Identification and biochemical characterisation of genes involved in mycobacterial cell wall assembly

by

Vijaya Shankar Nataraj

A thesis submitted to

The University of Birmingham

For the Degree of Doctor of Philosophy

Institute of Infection and Microbiology

School of Biosciences

College of Life and Environmental Sciences

The University of Birmingham

Novermber 2014

Abstract

The cell wall of *Mycobacterium tuberculosis* has a very complex ultrastructure, consisting of mycolic acids, an array of polysaccharides and surface exposed antigenic glycolipids. These cell wall components play a vital role in pathogenicity and virulence and hence, are attractive drug targets. Recent advances in TB drug discovery have produced a plethora of candidate drug molecules, Many of them target mycolic acid biosynthesis and transport.

A major part of the thesis work consists of biochemical and structural characterisation of proteins involved in mycolic acids biosynthesis and transport using a combination of genetic and biophysical tools. Through deletion of *fabH,* its non-essentiality for growth and survival of mycobacteria was demonstrated, and suggested the possibility of a functional substitute.

Using an array of biophysical tools, MmpL3, the protein involved in the transport of mycolic acids, was characterised and preliminary structural studies revealed that MmpL3 exists as a dimer. The possibility of the formation of protein complex with MmpL3 at the core, to achieve the transport of mycolic acids was also investigated. Although no facilitator protein was identified, our studies characterised the genes co-localised in *mmpL3* cluster.

The latter part of the thesis deals with the identification and characterisation of glycosyltransferases involved in the biosynthesis of lipooligosaccharides (LOS) in *M. kansasii*. LOS's are surface exposed, highly polar antigenic glycolipids, present in several mycobacterial species. Using targeted gene deletion, a mutant strain defective in LOS production was obtained that has provided the first insights into LOS biosynthesis in *M. kansasii*.

Declaration

The work presented in this thesis was carried out in the School of Biosciences at the University of Birmingham, U.K., B15 2 TT during the period September 2011 to September 2014. The work in this thesis is original except where acknowledged by references.

No portion of the work is being, or has been submitted for a degree, diploma or any other qualification at any other University.

Acknowledgement

The most cherished part of thesis writing is acknowledging those who helped to make this happen. I have several people to thank for giving me the three amazing, most rewarding years of life. I thank Professor Gurdyal S Besra, my supervisor for granting me the opportunity to pursue my PhD studies. I sincerely thank him for his support, guidance and help. I would like to express my gratitude to my supervisor, Dr. Apoorva Bhatt, who has helped me through my difficult times in research and personal life. I thank him for all his support, kindness, guidance and patience. I thank him for making my PhD a pleasant experience and instilling confidence in me where required. I also thank both Del and Apoorva for making "Thesis writing", easier than many explained. I would like to express my gratitude to the Darwin Trust of Edinburgh, whose funding made this thesis work possible. I would also like to thank Professor David E Minnikin for his valuable guidance.

The time spent in 'Besra lab' has been very fulfilling both professionally and personally. I would like to thank Natacha for her help in chemistry related aspects and for advices she gave, which made decision making a lot easier. I also take this opportunity to thank Usha, who never said 'no' to share her expertise in biochemistry. I would like to thank Albel, for his help with photography that appear in this thesis and helping me with several experiments. I take this opportunity to thank Sid, Luke and Klaus for teaching me several techniques. I would like to thank my Cat#3 partner Monika; together we killed so much of TB cultures. I would like to thank our badminton group members, Sid, Albel, Cristian, Pete, Oona and Houdini for their help to keep me refreshed. I would also like to thank other members of the lab Jon, Kat, Sarah, George, Asma, Panchali and Podric. I have had some wonderful moments with you guys. I also take this opportunity to thank Shipra and Mimi.

I would like to express thanks to Dr. Vassily Bavro (School of Biosciences, University of Birmingham) for his help with modelling and bioinformatics studies. I would like thank Dr. Klaus Futterer (School of Biosciences, University of Birmingham) for his help in protein crystallography. I also take this opportunity to thank our collaborator Prof. Anne Dell (Imperial College London, London) and her group for their help in MALDI-MS. I would like to thank Prof. Tim Dafforn, Dr. Sarah Lee and Rosemary Parlow for their help in protein purification and biophysical studies. I acknowledge the help from Dr. Cleidi Zampronio for protein sequencing. I also thank Prof. Jeff Cole for his guidance.

I have list of people to remember here, without them I wouldn't have even completed my secondary education. I must thank Gopalakrishna uncle (Thermo traders) and Bhaskar uncle (Ferrotech) who financially supported my secondary schooling during the difficult times. I also thank my teachers, in particular Head master (Peenya School), G. Premalatha More, S. Shanthabai, V.P. Dwarukaradhya, B.K. Shivamurthy and B.M. Matapathi, who went out of their way to help me when required. I also thank C. Shyamasundar sir and Y.N Chowdegowda sir, for their help to complete Class 12 studies. I thank, V.R. Devraj sir, who still remains as one of my well-wisher, guide and mentor. I take this opportunity to thank Dr. Umender K Sharma, my first supervisor at AstraZeneca and a great personality. His guidance has been instrumental in achieving my dream and I owe him a lot. I thank Vijaya Bank and Mr. Krishnayya Shetti for providing me the initial financial support to get started with my PhD.

Finally, I thank all my friends for their support and encouragement. I would like to thank Deepthi, for her constant encouragement and motivation. I express my gratitude towards my brother Guru and sister-in-law Kavitha for their support and encouragement. I thank Sujatha, Nalini and Jaishree aunty for their help and support. My grandfather, Narayana Rao, a big inspiration in my life and I thank him and Ajji for unconditional love and support.

Last but not the least, this thesis would not have been possible without the support and encouragement of my wife. You have been my spine during the difficult times and I thank you- Gayathri, for your unconditional love, patience and support.

This thesis is dedicated to my parents

Uma N and K.V. Nataraj

Content

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 *Mycobacterium* species

Phylogenetically, mycobacteria belong to the eubacteria or actinomycete group. These are a large group of Gram positive bacteria and are characterised by their high GC content.Within this group, mycobacteria belong to one branch, termed the CMN, which stands for *Mycobacterium*, *Corynebacterium* and *Nocordia*, owing to their close similarity in cell wall ultra structure (Brennan & Nikaido 1995). The cell wall of this group of bacteria is characterised by the presence of *meso*-diaminopimellic acid and N-glycolylated muramic acid in peptidoglycan. In addition, the presence of the cell wall polysaccharide arabinogalactan (AG) and mycolic acids distinguishes this group of bacteria from others (Lederer *et al*. 1975; Brennan & Nikaido 1995). To put into perspective the complexity of the mycobacterial cell wall, the aforementioned unique components make up majority of the cell wall.

Mycobacteria are typically unicellular, aerobic, rod shaped to curved rods or cocco-bacilli. They vary between 1 to 4 μm in length and 0.2 to 0.6 μm in width (Wayne and Kubica, 1986). The cells take up Carbol fuchsin when stained by the Ziehl-Neelsen method due to the presence of high amounts of mycolic acids in their cell wall and are hence grouped under acid-fast bacteria (Goren *et al*. 1978). Mycolic acids along with other unique complex lipids constitute 60% of the cell wall, rendering it highly hydrophobic, endowing the organism with a permeability barrier and innate resistance to therapeutic agents and host defences (Acharya & Goldman 1970; Minnikin *et al*. 2002; Karakousis *et al*. 2004).

The mycobacterial cell wall core is composed of peptidoglycan which is covalently linked to the mycolyl arabinogalactan to form the mycolyl-arabinogalactan- peptidoglycan (mAGP complex) (McNeil *et al*. 1990; Kremer *et al*. 2003; Dover *et al*. 2004; Brennan 2003; Besra *et al*. 1995) and the outer region consists of glycolipids, polysaccharides, lipoglycans, which intercalate within the mycolic acid layer (Brennan &Nikaido 1995; Brennan & Crick

2007). Thus, this unique cell wall is also an attractive drug target for antimycobacterials (Banerjee et al. 1994; Forbes et al. 1962).

The unique cell wall of the mycobacteria renders pathogenicity to the genera; however, not all the species are pathogenic. Well known human pathogens of the genera include *M. tuberculosis* that causes TB and *M. leprae* that causes Leprosy. The genera also include opportunistic human pathogens that cause TB like disease in humans. These include *M. avium*, *M. canetti* and *M. kansasii.* Further, mycobacterial species are divided into fast and slow growers depending on their generation time. Table 1-1 lists fast growing and slow growing mycobacterial species.

Table 1-1 List of fast and slow growing Mycobacteria

Fast growing mycobacteria	**Slow growing mycobacteria**
Mycobacterium smegmatis	*Mycobacterium tuberculosis*
Mycobacterium aurum	*Mycobacterium bovis* BCG
Mycobacterium chelonae	*Mycobacterium canetti*
Mycobacterium duvalii	*Mycobacterium kansasii*
Mycobacterium flavescens	*Mycobacterium avium*
Mycobacterium fortuitum	*Mycobacterium malmoense*
Mycobacterium gadium	*Mycobacterium ulcerans*
Mycobacterium komossense	*Mycobacterium gastri*
Mycobacterium phlei	*Mycobacterium genavense*
Mycobacterium porcinum	*Mycobacterium haemophilum*
Mycobacterium rhodesiae	*Mycobacterium intercellulare*
Mycobacterium shinshuense	*Mycobacterium leprae*
Mycobacterium vaccae	*Mycobacterium africanum*
	Mycobacterium marinum
	Mycobacterium microti
	Mycobacterium paratuberculosis
	Mycobacterium scrofulaceum
	Mycobacterium xenopi
	Mycobacterium simiae
	Mycobacterium szulgai
	Mycobacterium farcinogenes

1.2 Tuberculosis

Tuberculosis (TB) is the oldest known infectious disease and is the second leading cause of death from an infectious disease worldwide, after the human immune deficiency virus (HIV). TB is caused by the bacillus *M. tuberculosis.* It typically affects the lungs (pulmonary tuberculosis), but can also affect other sites as well (extra pulmonary tuberculosis) (WHO, 2013). Pathogenesis of TB will be discussed in detail in section 1.4.

Molecular and biochemical tools have helped to identify TB infection from various archaeological samples. Early studies were carried out using isolation and identification of *M. tuberculosis* specific DNA from the Egyptian and Peruvian mummies (Salo et al. 1994; Crubézy et al. 1998). Later, use of cell wall lipids as biomarkers has helped identify TB infection from 9000 year old mummy and a 17,000 year old bison skeleton (Donoghue 2009; Hershkovitz et al. 2008; Rothschild et al. 2001; Lee et al. 2012).

Although molecular studies points at existence of TB from 17,000 years ago, it was only in 1882 that the etiological agent of TB was first identified. German physician Robert Koch discovered the etiological agent of TB and also introduced 'sanitoria' as a preventive measure (Koch 1882; Cox 1923).

1.3 Epidemiology

Although effective chemotherapy has helped control TB over the years, TB stills remains a global burden due to high mortality rates. 9.3 million new TB cases were reported in 2012 with 1.3 million deaths due to TB with lowest rates of incidence (0-4%) reported from developed countries including North America, Western Europe and Asia pacific countries. The highest rates of incidence were reported from sub-saharan Africa (>50%), India and China (20-49%). Intermediate rates of incidence (5-19%) were reported from Central and South America and Eastern Europe (WHO 2013).

TB co-infection with HIV remains a underlying threat in African countries, which account for nearly 75% of the estimated TB-HIV co-infection incidence globally (WHO 2013). HIV has raised the risk of developing latent infections into active infection by 10-fold which has a direct impact on mortality due to both the diseases (Cole *et al*. 1998; Corbett *et al*. 2003)

1.4 Pathogenesis of tuberculosis

Pathogenic bacteria adopt different ways to stably infect their host and one of the effective ways is to reside intracellularly, which protects the pathogen from a variety of humoral defense mechanisms (Kaufmann, 1993). Our early understanding of *M. tuberculosis* pathogenicity comes from work conducted on rabbit models (Lurie, 1964). *M. tuberculosis* primarily infects the lungs and resides in alveolar macrophages (Ernst, 1998). The infection is transmitted through aerosols and post inhalation, the bacilli could face a number of fates; a strong host immune system may clear the bacilli resulting in no infection or in immunocompromised individuals, the pathogen could establish disease or could evade the immune response and enter a dormant stage resulting in latency (Long & Schwartzman 2014).

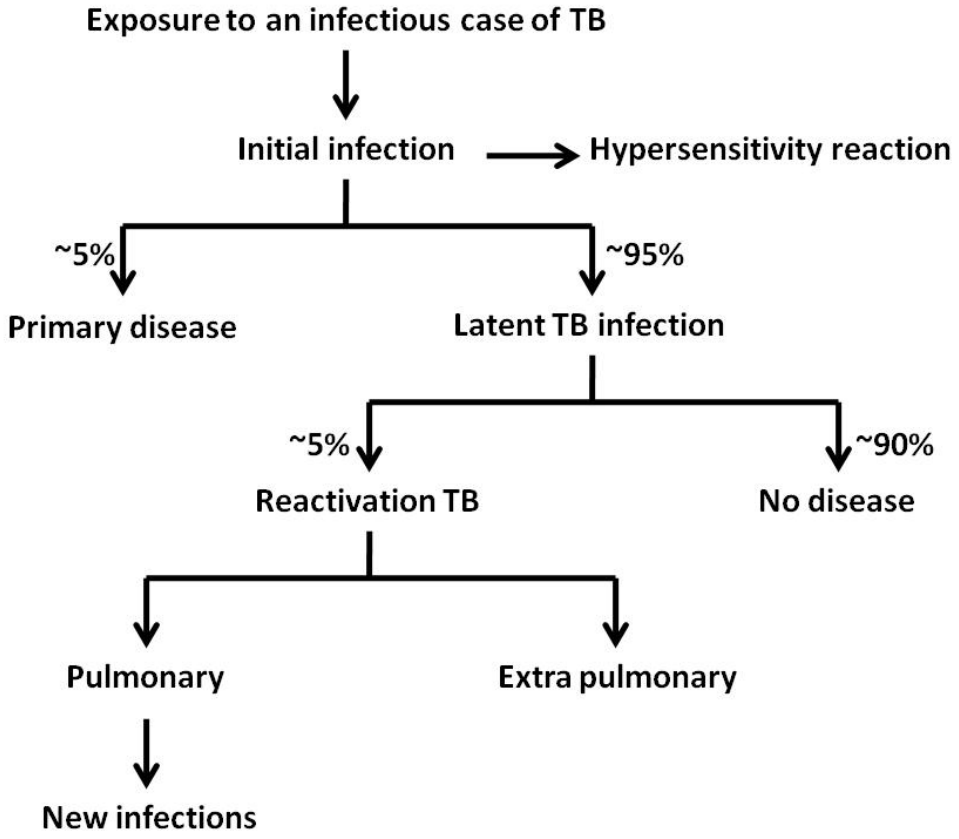


Figure 1-1Pathogenesis of Tuberculosis in the infected host

In the early stages of infection, the pathogen multiplies within an infected macrophage until the macrophage bursts and releases the bacilli (Long & Schwartzman 2014; Russell, 2007). The released bacilli attract more macrophages causing a localised pro-inflammatory response recruiting mononuclear cells and blood vessels to form a granuloma, a characteristic of *M. tuberculosis* infection (Russell, 2007; Podinovskaia *et al*. 2013; Long & Schwartzman 2014). The granuloma consists of three layers with a core consisting of infected macrophages, surrounded by foamy macrophages and an outer layer composed of lymphocytes and extracellular components (Russell, 2007). The bacilli can remain in this condition to cause latent infection or within a weakened host, could result in rupture of the

granuloma and establish an infection. The eventual spread of the bacilli to the rest of the lungs or other parts of the body is the final stages of the disease (Dannenberg & Rook 1994).

Phagosome maturation and subsequent fusion with lysozome has been shown to be modulated by several factors. These include some cell wall components, such as lipo-arabinomannan (LAM) (Sweet *et al*. 2010), phenolic glycolipid (PGL-1) (Robinson *et al*. 2007) and trehalose dimycolate (TDM) (Indrigo *et al*. 2003) and secreted proteins, such as the serine/threonine kinase PknG (Walburger *et al*. 2004). This process of arresting phagosome maturation is vital for the pathogen to evade the natural immunological response and establish a stable infection.

1.5 Anti TB drugs

1.5.1 History of TB drug discovery

In the early 1940's streptomycin was first identified to have a bacteriostatic effecton bacilliby Feldman and Hinshaw (Anon 1946; Jones *et al*. 1944) and was used for treating TB. Later, *p*-aminosalicylic acid (PAS) was used as a drug in isolation and in combination with streptomycin to treat TB (Graessle & Pietrowski 1949; Bernheim 1940; Yegian & Vanderlinde 1948; Anon 1950). The major breakthrough in TB treatment came with the discovery of isoniazid (INH) in the early 1950's (Steenken & Wolinsky 1952). The next decade saw the development of several TB drugs including pyrazinamide (Muschenheim *et al*. 1954; McDermott *et al*. 1954; Tompsett *et al*. 1954; Yeager *et al*. 1952), cycloserine (Albouy *et al*. 1955), ethambutol (Thomas *et al*. 1961; Forbes *et al*. 1962) and rifampin (Sensi 1983). The current treatment involves administration of INH in combination with ethambutol, pyrazinamide, rifampin and in some cases streptomycin (Espinal *et al*. 2000; WHO 2012).

Over the years the mechanisms of action of these drugs have been elucidated and also the resistance have been mapped, which is discussed briefly below. INH and ethionamide target the biosynthesis of mycolic acids by inhibiting the enoyl-ACP reductase (InhA) which is part of fatty acid synthase-II (FAS-II) (Winder & Collins 1970; Banerjee *et al*. 1998; Kremer *et al*. 2003; Quémard *et al*. 1995). However, both are *pro*-drugs and need to be acted up on by catalase-peroxidase (*katG*) and a monooxygenase (*ethA*), respectively for conversion into an active metabolite (Zhang *et al*. 1992; Baulard *et al*. 2000; Khasnobis *et al*. 2002). INH is converted to an isonicotinic acyl radical by KatG, which then reacts with nicotinamide dinucleotide (NAD) to form an INH-NAD adduct (Broussy *et al*. 2003; Rozwarski *et al*. 1998). Similarly, prothionamide (*pro*-drug of ehionamide) is converted to its active form 2-ethyl-4-aminopyridine by EthA, which abolishes mycolic acid biosynthesis through inhibiting InhA (Banerjee *et al*. 1994). The resistance to these compounds have been mapped using clinical isolates to mutations in *inhA, katG* and *ndhII* in case of INH and *ethA* and *ethR* in case of ETH (transcriptional regulator of *ethA*) (Zhang *et al*. 1992; Banerjee *et al*. 1994; Morlock *et al*. 2003; Dover *et al*. 2004). Although INH and ETH are structural analogues and target the same enzyme, cross-resistance is relatively low which is explained by divergent activation mechanisms (Fattorini *et al*. 1999; Banerjee *et al*. 1998; Baulard *et al*. 2000). The low cross-resistance observed was initially thought due to altered NADH/NAD ratios but was later surprisingly mapped to the mycothiol biosynthesis pathway (Vilchèze *et al*. 2005; Vilchèze *et al*. 2008; Vilchèze *et al*. 2011).

Rifampin is a semi-synthetic derivative of rifamycin B and inhibits the β-subunit of RNA polymerase (RpoB), and is bactericidal against both active and slow-growing bacilli. Rifamycin was shown to sterically inhibit translocation step in the synthesis of nascent mRNA by binding to RpoB subunit, bringning an abrupt end after transcribing two nucleotides (Kohanski *et al*., 2010; McClure *et al*., 1978). Resistance to RIF has been

mapped to *rpoB* (Taniguchi *et al*. 1996). Ethambutol is a synthetic drug that was initially shown to inhibit the transfer of mycolic acids to the cell wall with a concomitant increase of secreted trehalose monomycolate (TMM) and trehalose dimycolate (TDM) (Kilburn & Takayama 1981). Later it was shown that ethambutol primarily affects arabinogalactan biosynthsis by blocking the enzymatic activity of EmbB, which is involved in the formation of the terminal hexaarabinofuranoside motif of arabinogalactan, thus affecting mycolylation (Escuyer *et al*. 2001; Belanger *et al*. 1996). Resistance to ethambutol was mapped to mutations in *embB* (M306V/L) using drug resistant *M. tuberculosis* isolates (Ramaswamy *et al*. 2004).

Another first line drug pyrazinamide (PZA),which is also a *pro*-drug is converted into pyrazinoic acid (POA) by pyrazinamidase (Pzase) (Frothingham *et al*. 1996). PZA is most active against older cultures at an acidic pH and is believed to bring about cell death through affecting membrane potential by destructing the proton motive force (Zhang *et al*. 2003). Recent studies have shown that pyrazinamide brings about cell death by inhibiting trans-translation process by blocking the action of ribosomal protein S1(RpsA) and pyrazinamide but not pyrazanoic acid, is shown to competitively bind FAS-I and inhibiting mycolic acid biosynthesis (Shi *et al*., 2011; Sayani *et al*., 2011). However, the resistance to pyrazinamide is mapped to *pncA* geneTable 1-22 details the anti-mycobacterial drugs and their target genes, and genes involved in resistance.

As described above these drugs are used in combination to treat modern day TB. The standard regimen involves a two phase therapy designed to combat actively growing and dormant bacilli. The first phase employs treatment with first line anti-TB drugs for two months followed by four months of a continuous phase. The first line anti-tubercular drugs are INH, EMB, PZA and RIF (Table 1-2). The second line drugs include aminoglycosides

(e.g.kanamycin), polypeptides (e.g. capreomycin), fluoroquinolones (e.g. ciprofloxacin), thioamides (e.g. ethionamide), cycloserines and *p*-aminosalicylic acid. Rifabutin and marcolides constitutes third line drugs which are sparsely used and are less effective (Janin 2007).

Table 1-2Anti-TB drugs and their target pathway.

Drug	**MIC (µg/ml)**	**Effects on Bacterial cell**	**Target pathway**	**Genes involved in resistance**
Isoniazid (INH)	≤0.25	Bactericidal	Mycolic acid biosynthesis	*katG*/*inhA*/*ndh*
Ehionamide	0.6-2.5	Bacteriostatic	Mycolic acid biosynthesis	*inhA/ethA*
Rifampin (RIF)	≤0.125	Bactericidal	RNA synthesis	*rpoB*
Ethambutol (EMB)	2.0	Bacteriostatic	Cell wall arabinan biosynthesis	*embCAB*
Pyrazinamide (PZA)	50-100	Bactericidal/ Bacteriostatic	Disrupts membrane transport and energy depletion	*pncA*

1.5.2 DOTS

In 1991, World Health Organization (WHO) introduced "Directly Observed Therapy, Short-course" (DOTS) to combat TB in an efficient way (Obermeyer *et al*. 2008). DOTS focus on the logistics of TB treatment along with actual chemotherapy. The logistics involve commitment from Government policy towards controlling TB, the supply of drugs and standardised reporting and recording of cases and outcome of treatment. Chemotherapy involves diagnosis based on sputum smear microscopy and DOTs administration. In 2006, the "Stop TB" campaign was initiated to focus on aneffective implementation of DOTS and to increase awareness of drug resistance. DOTS chemotherapy involves administration of a cocktail of drugs for a minimum of 6 months for drug susceptible TB cases and up to 2 years for drug resistant TB cases. The current DOTS treatment consists of administration of INH,

EMB, RIF and PZA daily for two months followed by INH and RIF thrice weekly for another 4 months (Cox *et al*. 2006).

