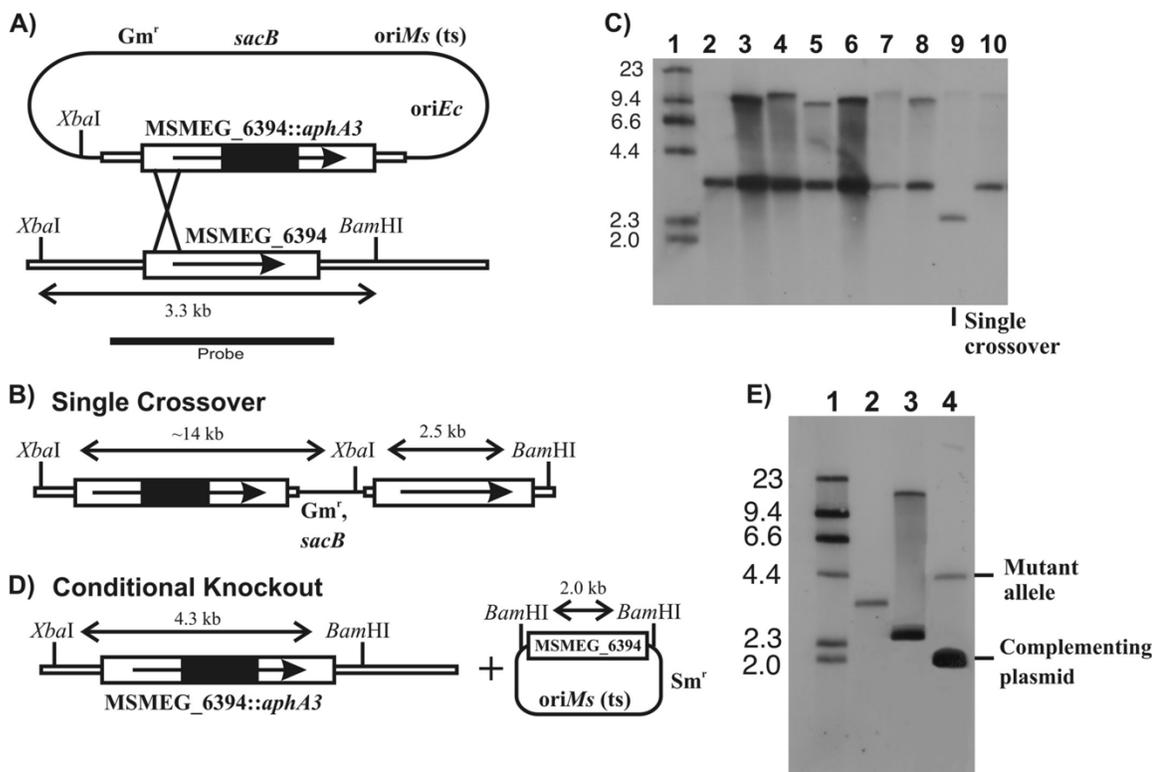


FIGURE 5. Construction of a conditional knock-out of *MSMEG_6394*. *A*, recombination plasmid contained a cloned copy of *MSMEG_6394* interrupted by a nonpolar kanamycin resistance cassette (*MSMEG_6394::aphA3*), a gentamycin resistance marker (Gm^r), a temperature-sensitive replication origin for *M. smegmatis* ($oriMs (ts)$), a replication origin for *E. coli* ($oriEc$), and a counterselectable marker encoding sucrose sensitivity ($sacB$). The construct was introduced into *M. smegmatis* at the permissive temperature (30 °C). Integration of the plasmid by a single crossover at the position indicated was detected by growing the cells at the nonpermissive temperature (42 °C) in the presence of kanamycin. *B*, genetic map of the single crossover, showing key restriction sites. *C*, Southern blot of $XbaI/BamHI$ digests of genomic DNA from potential single crossover strains, probed with the fragment indicated in *A*. *Lane 1*, DNA molecular weight DNA markers of the sizes indicated in kilobases; *lanes 2–10*, potential single crossover strains. *Lane 9* shows the single crossover strain that was selected for further manipulation. *D*, culturing the single crossover strain containing a complementation plasmid gave rise to a disrupted copy of *MSMEG_6394* in the chromosome, producing the conditional knock-out (6394CKO). Because the disruption of *MSMEG_6394* coincided with the loss of the $sacB$ gene, the conditional knock-out strain could be selected on sucrose plates. *E*, Southern blot of $XbaI/BamHI$ digests of genomic DNA showing isolation of the conditional knock-out strain. *Lane 1*, DNA molecular weight DNA markers of the sizes indicated in kilobases; *lane 2*, wild-type *M. smegmatis* mc²155; *lane 3*, single crossover strain; *lane 4*, conditional knock-out of *MSMEG_6394* (6394CKO).