1.5.3 Drug resistance

Multi drug resistant (MDR) TB exhibits drug resistance towards at least two of the first line drugs INH and RIF. MDR-TB could result from either through primary infection with resistant bacteria or developed during the treatment phase. Non-compliance with the treatment regime and/or short drug supply, are the two important reasons for developing MDR-TB. Recent studies have shown that the pathogen could develop antibiotic resistance through mutations in the chromosomal targets rendering anti-TB agents ineffective (Cole 1994). Globally, 3.7% of new cases and 20% of previously treated cases are estimated to be MDR-TB, of which 60% coming from India, China and the Russian Federation. These countries have reported a worrying 9-32% of new cases and nearly 50% of the previously treated cases to be MDR-TB. Treatment of MDR-TB is both expensive and for prolonged

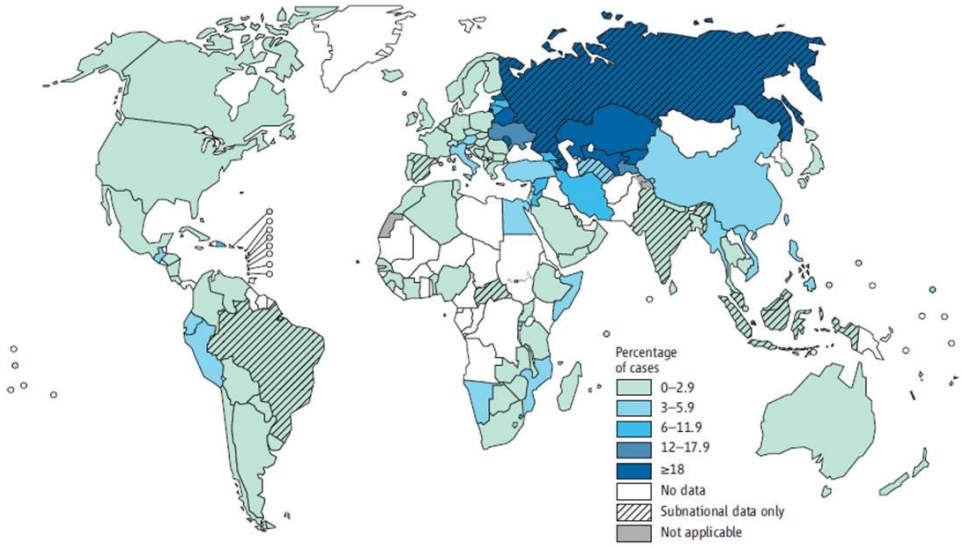


Figure 1-2 Map showing percentage of new cases with MDR-TB in 2012, by country (*adopted from* WHO 2013).

duration, and requires more toxic second line anti-TB drugs. The treatment regimen (DOTS-Plus-an alternative to DOTS strategy) lasts close to 18 months post smear negative using 4-6 second line drugs selected on their efficacy and susceptibility testing (WHO 2013; Shin *et al*. 2006).

Extensively drug resistant TB (XDR-TB) is an additionally challenging and a more dangerous variant. XDR-TB strains are resistant to INH, RIF and any of the fluoroquinolones along with resistance to at least one of amikamycin, kanamycin and capreomcyin, which are the injectable second line anti-TB drugs (Raviglione 2006). XDR-TB has been reported from 92 countries globally and an estimated 9.6% of MDR-TB cases have XDR-TB, with a mortality rate of nearly 50% (WHO 2013). In December 2011, a new strain was reported from India which exhibited resistance to all of the first-line and the second-line drugs and is termed Totally Drug Resistance TB (TDR-TB), which poses further challenges to TB treatment.

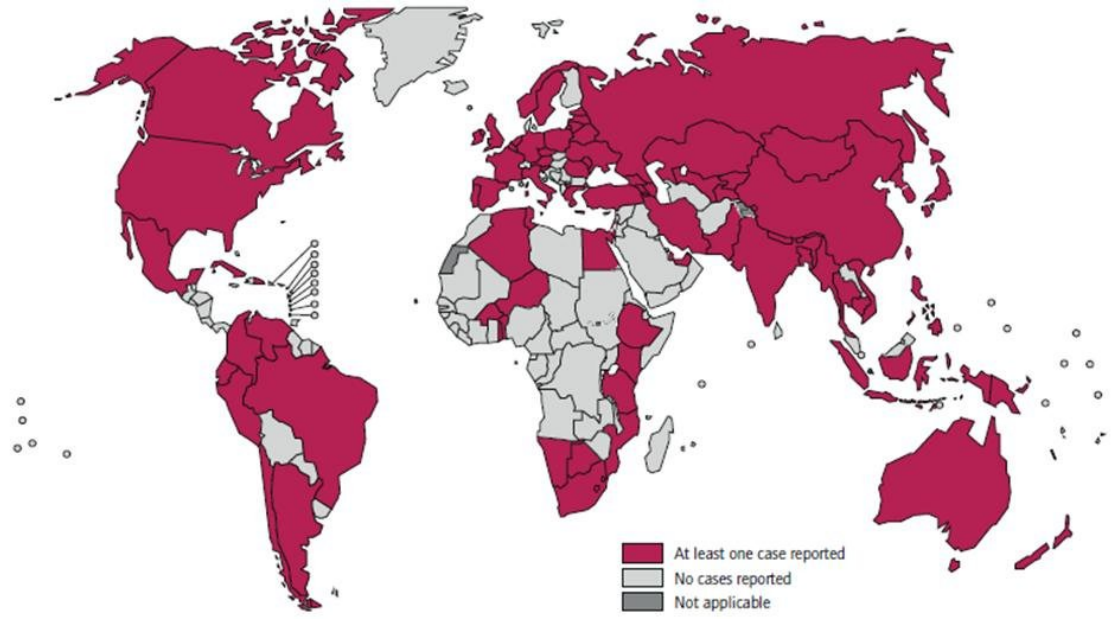


Figure 1-3Map showing distribution of countries that have reported at least one case of XDR-TB (*adopted from* WHO, 2013)

1.5.4 New drugs and combination therapies in the pipeline

Research over the years has shown that the unique mycobacterial cell wall is one of the most attractive drug targets and several candidate drugs have now been identified. Apart from the development of new drugs, a new combination regimen NC-001 involving PA-824, moxifloxacin and pyrazinamide is in phase-II and has showed encouraging results (WHO 2012). Currently there are 4 drugs in phase-III and 11 in phase-II clinical trials. Fluoroquinalones, gatifloxacin and moxifloxacin are all in phase-III and are shown to inhibit DNA supercoiling in mycobacteria by inhibiting DNA gyrase (Tomioka *et al*. 1993; Sriram *et al*. 2006; Burman 2010). OPC-67683 is a *pro*-drug and is acted up on by Rv3547 to form an active compound that inhibits mycolic acid biosynthesis and thereby bringing about cell death. OPC-67683 has a MIC of 0.006-0.024 μM against drug susceptible and MDR-TB strain (Matsumoto *et al*. 2006). This compound is currently in phase-III clinical trials (WHO 2012). The diarylquinoline drug TMC-207 is currently being used in monotherapy (provisional FDA licence), kills the bacteria by targeting proton pump of the adenosine triphosphate (ATP) synthase (Andries *et al*. 2005; Sloan *et al*. 2013). Chemically SQ109 has the same core as that of ethambutol and targets mycolic acid transport by binding to MmpL3 (Mycobacterial Membrane Protein Large 3) (Protopopova *et al*. 2005; Tahlan *et al*. 2012). Recently, a sulfur containing compound has shown bactericidal activity against mycobacteria with a high specificity and potency. Benzothiozionone (BTZ) has a MIC value of 1-30 ng/ml and is currently in the pre-clinical phase. BTZ binds to DprE1 and consequently blocks arabinan biosynthesis and cell death is brought about by blocking the recycling of decaprenyl phosphate (Makarov *et al*. 2011; Batt *et al*. 2012; WHO 2012; Grover *et al*. 2014). Phase-III clinical trials testing gatifloxacin and moxifloxacin were expected to finish by the end of 2013 and data will be made available shortly which should these new drugs available to treat TB (WHO 2013). Importantly all of the above discussed drugs have shown good potency

against even the drug resistant strains of TB (Makarov *et al*. 2011; Andries *et al*. 2005; Protopopova *et al*. 2005; Matsumoto *et al*. 2006; Sriram *et al*. 2006; WHO 2012; WHO 2013). Apart from new drugs, there are number of vaccines have been proposed and are currently in different phases of vaccine development (WHO 2012; WHO 2013).

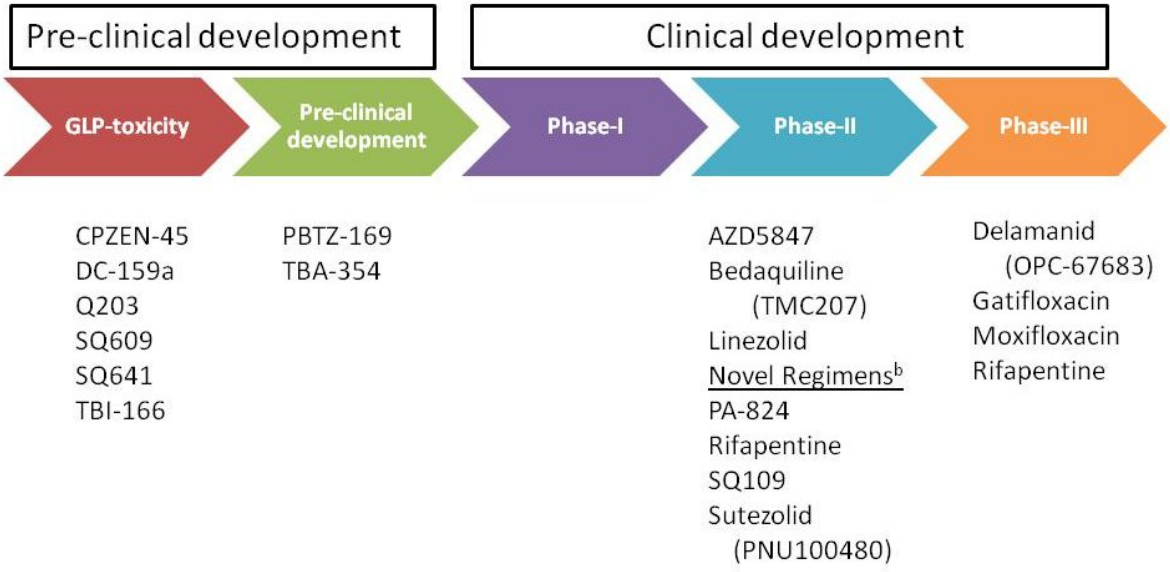


Figure 1-4 Development pipeline for new TB drugs.

1.6 The mycobacterial cell wall

The mycobacterial cell wall is composed of unique lipids and polysaccharides. Although mycobacteria are grouped under Gram positive bacteria the cell envelope differs from those of others in the group. The cell wall is unusually lipid rich, extremely hydrophobic and contributes low permeability to hydrophilic drugs and rendering the organism resistant to therapeutic agents (Brennan & Nikaido 1995). The cell wall of mycobacteria is very complex in its arrangement and for the ease of understanding could be split into three main domains (Figure 1-5);

- Cytoplasmic membrane
- Cell wall matrix composed of a mycolyl-arabinogalactan-peptidoglycan (mAGP) complex
- Other non-covalently associated polysaccharides, phospholipids, glycolipids, waxes and proteins.

The cell wall skeleton is made up of peptidoglycan (PG) covalently linked to arabinogalactan (AG) which is esterified at the distal end to mycolic acids (Brennan 2003) forming the mAGP complex mesh like structure. Solvent extractable lipids intercalate the mAGP core. In some cases the lipid rich cell wall is enveloped by a sugar rich capsule layer. The composition of the capsule varies from species to species. For example, capsule found in *M. tuberculosis*, *M. kansasii* and *M. gastri* are polysaccharide rich whereas capsules of *M. smegmatis* and *M. phlei* are mainly proteins. The capsule forms a interface between the bacteria and its environment (Daffé & Etienne 1999; Brennan & Nikaido 1995; Minnikin *et al*. 2002; Daffé & Lanée 2001).

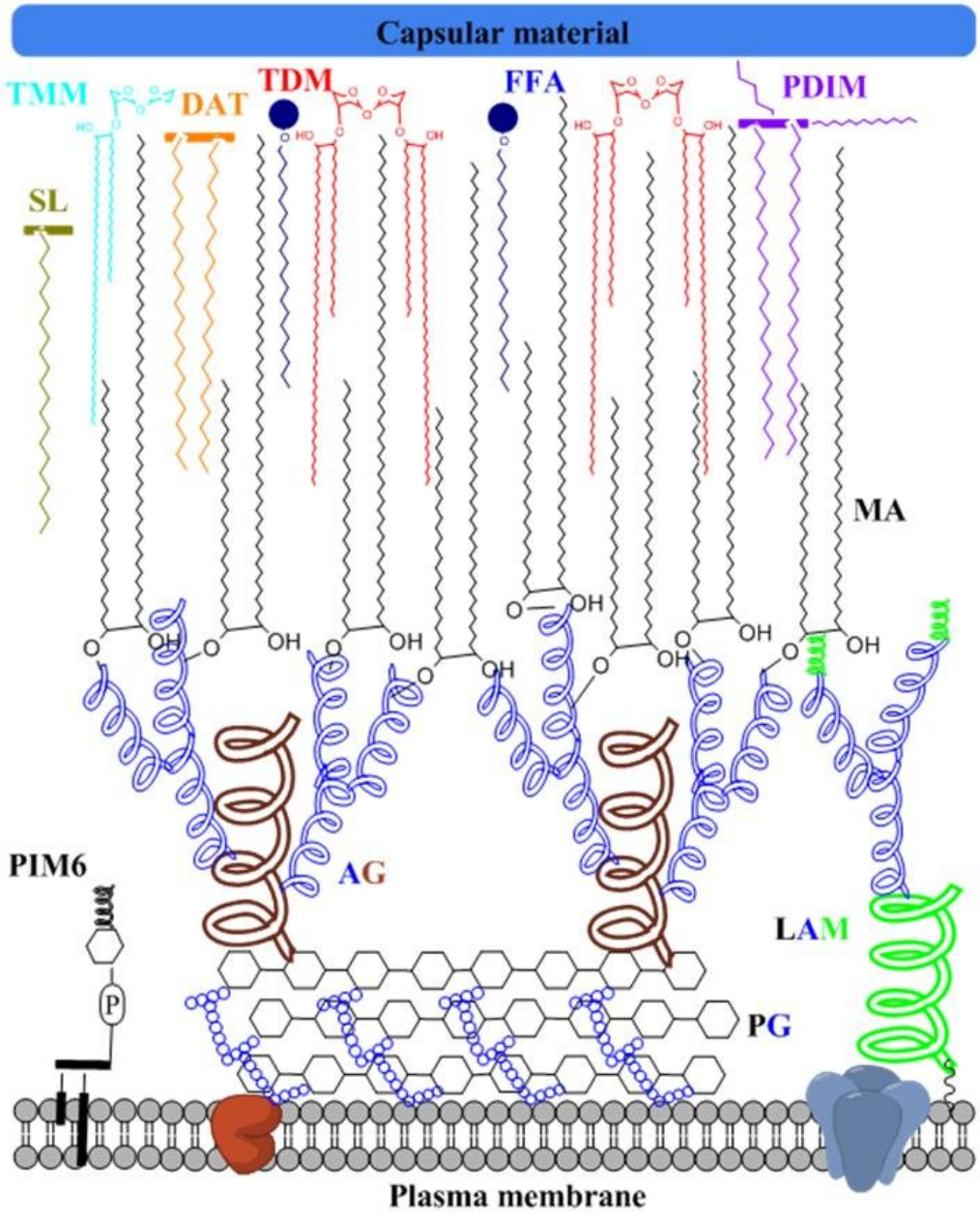


Figure 1-5 Arrangement of structural components of cell envelope of *M. tuberculosis*.PG-Peptidoglycan; AG-Arabinogalactan; LAM-Lipoarabinogalactan; PIM6-Phospho inositol hexa mannose; MA-Mycolic acids; TMM-Trehalose monomycolate; TDM-trehalose dimycolate; PDIM-Pthiocerol dimycocerosate; SL-Sulfolipid; DAT-Diacyl trehalose; FFA- Free fatty acid.

The Figure 1-5 presents the architecture of mycobacterial cell wall. This was first proposed by Minnikin D E (1982) and further modified by McNeil and Brennan in 1991 and

Minnikin D E *et al*. in 2002 (Davidson *et al*. 1982; McNeil & Brennan 1991; Minnikin *et al*. 2002). According to the model proposed, the core cell wall contains covalently linked peptidoglycan, arabinogalactan and mycolic acid and other structural lipids forming the outer layer of the cell wall. Peptidoglycan associated with the plasma membrane on one side provides a polysaccharide anchor on the outer surface to which mycolic acids are attached (McNeil & Brennan 1991; Chatterjee *et al*. 1991). A highly ordered galactan is covalently linked to peptidoglycan and attached to a branched arabinan, which is esterified to mycolic acids. The compact structure of the mAGP complex results in a highly hydrophobic layer of extremely low fluidity (Besra & Brennan 1997). Phenolic glycolipids (PGL), phthioceroldimycocerosates (PDIM), trehalose dimycolate also known as cord factor (TDM), sulfolipids (SL) and phosphotidyl-inositolmannosides (PIM) constitutes outer layer of the membrane. The cell wall also consists of lipoarabinomannan (LAM) and lipomannan (LM) originating from plasma membrane extending out into the outer membrane. In the following section these cell wall components will be discussed in detail.

1.6.1 Plasma membrane

The plasma membrane of mycobacteria is fairly less complex and is composed of a phospholipid bilayer. The presence of unknown carbohydrates in the cell membrane has been reported, however these have not been characterised. This carbohydrates are believed to render the membrane uneven (Silva & Macedo 1983). One of the unique mycobacterial cell wall components phosphoinositol mannosides (PIMs) is found in the cell membrane. The inositol moiety is covalently linked to a glycerophospholipid in the cell membrane and is attached to mannose units through glycosidic linkages. The structures of higher homologues, PIM3, PIM4, PIM5 and PIM6 have all been elucidated (Lee & Ballou 1964; Brennan & Ballou 1967; Brennan & Nikaido 1995). Further PIMs also form the lipid base for lipoarabinomannan (LAM) and lipomannan (LM) (Brennan & Nikaido 1995). These will be

discussed in detail below. Other components associated with the plasma membrane include polyterpene based 'carotinoids' (which give a characteristic yellow-orange colour for photochromogenic mycobacteria, such as *M. gordonae* and *M. kansasii*) and menaquinones involved in electron transport, glycosylphosphopolyprenols, such as β-D-mannopyranosyl phosphodecaprenol and β-D-arabinofuranosyl phosphodecaprenol (Takayama & Goldman 1970; Wolucka *et al*. 1994; Brennan & Nikaido 1995).

1.6.2 Peptidoglycan

Mycobacterial peptidoglycan consists of a linear chain consisting of N-acetyl glucosamine (GlcNAc) and muramic acid (Mur) residues linked *via* β-(1→4) linkages (Schleifer & Kandler 1972). The muramic acid is branched with a pentapeptide side-chain consisting of L-alanyl-D-isoglutaminyl-*meso*-diaminopimelyl-D-alanine and individual monomers cross-linked through covalent linkage between *meso*-diaminepimelic acid (mDAP) and D-alanine residue. This type of cross-linking is observed in A1-type peptidoglycan and the presence of amidated muramic acid puts it into the A1γ class (Lederer *et al*. 1975; Schleifer & Kandler 1972; Brennan & Nikaido 1995).

Further studies have reported additional modifications in the peptidoglycan structure. Oxidation of N-acetyl to N-glycolyl forms of muramic acid is interesting, as it allows for extensive hydrogen bonding and helps in tightening the peptidoglycan structure (Brennan & Nikaido 1995; Azuma *et al*. 1970; Raymond *et al*. 2005). However, it is the mDAP that plays a key role in extensive cross-linking to form a mesh like structure. Apart from cross linking between mDAP and D-alanine, an additional cross-link is observed between two mDAP residues (Ghuysen *et al*. 1968; Wietzerbin *et al*. 1974). Further, these mDAP-mDAP linkages are extensive as the organism enters stationary phase (Wietzerbin *et al*. 1974). This results in a very high percentage of cross-linking (70-80%) compared to *E.coli* (20-30%) (Usha *et al*.

2012). The extensively cross linked peptidoglycan forms the principal structural framework of the mycobacterial cell wall on which a complex lipophillic matrix is built (Alderwick *et al*. 2007).

1.6.3 Arabinogalactan

Arabinogalactan (AG) is a complex heteropolysaccharide structure that covalently anchors the outer lipid layer to the inner peptidoglycan. Mass spectroscopic studies had helped early understanding of the sugar composition of arabinogalactan, however it was not until the late 90's that the complete structure of this polysaccharide was established (Figure 1-6) (Azuma *et al*. 1970; Kanetsuna *et al*. 1968; McNeil *et al*. 1987; Besra *et al*. 1995). Arabinogalactan is attached to peptidoglycan through a linker unit consisting of L-Rha*p*-(1→4)-α-D-GlcNAc (McNeil et al. 1990; Besra et al. 1995). Arabinogalactan is synthesised on a decaprenyl phosphate (C_{50}-P) lipid carrier prior to its transfer to the muramic acid residue of peptidoglycan. N-acetyl glucosamine and rhamnose residues are sequentially added on to decaprenyl phosphate (C50-P-P-GlcNAc-Rha) to provide a base for assembly of the linear galactan chain (Mikusová *et al*. 1996; Mills *et al*. 2004; Alderwick *et al*. 2006; Jankute *et al*. 2012). The linear galactan chain is composed of β-D-galactofuranose residues linked through alternative β(1→5) and β(1→6) linkages (Belánová *et al*. 2008; Szczepina *et al*. 2010). The galactan chain is primed at position 8, 10 and 12 with a arabinofuronosyl residue, which is further elongated (α-1, 5) and branched (α-1, 3) (Besra *et al*. 1995; Birch *et al*. 2008). Chain termination occurs with a unique hexaarabinofuranosyl (Ara_6) motif with 2-β-D-arabinofuranose and α-D-arabinofuranose residues providing the sites for mycolylation (Alderwick *et al*. 2006; Jankute *et al*. 2012; McNeil *et al*. 1994a).

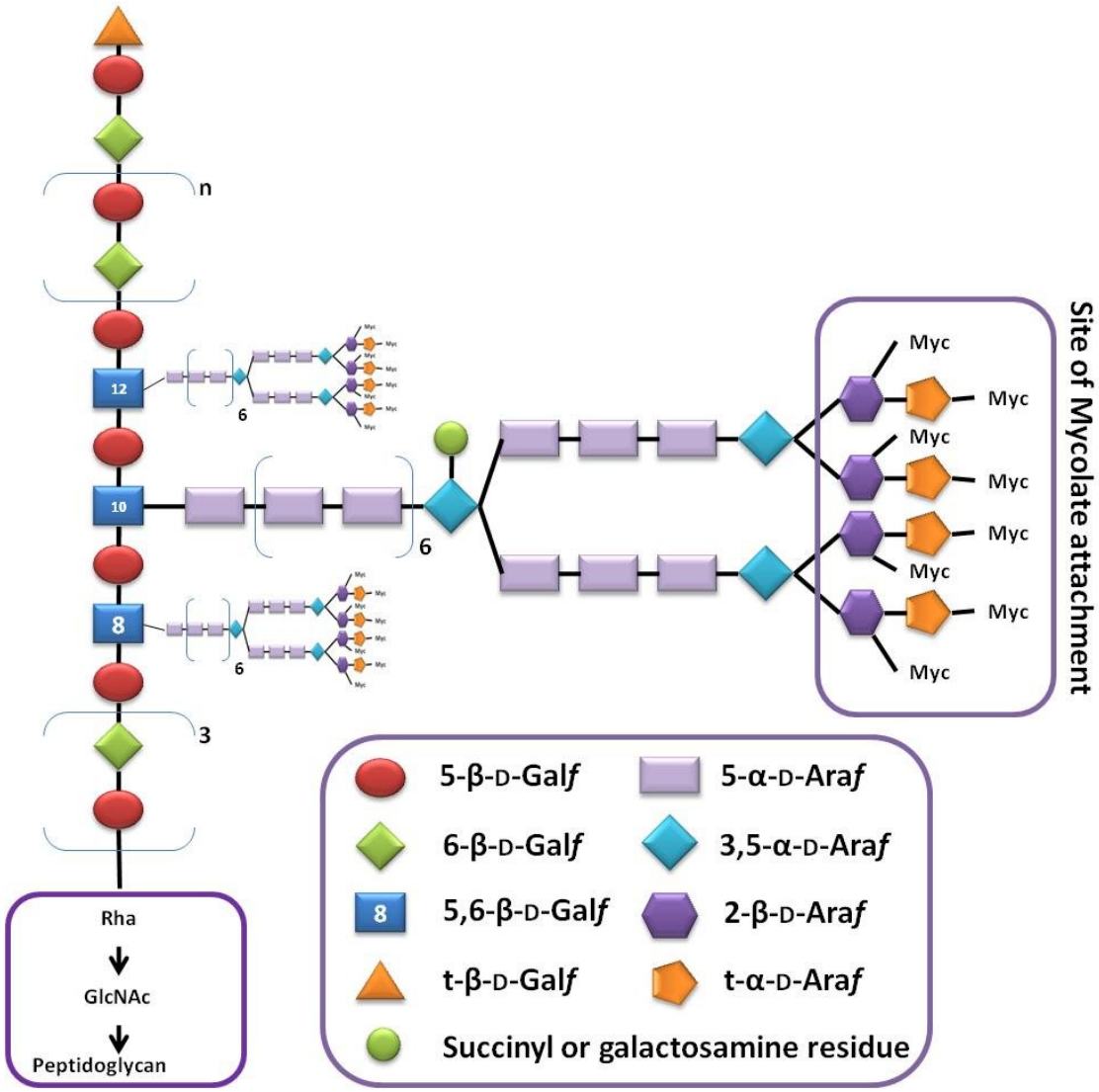


Figure 1-6 Essential structural topography of *M.tuberculosis* arabinogalactan-peptidoglycan.

1.6.4 Mycolic acids

As defined by Asselineau and Lederer, mycolic acids are α-alkyl, β-hydroxy fatty acids (C_{70}-C_{90}) and are essential components of the mycobacterium cell wall (Asselineau & Lederer 1950; Takayama *et al*. 2005). The structure of mycolic acids and functional groups associated with it, were elucidated during the 1960's by concerted efforts of several research groups (Lannelle 1963). Mycolic acids could be classified depending on the chemical modifications

observed in the meromycolate chain or the main chain of the mycolic acid. The non-oxygenated mycolates are termed as α-mycolates. This is the abundant form (>70%) found in the cell wall of *M. tuberculosis*. The oxygenated mycolates sub-divided into methoxy and keto-mycolates, are less abundant form (10 to15%). Presence of a methyl branch adjacent to cyclopropane ring, *cis* or *trans* double bonds, provide additional modifications to the oxygentated and non-oxygenated mycolic acids (Minnikin *et al*. 1982). The functional group forms the basis of classification of mycolic acids and is useful in taxonomic identification. For example, the *M. tuberculosis* cell wall has α, methoxy and ketomycolates while *M. smegmatis* has α, α' and epoxy mycolates (Davidson *et al*. 1982; Yuan & Barry 1996).

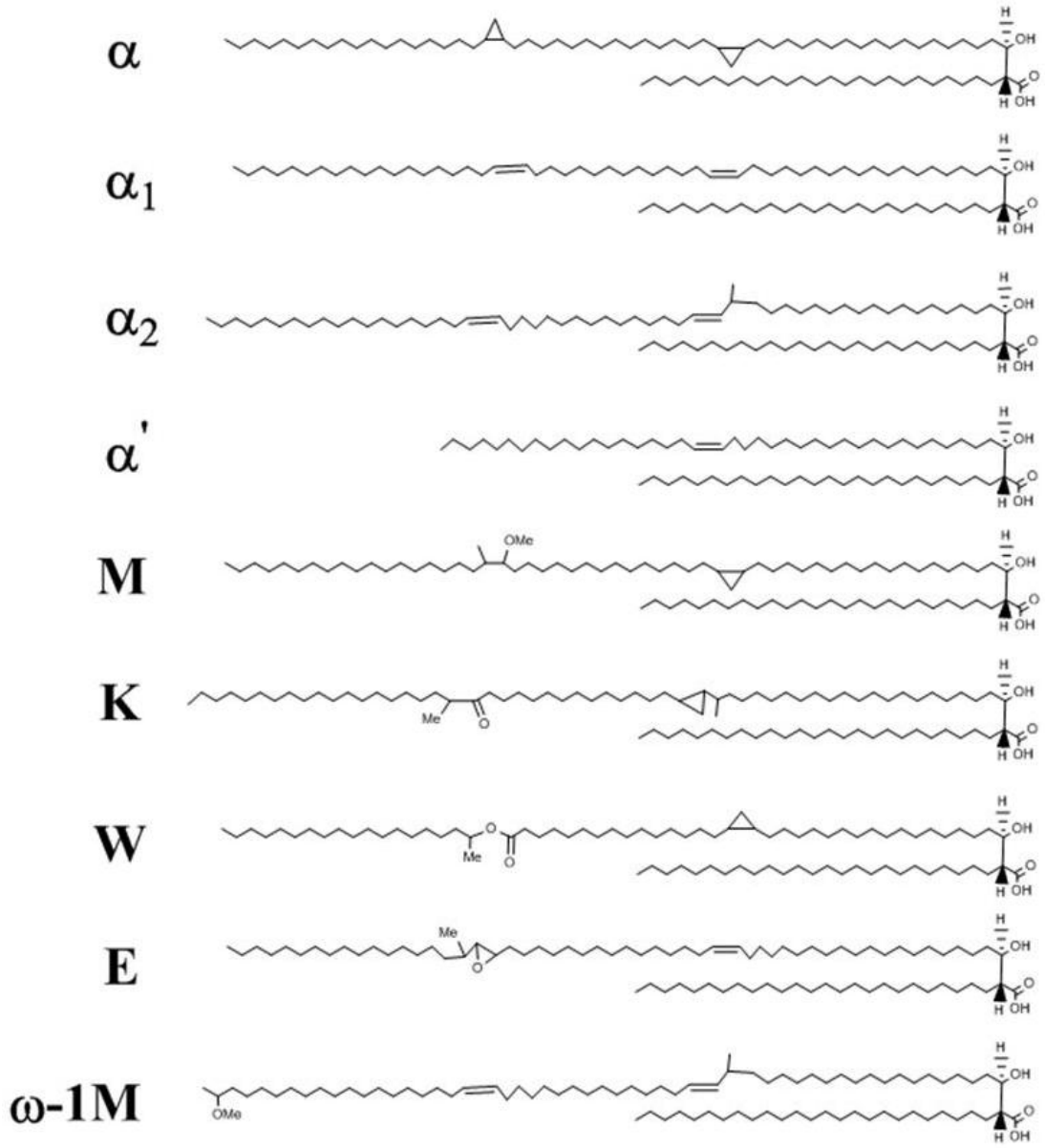


Figure 1-7 Representative structures of different types and classes mycolic acids from mycobacterial species. αsubtypes α, α1, α2 and α' all lacking oxygen functions. M-methoxy, K-keto, W-wax ester, E-epoxy and ω-1M- ω-1 methoxy are all oxygenated.

The evidence for association of mycolic acids with arabinogalactan came from chemical isolation of D-arabinose-5-mycolate fragments (Azuma *et al*. 1968; Kanetsuna *et al*. 1969). Later it was confirmed that mycolic acids are esterified to the non-reducing termini of arabinogalactan at all available *O*-5 positions of the Ara_6 motif (as discussed in chapter 1.6.3)

(McNeil *et al*. 1991).Two thirds of this terminal Ara_6 motifs are mycolylated and it is presumed that mycolic acids are arranged perpendicular to the cell membrane and parallel with respect to each other, forming a compact structure that serves as a permeability barrier while rendering structural rigidity to the cell wall (Minnikin *et al*. 2002). Protruding through regular gaps that exist in the mycolic acid layer are lipoglycans, *e.g.* LAM, that presents themselves to the cell surface (Chatterjee *et al*. 1991).

1.6.5 Lipomannan and Lipoarabinomannan

Interspersed within the mycobacterial cell wall, are found complex, structurally related glycolipids and lipoglycans. These are anchored to the cytoplasmic membrane through a common phosphotidyl-*myo*-inositol (PI) anchor (Hunter & J.Brennan 1990). Chemically the PI unit is based on an acylated *sn*-glycero-phosphate and is attached to *myo*-inositol (Lee & Ballou 1964; Ballou *et al*. 1963). Phosphotidyl-*myo*-inositol is mannosylated at 2-OH position to generate PIM_1. Acylation of the mannose residue at 3-OH or 6-OH position yields $AcPIM_1$. Acylation of mannose at both 3-OH and 6-OH position would yield Ac_2PIM_1. Further mannosylation at 6-OH position of inositol produces $AcPIM_2$/Ac_2PIM_2 (Korduláková *et al*. 2002; Korduláková *et al*. 2003; Jackson *et al*. 2000). Majority of PIMs exist as dimannosides (PIM_2) or hexamannosides (PIM_6).

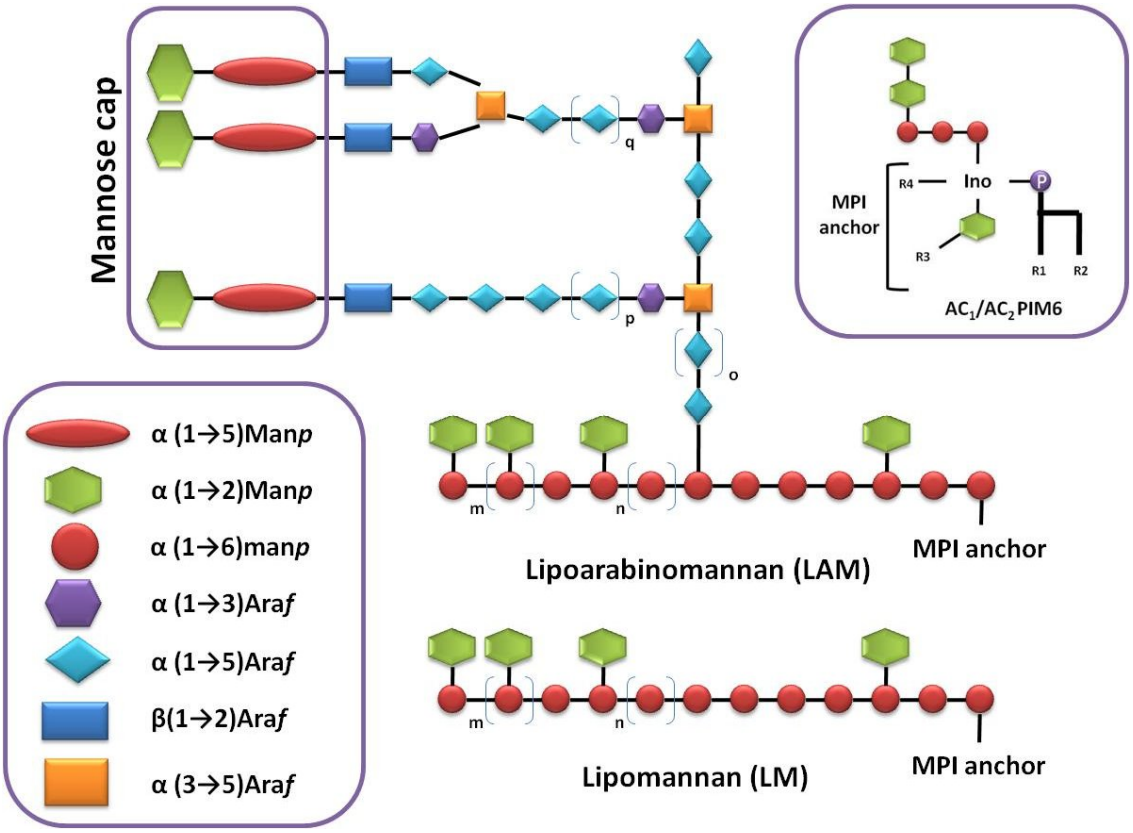


Figure 1-8 Essential structural topography of *M.tuberculosis*lipomannan (LM) and lipoarabinomannan (LAM). R_1 and R_2on MPI anchor indicate acylation at 1-OH and 2-OH position onthe glycerol of the MPI anchor, R_3-acylation at 3-OH of the inositol and R_4- acylation at 6-OH position of Man*p*. Sites of species specific glycosylation of the LM and LAM are indicated with m, n, o, p and q.

A linear mannan backbone extends from the mannose linked to the *O*-6 position of inositol in PIM_2 to produce lipomannan (LM). The extending Man*p* units are linked *via* α-(1→6) linkages with intermittent α-(1→2) branches (Chatterjee & Khoo 1998; Besra & Brennan 1997).

LAM is synthesised on a LM backbone with the addition of a branched arabinan motif. The point of arabinan attachment has not been identified to date (Jankute *et al*. 2012). The arabinan polymer consists of up to 60 α-(1→5)-Ara*f* units composed of branched hexaarabinofurano-sides (Ara_6) and linear tetraarabinofuranosides (Ara_4) (Chatterjee *et al*. 1991; Chatterjee *et al*. 1993; McNeil *et al*. 1994b). In *M. tuberculosis*, termini of the non-

reducing arabinan are capped with mannose residues linked through α-(1→2) linkage forming mannose-capped LAM (ManLAM) (Khoo *et al*. 2001). In other species, the arabinan is either capped with inositol (PILAM) or not capped (AraLAM) (Khoo *et al*. 1995).

1.6.6 Solvent extractable lipids in the mycobacterial cell wall

While the inner membrane cell wall components are essential for mycobacterial survival, components of the outer membrane have been demonstrated to play a critical role in pathogenicity. In a recent review, Minnikin *et al*. (2015) drew up a link between pathogenicity and components of the outer membrane components such as lipooligosaccharides (LOSs), sulfolipids (SL) and phenolic glycolipids (PGLs). The mycolic acids of the mAGP complex provide a hydrophobic environment for the non-covalent assembly of this vast variety of lipophillic, amphiphillic and polar lipids. The composition of the outer cell membrane components varies among species of mycobacteria.

1.6.6.1 Trehalose monomycolate and dimycolate

While mycobacterial specific mycolic acids are found esterified to arabinogalactan, they are also found non-covalently assembled in the outer membrane bound to trehalose. Trehalose is a non-reducing disaccharide of glucose (α-D-glucopyranosyl-α-D-glucopyranoside), and is found in acylated forms with fatty acyl chains and mycolic acids. Mono-mycolylated trehalose was isolated and characterised and was postulated to be the form in which they are transported across the membrane (Takayama & Armstrong 1976; Takayama & Armstrong 1977; Takayama *et al*. 2005). Later, this hypothesis was confirmed through genetic studies (Varela *et al*. 2012). Trehalose monomycolate (TMM) serves as substrate for mycolylation of arabinogalactan and also for the formation of 'cord factor' trehalose dimycolate (TDM) (Belisle *et al*. 1997; Puech *et al*. 2002). Mycolic acids are also found in free form and bound

to other residues such as glucose (glucose monomycolate) and glycerol (glycerol monomycolate). The role of these in pathogenicity is discussed following sections.

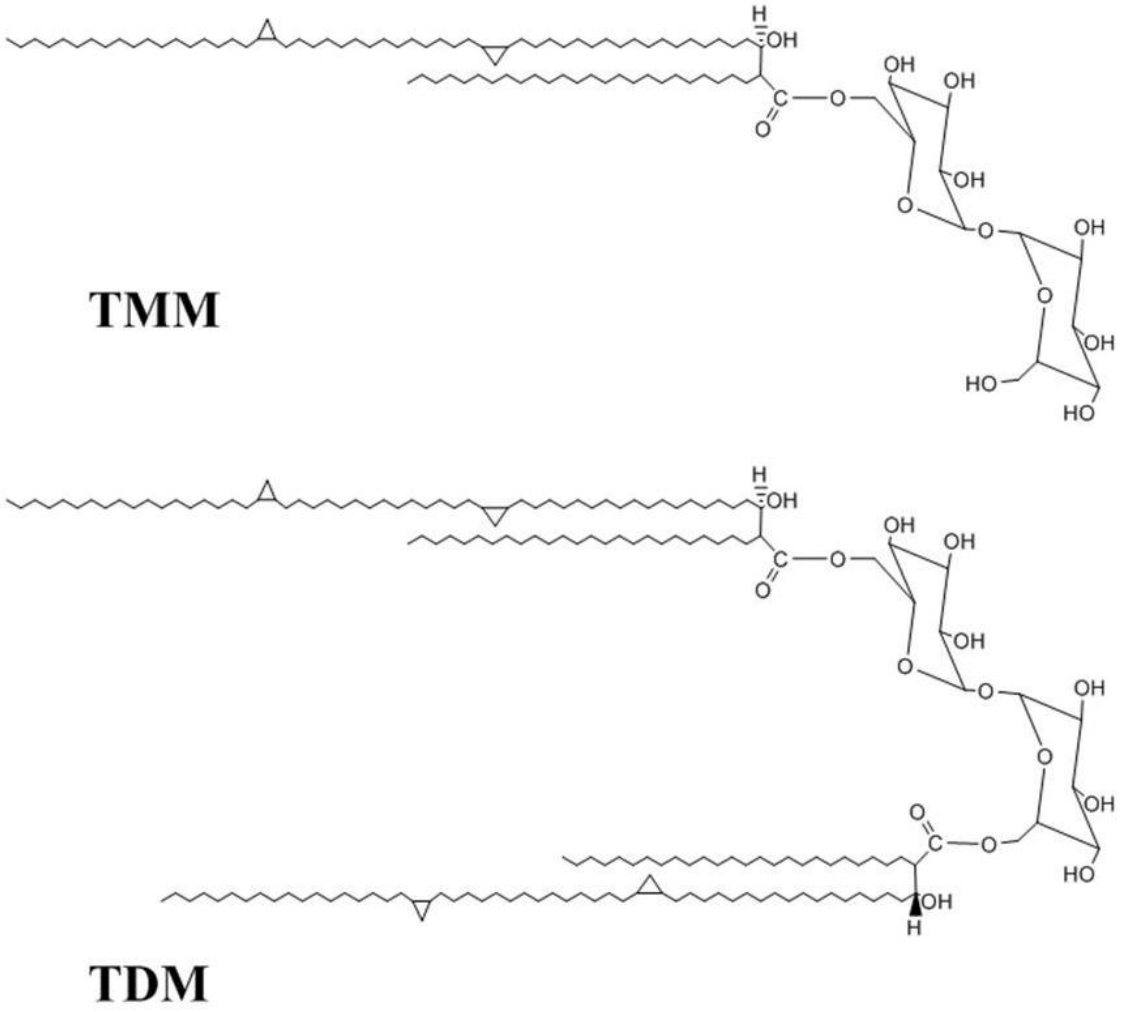


Figure 1-9 Structures of trehalose bound mycolates, TMM and TDM.

1.6.7 Methyl branched acyl trehaloses

The mycobacterial cell wall is characterised by an array of complex glycolipids based on trehalose. However, methyl-branched acyltrehaloses are exclusively present in MTBC organisms suggesting their possible role in virulence. Diacyltrehaloses (DATs) were first isolated and characterised by Besra *et al.* (1992) and showed that these are based on

mycosanoic (DAT1) or mycolipanolic acids (DAT2). More structurally related and heavily acylated lipids occur in *M. tuberculosis*, the pentaacyl trehaloses (PATs) and triacyl trehaloses (TATs) (Minnikin *et al*. 1985; Muñoz *et al*. 1997). These methyl-branched acyl trehaloses are amphillic in nature and suggest to act as a conduit between the hydrophobic mycolic acid layers and the polar capsule material (Minnikin *et al*. 2002), and has been shown to be effective diagnostic molecules (Muñoz *et al*. 1997).

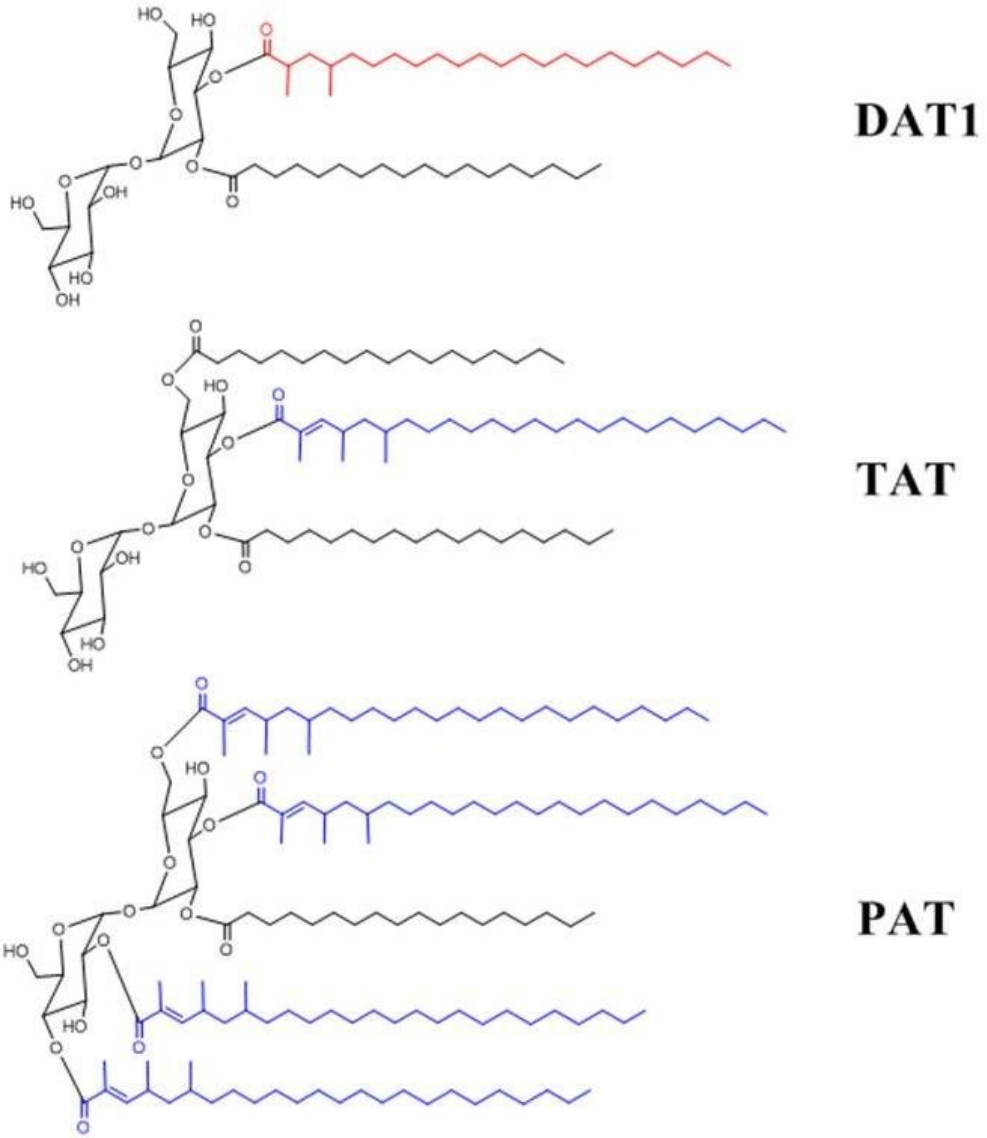


Figure 1-10 Structures of methyl branched acyl trehaloses, diacyl trehalose -1(DAT1), triacyl trehalose (TAT) and pentaacyl trehalose (PAT). Acyl chain in red- mycosanoic acid and acyl chain in blue is mycolipenic acids.