Conditional Disruption of *MSMEG_6394* Proves Gene Essentiality in *M. smegmatis*—To gain insights into the function of Rv3802c, we attempted to make a mutant strain of *M. smegmatis* in which the *MSMEG_6394* gene was disrupted with a drug resistance cassette. Despite several attempts, we were unable to generate this mutant, indicating that the enzyme might be essential to the viability of *M. smegmatis* and a potential drug target in mycobacteria. To investigate this further, we devised a genetic approach to assess the essentiality of *MSMEG_6394* (see under “Materials and Methods”). Briefly, a homologous recombination strategy was used to disrupt *MSMEG_6394* in the *M. smegmatis* chromosome in the presence of a rescue plasmid carrying an intact *MSMEG_6394* gene (Fig. 5). This involved isolation of a single crossover strain (Fig. 5, A–C) followed by initiation of a second crossover event in the presence of a rescue plasmid encoding the *MSMEG_6394* gene (Fig. 5, D and E). Allelic replacement of the chromosomal *MSMEG_6394* by the disrupted copy

was successfully achieved in the presence of the plasmid, giving rise to a “conditional” knock-out strain that we designated 6394CKO (Fig. 5E, lane 4).

If *MSMEG_6394* is essential, then 6394CKO should be fully reliant on the rescue plasmid for its survival. This plasmid has a temperature-sensitive origin of replication and can replicate at the permissive temperature (30 °C) but not at the nonpermissive temperature (42 °C) and is cured from the bacterial population when cells are grown at 42 °C (41). To confirm that 6394CKO is reliant on the rescue plasmid, the strain was cultured at 30 °C, then diluted into fresh medium at 30 and 42 °C, and sampled regularly to determine the number of viable bacteria as colony-forming units/ml. As shown in Fig. 6, 6394CKO continued to grow at 30 °C (at which the plasmid can replicate) but showed a dramatic loss of viability at 42 °C (at which the plasmid cannot replicate). By contrast, a culture of wild-type *M. smegmatis* mc²155 carrying the kanamycin resistance plasmid pMV261 grew

Structure/Function Analysis of a Mycobacterial Lipase

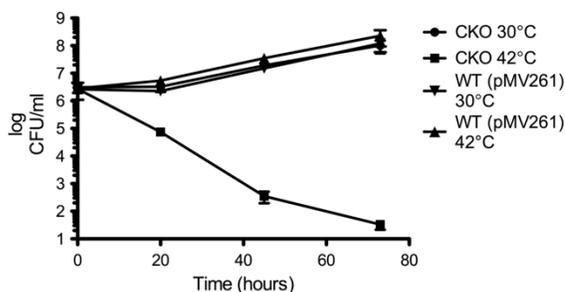


FIGURE 6. *MSMEG_6394* is essential to the viability of *M. smegmatis*. The conditional knock-out strain 6394CKO was cultured at 30 °C in LB containing kanamycin and streptomycin. At saturation, 5 ml was used to inoculate 200 ml of LB/kanamycin medium that had been prewarmed at the permissive (●, 30 °C) or nonpermissive (■, 42 °C) temperature. Incubation was continued at the two temperatures, and both cultures were sampled regularly with serial dilutions plated on LB plates containing kanamycin to determine colony forming units (CFUs) per ml. A wild-type *M. smegmatis* mc²155 strain containing the kanamycin resistance plasmid pMV261 was included as a control (▼, 30 °C; ▲, 42 °C).

well at both temperatures and at a similar rate as 6394CKO at 30 °C, showing that the loss of viability of 6394CKO at 42 °C was not just due to the temperature shift. Overall, these data confirmed that *MSMEG_6394* is essential for the growth and survival of *M. smegmatis*.