1.6.8 Phthiocerol dimycocerates (PDIMs) and phenolic glycolipids (PGLs)

PDIMs and PGLs are complex cell wall associated lipids that play a pivotal role in pathogenesis. These are very large (C>90) hydrophobic molecules and are constituents of the outer membrane of the mycobacterial cell wall (Minnikin *et al*. 2015, In press). Chemically these are long chain diols esterified by multi-methyl branched mycocerosic acids (Minnikin *et al*. 1985; Daffe & Laneelle 1988). PDIMs are based on three closely related long chain diols- phthiocerol A, phthiocerol B and phthiodiolones. Phthiocerols linked to phenol residue at the distal end forms the base for phenolphthiocerol dimycocerosates and characterised by the species-specific antigenic oligosaccharides attached to phenol and hence are commonly termed phenolic glycolipids (PGLs) (Daffe & Laneelle 1988; Minnikin *et al*. 2002). PDIMs and PGLs are shown to be modified in highly virulent Beijing strains of ***M. tuberuculosis***, where phthiotriols and phenolphthiotriol replace phthiocerols and phenolphthiocerols (Huet *et al*. 2009). Further, the length of the oligosaccharide attached to phenol residue varies among PGLs found in different species of mycobacteria.

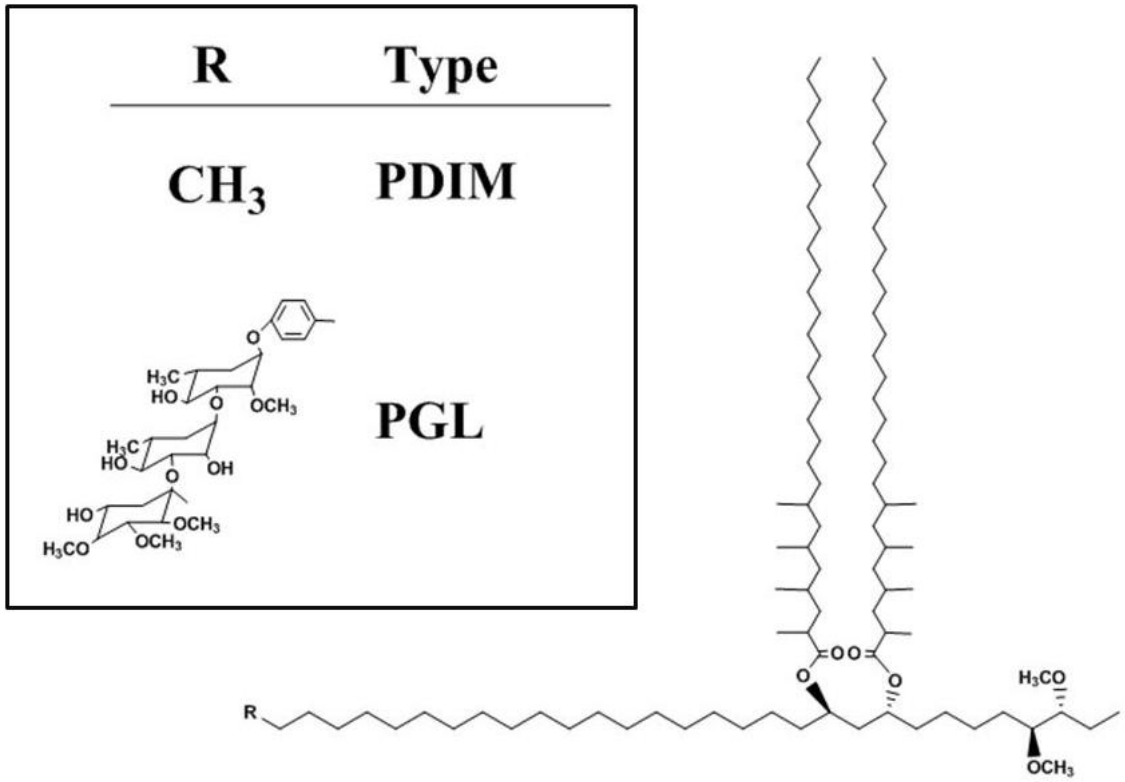


Figure 1-11 Structures of phthiocerols dimycocerosates (PDIMs) and phenolic glycolipids (PGLs). Different 'R' groups found in PDIMs and PGLs are shown inside the box.

1.6.9 Sulfolipids

Sulfolipids belong to a group of lipids termed acyltrehaloses along with TMMs, TDMs, DATs and PATs, however, they are uniquely characterised by the presence of a sulfated trehalose. Chemically, sulfolipid is a 2,3,6,6-tetraacyl 2' sulfotrehalose. Tetraacyl chains composed of one-short chain saturated fatty acid, usually palmitic or strearic acid, and a different long-chain multi-methyl branched fatty acids, termed as phthioceranic acid and hydroxyl phthioceranic acids are attached to the sulfated trehalose (Goren *et al*. 1976). The sulfolipids are exclusively found in virulent strains of *M. tuberculosis* and SL-1 is the most abundant species observed.

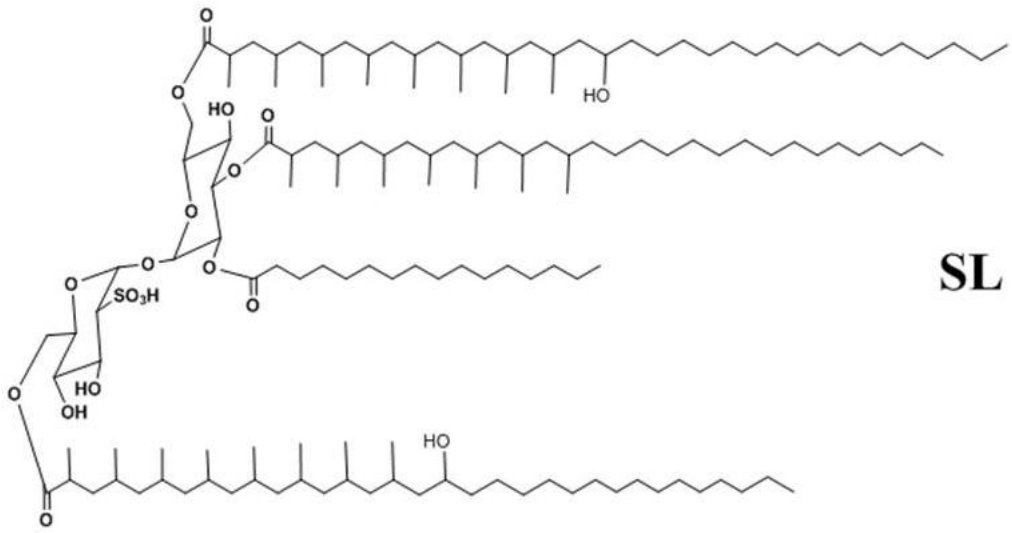


Figure 1-12 Structure of sulfated tetracyl trehalose (SL) from *M.tuberculosis.*

1.6.10 Lipooligosaccharides

Several mycobacterial species are characterised by the presence of highly polar oligosaccharide containing lipids in their outer membrane and these include phenolic glycolipids (PGLs), lipooligosaccharides (LOSs) and glycopeptidolipids (GPLs) (Asselineau & Laneelle 1998). A detailed account of PGLs was given in 1.6.8. This section gives a brief account of GPLs and LOSs. GPLs are composed of a fatty acyl chain and a peptide core, which is made up of D-phenylalanine, D-*allo*-threonine, D-alanine and L-alanine. Glycosylation varies across pathogenic and non-pathogenic bacteria. GPLs have been shown to play a role in sliding motility and pathogenesis (Recht & Kolter 2001; Villeneuve *et al*. 2003).

Lipooligosaccharides (LOSs) are highly polar, trehalose associated surface glycolipids and are present in several species of mycobacteriaand restricted to opportunistic pathogens and non-pathogens. LOSs have been identified and characterised from various mycobacterial species, which include the opportunistic pathogen *M. kansasii* (Hunter *et al*.

1985), *M. gastri* (Gilleronl & Vercauterent 1993), *M. szulgai* (Hunter *et al*. 1988), *M. malmoense* (M McNeil *et al*. 1987), *M. gordonae* (Besra *et al*. 1993), *M. marinum* (Minnikin *et al*. 1989) and the *M. tuberculosis* complex strain *M. canetti* (Daffe *et al*. 1991). Obligate human pathogens *M. tuberculosis* and *M. leprae* do not produce LOSs.

Structures of LOSs have been elucidated from many of these species which revealed a common polyacylated trehalose core with further glycosylations. The nature of the acyl-chain and position of linkage varies from species to species. Similarly, the oligosaccharide is either linked to C_3, C_4 or C_6 of trehalose and the length of the oligosaccharide could vary between 2 to 13 sugar residues (Besra *et al*. 1993; Hunter *et al*. 1983).The nature of the sugar residues that make up oligosaccharide component varies largely among different species and species specific sugar residues have been identified to date (Minnikin *et al*. 1989; Hunter *et al*. 1984; Rombouts *et al*. 2009). Although, a definitive role of LOSs is not yet completely understood, they have been shown to play a role in virulence and sliding motility. A detailed account of LOS biosynthesis and its role in pathogenesis is discussed in Chapter 6.

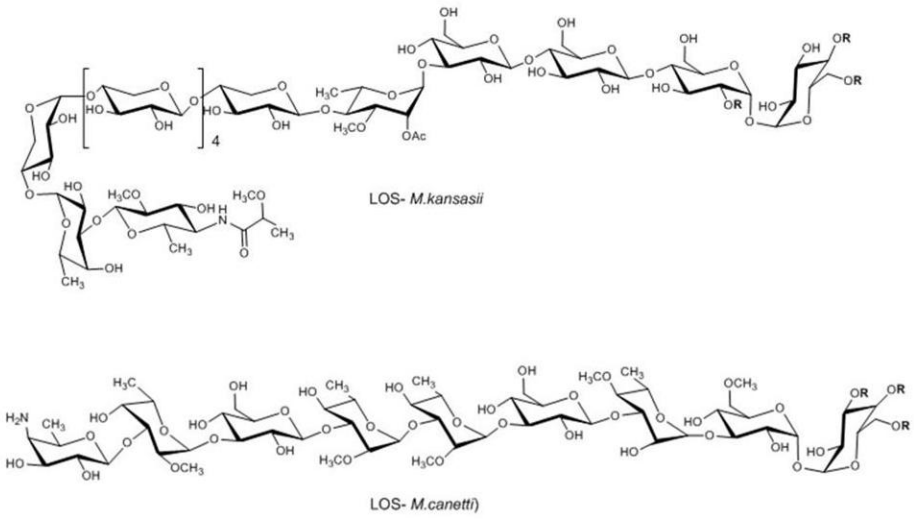


Figure 1-13 Structure of lipooligosaccharides from different mycobacterial species. Oligosaccharide compositions from two representative species are shown in the figure. Classes of LOSs produced by different mycobacterial species is tabulated below.

Table 1-3 Composition of LOSs from different mycobacterial species. Multiple numbers of identical residues are denoted by 'n'. (Adopted from Table 1-3 in PhD thesis titled "Mycobacterial Glycolipids – Pathways to Synthesis and role in Virulence"submitted by Debasmita Sarkar to the University of Birmingham).

M. kansasii	LOS-I (n=1) LOS-II (n=2) LOS-III (n=3)	(β-**D**-Xyl*p*)$_n$ (1→4)-3-*O*-Me-α-Rha*p* (1→3)-β-**D**-Glc*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
	LOS-IV (n=4) LOS-V (n=5) LOS-VI (n=6) LOS-VII (n=7)	Kan*N*acyl (1→3)-Fuc*p* (1→4)-(β-**D**-Xyl*p*)$_n$ (1→4)-3-*O*-Me-α-Rha*p* (1→3)-β-**D**-Glc*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
M. malmoense	LOS-II	α-**D**-Man*p*(1→3)-α-**D**-Man*p*(1→2)-α-**L**-Rha*p*(1→2)-[α-**L**-3-*O*-Me-Rha*p*(1→2)]$_2$-α-**L**-Rha*p*(1→3)-α-**D**-Glc*p*(1→1)-α-**D**-Glc*p*
M. szulgai	LOS-I	α-**L**-2-*O*-Me-Fuc*p*(1→3)-α-**L**-Rha*p*(1→3)-β-**L**-Rha*p*(1→3)-β-**D**-Glc*p*(1→6)-α-**D**-Glc*p*(1→1)-α-**D**-2-*O*-Me-Glc*p*
M. tuberculosis "Canetti"	LOS-I	N-acyl-4-amino-4,6-dideoxy-Gal*p*(1→4)-2-*O*-Me-α-**L**-Fuc*p*(1→3)-β-**D**-Glc*p*(1→3)-[2-*O*-Me-α-**L**-Rha*p*(1→3)]$_2$-β-**D**-Glc*p*(1→3)-4-*O*-Me-α-Rha*p*(1→3)-6-*O*-Me-α-**D**-Glc*p*(1→1)-α-**D**-Glc*p*
M. gordonae(989)	LOS-I	N-acyl-4-amino-4,6-dideoxy-2,3-*O*-Me-α-Gal*p*(1→3)-2-*O*-Me-4-*O*-Ac-α-**L**-Fuc*p*(1→3)-β-**D**-Glc*p*(1→3)-2-*O*-Me-α-**L**-Rha*p*(1→3)-β-**D**-Xyl*p*(1→2)-α-**L**-Rha*p*(1→3)-β-**D**-Glc*p*(1→3)-α-**L**-Rha*p*(1→3)-6-*O*-Me-α-**D**-Glc*p*(1→1)-α-**D**-Glc*p*
M. gordonae (990)	LOS-I	α-**L**-Rha*p*(1→2)-3-*O*-Me-Rha*p*(1→3)-β-**D**-Xyl*p*(1→2)-α-**L**-Rha*p*(1→3)-β-**D**-Glc*p*(1→3) -β-**D**-Glc*p*(1→3)- α-**L**-Rha*p*(1→3)-6-*O*-Me-α-**D**-Glc*p*(1→1)-α-**D**–Glc*p*
M. gastri	LOS-I (n=1) LOS-II (n=2)	[β-**L**-Xyl*p*-(1→4)]$_n$-3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**–Glc*p*
	LOS-III (n=6) LOS-IV (n=7)	3,6-dideoxy-4-*C*-(1,3-dimethoxy-4,5,6,7-tetrahydroxy-hepty)-α-xylo-Hex*p* (1→3)-[β-**L**-Xyl*p*-(1→4)]$_n$-3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
M. marinum	LOS-I	3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
	LOS-II	[3,6-dideoxy-4-*C*-(**D**-*altro*-1,3,4,5-tetrahydroxyhexyl)-**D**-xylo-hexapyranose-(1→4)]$_n$-[β-**L**-Xyl*p* (1→4)]-3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
	LOS-III	α-Car (1→4)-[β-**L**-Xyl*p*-(1→4)]-3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*
	LOS-IV	α-4-amino-4,6-dideoxy-Gal*p* (1→c)-α-Car (1→c)-α-Car (1→4)-[β-**L**-Xyl*p*-(1→4)]-3-*O*-Me-Rha*p* (1→3)-β-**D**-Gal*p* (1→3)-β-**D**-Glc*p* (1→4)-α-**D**-Glc*p* (1→1)-α-**D**-Glc*p*

1.7 Role of cell wall components in Virulence

The mycobacterial cell wall is complex and composed of several unique components. While some of these components are essential for its survival, some play a vital role in pathogenesis. This section attempts to shed light on factors that play an important role establishing infection and the progression of disease.

While mycolic acid biosynthesis is indispensable, alteration in its structure or modifications is tolerated, but shown to affect virulence. Strains lacking the distal cyclopropanation on α-mycolates were reported to have growth defects in mice lung and failed to persist during infection. Cyclopropanation modifications have also been shown to affect macrophage activation (Glickman *et al*. 2000; Rao *et al*. 2006; Rao *et al*. 2005). Recent studies have shown that the distal oxygen containing modifications on the meromycolate chain are important in suppressing IL-12 production (Dao *et al*. 2008). Other mycolic acid containing cell wall components also play pivotal role in infection. TDM, also known as cord factor, plays an important role in the recruitment of cells in granuloma formation and also modulates expression and production of cytokines (Indrigo *et al*. 2003; Hunter *et al*. 2006; Lima *et al*. 2001). Although, no direct evidence shows that DATs and PATs play a role in virulence, these are exclusively present in virulent strains, which points at possible role in virulence (Minnikin *et al*. 2015 In press).

Cox *et al.*(1999) reported that the loss of PDIMs leads to aattenuation of *M. tuberculosis* in lungs, but not in other organs *viz* liver and spleen, however the molecular mechanism was not determined. PDIM deficient cells have been shown to be hypersensitive to reactive nitrogen oxide released by activated macrophages with an associated elevation in TNF-α and IL-6 release (Rousseau *et al*. 2004; Reed *et al*. 2004). Other important classes of lipids found exclusively in virulent strains of mycobacteria are the sulfolipids (Middlebrook

et al. 1959). Domenech *et al* (2004) have shown that the transport of sulfolipids across the membrane is important for virulence and a reduced ability to survive intracellularly (Domenech *et al*. 2004; Domenech *et al*. 2005). This was supported by further studies with a Δ*phoP* attenuated strain that also showed a role for sulfolipids along with DATs and PATs in virulence (Asensio *et al*. 2006). However, a recent study has contradicted the above observations, which claims PDIMs and not sulfolipids are crucial for attenuation (Chesne-Seck *et al*. 2008).

The mycobacterial cell wall consists of several antigenic glycolipids of which LM and LAM are present universally across mycobacterial species. However, the LAM/LM ratio which varies from highly pathogenic to facultative pathogenic strains, may play a role in virulence (Briken *et al*. 2004). LM of mycobacteria are strongly proinflamatory and induce the production of IL-8, IL12 and TNF-α (Vignal *et al*. 2003; Briken *et al*. 2004). Also, PI-capped LAM has shown cytokine induction in human macrophages (Guerardel *et al*. 2002). However, LAM, Ara-LAM and Man-LAM do not show any pro-inflamatory activities (Guerardel *et al*. 2002; Dao *et al*. 2004; Vignal *et al*. 2003; Briken *et al*. 2004). Recently, LAM was shown to play important role in reducing/arresting phagosome maturation by incorporation into membrane rafts of macrophages through its phosphor-*myo*-inositol anchor (Welin *et al*. 2008).

Lipooligosaccharides are surface exposed highly polar antigenic glycolipids. Although, comprehensive information is not available for role of LOSs in pathogenicity, there have been attempts to study these using purified LOSs and LOS deficient mutants. LOS deficient mutants have been reported to exhibit increased virulence in Zebrafish embryos (van der Woude *et al*. 2012). Further, purified LOSs from *M. marinum* were shown to inhibit TNF-α secretion from LPS activated human macrophages (Rombouts *et al*. 2009). Recently,

a relationship between LOS production and uptake of cells by macrophages was established showing LOS deficient strains were efficiently phagocytosed (Alibaud *et al*. 2014). However, further studies are required to understand a clear role of LOSs in pathogenicity.

1.8 Aims and objectives

TB accounts for nearly 1.3 million deaths annually and is a massive burden both on humanitarian and on the economy. Treating TB is a big challenge and is complicated by the advent of drug resistance strains, including the recently isolated, what is termed as 'totally drug resistance' (TDR) strains (Udwadia *et al*. 2012; WHO 2013). Development of new anti-TB drugs is of paramount importance to fight the deadly disease.

Biochemical knowledge of mycobacterial cellular processes is a key aspect in drug development cascade. The unique cell wall is both a friend and a foe for mycobacteria, while it protects the cell by acting as a permeability barrier against therapeutic agents; it also provides high value drug targets. Three out of four front line drugs target the biosynthesis of cell wall components and in recent years several new drug targets have been identified, also are involved in the biosynthesis and transport of cell wall components (Banerjee *et al*. 1994; Telenti *et al*. 1997; Batt *et al*. 2012; Tahlan *et al*. 2012). One of the vital cell wall components is mycolic acids, which along with arabinogalactan forms the core of the cell wall. Biochemical knowledge of biosynthesis and transport of mycolic acids have provided useful insights on its essentiality for survival and pathogenesis apart from offering novel drug targets (Banerjee *et al*. 1994; Bhatt *et al*. 2005; Banerjee *et al*. 1998; Lai & Cronan 2003; Varela *et al*. 2012; A. Bhatt, Fujiwara, *et al*. 2007; Scarsdale *et al*. 2001; Sassetti *et al*. 2001). However, there are some unanswered questions with respect to our current understanding of mycolic acids biosynthesis, processing and transport. One of the objectives of this thesis is to

test the essentiality of *fabH* which is thought to play a crucial role in initiating the FAS-II cycle of fatty acid/mycolic acid biosynthesis. *fabH,* is thus believed to be essential for the survival of the organisms although TraSH analysis suggest that it may be non-essential for the growth and survival of *M. tuberculosis*. We seek to answer this question regarding its essentiality unambiguously and further, we have also attempted to evaluate functional redundancies and/or possibility of alternative pathways that could shunt the substrates into FAS-II for synthesis of fatty acid/mycolic acids.

Although mycolic acid biosynthesis has been studied for many years, research into its transport gained speed only recently. Several new class of antibiotics have been shown to target MmpL3 mediated transport of mycolic acids (Varela *et al*. 2012; Tahlan *et al*. 2012; La Rosa *et al*. 2012; Remuiñán *et al*. 2013; Brown *et al*. 2011). Although genetic evidence points at MmpL3 as the drug target, direct evidence in terms of protein-inhibitor interactions, binding studies or co-crystallisation are still elusive. To answer these questions, we sought to express, purify and characterise the MmpL3 protein and carry out biophysical and structural studies and explore possibilities of ligand binding and co-crystallisation studies.

Previous studies have shown that the transport of complex lipids requires the coupling of biosynthesis, processing and transport. Other MmpLs involved in transport of PDIMs and siderophores have shown to form protein scaffolds to couple both biosynthesis and transport (Jain & Cox 2005; Zheng *et al*. 2011; Wells *et al*. 2013). To find out whether MmpL3 also adopts a similar strategy to achieve transport of mycolic acids, we sought to investigate the function of uncharacterised genes in the *mmpL3* cluster and also use MmpL3 as a bait protein to identify interacting partners.

Lastly, the outer membrane of mycobacteria consists of several antigenic lipids that play critical role in virulence and immunomodulation. Of these outer membrane lipids, role

of LOSs has not been studied in detail and reasons for its absence in *M.tuberculosis* are still unknown. As a first step towards this, we seek to study biosynthesis and transport of LOSs in *M. kansasii* and *M.canetii*, which are genetically close to LOS deficient *M. tuberculosis.* A brief summary of project objectives outlined below;

- Deletion of *fabH* in *M. bovis* BCG to probe its essentiality and identify any functional redundancies or alternative pathways that may substitute FabH catalysed reaction.
- Biophysical and structural studies of MmpL3.
- Identification of proteins associated with MmpL3 in assisting the transport of mycolic acids.
- Deletion of glycosyl transferases, polyketide synthases and *mmpL* genes in *M. kansasii* and *M. canetti* to study biosynthesis and transport of LOSs.