Electron Microscopy Analysis of an *MSMEG_6394* Conditional Knock-out—To examine the effects of the loss of *MSMEG_6394* on cell structure and integrity, 6394CKO was examined by TEM and scanning electron microscopy following growth at the permissive and nonpermissive temperatures. 6394CKO was cultured for 5 days on LB agar containing kanamycin and streptomycin and then subcultured onto LB/Kan plates at 30 and 42 °C for a further 5 days to allow curing of the rescue plasmid to occur from the 42 °C samples. The cells were fixed in glutaraldehyde and then scraped from the plates and processed for microscopy (see under “Materials and Methods”). Scanning electron microscopy revealed that CKO6394 cells grown at the nonpermissive temperature (Fig. 7D) were elongated and had a rough surface relative to those grown at the permissive temperature (Fig. 7C) and wild-type controls (Fig. 7, A and B), and many appeared to have lysed. TEM was then applied to examine the cell walls and internal details (Fig. 8). TEM revealed that CKO6394 cells grown at the permissive temperature were intact with regular and typical internal compartments visible (Fig. 8C). However, CKO6394 cells grown at the nonpermissive temperature fell into one of two classes, and typical examples of each are shown in Fig. 8, D–G. Members of class 1 were elongated and showed a loss of cell wall integrity and internal structure (Fig. 8, D and E) and appeared to have lysed. Class 2 had retained an intact cell wall and internal structure (Fig. 8, F and G) but contained several electron transparent zones in their cytoplasm. In some cases, these zones were huge and dominated most of the internal space. In contrast, a wild-type strain carrying pMV261 included as a control appeared normal and intact at both temperatures (Fig. 8, A and B), showing that the cellular phenotype was not just due to the temperature shift. Combining all our findings, it is clear that *MSMEG_6394* is essential for survival, playing a critical role in maintaining the cellular integrity of the bacterium. Our results have important implications regarding the role of the homologous proteins in pathogenic mycobacteria such as *M. tuberculosis*.

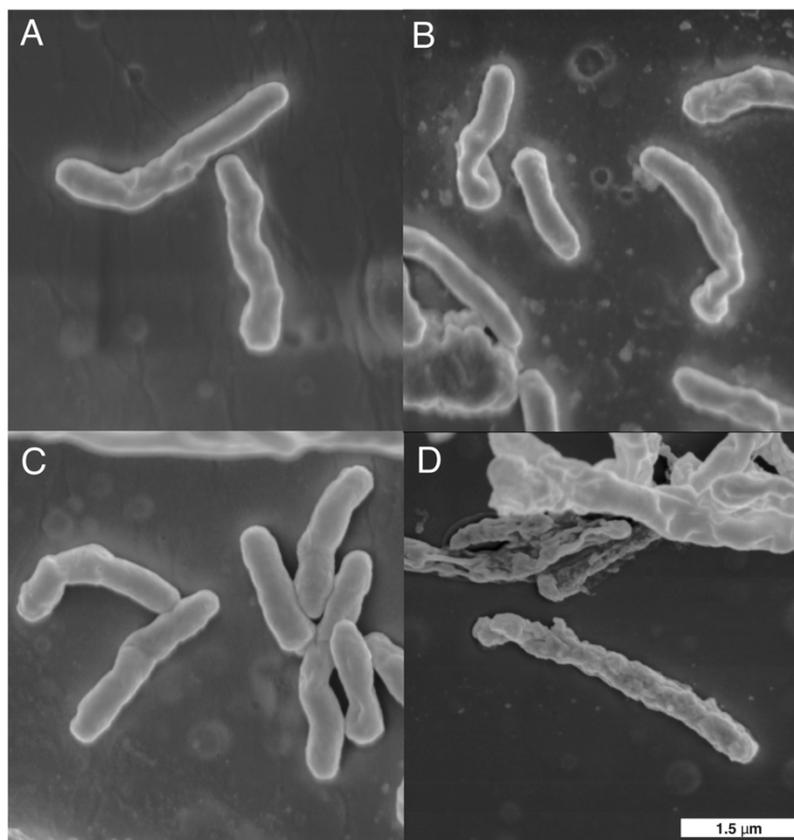


FIGURE 7. Scanning electron microscopy of an *MSMEG_6394* conditional knock-out. Bacteria were cultured on LB agar for 5 days prior to processing for scanning electron microscopy (see “Materials and Methods”). Wild-type *M. smegmatis* mc²155 strain containing the kanamycin resistance plasmid pMV261 was grown at 30 °C (A) or 42 °C (B). Conditional knock-out strain 6394CKO was grown at 30 °C (C) or 42 °C (D). All panels are the same magnification, and a scale bar is in the bottom right corner.