Chapter 2

Evaluating the essentiality of *fabH* for the growth and survival of mycobacteria

2.1 Introduction

Mycolic acids are a major component of the cell wall and associated free lipids found in mycobacteria (Brennan & Nikaido 1995). These long-chain unique lipids are an integral part of the mycobacterial cell wall. Mycolic acids are covalently linked to arabionogalactan to form the mycolyl-arabinogalctan-peptidoglycan (mAGP) complex, the integrity of which is vital for *in vitro* and *in vivo* survival. The biosynthesis of mycolic acids is well studied and several genes have been shown to be essential for the survival of the bacteria. These include *inhA* (Vilchèze *et al*. 2000), *fabG* (Parish *et al*. 2007), *kasA* (Bhatt *et al*. 2005), *pks13* (Portevin *et al*. 2004), *mmpL3* (Varela *et al*. 2012), and are targeted by several anti-TB drugs (Winder & Collins 1970; Banerjee *et al*. 1994; Quémard *et al*. 1995; Kremer *et al*. 2000; Tahlan *et al*. 2012). While the absence of mycolic acids is detrimental to cell viability, any alteration either in its length or its fine structure is tolerated but is shown to affect cell wall permeability, thereby making the cell susceptible to lipophillic antibiotics and host cell antimicrobial molecules, such as defensin and lysozyme (Gao *et al*. 2003; Corrales *et al*. 2012). It is for both of the above reasons that the biosynthesis of mycolic acids is seen as a potential target for drug development.

The biosynthesis of mycolic acids is accomplished by the concerted efforts of two fatty acyl synthase (FAS) systems (Figure 2-1). α-alkyl and β-hydroxy (meromycolate) chains are synthesised separately by FAS-I and FAS-II respectively. Further modifications on meromycolate chain followed by the condensation of α-alkyl and modified meromycolate chains results in mature mycolic acids (Brennan & Nikaido 1995).

FAS-I plays a housekeeping role by synthesising fatty acids *de-novo*, which are either used to produce membrane lipids or for further elongation into mycolic acids by FAS-II. FAS-I is a multifunctional, multi-domain enzyme that synthesises C_{16-18} CoA primers for

further elongation by FAS-II and C_{26} substrates for condensation with meromycolate chain to produce mycolic acids (Takayama *et al*. 2005; Bhatt *et al*. 2008).

FAS-II consists of array of enzymes elongating the fatty acyl-CoA substrates into a C_{56} mero-mycolate. While FAS-II from other bacteria, chloroplasts and mitochondria can initiate fatty acid synthesis *de novo*; mycobacterial FAS-II cannot initiate *de novo* synthesis and can only extend the substrates generated by FAS-I (Takayama *et al*. 2005; Barry *et al*. 1998) . The fatty acyl Co-A generated by FAS-I are sequentially processed by β-ketoacyl AcpM synthase (KasA/KasB), β-ketoacyl AcpM reductase (MabA), β-hydroxyacyl AcpM dehydratase and enoyl AcpM reductase (InhA) (Takayama *et al*. 2005). The mero-mycolate chain produced by FAS-II undergoes species specific modifications and is transferred to condensing enzyme PKS13(Portevin et al. 2004), which condenses it with α-alkyl chain to produce mature mycolic acids. The post FAS-II modifications include cyclopropanation, introduction of *cis*/*trans* double bonds, methoxy, keto and/or epoxy groups. These modifications are essential to produce mature and functional mycolates (Kolattukudy *et al*. 1997; Zimhony *et al*. 2004; Takayama *et al*. 2005).

Acyl CoA bound substrates generated by FAS-I are channelled to FAS-II. The transitions through a set of enzymes AcpM, mtFabD and mtFabH (Wong *et al*. 2002; Choi *et al*. 2000a; Brown *et al*. 2005; Ghadbane *et al*. 2007). Since mycolates are shown to be vital for the survival and the genes involved in the pathway are shown to be essential, *fabH* was also believed to be essential and play critical role (Choi *et al*. 2000b). Previously, *fabH* orthologues from other Gram positive (*Lactococcus lactis*) and Gram negative (*E. coli*) bacteria, where it plays a critical role in the synthesis of short-chain fatty acid primers for the initiation of FAS-II biosynthesis, were shown to be essential supporting this hypothesis (Lai & Cronan 2003). Transposon site hybridization (TraSH) is a genetic tool that allows rapid functional characterisation of genes required for grown under different conditions. TraSH

analysis of mycobacterial genes lists *fabH* as a non-essential gene as opposed to earlier observations made in other bacteria (Sassetti *et al*. 2001; Sassetti *et al*. 2003) and additionally *M. leprae* which has minimum 'essential' genome and has a intact cell wall does not possess *fabH* (Cole et al. 2001; Kapopoulouet al. 2011). Also there have been no successful reports of a *fabH* deletion mutant. However, some of the genes listed as non-essential in the TraSH analysis have been later shown to be essential, for example *mmpL3* (Varela *et al*. 2012).

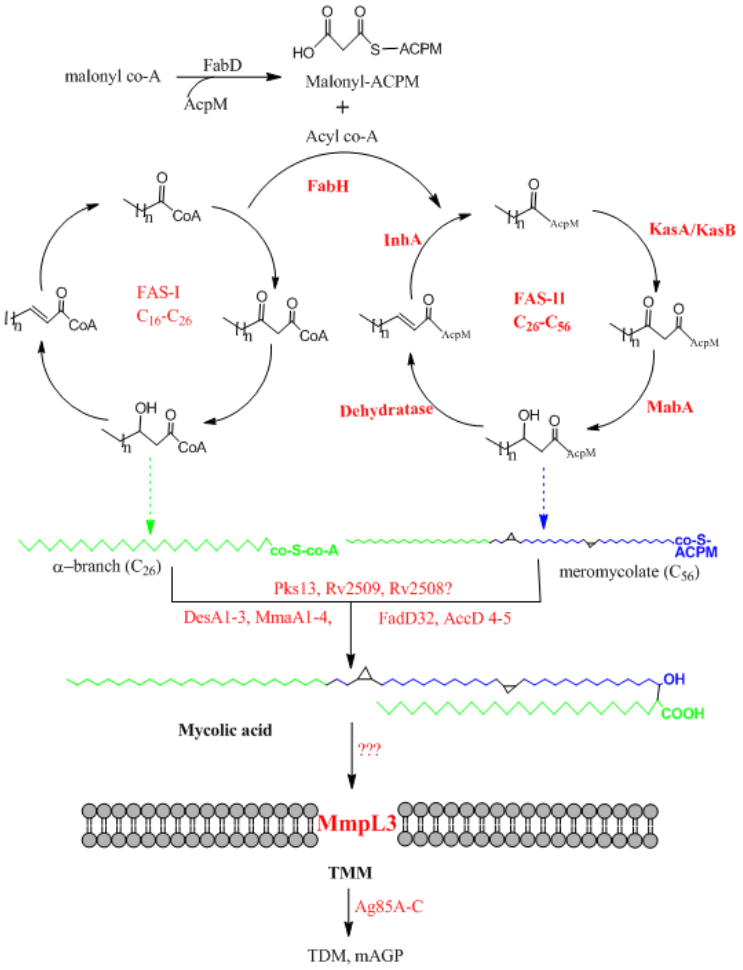


Figure 2-1 Schematic presentation of pathway for the biosynthesis of Mycolic acid biosynthesis in *M. tuberculosis*. The C_{16-26} CoA substrates produced by FAS-I are condensed with malonyl- AcpM by *mt*FabH and channelled into FAS-II for the generation of meromycolate chain. α-fatty acyl chain and the meromycolate chains are condensed by a polyketide synthase (Pks13) to produce nascent mycolic acids. Malonyl-AcpM also acts as a source of C2 donor for the β-ketoacyl-ACP synthases, KasA ad KasB. Several processing enzymes are involved in processing of mycolic acids to produce mature, trehalose bound form, which is transported across the membrane by MmpL3 for cell wall mycolylation and other functions.

FabH is a type-III β-ketoacyl synthase, that catalyses the reaction that bridges FAS-I and FAS-II. The acyl CoA primers synthesised by FAS-I are condensed through a decarboxylative condensation with malonyl-AcpM and feed into FAS-II. Substrate specificity experiments have shown that mtFabH has a distinct preference for long-chain acyl Co-A primers over short-chain CoA primers, and over AcpM primers (Choi *et al*. 2000b; Laurent Kremer *et al*. 2002). The catalytic triad of Cys-His-Asn is conserved across β-ketoacyl synthases, however the key amino acid changes in the binding pocket allows binding of longer acyl CoA chains (Scarsdale *et al*. 2001).

As discussed, the present understanding of the essentiality of *fabH* is ambiguous. Although, structural and enzymatic studies strongly suggest a key role in mycolic acid biosynthesis, there is no functional evidence to support the hypothesis. In this chapter, we have probed the essentiality of *fabH*. Further we also probed for functional redundancies in the β-ketoacyl synthase activity. Initial bioinformatics studies showed the presence of polyketide synthase (PKS) in the *M. tuberculosis* genome, the function of which is not yet known. PKS18 is a type-III polyketide synthase which is evolutionarily related to plant chalcone synthases (CHS), and contains an active site characteristic of mycobacterial β-ketoacyl synthases. Although PKS18 shares very low sequence homology with FabH, they are structurally related. Further, *in*-vitro substrate specificity studies also indicate that Pks18 could possibly replace the function of FabH. Interestingly, *M. leprae* which also produces mycolic acids, lack both *fabH* and *pks18* and it is likely that *M. leprae* KasA/KasB could perform FabH role. However, *M. tuberculosis* KasA/KasB have been characterised but are not shown to substitute FabH function in *M. tuberculosis.* In the light of this we decided to probe the function of FabH and Pks18 using genetic tools.

Our data suggest that *fabH* is non-essential for the growth of *M. bovis* BCG in agreement with TraSH data. This prompted us to look at other alternative routes of initiation of FAS-II biosynthesis. We have probed the function of *pks18* as a functional substitute for *fabH*.

2.2 Materials and Methods

2.2.1 *In-silico* analysis

The gene sequences for *fabH (Rv0533c*) and*pks18 (Rv1372*) were obtained from TB genome database Tuberculist (http://genolist.pasteur.fr-/TubercuList/). Structural analysis of Pks18 for β-ketoacyl synthase activity was cited from Saxena *et al*. (2003).

2.2.2 Plasmids, strains and DNA manipulation

Table 2-1lists the plasmids, bacterial strains and mycobacteriophages used in this study. *E. coli* strains were routinely cultured in LB broth at 37°C. *M. smegmatis* strain mc^2155 was cultured at 37°C in tryptic soy broth (TSB) with 0.05% Tween 80. The generation and propagation of mycobacteriophages was done on a lawn of *M. smegmatis* at 30°C in 7H10 basal plates overlaid with 7H9 soft agar as described in Chapter 7. High titres of mycobacteriophages were generated using protocols described in Chaper 7 (Larsen *et al*. 2007). *M. bovis* BCG and *M. tuberculosis* were grown in 7H9 broth supplemented with 10% OADC (oleic acid, albumin, dextrose and catalase) and 0.05% Tween 80, at 37°C. For plate growth, 7H10 media was supplemented with 10% OADC and incubated at 37°C. Antibiotics were routinely used for selection; hygromycin was used at 150µg/ml for *E. coli*, and at 100µg/ml for *M. smegmatis* and 50µg/ml for *M. bovis* BCG; kanamycin was used at 50µg/ml for *E. coli* and at 25µg/ml for *M. smegmatis* and *M. bovis* BCG; apramycin was used at 50µg/ml for *E. coli* and *M. bovis* BCG.

Table 2-1 Bacterial strains, plasmids and mycobacteriophages used in this studies described in this chapter

Plasmids, phages and strains	Description	Reference
Plasmids		
p0004S	Cosmid vector containing Hyg-sacB cassette	Larsen *et al*., 2007
p∆*fabH*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *mt.fabh* (Rv0533c)	Spivey V
p∆*pks18*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *mt.pks18* (Rv1372)	Spivey V
p∆*fabH*_Apra	Derivative of p0004s_Apramycin cosmid obtained by cloning upstream and downstream flanks of *mt.fabh* (Rv0533c)	Spivey V
p∆*pks18*_Apra	Derivative of p0004s_Apramycin cosmid obtained by cloning upstream and downstream flanks of *mt.pks18* (Rv1372)	Spivey V
pMV261	Kan^R, *E.coli*-mycobacterial shuttle vector (ColE1-MP*hsp60*)	Stover *et al*., 1991
pMV261- *fabH*	Functional copy of *mt.fabH* cloned into pMV261 shuttle vector	Spivey V
pMV261- *pks18*	Functional copy of *mt.pks18* cloned into pMV261 shuttle vector	Spivey V
Phages		
phAE159	Thermo sensitive, conditionally replicating shuttle phasmid derived fromlytic mycobactiophage TM4	Larsen *et al*., 2007
ph∆*fabH*	Derivative of phAE159 obtained by cloning p∆*fabH*into its PacI site	This work
ph∆*pks18*	Derivative of phAE159 obtained by cloning p∆*pks18*into its PacI site	This work
ph∆*fabH*_Apra	Derivative of phAE159 obtained by cloning p∆*fabH*_Aprainto its PacI site	This work
ph∆*pks18*_Apra	Derivative of phAE159 obtained by cloning p∆*pks18*_Aprainto its PacI site	This work
Bacterial strains		
E. coli Top 10	F– *mcr*A ∆(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z∆M15 ∆*lac*X74 *rec*A1 *ara*D139 ∆(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen

E. coli HB101	*E.coli* K-12 _F- mcrB mrr hsdS20(rB- mB-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(SmR) glnV44 λ-	Stratagene
M. smegmatis mc^2155	Wild type strain, E*pt*mutant of *M. smegmatis* strain mc^26	Snapper *et al*., 1990
M. bovis BCG	Wild type strain of *M. bovis* BCG *var* Pasteur	
M. tuberculosis H37Rv	Wild type strain of *M. tuberculosis*	
Δ*fabH*	*M. bovis* BCG strain with *mt.fabH* replaced with Hyg-sacB cassette	This work
Δ*pks18*	*M. bovis* strain with *mt.pks18* replaced with Hyg-sacB cassette	Bhatt A, University of Birmingham
Δ*fabH*_Apra	*M. bovis* BCG strain with *mt.fabH* replaced with Apramycin cassette	This work

2.2.3 Construction of knockout phage for deletion of *mt.fabH and mt.pks18*

The allelic exchange substrate plasmids listed in Table 2-1(gift from V. Spivey, University of Birmingham) were cloned into phAE159 at the *PacI* site and packaged into empty λ-phage heads. The packaged phasmids were transduced into *E. coli* HB101 cells. The positive phasmid for each gene was transformed by electroporation into *M. smegmatis* mc^2155 and recovered at 30°C for one generation time (3-4 hrs) in 1 ml of TSB containing 0.05% Tween 80. The recovered cells were split into 200 µl and 800 µl and mixed with 100 µl of log phase *M. smegmatis* and 5 ml of molten 7H9 soft agar. This mix was overlaid on 7H10 basal plates and incubated at 30°C for 2-3 days for plaques to be formed. The plates were soaked in mycobacteriophage (MP) buffer for 4-6 hrs and syringe filtered and either used straight away or stored at 4°C.

2.2.4 Generation of null mutants

Specialised transduction of *M. bovis* BCG to generate Δ*fabH* was performed as described for other mycobacterial species (Bardarov *et al*. 2002). *M. bovis* BCG was grown in 7H9 media containing 10% OADC and 0.05% Tween 80 to an OD_{600nm} of 0.8 and harvested by

centrifugation. The cell pellet was washed twice with an equal volume of MP buffer. Finally, the cells were suspended in a small volume of MP buffer (2 ml). High titre (10^8-10^{10}) phage lysate (1 ml) was mixed with 1 ml of cells for transduction and 1 ml of MP buffer mixed with 1 ml of cells for control reaction and incubated at 37°C, harvested and recovered in 7H9 media containing 10% OADC and 0.05% Tween 80 for 24hrs at 37°C. The recovered cells were harvested and resuspended in 400-600 µl of 7H9 media and plated on 7H10 agar containing 10% OADC and 50 µg/ml hygromycin (or 50 µg/ml apramycin where applicable) as a selection marker. Upon 3-4 weeks of incubation, hygromycin (or apramycin) resistant colonies that appear on the plates were inoculated into 10 ml of 7H9 media containing 10% OADC, 0.05% Tween 80 and 50 µg/ml of hygromycin B (or 50 µg/ml apramycin where applicable) for genomic DNA extraction and further characterisation. Allelic exchange of each gene with the *hyg* cassette (or *apramycin* cassette) in transductants was confirmed by southern blot.

2.2.5 Generation of double knock out of *fabH andpks18*

Specialised transduction of Δ*fabH* and Δ*pks18* were carried out in an attempt to generate Δ*fabH::pks18* double knock out. Δ*fabH,* Δ*fabH*_apra and Δ*pks18* cultures were grown in 7H9 media containing 10% OADC and 0.05% Tween 80 to an OD_{600nm} of 0.8 and harvested by centrifugation. The cell pellet was washed twice with an equal volume of MP buffer. Finally, the cells were resuspended in small volume of MP buffer (2 ml). High titre (10^8-10^{10}) phage lysate (1 ml) was mixed with 1 ml of cells for transduction; the combination of double mutants attempted is listed in Table 2-2. Control reactions were set up by adding 1ml of MP buffer to 1ml of cells. Transduced cells were incubated at 37°C for 24 hrs, followed by harvesting and recovering in 7H9 media containing 10% OADC and 0.05% Tween 80 for 24 hrs at 37°C. The recovered cells were harvested and resuspended 400-600 µl of 7H9 media and plated on 7H10 agar containing 10% OADC and 50 µg/ml hygromycin B (or 50 µg/ml

apramycin where applicable) as selection marker. Upon 3-4 weeks of incubation, hygromycin (or apramycin) resistant colonies that appear on the plates were inoculated into 10 ml of 7H9 media containing 10% OADC, 0.05% Tween 80 and 50 µg/ml of hygromycin B (or 50 µg/ml apramycin where applicable) for genomic DNA extraction and further characterisation. Allelic exchange of each gene with the *hyg*cassette (or *apra* cassette) in transductants was confirmed by Southern blot.

Table 2-2 Generation of Δ*fabH::pks18* double knock out

Single knock out	Mycobacteriophage	Double knockout
Δ*fabH*	phΔ*pks18*_apra	Δ*fabH::pks18*
Δ*fabH_apra*	phΔ*pks18*	Δ*fabH*_apra*::pks18*
Δ*pks18*	phΔ*fabH*_apra	Δ*pks18::fabH*_apra

2.2.6 Generation of complemented strains of *mt.fabH* and *mt.pks18*

pMV261*_fabH* and pMV261*_pks18* were electroporated into Δ*fabH*, Δ*pks18* and Δ*fabH*_apra strains and selected on 7H10 agar plates containing 50 µg/ml hygromycin (or 50 µg/ml apramycin where applicable) and 25 µg/ml kanamycin. Table 2-3 lists complemented strains used in this study.

Table 2-3 Complemented strains used in this study

Plasmids	**Description**	**Reference**
Δ*fabH-C-fabH*	Δ*fabH* complemented withpMV261-*fabH*	This work
Δ*fabH-C-pks18*	Δ*fabH* complemented withpMV261-*pks18*	This work
Δ*fabH*_apra_*C-fabH*	Δ*fabH _Apra* complemented withpMV261-*fabH*	This work
Δ*fabH*_apra_*C-pks18*	Δ*fabH _Apra* complemented withpMV261-*pks18*	This work
Δ*pks18_C-fabH*	Δ*pks18* complemented withpMV261-*fabH*	This work
Δ*pks18_C-pks18*	Δ*pks18* complemented withpMV261-*pks18*	This work

2.2.7 Extraction and analysis of lipids from *M. bovis* BCG strains

M. bovis BCG, Δ*fabH* and Δ*pks18* cultures were grown at 37°C, to mid-log phase in 10 ml of 7H9 medium supplemented with 10% OADC and 0.05% Tween 80. [^{14}C]-acetate, (50 µCi, 57 mCi/mmol, GE Healthcare, Amersham Biosciences) was added to a mid-log phase and grown for a further 24 hrs to label lipids. The labelled bacterial cultures were harvested and washed with PBS. The washed pellet was resuspended in 2 ml of methanolic saline (methanol: 0.3% NaCl, 10:1 v/v) and apolar lipids extracted using 2 ml of petroleum ether (60-80 bp). Apolar lipids was analysed by 2-D TLC (Dobson *et al*. 1985) and visualised by autoradiography by exposing a Kodak MR film to the TLC plates overnight.

2.2.8 Fatty acid methyl esters (FAMEs) and mycolic acid methyl esters (MAMEs) extraction from defatted cells and whole cells

Alkaline hydrolysis was performed on de-fatted cells after the extraction of polar lipids as described in Chapter 7. Briefly, 2 ml of 5% aqueous tetrabutylammonium hydroxide (TBAH) was added to defatted cells and incubated at 100°C, overnight. Base hydrolysed fatty acids and mycolic acids were extracted as methyl esters by mixing with 4 ml of dichloromethane (CH_2Cl_2), 300 µl of methyl iodide (CH_3I), and 2 ml of water for 30 min. Lower layer organic phase containing FAMEs and MAMEs were washed three times with water and dried, and re-dissolved in diethyl ether. The organic layer was taken into separate tube and dried under air. The samples were dissolved in CH_2Cl_2 (200 µl), analysed by single dimension TLC using petroleum ether: acetone (95:5) and visualised by autoradiograph.

2.3 Results

2.3.1 In silico analysis of *fabH* and *pks18*

fabH has been previously listed as a non essential gene for *M. tuberculosis* growth in TraSH analysis of *M. tuberculosis* H37Rv (Sassetti *et al*. 2001; Sassetti *et al*. 2003), however there have been no successful reports of a *fabH* deletion mutant, and no functional redundancies have been identified. Also, FabH is indispensable in other Gram positive and Gram negative bacteria (Lai & Cronan 2003). These observations strongly suggest at possible essentiality of *fabH*. This led us to validate the essentiality of *fabH* in mycobacteria. Further, we probed redundancies in the β-keto acyl synthase function in case of non-essentiality of *fabH*. PKS18, a Type-III polyketide synthase was shown to share structural similarity with FabH and exhibit preference for short-chain fatty acids in a condensation reaction (Saxena *et al*. 2003). Interestingly, PKS18 shares only 27% identity with FabH.

PKS18, a type-III polyketide synthase is evolutionarily related to plant chalcone synthases (CHS), with fusion active sites from CHS and mycobacterial β-ketoacyl synthases. Pks18 predicted to contain a classical thiolase fold (αβαβα), His-Cys-Asn catalytic triad, characteristic of β-ketoacyl synthases and CHS. CHS are typically involved in the synthesis of α-pyrones through decarboxylative condensation (Saxena *et al*. 2003). However, these metabolites have not been isolated in mycobacteria (Rukmini *et al*. 2004; Gokulan *et al*. 2013). The structural similarity of the Pks18 to FabH and its preference for long chain CoA substrates encouraged us to probe its role as functional substitute for FabH. Additionally, functional redundancy in *fabH* was recently reported in *Pseudomonas*. An alternative β-ketoacyl synthase encoded by *PA3286*, was shown to functionally substitute FabY (pseudomonas equivalent of FabH) (Yuan, *et al*. 2012a; Yuan, *et al*. 2012b). PA3286, shares low sequence identity with FabH (29%) and PKS18 (22%).