Structure/Function Analysis of a Mycobacterial Lipase

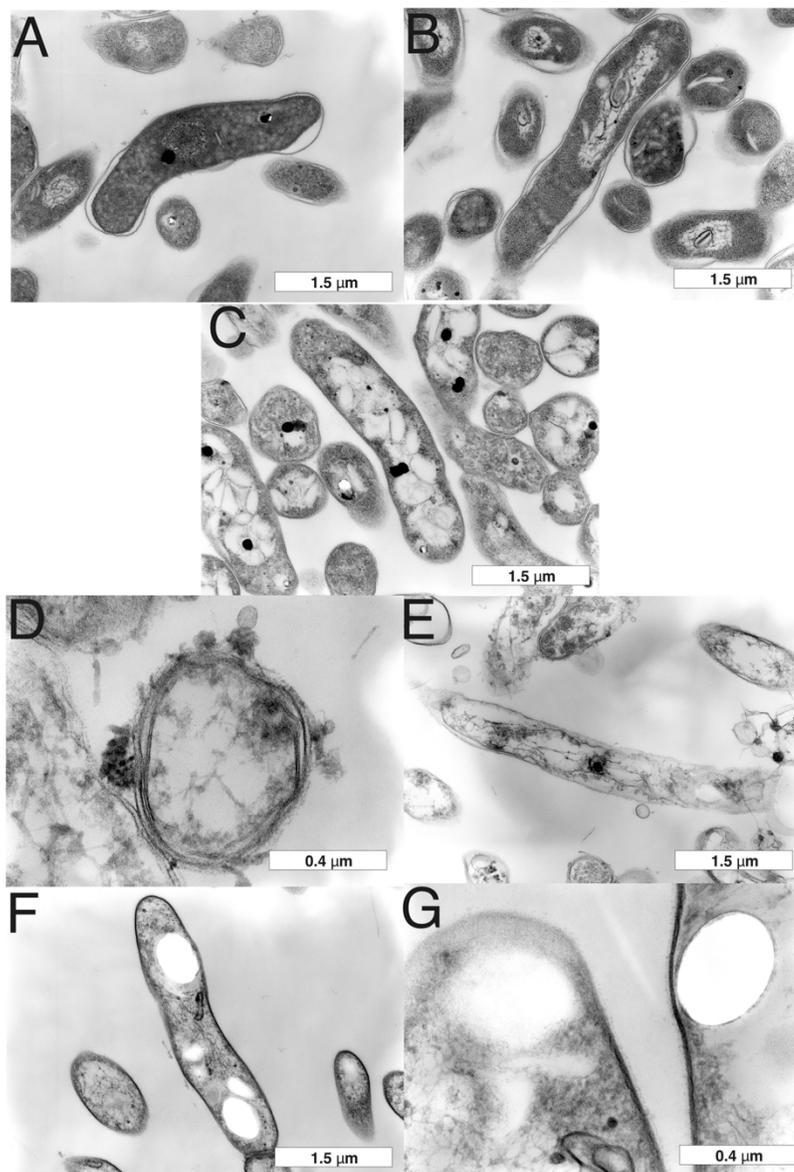


FIGURE 8. Transmission electron microscopy of an MSMEG_6394 conditional knock-out. Bacteria were cultured on LB agar for 5 days prior to processing for TEM (see "Materials and Methods"). Wild-type *M. smegmatis* mc²155 strain containing the kanamycin resistance plasmid pMV261 was grown at 30 °C (A) or 42 °C (B). Conditional knock-out strain 6394CKO was grown at 30 °C (C) or 42 °C (D–G). Two classes of cells were observed for CKO6394 at 42 °C as follows: class 1 cells had lost cell wall integrity and lysed (D and E) although class 2 cells were intact but contained large electron transparent zones (F and G). Scale bars are shown at the bottom right of each panel.