2.3.2 Generation of null mutants and Southern blot analysis

Specialised transduction, a highly efficient mycobacterial knockout method, was employed to generate null mutants in *M. bovis* BCG. The *fabH* gene was replaced with *hyg* and *apramycin* cassette separately. The genomic DNA from the hygromycin and apramycin resistance colonies was digested with *KpnI* and subjected to Southern blot analysis (Appendix 1) as described in Chapter 7. One confirmed knockout with a *hyg* cassette was labelled Δ*fabH* and one with an *apramycin* cassette was labelled Δ*fabH_apra*. Here, we report that *fabH* is not essential as suggested previously in the TraSH analysis by Sassetti *et al.* (2001). Δ*fabH* was used for all the lipid analysis.

Δ*pks18* was a gift from Dr. A. Bhatt (University of Birmingham), where *pks18* was replaced by a hygromycin cassette and was used for lipid characterisation.

Multiple attempts to generate a double mutant Δ*fabH::pks18* were unsuccessful. Further, attempts to generate a double mutant, in the presence of functional copy of one of the genes, are in progress.

2.3.3 Lipid analysis

To assess the effects of loss of *fabH* and *pks18* function on the biosynthesis of mycolic acids, cultures of *M. bovis* BCG WT, Δ*fabH* and Δ*pks18* strains were grown in 7H9 medium supplemented with 10% OADC and 0.05% Tween-80 and labelled with [^{14}C] acetate. Apolar lipids were extracted and analysed by 2D-TLC using systems A-D, designed for the analysis of apolar lipids, and visualised by autoradiography (Dobson *et al*. 1985).

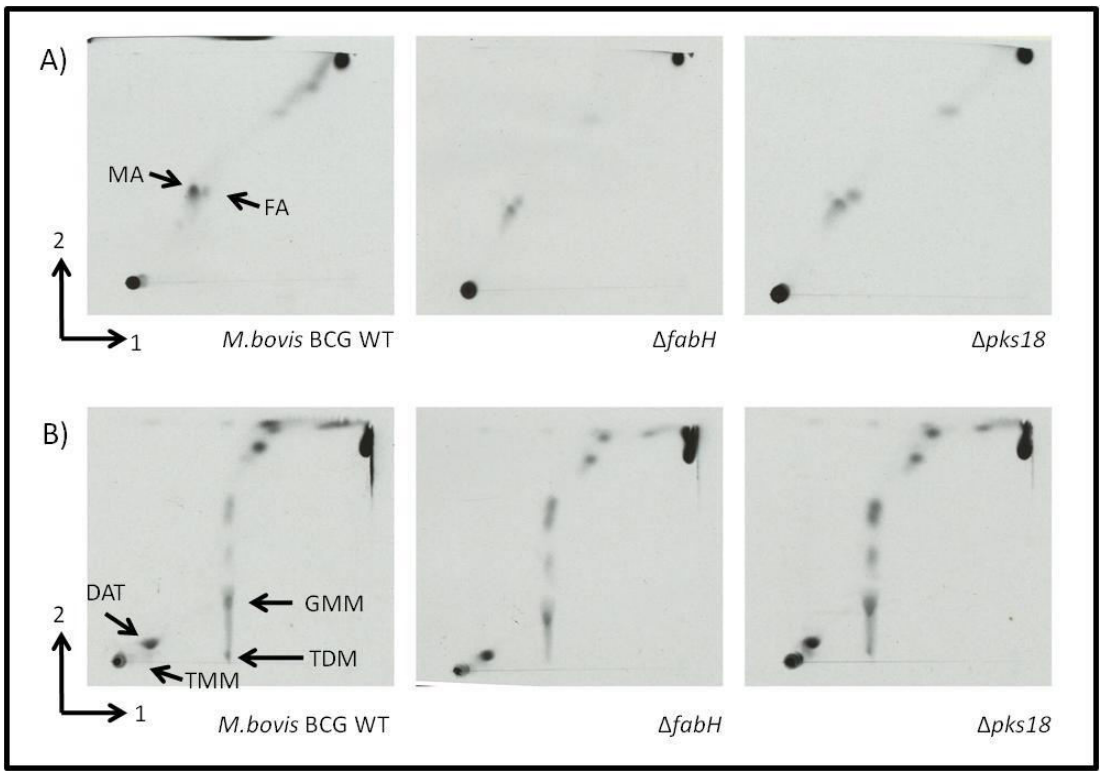


Figure 2-2 2D TLC analysis of apolar lipids from *M.bovis* BCG WT, Δ*fabH* and Δ*pks18* strains.(A) Direction 1- chloroform: Methanol (96:4 v/v); Direction 2- toluene: acetone (80:20 v/v); FA-fatty acid; MA-mycolic acid. (B) Direction 1- Chloroform: methanol: water (100:14:0.8 v/v/v); Direction 2- chloroform: acetone: methanol: water (50:60:2.5:3 v/v/v/v); TMM-trehalose monomycolate; TDM-trehalose dimycolate; GMM-glucose monomycolate; DAT- diacyl trehalose.

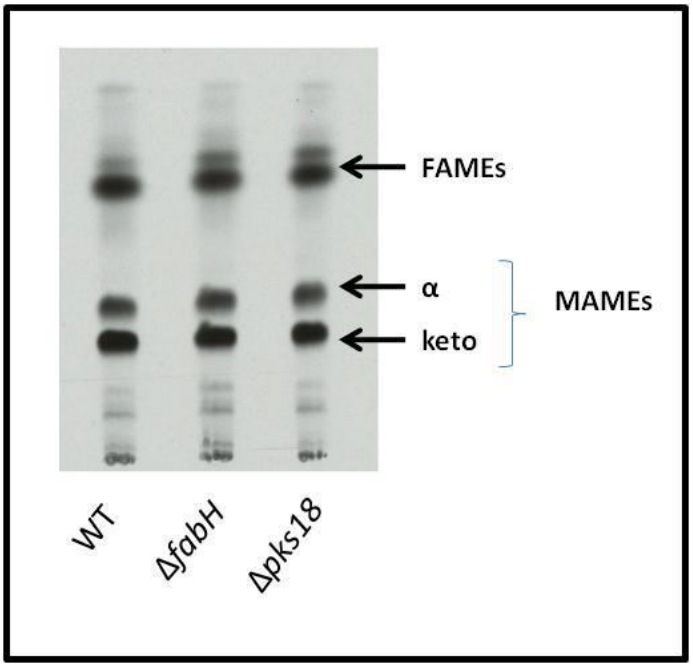


Figure 2-3TLC analysis of FAMEs & MAMEs extracted from *M.bovis* BCG WT, Δ*fabH* and Δ*pks18* strains, developed using petroleum ether: acetone (95:5 v/v) MAMEs- Mycolic acid methyl esters; FAMEs-Fatty acid methyl esters.

System C (Figure 2-2, A) used for the analysis of free fatty acids and free mycolic acids, revealed biosynthesis of either is not affected in the Δ*fabH* and Δ*pks18* knock out strains. System D (Figure 2-2, B), used for the analysis of trehalose bound mycolic acids namely trehalose monomycolates and trehalose dimycolates, revealed no significant difference between the WT and mutant strains. Further to probe any alteration in mycolic acid biosynthesis, cell wall bound mycolates were analysed. Cell wall bound mycolic acids were hydrolysed using TBAH and esterified to form mycolic acid methyl ester and analysed by TLC (Figure 2-3). The Δ*fabH* and the Δ*pks18* knock out strains did not show any defect in mycolic acid biosynthesis. Thus, *fabH* is not essential for bridging FAS-I and FAS-II, which was previously thought to be an essential process, and *fabH* and *pks18* may be functionally redundant.

2.4 Discussion

Mycolic acids are vital components of the mycobacterial cell wall and are critical for cell wall permeability and pathogenicity. The biosynthesis and transport of mycolic acids are of particular interest due to the enzymes involved in these process are essential for the survival of mycobacteria, and hence attractive anti-TB drug targets (Brennan & Nikaido 1995; Steenken & Wolinsky 1952; Thomas *et al*. 1961; Protopopova *et al*. 2005). Mycolic acids are synthesised by the concerted effort of two FAS synthases. The short-chain (C_{20-26}) CoA substrates are synthesised by FAS-I are channelled into FAS-II for reiterative steps of reductive decarboxylation. This is a critical step in the biosynthesis of mycolic acids as FAS-II cannot initiate *denovo* synthesis of mycolic acids (Bloch & Vance 1977; Mdluli *et al*. 1998; Choi *et al*. 2000b; Laurent Kremer *et al*. 2002). FabH is a β-ketoacyl synthase that catalyses decarboxylative condensation of long-chain acyl CoA substrates with the acyl carrier protein AcpM, which is used as a starting substrate by FAS-II (Choi *et al*. 2000).

FabH from *M. tuberculosis* is biochemically and biophysically well characterised. The crystal structure of mt.FabH has assisted development of several inhibitors targeting mycolic acid biosynthesis (Lv *et al*. 2009; Liu *et al*. 2012; Scarsdale *et al*. 2001).

In spite of biochemical and structural insight into the mtFabH protein, the functional role of FabH is still not clear. Given that FAS-II cannot synthesise a fatty acyl chains *de-novo* and the essentiality of mycolates, mt.FabH or a similar protein should have a definite role in mycolate biosynthesis.

TraSH analysis has listed mt.*fabH* as a non-essential for the growth of *M. tuberculosis*. However, TraSH predictions have exceptions, for example *mmpL3* which was predicted as non-essential was later show to be essential (Sassetti *et al*. 2003; Varela *et al*. 2012). Also, biochemical evidence suggests that the essential bridge reaction is catalysed by FabH. In the light of these observations, we sought to probe the essentiality of *fabH*. We employed specialised transduction method to generate a *fabH* null mutant in *M. bovis* BCG, showing *fabH* is non-essential for the survival of *M. bovis* BCG. Additionally, deletion of *fabH*, did not affect the biosynthesis of mycolic acids in a significant manner.

This led us to hypothesise that there could be a functional substitute for FabH for the above stated reasons. Although, a sequence homology search did not yield any obvious hits, a structural homology search indicated that PKS18 as the best hit (Saxena *et al*. 2003). A β-ketoacyl synthase (PA3286) that substitutes for FabY in *Pseudomonas* shared similar homology between FabH and PKS18, suggesting that PKS18 is the most likely replacement for FabH. *Invitro* substrate specificity experiments with purified PKS18 showed its preference for long-chain acyl CoA substrates over short-chain substrates. *In vitro* data is validated by preliminary structural data and modelling studies. Although, PKS18 is shown to produce fatty acid derived alkylpyrones *in vitro,* these metabolites have not yet been isolated

in mycobacteria (Saxena *et al*. 2003; Gokulan *et al*. 2013; Rukmini *et al*. 2004). Further, mycobacterial PKS proteins are not implicated in the production of cyclic polyketide products unlike PKS proteins from other sources (Austin & Noel 2003; Gokulan *et al*. 2013). Structural similarity of PKS18 with FabH, coupled with its substrate preference for long-chain substrates (C_{16} -CoA), prompted us to probe its function *in vivo*. However, the *pks18* deletion mutant generated suggests that it is a non-essential gene and did not affect the biosynthesis of mycolic acids or other apolar lipids. We thus, sought to probe functional redundancy of FabH and PKS18. Our attempts to generate a double mutant of *fabH* and *pks18* have not yet been successful. This suggests that, the function of one of the gene is essential for the survival of the mycobacteria. Work to confirm this by generation of a double mutant in the presence of a second copy of each gene is underway but could not be completed during this thesis.

Chapter 3

Structural and Biophysical characterisation of MmpL3; mycolic acid transporter protein

3.1 Introduction

The key enzymes involved in the biosynthesis of mycolic acids were discussed in the previous chapter. This chapter particularly focuses on the transport of mycolic acids in mycobacteria. The current knowledge about mycolic acid biosynthesis and function stems from genetics, biochemical and structural studies of the enzymes involved in the mycolic acid biosynthesis (Vilchèze *et al*. 2000; Bhatt *et al*. 2005; Bhatt *et al*. 2007; Banerjee *et al*. 1998; Choi *et al*. 2000). While the biosynthesis of mycolic acids is thoroughly understood, the information about late stage processing and transport of mycolic acids is still rudimentary.

An early insight into the transport of mycolic acids came from an identification of Myc-PL (6-O-mycolyl-β-D-mannopyranosyl-mono-phospho-heptoprenol), a polyprenol carrier-based intermediate accumulating in *M. smegmatis* (Besra *et al*. 1994). Following this, a hypothesis (Figure 3-1) was put forth by Takayama *et al*. (2005) who proposed that the nascent mycolic acid bound to the PPB domain (phosphopentotheine-binding; signature motif of acyl carrier proteins) of PKS13 is transferred onto polyprenol carrier by a yet unidentified mycolyltransferase (mycolyltransferase I). The mycolic acid is then transferred onto trehalose by a second mycolyltransferase (mycolyltransferase II), and subsequent dephosphorylation results in trehalose monomycolate (TMM), which is transported across the membrane by a transporter for cellwall mycolylation (Takayama *et al*. 2005).

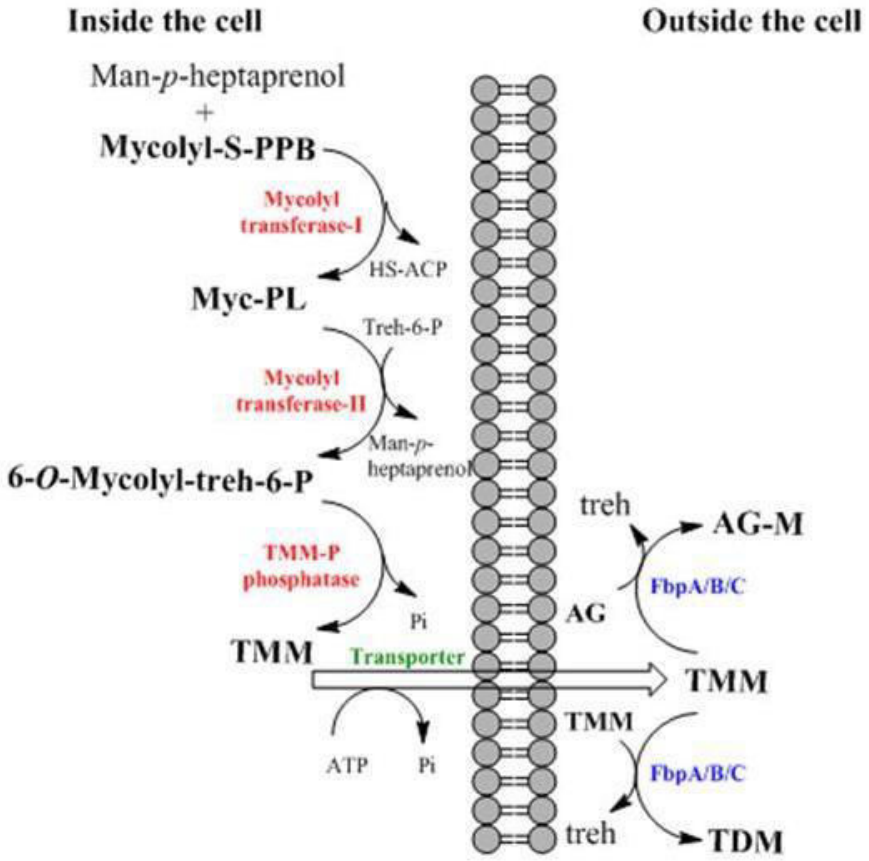


Figure 3-1 Schematic presentation of early hypothesis for transport of nascent mycolic acid for cell wall mycolylation. Nascent mycolic acid transferred from processing enzymes on to prenol carrier which is eventually transferred to trehalose to form TMM. TMM is then transferred across the membrane, where it acts as a substrate for mycolylation of cell wall AG. Man-p-heptaprenol- mannosylphosphorylheptaprenol; treh-trehalose; AG-arabinogalactan; AG-M-arabinogalactan-mycolate; TMM-trehalose monomycolate and TDM-trehalose dimycolate. Mycolyl transferases and the phosphotase involved in the late processing of mycolic acid to convert them into TMM, are yet to be identified, and indicated in red. The transporter is now identified as MmpL3 and is indicated in green. Mycolyl transferases FbpA/B/C involved in the cell wall mycolylation and the synthesis of TDM, are indicated in blue.

Advances in TB drug discovery have yielded several novel candidate drugs, that include the adamantyl ureas, THPP, SQ109 and BM212 (Protopopova *et al*. 2005; Deidda *et al*. 1998; J.R.Brown *et al*. 2011; Remuiñán *et al*. 2013). Target identification efforts by generating spontaneous durg-resistant mutants generated single nucleotide polymorphisms (SNP's) in *mmpL3*. MmpL are a class of transporter proteins that are involved in the transport of complex lipid molecules in mycobacteria and related species. Biochemical characterisation of the resistant mutants showed accumulation of TMM inside the cell indicating that *mmpL3*

is likely to encode a transporter protein responsible for mycolic acid transport (Tahlan *et al*. 2012; Remuiñán *et al*. 2013; La Rosa *et al*. 2012; Grzegorzewicz *et al*. 2012). This evidence coupled with earlier reports on other MmpL proteins, which were shown to be involved in transport of various metabolites including complex lipid molecules, highlighted the possibility of its role in mycolic acid transport (Cole *et al*. 1998; Converse *et al*. 2003; Cox *et al*. 1999; Wells *et al*. 2013). Definitive genetic evidence to this hypothesis came when TMM accumulation was shown to occur following conditional depletion of MmpL3, in *M. smegmatis mmpL3* conditional knock out strain (Varela *et al*. 2012).

Although, several MmpL proteins have been implicated in the transport of lipid molecules, the underlying mechanism is still unclear. MmpL proteins belong to the RND (Resistance, Nodulation and Division) superfamily of proteins (Domenech *et al*. 2004). The RND superfamily is one of the five major families of bacterial efflux pumps. Although, RND proteins were initially identified and characterised from Gram negative bacteria, they are found in both Gram positive and Gram negative bacteria (Paulsen *et al*. 1996). The predicted topology of MmpL proteins suggests they share structural features of RND proteins and are characterised by the presence of twelve transmembrane (TM) helices with two non-TM domains. But unlike other RND proteins, such as AcrB, MexA, CusA and SecDF, MmpL proteins exhibit high substrate specificity (Converse *et al*. 2003; Cox *et al*. 1999; Sondén *et al*. 2005; Varela *et al*. 2012).

RND protein mediated transport is shown to be driven by proton motive force (PMF) in other bacteria and the aminoacid residues involved have also been identified (Long *et al*. 2010; Seeger *et al*. 2009; Sennhauser *et al*. 2009; Tsukazaki *et al*. 2013). Recent studies using a known PMF inhibitor carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) affected

MmpL3 function, suggesting the possibility of PMF assisted transport by MmpL proteins (Yang *et al*. 2014).

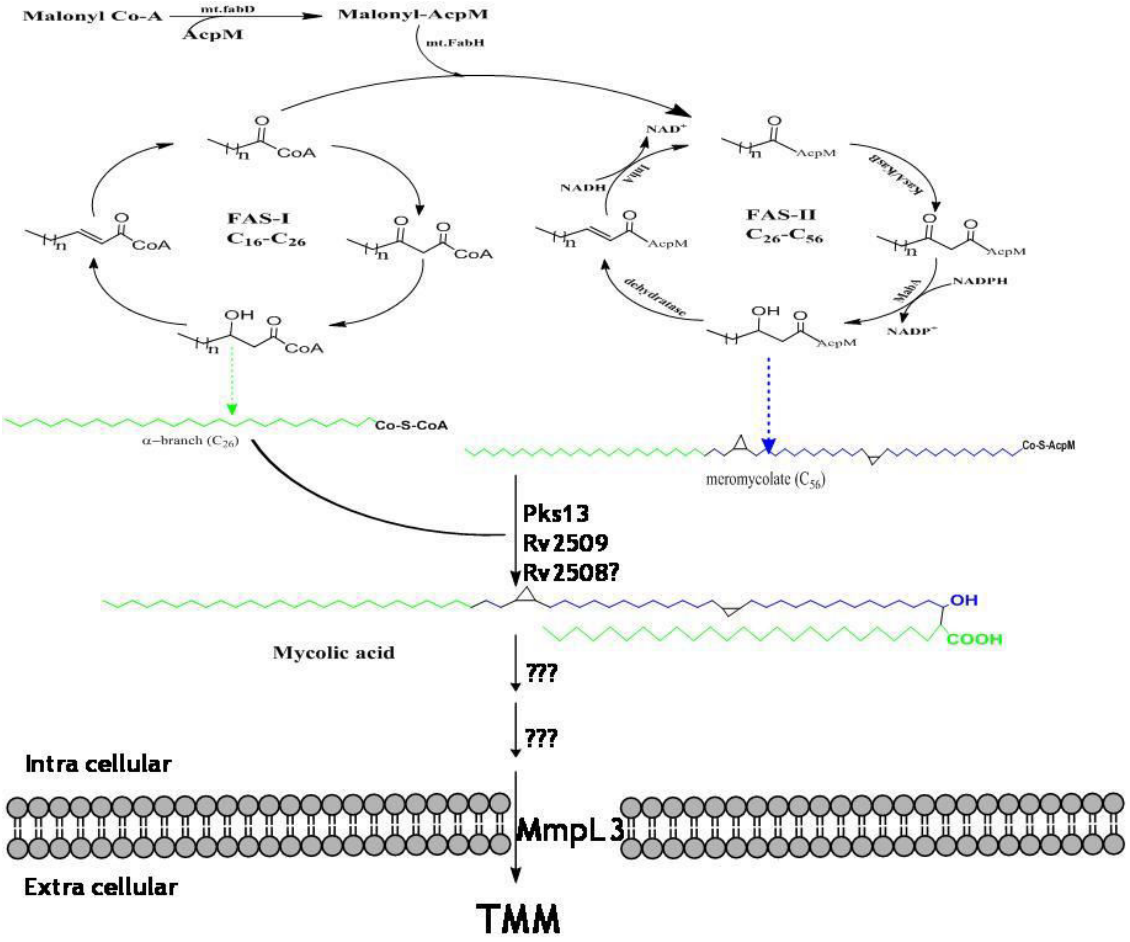


Figure 3-2 Schematic representation of mycolic acid processing highlighting the role of MmpL3

While the above evidence draws similarity between other RND and MmpL proteins, RND proteins from other bacteria exhibit broad substrate specificity and the two non-TM domains have been shown to play a vital role in substrate recognition. The two non-TM domains occur between TM#1 and TM#2 and TM#7 and TM#8 in almost all the RND transporters. The non-TM domains from different RND transporters have been structurally and functionally characterised (Tsukazaki *et al*. 2013; Elkins & Nikaido 2002). Non-TM domains were experimentally shown to contribute to substrate recognition and substrate

binding in AcrB by loop replacement studies (Elkins & Nikaido 2002). Bacterial two hybrid system studies with MmpL7 cytoplasmic domains showed interaction with late biosynthetic enzymes involved in PDIM biosynthesis (Jain & Cox 2005). These reports implicate independent and diverse functional roles for non-TM domains.

Crystal structures of several RND transporters have been solved over the years and have led to a better understanding of a structure-function-relationship. These studies have also helped reveal important amino acid residues involved in both proton relay and substrate recognition and binding (Tsukazaki *et al*. 2013; Long *et al*. 2010; Yu *et al*. 2003; Seeger *et al*. 2009; Sennhauser *et al*. 2009). Most of our knowledge regarding the biochemical function and substrate specificity of MmpLs comes from genetic studies (Varela *et al*. 2012; Converse *et al*. 2003; Jain & Cox 2005; Sondén *et al*. 2005). No direct evidence of substrate binding or structural information has been reported. Structural and biophysical studies would help explain substrate specificity of the MmpLs and understand the mechanism of transport processess. Dependence of MmpLs on PMF for transport process still needs to be established and the amino acid residues involved are yet to be identified.