DISCUSSION

The unique and highly impermeable mycobacterial cell wall is a key virulence factor that forms the interface between host and pathogen. It enables the bacterium to resist destruction by the host and also contains unusual molecules that promote uptake by host macrophages and modify host

responses to create a favorable environment for bacterial survival and replication. As a result, the biosynthetic processes involved in the synthesis of the mycobacterial cell wall have been the subject of intensive research, and many of the key cell wall enzymes are now known. For example, nearly all genes within an ~20-kb genetic locus dedicated to cell wall biosynthesis have now been characterized. In this study, we have focused on one incompletely characterized gene of this cluster, *Rv3802c* from *M. tuberculosis*.

Homology searches using *M. tuberculosis* *Rv3802c* revealed that this gene is very well conserved in *Actinomyces* genomes, and we chose to focus on previously uncharacterized homologs from *M. smegmatis* and *C. glutamicum*. Previous studies using *p*-nitrophenyl butyrate substrates had revealed a significant lipase activity of *Rv3802c* (14), and THL-inhibited phospholipase A/thioesterase activities have also been reported (12). Here, we have shown that all three enzymes have significant lipase activities with similar affinity for the substrate and turnover rate. In addition, all three enzymes are inhibited by micromolar concentrations of THL. This suggests that the active sites of all three enzymes are relatively similar. Although Parker *et al.* (12) reported a lack of THL inhibition against an *M. smegmatis* enzyme, the enzyme tested (MSMEG_1403) is not the homolog of *Rv3802c* but rather a culture supernatant enzyme with phospholipase A activity. They suggested that a lack of MSMEG_1403 inhibition by THL is significant given that *M. smegmatis* growth is also not inhibited by the drug. However, our finding that MSMEG_6394 is inhibited by THL as well as being an essential enzyme in *M. smegmatis* would suggest that growth should be inhibited. We propose that the lack of growth inhibition in this organism is due to the inability of THL to reach its target(s) rather than a lack of activity against any particular enzyme. Because THL is active against *M. tuberculosis*, which shares a very similar cell wall architecture with *M. smegmatis* (42), lack of cell entry seems an unlikely possibility. We suggest that efflux of the drug

Structure/Function Analysis of a Mycobacterial Lipase

via one of the many transporters in *M. smegmatis* is the most likely explanation for the lack of activity against this organism.

The annotation of Rv3802c as a putative cutinase and the observed enzyme activities of Rv3802c and its orthologs are entirely consistent with the three-dimensional structure that showed MSMEG_6394 to be a member of the cutinase family of α/β hydrolases. Interestingly, one of the closest structural matches was to mycobacteriophage lysin B (37), a novel mycolylarabinogalactan esterase that cleaves the mycobacterial cell wall, releasing free mycolic acids from the arabinogalactan layer. LysB has been proposed to facilitate mycobacterial lysis by cleaving the outer mycolate-containing layer of the cell wall from the peptidoglycan-arabinogalactan complex (37), promoting bacteriophage release from its host. The structural similarity raises the possibility of a similar function for Rv3802c, and points to a degradative role for this enzyme during mycobacterial growth. Because the deposition of new cell wall material requires cleavage and opening of the existing structure, a lipase-like Rv3802c could theoretically fulfill this role. Indeed, the unusual cell wall of members of the *Actinomycetes* would require the presence of a dedicated enzyme restricted to this group of bacteria. This role is consistent with the retention of Rv3802c in the cell wall, although other cutinase-like proteins are secreted into the culture filtrate of *M. tuberculosis* (14). This function would also explain the proposed essentiality of Rv3802c and the elongated nature of our conditional knock-out in *M. smegmatis* grown at the nonpermissive temperature, as described below.