This chapter aims to answer some of these questions. Protein structure determination was one of the key aims of the project and for this; several constructs were designed using bioinformatics tools. Protein crystallography and cryo-electron microscopy was used to get structural insights of the protein. The different constructs generated were also used in stubstrate binding studies. Although, genetic studies have pointed that different MmpL protein specifically transport different substrates, the mechanism underlying substrate specificity is not yet known. In this project we used substrate analogues to evaluate substrate binding to MmpL3 protein and validated specificity of binding using specific inhibitors. Further, using gel permeation chromatography in combination with analytical

ultracentrifugation, we attempted to evaluate the oligomeric status of the protein since other RND proteins have been shown to work as multimeric complexes. We also studied the mechanism underlying the transport mechanism through mutagenesis studies. Following section gives a detailed description of the selection of amino acid residues for mutagenesis and also other bioinformatics studies along with project pipeline.

3.2 Project pipeline

This section details the different phases of the project and describes the flow of the project. The initial phase of the project involved bioinformatics studies carried out in collaboration with Dr. Vassily Bavro (School of Bioscinces, University of Birmingham). Bioinformatics studies helped to establish a structural link between MmpL3 and other RND proteins. The available crystal structures of RND proteins were used to build MmpL3 model, which helped to design various constructs for biophysical and structural characterisation of MmpL3. The different constructs designed is schematically presented in Figure 3-3 (aminoacid sequences and molecular weight details in Appendix 4).

The different constructs designed include cytoplasmic domains (non-TM region 1- ML1, non-TM region 2- ML2) individually and the two cytoplasmic domains (ML) linked via artificial linker peptide mimicking the native spatial distance. The two cytoplasmic domains were linked using a linker peptide since there is no literature evidence available to show if these domains work individually or coherently. The rationale behind designing these constructs is that the cytoplasmic domains of other RND proteins and also other MmpL proteins have been shown to function independently. Further, the cytoplasmic domains across the RND family and MmpL proteins share similar structural features with a conserved folds. Additionally, working with large membrane protein such as MmpL3 could pose practical

difficulties in expression and purification and hence, other possibilities using cytoplasmic domains were also explored.

RND proteins have been shown to utilise proton motive force to drive the transport and the amino acid residues involved in generating the proton motive studies have identified by mutational studies. By comparing MmpL3 model with available crystal structures of other RND proteins, amino acids that are likely to be responsible generating proton motive force were identified and mutational studies were planned to evaluate structure-function relationship.

While, early phases of the project focus on characterisation of cytoplasmic domains, efforts to optimise the expression and purification of full length protein was carried out simultaneously and was prioritised once the conditions were optimised. Thus the later stages of the project focus on characterisation of full length protein. Figure 3-4 outlines the flow of the project highlighting key points at different phases of the projects.

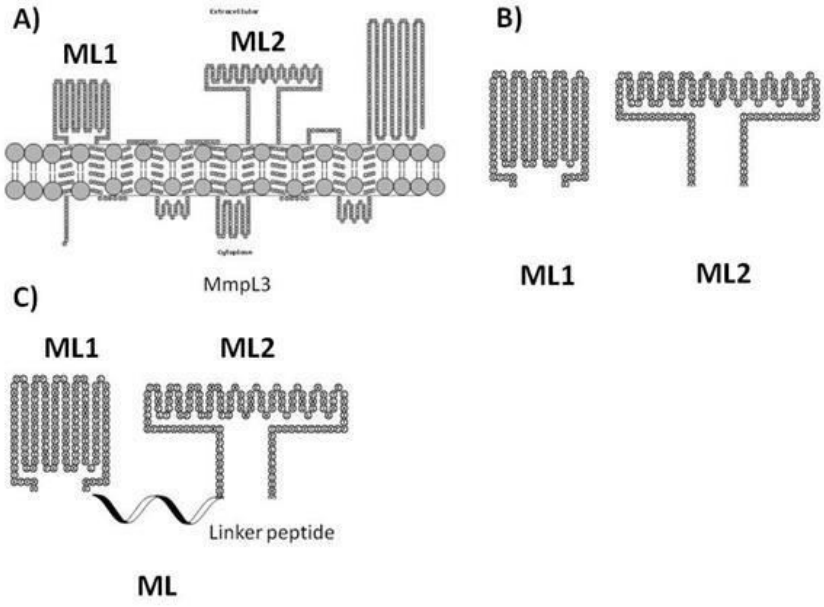


Figure 3-3 Schematic representation of different constructs of MmpL3 used for structural and biophysical studies. A) Full length MmpL3 with its predicted topology, B) individual cytoplasmic domains and C) cytoplasmic domains linked using a peptide linker to mimick the native spatial distance.

Figure 3-4 Flowchart of the project highlighting key points at different phases.

3.3 Materials and Methods

3.3.1 *In silico* analysis

The gene sequence for *mmpL3* was obtained from tuberculosis genome database (http://genolist.pasteur.fr/TubercuList/).The protein sequence alignments were done using blastp, two sequence alignments on (http://blast.ncbi.nlm.nih.gov/Blast). Structural homology studies were done by our collaborators Dr. Vassily Bavro. Topology prediction was done using TMpred (http://www.ch.embnet.org/software/TMPRED_form.html).

3.3.2 Plasmids, strains and DNA manipulations

Plasmids and bacterial strains used in this study are listed in Appendix 2. *E. coli* strains were routinely cultured in LB broth at 37°C unless mentioned otherwise. Antibiotics were added at following concentrations where required, ampicillin was routinely used at 100 µg/ml for *E. coli* strains and kanamycin was used at 50 µg/ml for *E. coli* strains unless mentioned otherwise.

3.3.3 Generation of expression constructs

The gene *mmpL3* (Rv0206c) was amplified to obtain either full length, truncated or specific regions of the gene by PCR using *M. tuberculosis* H37Rv genomic DNA. The primers used for the amplification of the gene are listed in Appendix-3. PCR amplified products were purified using Qiagen gel extraction kit and digested using restriction enzymes. The digested PCR products were then ligated using the respective vectors also cut with same restriction enzymes. The ligation mix was transformed to *E. coli* Top10 cells as described in Chapter 7. The transformants were selected on antibiotic selection plates. The transformants were analysed for the presence of the gene of interest through restriction digestion analysis and confirmed through sequencing.

To generate pET28b:ML (MmpL-Loops hereafter referred to as ML) construct, non-TM region 1 (here after referred as ML1) and non-TM 2(here after referred to as ML2) were first identified using TMpred and amino acid boundaries of the region was confirmed through model predictions (done in collaboration with Dr. Vassily Bavro, School of Biosciences, University of Birmingham). The two regions were linked through flexible poly glycine-serine peptide linker amounting 10Å distance to mimic the native spatial distance. The fusion protein was then codon optimised for expression in *E. coli*. The gene was synthesised at Genscript.Inc and cloned into pET28b.

pET41c_MmpL3-CB construct was generated by adopting the codon alteration method where the codon at the 5' end of the gene was altered for GC content to overcome possible secondary structure of the transcript that could hinder translation, a technique that has been successfully employed in mycobacteria (Usha *et al*. 2006). For this, the GC content in the forward primer (54mer) was optimised and full length *mmpL3* is PCR amplified and cloned into pET41c.

To generate pET28b: MmpL3 expression construct, *mmpL3* gene was codon optimised for expression in *E. coli* and the gene was synthesised at Genscript.*Inc* and cloned into pET28b with c-terminal histidine affinity tag.

3.3.4 Protein expression studies

The expression constructs listed in Appendix 2 were introduced into an *E. coli* expression hosts BL21 (DE3), C41 (DE3), or Tuner cells by transformation. Recombinant strains were inoculated for overnight cultures (5 ml LB, 50 µg/ml kanamycin) which were fed into bulk medium (6 X 1L of LB, 50 µg/ml kanamycin) and grown under shaking at 37°C to A_{600}~0.6, then induced with IPTG (1 mM) and incubated for further 3 hrs before harvesting the cells.

For optimising the expression of full length MmpL3 protein, pET28b:MmpL3 was introduced into *E. coli* C41 (DE) or BL21 (DE3) cells by transformation. Recombinant strains were used to inoculate overnight cultures (5 ml LB, 50 µg/ml kanamycin). 25 ml LB containing 50 µg/ml kanamycin was aliquoted in 250 ml flasks were inoculated with 1% inoculums from overnight culture and incubated at different temperatures (37°C, 30°C and 25°C) and grown under shaking to an OD_{600}~ 0.5 then each set of cultures are induced with different concentrations of IPTG (10 µM, 20 µM, 50 µM, 100 µM, 250 µM and 1 mM). The growth was monitored over 20-36 hrs and cells were harvested at each time interval. The expression of the protein was monitored by Western blot.

Western blot was performed routinely to monitor the expression levels of MmpL3 and purification. Anti-His IgG and anti-MBP IgG (mouse) were used as primary antibodies respectively against Histidine and Maltose binding protein fusion tags and anti-IgG (goat) coupled to alkaline phosphatase was used as secondary antibody. Nitro-blue tetrazolium and 5-bromo-4-chloro-3'-Indolyphosphate (NBT/BCIP) was used for chromogenic visualisation.

For overexpression of MmpL3 in *M. smegmatis*, pSD26: MmpL3 and pVV16: MmpL3 constructs were introduced by electroporation. The transformants were grown in 10 ml TSB containing 100 µg/ml hygromycin or 25 µg/ml kanamycin at 37°C which were fed into bulk cultures (2 X 1L TSB, 100 µg/ml hygromycin or 25 µg/ml kanamycin) and grown at 37°C. The culture harbouring pSD26: MmpL3 was induced with 0.2% w/v acetamide and culture harbouring pVV16: MmpL3 expresses the protein constitutively.

3.3.5 Protein purification

3.3.5.1 Purification of non-TM regions of MmpL3 and non-TM fusion construct (ML1, ML2 and ML)

Cells (10 g wet weight) were resuspended into 50 mM Tris.Cl (pH 8.0), 500 mM NaCl and10 mM imidazole supplemented with 1 mM PMSF and Protease inhibitor cocktail (Roche) (1 tablet/50 ml). Cells were subjected to probe sonication (MSE Soniprep 150, 12 micron amplitude, 20s ON, 30s OFF for 12 cycles on ice). The cell slurry was centrifuged at 27,000 x g at 4°C for 40 min and the resulting clarified supernatant applied to a Ni^{2+}NTA HiTrap column (GE healthcare, Amersham, UK) pre-equilibrated with 50 mM Tris.Cl (pH 8.0), 500 mM NaCl and 10 mM imidazole supplemented with 1 mM PMSF. The column was eluted with 5 column volumes of equilibration buffer (5 ml) containing 10 mM, 20 mM, 40 mM, 60 mM, 80 mM, 100 mM and 250 mM imidazole. The fractions were analysed by SDS-PAGE and fractions containing protein of interest were pooled and dialysed against 50 mM Tris.Cl (pH 8.0) and 150 mM NaCl and concentrated. The concentrated protein (1 mg/ml) is applied to Superdex-200 size exclusion column. The column was developed using Tris.Cl pH 8.0, 150 mM NaCl. The fractions were collected and analysed by SDS-PAGE. The fractions containing protein of interest were pooled and concentrated.

3.3.5.2 Purification of MmpL3 using SMALPs

SMA is chemically styrene-co-maleic acid polymer that is amphipathic in nature that is used to extract membrane protein by encapsulating the membrane protein along with membrane lipid molecules. The principle and working of SMALP is schematically presented in Appendix 5. Cells (10 g wet weight) were resuspended into 50 mM Tris.Cl (pH8.0), 2mM EDTA, 5% glycerol supplemented with 1mM PMSF and protease inhibitor cocktail tablet. Cells were lysed by French press (10,000 psi 4 cycles). The cell slurry was centrifuged at

10,000g at 4°C for 30 min and the resulting clarified supernatant containing the membrane fraction was processed further to purify the MmpL3 protein. The membrane fraction was separated by centrifuging at 100,000g at 4°C for 45 min. The membrane was homogenised into 50mM Tris.Cl pH 8.0, 500 mM NaCl, 10% NaCl supplemented with 1 mM PMSF and protease inhibitor cocktail tablet. The homogenised membrane was solubilised using styrene-co-maleic acid polymers to generate SMALPs [poly (styrene-co-maleic acid) lipid particles]. For this 2.5% SMA polymer was added to homogenised membrane and incubated for 2 hrs at room temparature with gentle shaking. At the end of 2hrs the mix was centrifuged at 100,000g for 45 min to remove insoluble debris. The supernatant was applied to a Ni^{2+}-NTA HiTrap column (GE healthcare, Amersham, UK) pre-equilibrated with 50 mM Tris.Cl (pH 8.0), 500 mM NaCl, 10% glycerol. The column was eluted with 5 column volumes of equilibration buffer (5 ml) containing 10 mM, 20 mM, 40 mM, 60 mM, 80 mM, 100 mM and 250 mM imidazole. The fractions were analysed by SDS-PAGE and fractions containing protein of interest were pooled and dialysed against 50 mM Tris.Cl (pH 8.0) and 150 mM NaCl and concentrated. The concentrated protein (1 mg/ml) applied to a Superdex-200 size exclusion column. The column was eluted using Tris.Cl pH 8.0, 150 mM NaCl and the fractions were collected and analysed by SDS-PAGE. The fractions containing protein of interest were pooled and concentrated.

3.3.6 Biophysical characterisation of proteins

3.3.6.1 Analytical Ultra Centrifugation (AUC)

AUC was performed to evaluate the oligomeric states of MmpL3. The experiment was performed at central facility using a Beckman Coulter XL ultracentrifuge. In a sedimentation velocity experiement sediment coefficients are calculated which depends directly on the mass of the particle. Briefly, the purified protein sample (0.8 mg/ml) was centrifuged at 40,000g at

20°C using buffer as reference. The sedimentation was monitored by absorbance overnight and the data was analysed using Sedfit software.

3.3.6.2 Circular Dichroism (CD)

CD was performed using Jasco (J810) spectropolarimeter. To perform CD the protein was dialysed against 20 mM KH_2PO_4 buffer pH7.5, 100 mM NaCl and concentrated to 1 mg/ml. 60 µl sample was taken in 0.1 mm cuvette, scanned between 280nm to 190nm and ellipticity was measured (in degrees). Ellipticity was plotted against wavelength to get a CD spectrum.

3.3.6.3 Protein cross linking studies

Protein cross linking studies were carried out using glutaraldehyde and EGS (ethylene glycol bis[succinimidylsuccinate]). The cross linking was done both in presence and absence of trehalose, trehalose monomycolate (C_{24}) and trehalose monopalmitate. The protein (+/- substrate) is incubated with cross-linker and the reaction was stopped at different time intervels (0 to 30 min) using buffer containing amines at high pH (10µl of 1M Tris.Cl, pH 8.0). The samples were warmed at 70°C for 2 min after adding SDS-loading dye and analysed on SDS-PAGE.

3.3.7 Structural studies

3.3.7.1 Protein crystallography

Purified protein samples, MmpL3 non-TM region constructs and non-TM fusion protein, ML1, ML2 and ML proteins were concentrated to 10 mg/ml and 20 mg/ml using a centriprep centrifugal device at 4°C. Crystals were grown by hanging drop vapour diffusion method. 30 µl reservoir solution was dispersed into wells and mixing 1 µl of protein with 1µl of reservoir solution in crystallisation wells and incubated at 18°C to facilitate crystal growth and monitored by observing the plate under microscope daily for a week. Different reservoirs were used to screen the condition for crystallisation and these include commercially available

crystallisation screens namely, Structure Screen 1 & 2, Pact premier, JCSG, Midas, Morpheus, Hampton, Wizard 1 & 2 and Wizard 3 & 4.

3.3.7.2 Protein structure studies using of Nuclear Magnetic Resonance (NMR)

NMR was used to carry out initial studies on structural aspects of MmpL3 non-TM regions ML1 and ML2. Respective constructs were transformed to *E. coli* C41 (DE3) cells and transformants were grown overnight in M9 minimal media containing 25 µg/ml kanamycin, which was fed into bulk cultures (2 x 1L M9 + 25 µg/ml kanamycin) and labelled with N^{14}. Protein purification was done as described earlier. NMR experiment was done by our collaborator Dr. Mark Jeeves (School of Cancer Science, University of Birmingham).

3.3.8 Ligand binding studies

3.3.8.1 Intrensic Tryptophan Fluorescence (ITF) assay

Intrensic tryptophan fluorescence (ITF) assay exploits the fluorescing property of tryptophan aminoacid which emits between 300 to 400 nm with a emission maxima at 335 nm, when excited at 285 nm. The emission was measured between 300 nm to 400 nm using Perkin-Elmer Luminescence Spectrophotometer by exciting the sample at 285nm at 25°C. Spectra were recorded for each addition of substrate (trehalose monopalmitate) aliquot (1 µl) added to 500 µl of 6 µM non-TM fusion protein ML in 50mM Tris.Cl buffer pH 8.0, contanining 150mM NaCl, until reaching the final concentration of 3 mM (15 additions) and on addition of known inhibitors of MmpL3 (adamantyl urea (J.R.Brown et al. 2011), BM212 (Deidda et al. 1998) and THPP (Remuiñán et al. 2013) compounds) for each aliquot (1 µl) added to 500 ul of 1.5 µM MmpL3 (SMA-solubilised purified full length protein), until reaching the required final concentration. Data were collected using Hitachi FL Solutions 4.6 software and analysed in Prism 5 (Graphpad).

3.3.8.2 TMP-BODIPY binding assay

The substrate analogue trehalose mono palmitate tagged with fluorochrome BODIPY (*boron-dipyrromethene*) was used as a substrate for binding studies. The substrate analogue emits at 515 nm when it is in hydrophobic environment or bound to protein on excitation at 485 nm. Binding of substrate analogue to ML, ML1 and ML2 was tested by recording spectra for each addition of substrate analogue (1 µl) to 500 µl of 6 µM protein in 50mM Tris.Cl buffer pH 8.0, contanining 150mM NaCl, until reaching the final required concentration. Data were collected using Hitachi FL Solutions 4.6 software and analysed in Prism 5 (Graphpad).

3.3.9 Site directed mutagenesis

Site directed mutagenesis was performed using Q5 site directed mutagenesis kit and was done according to manufacturers recommendations. The primers were designed using NEBaseChanger™ program (http://nebasechanger.neb.com/). Briefly, pMV261 (*apra*):MmpL3 was amplified using primers carrying desired base change followed by removal of parental DNA from the mix using *DpnI*. The plasmid DNA carrying the mutation is transformed into *E.coli* NEB5-α strain. Plasmids isolated from transformants and sequenced to confirm the mutation.

3.3.10 Fermentation

pET28b:MmpL3 was introduced into *E. coli* BL21 (DE3) cells by transformation and transformants were inoculated (250 ml LB, 50 µg/ml kanamycin) prepare inoculum for initiating the fermentation. Fed batch fermentation was carried out in a Infors fermentor (3 L capacity) feeding inoculum into 2.5 L of LB supplemented with 1% glucose, 50 µg/ml kanamycin and 0.5 ml/L antifoam (25°C, 300rpm, pH 7.0). Protein expression was induced at A600~0.4 with IPTG and culture was fed with feeding medium (10X LB containing 500

µg/ml kanamycin and 20% glucose, 10 ml/hr) for 20 hrs at the end of which the cells were harvested.

3.4 Results

3.4.1 MmpL3 bioinformatics studies

The bioinformatics studies were carried out in collaboration with Dr. Vassily Bavro, School of Biosciences, University of Birmingham. Although, MmpL proteins have been catagorised into RND family of proteins, no comprehensive studies was carried out. Our colloborators, compared sequence of MmpL3 protein to other RND protein sequences from different organisms. The sequence comparision studies did not show significant similarity, however shared a structural similaty. Our colloborators built a homology model by threading the aminoacid sequence through available crystal structures of other RND proteins, namely AcrB (1OY6.pdb) (Yu *et al*. 2003) from *E. coli,* MexB (2V50.pdb) (Sennhauser *et al*. 2009) from *P. aeureginosa,* and CusA (4DOP.pdb) (Su *et al*. 2012) from *E. coli.* Structural homology comparisions revealed few unexpected informations. While, as expected transmembrane regions aligned tightly, cytoplasmic domains exhibited variance in alignement, a detailed study, suggested a common fold shared among all the RND proteins (Figure 3-5). The variance in the cytoplasmic domains comes from insertions at different places in different protein and in some cases due to duplication of cytoplasmic domain as in AcrB. These Cytoplasmic domains have been shown to be involved in substrate recognition and substrate binding (Tsukazaki *et al*. 2013; Long *et al*. 2010; Yu *et al*. 2003; Seeger *et al*. 2009; Sennhauser *et al*. 2009). This prompted to us to evaluate the functional role of cytoplasmic domains in MmpL3 by constructing aforementioned cytoplasmic domain construct (detailed earlier in section 3.2).

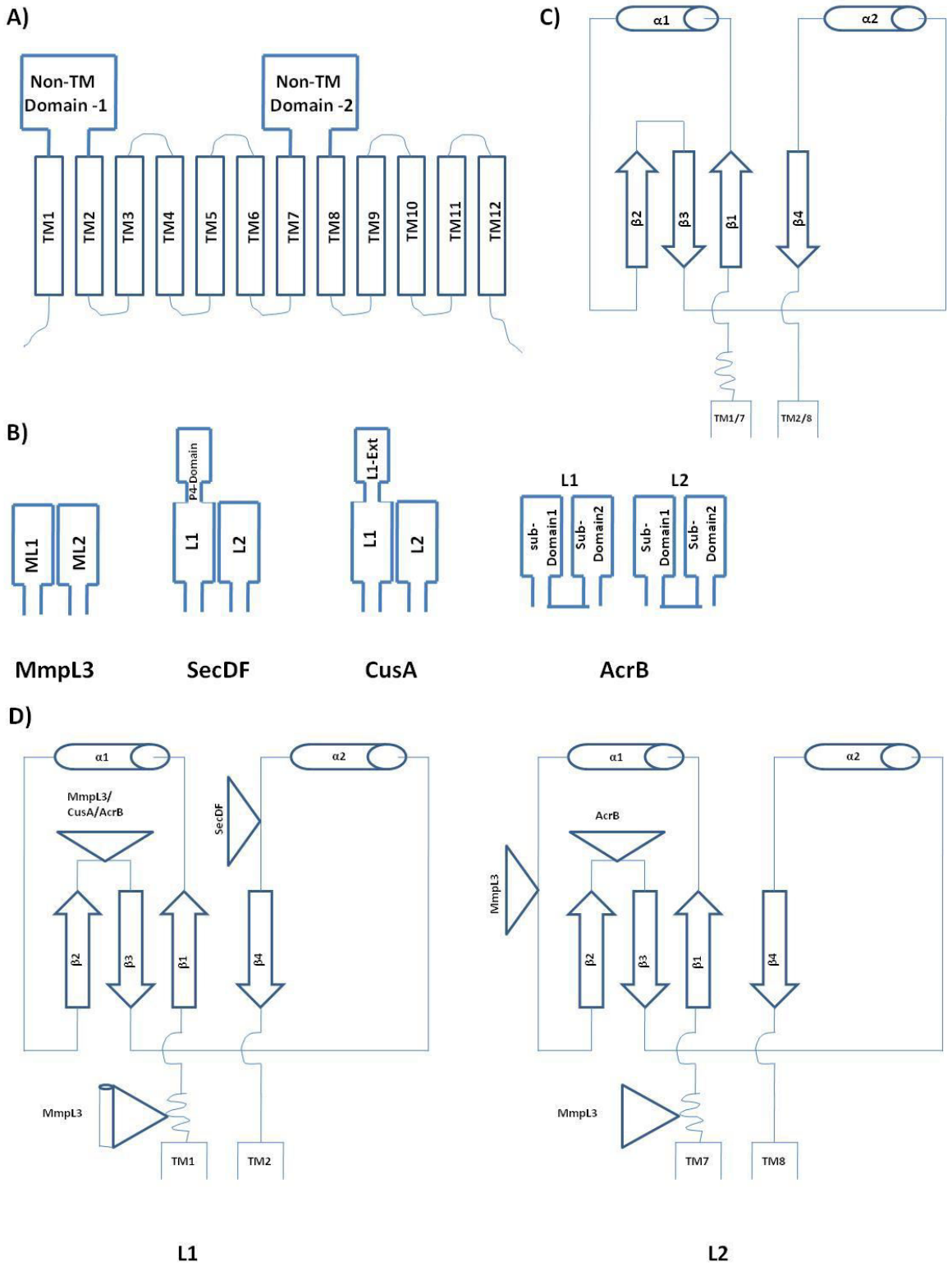


Figure 3-5 Schematic depiction of general topological structure of RND family of proteins and their non-TM domains. **(A)** Topology of MmpL3 compared to other known RND transporter proteins shows a shared general architecture, **(B)** non-TM domain organisation in MmpL3, SecDF(with P4 domain), CusA (with L1-extension) and AcrB (sub-domains 1 and 2 in each of the non-TM domains) proteins, **(C)** genreal fold of non-TM domains showing mixed α-β arrangements and **(D)** variations in the general fold due to insertions in different RND transporter non-TM region-1 (L1) and non-TM region-2 (L2). ML1 and ML2 correspond to non-TM region-1 and non-TM region-2 respectively.