Rv3802c is thought to be essential for life in *M. tuberculosis* because of the failure of the *Rv3802c* gene to accumulate transposon insertions in saturation mutagenesis experiments (10), although direct evidence of essentiality is lacking. Our inability to disrupt the homologous gene in *M. smegmatis* (MSMEG_6394) suggested essentiality in this species as well, and we confirmed this by creating and analyzing CKO6394, a conditional knock-out strain. CKO6394 was found to be reliant on the plasmid-encoded copy of MSMEG_6394, and curing the plasmid correlated with the appearance of surface roughness, loss of cell wall integrity, and appearance of large electron transparent zones in the cytoplasm of otherwise intact cells. Our interpretation of these observations is that the loss of the MSMEG_6394 gene results in cell death via the formation of the electron transparent zones followed by progression to the lysis stage over time. Given the suggested role of Rv3802c in mycolic acid biosynthesis (12), we tried to detect lipid bodies with a fluorescent lipid stain, but the results were inconclusive. The composition of these zones is not known, but they seem to be composed of a hard substance that resists the resin added during processing for EM. Interestingly, Kremer *et al.* (19) reported minor surface changes and the presence of electron-translucent bodies in THL-treated *Mycobacterium kansasii*, although these were confirmed to be lipid bodies and were not as well defined as the electron transparent zones described here. We also attempted to extract cell wall components following loss of the plasmid to try and detect an accumulating species, but cell death resulted in a spectrum of cell wall changes, most associated with cellular disintegration and not directly related to the loss of Rv3802c.

A recent study has suggested a role for Rv3802c and its orthologs in regulating outer membrane lipid composition under stress conditions (15). The authors found that the *C. glutamicum* ortholog NCgl2775 could be disrupted to give a viable mutant with no obvious alteration in cell wall composition, and we have confirmed these findings as our own mutant of *C. glutamicum* NCgl2775 was viable (data not shown). This study provided evidence of transcriptional induction of NCgl2775 during heat stress leading to an increase in mycolic acid biosynthesis and a decrease in phospholipid content (15). However, this finding does not account for the essential nature of MSMEG_6394 and Rv3802c in *M. smegmatis* and *M. tuberculosis*, respectively, under normal growth conditions. This discrepancy strongly suggests that a more fundamental role for Rv3802c and its orthologs remains unidentified. We propose a degradative role for the enzyme that is essential for the deposition of new cell wall material during active mycobacterial growth. The essentiality of Rv3802c and the fact its inhibitor, THL, also kills *M. tuberculosis* (19) make Rv3802c an attractive target to develop new antimicrobials to increase the arsenal against drug-resistant *M. tuberculosis* (1).

Acknowledgments—We thank Gunta Jaudzems for assistance with electron microscopy; Svetozar Kovacevic and David Lea-Smith for helpful discussions; Brigitte Gicquel for providing pPR27, and Helen Billman-Jacobe for providing pCG76.

REFERENCES

1. World Health Organization (2009) in *WHO Report 2009: Global Tuberculosis Control, Epidemiology, Strategy, Financing*, pp. 1–314, World Health Organization, Geneva, Switzerland
2. Brennan, P. J. (2003) *Tuberculosis* 83, 91–97
3. Brennan, P. J., and Crick, D. C. (2007) *Curr. Top. Med. Chem.* 7, 475–488
4. Portevin, D., De Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffé, M., and Guilhot, C. (2004) *Proc. Natl. Acad. Sci. U.S.A.* 101, 314–319
5. Lea-Smith, D. J., Pyke, J. S., Tull, D., McConville, M. J., Coppel, R. L., and Crellin, P. K. (2007) *J. Biol. Chem.* 282, 11000–11008
6. Jackson, M., Raynaud, C., Lanéelle, M. A., Guilhot, C., Laurent-Winter, C., Ensergueix, D., Gicquel, B., and Daffé, M. (1999) *Mol. Microbiol.* 31, 1573–1587
7. Seidel, M., Alderwick, L. J., Birch, H. L., Sahn, H., Eggeling, L., and Besra, G. S. (2007) *J. Biol. Chem.* 282, 14729–14740
8. Mikusová, K., Yagi, T., Stern, R., McNeil, M. R., Besra, G. S., Crick, D. C., and Brennan, P. J. (2000) *J. Biol. Chem.* 275, 33890–33897
9. Kremer, L., Dover, L. G., Morehouse, C., Hitchin, P., Everett, M., Morris, H. R., Dell, A., Brennan, P. J., McNeil, M. R., Flaherty, C., Duncan, K., and Besra, G. S. (2001) *J. Biol. Chem.* 276, 26430–26440
10. Sasseti, C. M., Boyd, D. H., and Rubin, E. J. (2003) *Mol. Microbiol.* 48, 77–84
11. Takayama, K., Wang, C., and Besra, G. S. (2005) *Clin. Microbiol. Rev.* 18, 81–101
12. Parker, S. K., Barkley, R. M., Rino, J. G., and Vasil, M. L. (2009) *PLoS One* 4, e4281
13. West, N. P., Wozniak, T. M., Valenzuela, J., Feng, C. G., Sher, A., Ribeiro, J. M., and Britton, W. J. (2008) *Vaccine* 26, 3853–3859
14. West, N. P., Chow, F. M., Randall, E. J., Wu, J., Chen, J., Ribeiro, J. M., and Britton, W. J. (2009) *FASEB J.* 23, 1694–1704
15. Meniche, X., Labarre, C., de Sousa-d'Auria, C., Huc, E., Laval, F., Tropis, M., Bayan, N., Portevin, D., Guilhot, C., Daffé, M., and Houssin, C. (2009) *J. Bacteriol.* 191, 7323–7332
16. Hadváry, P., Lengsfeld, H., and Wolfer, H. (1988) *Biochem. J.* 256, 357–361

Structure/Function Analysis of a Mycobacterial Lipase

17. Hadváry, P., Sidler, W., Meister, W., Vetter, W., and Wolfer, H. (1991) *J. Biol. Chem.* **266**, 2021–2027
18. Pemble, C. W., 4th, Johnson, L. C., Kridel, S. J., and Lowther, W. T. (2007) *Nat. Struct. Mol. Biol.* **14**, 704–709
19. Kremer, L., de Chastellier, C., Dobson, G., Gibson, K. J., Bifani, P., Balor, S., Gorvel, J. P., Locht, C., Minnikin, D. E., and Besra, G. S. (2005) *Mol. Microbiol.* **57**, 1113–1126
20. Jacobs, W. R., Jr., Kalpana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G., and Bloom, B. R. (1991) *Methods Enzymol.* **204**, 537–555
21. Anderberg, R. J., Strachan, J. A., and Cangelosi, G. A. (1995) *BioTechniques* **18**, 217–219
22. Southern, E. M. (1974) *Anal. Biochem.* **62**, 317–318
23. Kjer-Nielsen, L., Clements, C. S., Purcell, A. W., Brooks, A. G., Whisstock, J. C., Burrows, S. R., McCluskey, J., and Rossjohn, J. (2003) *Immunity* **18**, 53–64
24. Winkler, U. K., and Stuckmann, M. (1979) *J. Bacteriol.* **138**, 663–670
25. Collaborative Computation Project No. 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
26. Leslie, A. G. W. (1992) *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26
27. Evans, P. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
28. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861
29. Ioerger, T. R., Holton, T., Christopher, J. A., and Sacchettini, J. C. (1999) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 130–137
30. Ioerger, T. R., and Sacchettini, J. C. (2003) *Methods Enzymol.* **374**, 244–270
31. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
33. Richardson, J. S. (2003) *Methods Biochem. Anal.* **44**, 305–320
34. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103–119
35. Ghosh, D., Erman, M., Sawicki, M., Lala, P., Weeks, D. R., Li, N., Pangborn, W., Thiel, D. J., Jörnvall, H., Gutierrez, R., and Eyzaguirre, J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 779–784
36. Masaki, K., Kamini, N. R., Ikeda, H., and Iefuji, H. (2005) *Appl. Environ. Microbiol.* **71**, 7548–7550
37. Payne, K., Sun, Q., Sacchettini, J., and Hatfull, G. F. (2009) *Mol. Microbiol.* **73**, 367–381
38. Martinez, C., De Geus, P., Lauwereys, M., Matthyssens, G., and Cambillau, C. (1992) *Nature* **356**, 615–618
39. Holmquist, M. (2000) *Curr. Protein Pept. Sci.* **1**, 209–235
40. Schrag, J. D., and Cygler, M. (1997) *Methods Enzymol.* **284**, 85–107
41. Guilhot, C., Gicquel, B., and Martín, C. (1992) *FEMS Microbiol. Lett.* **77**, 181–186
42. Brennan, P. J., and Nikaido, H. (1995) *Annu. Rev. Biochem.* **64**, 29–63
43. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) *Nucleic Acids Res.* **34**, W116–W118
44. Risler, J. L., Delorme, M. O., Delacroix, H., and Henaut, A. (1988) *J. Mol. Biol.* **204**, 1019–1029
45. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23**, 2947–2948
46. Gouet, P., Courcelle, E., Stuart, D. I., and Métoz, F. (1999) *Bioinformatics* **15**, 305–308