Structural predictions of MmpL3 supported earlier reports suggesting that MmpL3 belongs to the RND family of proteins and recently MmpL3 was shown to use proton motive force (PMF) as the energy source for its activity (Yang *et al*. 2014). Comparing the MmpL3 structure prediction with available crystal structures of other RND transporters helped us to predict residues that are likely to be involved in generating a proton-motive force (Tsukazaki *et al*. 2013; Seeger *et al*. 2009). The amino acid residues that generate proton motive force resides on transmembrane helices TM#4 and TM#10 which aligns closely in 3-D space. The two negatively charged aspartate residues (D407, D408) on TM#4 facing a positively charged Lysine (K940) constitutes proton relay pump in AcrB. MmpL3 has a single aspartate residue (D251) on TM#4 in the same position however; do not possess a counter charged Lysine or arginine residue. However, it does possess two negatively charged glutamate residues (E262, E263), slightly lower on the TM#4 with a positively charged arginine (R653) on the TM#10, and is identical to one observed in SecDF proton pump (E351, E352 on TM#4 and R644 on TM#10). However SecDF also has another positively charged arginine (R671) on adjacent TM#11 which is not present in MmpL3. This prompted us to generate site directed mutants of these residues (I250T, Y252A, Y252D, D251A, E262A and R653A) to test whether these mutations affect proton relay and there by the transport of mycolic acids.

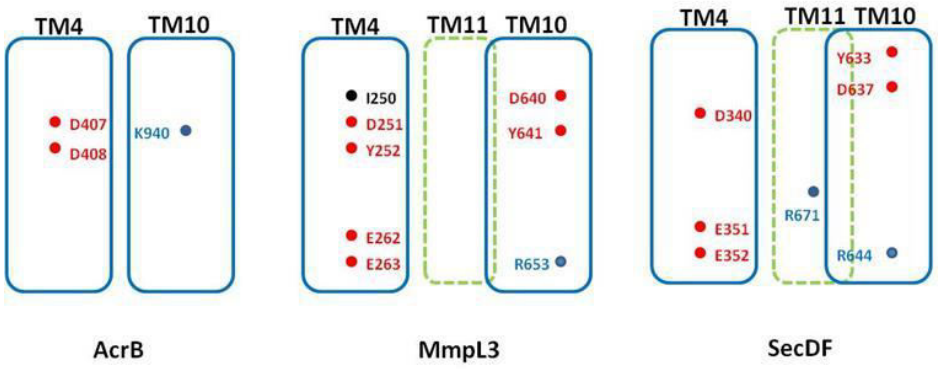


Figure 3-6 Schematic representation of transmembrane regions involved in proton relay. Residues involved in proton relay are indicated; acidic residues indicated in red and positively charged amino acids in blue. Transmembrane helices are indicated above.

3.4.2 Expression and purification of non-TM constructs

Non-TM domains were expressed and purified individually for structural and biophysical studies. ML constructs (ML-C_{HIS} and ML-N_{HIS}) were expressed in *E. coli* C41 (DE3). Figure 3-7 shows the expression of ML constructs (ML-C_{HIS} and ML-N_{HIS}). ML C_{HIS} protein was purified using Immobilised Metal Affinity Chromatography (IMAC) and was further purified using ion exchange chromatography (Figure 3-8). The purified protein was used for further biophysical and structural studies. ML-N_{HIS} protein was purified using affinity chromatography (Figure 3-9). Protein purified to homogeneity was used for structural studies.

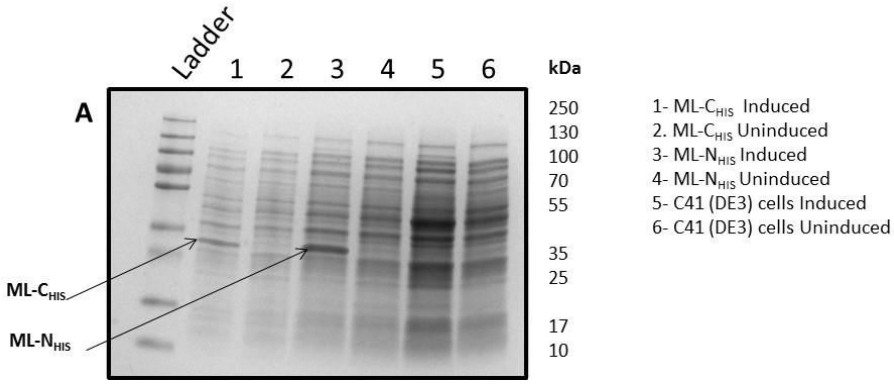


Figure 3-7 SDS-PAGE analysis of expression of ML. ML was expressed in *E. coli* C41 (DE3) cells and analysed by SDS-PAGE. Both N_{HIS} and C_{HIS} constructs expressed the protein significantly (indicated by arrows).

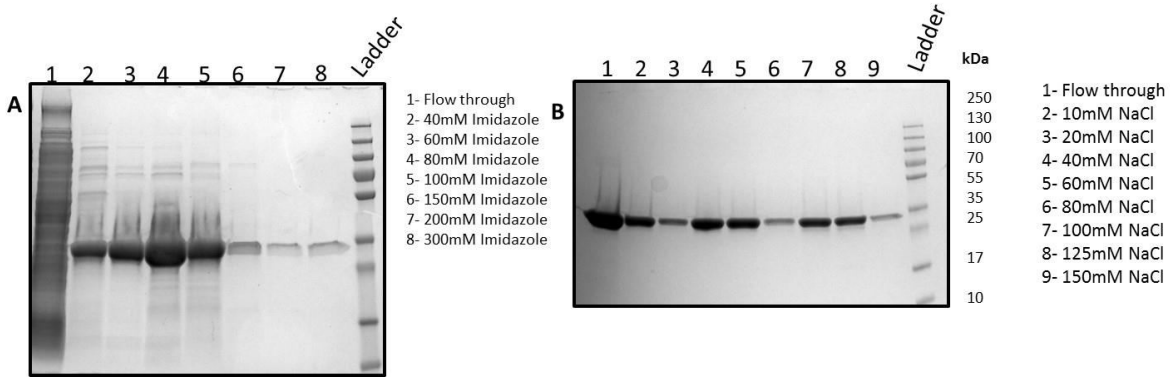


Figure 3-8 SDS-PAGE analysis of purification of ML-C_{HIS} using affinity and ion-exchange chromatography. ML was first purified using Ni^{2+}-NTA column (A) and fractions containing pure protein (60 mM-100 mM fractions) were concentrated and dialysed against buffer containing no salt. Dialysed protein was then purified using ion-exchange column developed using NaCl gradient (B), fractions containing pure protein (0 mM to 125mM fractions) were pooled, concentrated and used in downstream activities.

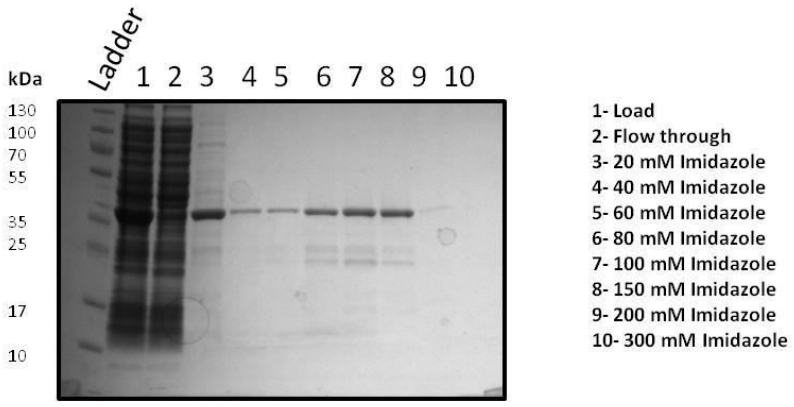


Figure 3-9 SDS-PAGE analysis of expression of ML construct and purification of ML-N_{HIS} protein using affinity chromatography. Fractions containing (40 mM to 150 mM fractions) pure protein were pooled, concentrated and dialysed. Purified protein was used for further structural studies.

Non-TM domains ML1 and ML2 were also expressed in *E. coli* expression strain C41 (DE3) as both N_{HIS} and C_{HIS} constructs. Both the constructs for ML1 expressed the protein significantly while only ML2-N_{HIS} expressed protein significantly (Figure 3-10). ML1 and ML2proteins were purified using IMAC as described previously. Fractions containing protein of interest (Figure 3-11, Figure 3-12) were pooled, concentrated and dialysed for use in downstream biophysical and structural studies.

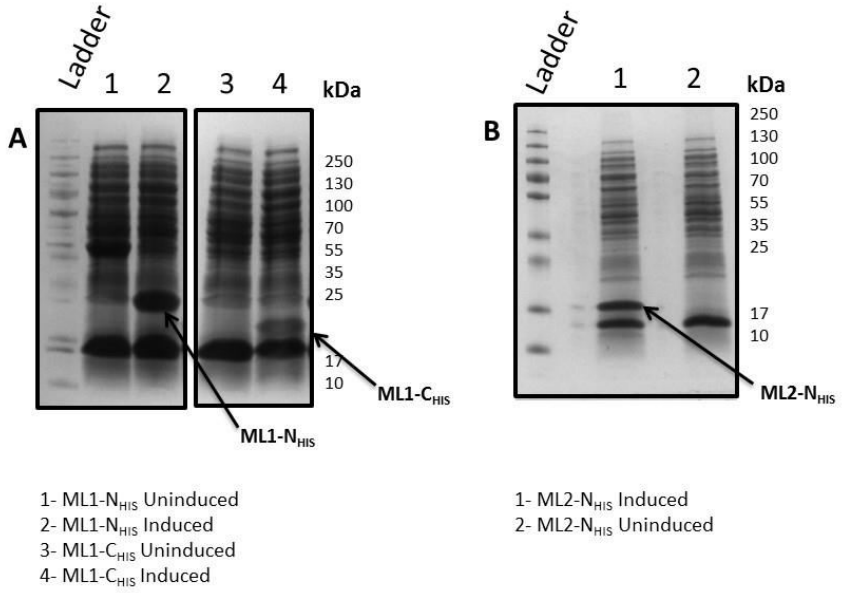


Figure 3-10 SDS-PAGE analysis of expression of ML1 and ML2. ML1 and ML2 were expressed in *E. coli* C41 (DE3) cells and analysed by SDS-PAGE. Both N_{HIS} and C_{HIS} constructs for ML1 expressed significantly (indicated by arrows) while only N_{HIS}construct of ML2 expressed significantly (indicated by arrow).

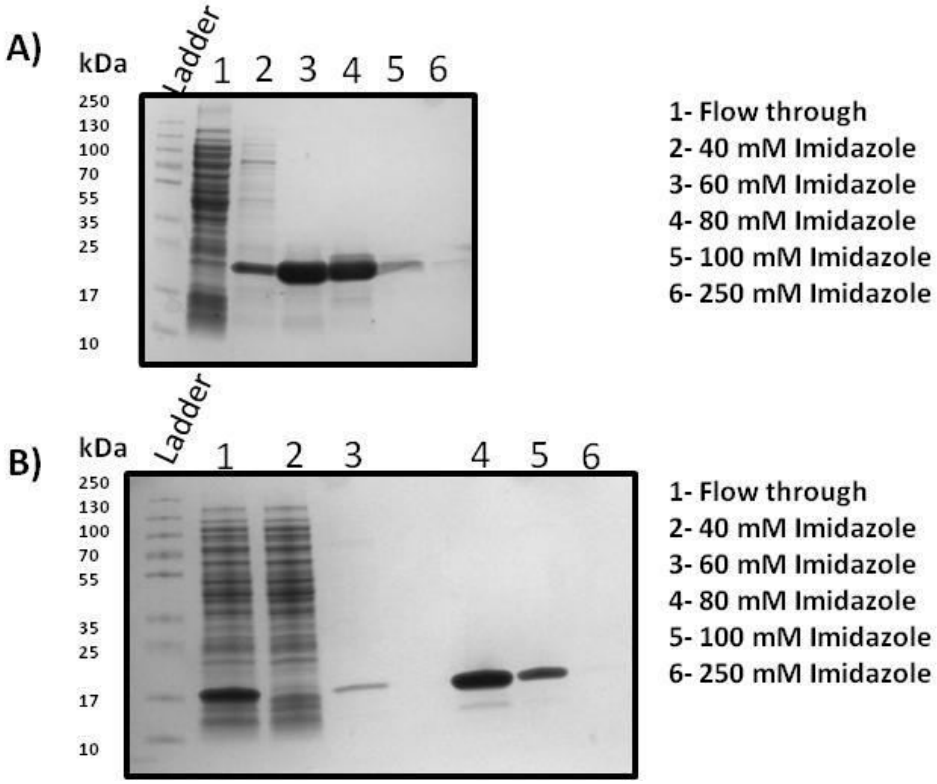


Figure 3-11 SDS-PAGE analysis of purification of MmpL3 non-TM domain ML1-N_{HIS} and ML1-C_{HIS} using affinity chromatography. ML1-N_{HIS} was purified using Ni^{2+}-NTA column (A) and fractions containing pure protein (60 mM-80 mM fractions) were concentrated and dialysed for downstream usage. ML1-C_{HIS} was also purified using Ni^{2+}-NTA column (B) and fractions containing pure protein (80 mM-100 mM fractions) were concentrated and dialysed for downstream usage.

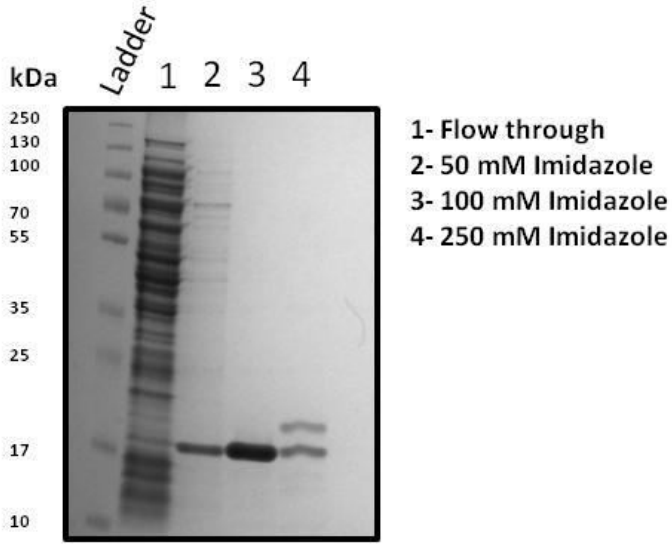


Figure 3-12 SDS-PAGE analysis of purification of MmpL3 non-TM domain ML2-N_{HIS} using affinity chromatography. ML2-N_{HIS} was purified using Ni^{2+}-NTA column and fractions containing pure protein (100 mM fraction) was concentrated and dialysed for downstream usage.

3.4.3 Structural and biophysical characterisation of MmpL3 non-TM domains

3.4.3.1 Size exclusion chromatography

One of the key aims of the study was to decipher the oligomeric states of the protein. Towards this, size exclusion chromatography (SEC) was used to study oligomeric states of the ML-C_{HIS}protein. The oligomeric state of the protein was predicted with the help of molecular weight standards corresponding to 30 kD and 60 kD (Appendix-3). The results (Figure 3-13) suggested that ML protein eluted as a single peak but over a range that corresponded to 30 kD and 60 kD, which means the protein could be a mix of monomer and a dimer but could not interpret confidently.

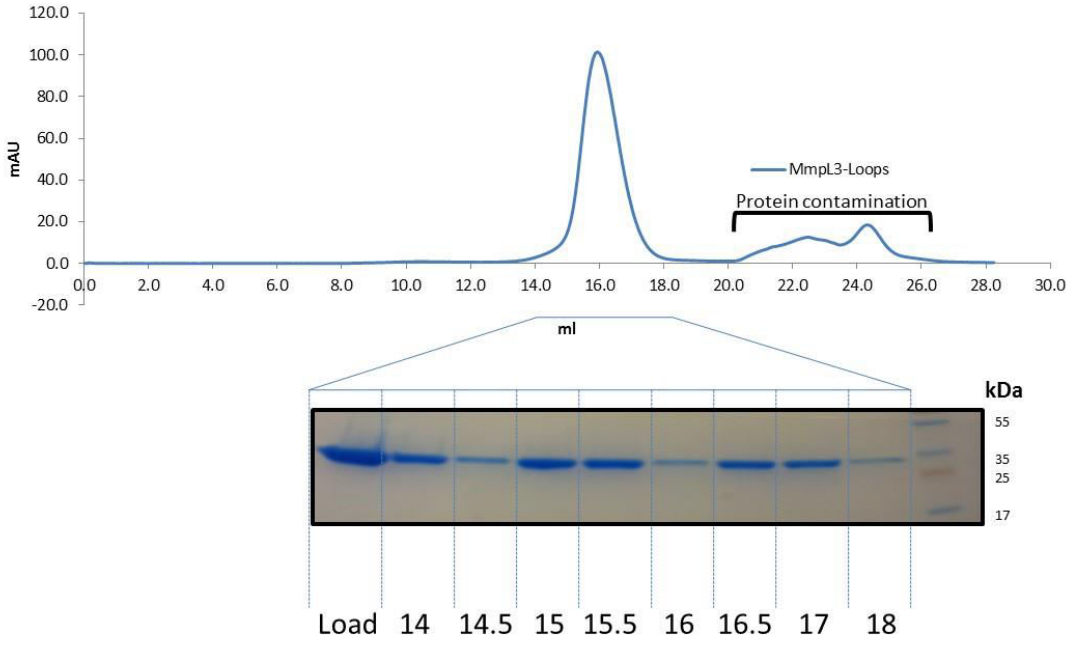


Figure 3-13 Size exclusion chromatogram of ML-C_{HIS}. ML chromatogram from size exclusion chromatography showing single peak corresponding to a mix of monomer and a dimer, analysis of protein by SDS-PAGE shown in the insert below.

3.4.3.2 Protein cross linking studies

Chemical cross linking of proteins in solution has been previously used to study the subunit stoichiometry of oligomeric proteins (Davies & Stark 1970). Oligomeric states of the protein

could not be deciphered conclusively using SEC and thus decided to follow up with protein cross linking studies and analytical ultracentrifugation studies. Chemical cross-linking agents glutaraldehyde and EGS (Ethylene glycol bis[succinimidylsuccinate])were used to test the oligomeric states of ML-C_{HIS}. The cross linking experiments with both glutaraldehyde (Figure 3-14, A) and EGS (Figure 3-14, B) suggested that a small fraction of the protein readily gets cross linked to form higher oligomers while a large fraction of the protein remain as monomers. It was also observed that the cross linking slightly improves in the presence of its cognate substrate TMM and trehalose (Figure 3-14, C). However, since the bands of monomeric protein were not depleting either with increasing concentration of cross linker or with increasing incubation time, we could not clearly establish the oligomeric states of the protein.

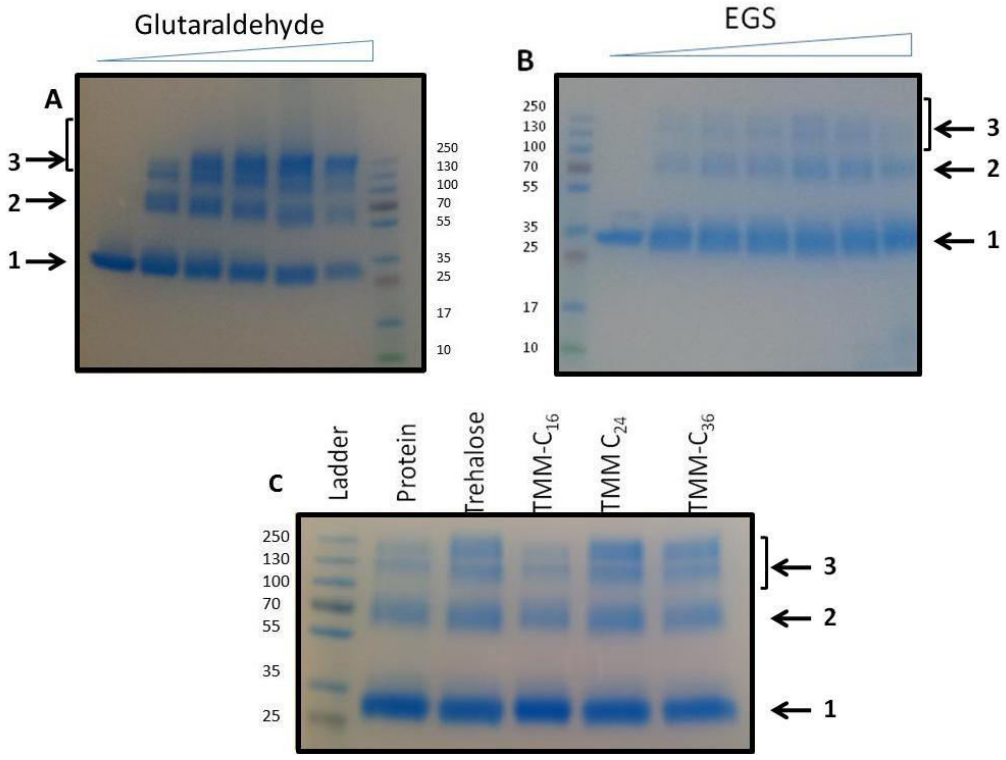


Figure 3-14 Chemical cross linking of ML to study oligomeric states of the protein. Protein crosslinking with glutaraldehyde revealed that higher oligomers were formed with increase in concentration of cross-linking agent (A) and results were reproducible with a different chemical cross linker (B). Further, extent of cross-linking improved in the presence of its cognate substrate TMM and trehalose (C), suggesting that a small fraction of the protein readily gets oligomerised. Oligomers are indicated with arrows 1-monomer, 2-dimer and 3- higher oligomers.

3.4.3.3 AUC

To get conclusive data on oligomeric status of the protein, AUC was performed with ML, ML1, ML2 and a stoichiometric mix of ML1 and ML2. Analysis of the AUC data revealed that ML1 and ML2 exist as monomers and a stoichiometric mix of ML1 and ML2 does not oligomerise. ML gives significantly sediments as two distinct species corresponding to 30 kD and 61 kD, respectively, which represents monomeric and dimeric species of ML (Figure 3-15). However, a large fraction of the protein behaves as monomer while only a small fraction forms dimer. This corroborates earlier observations made using protein cross-linking reagents.

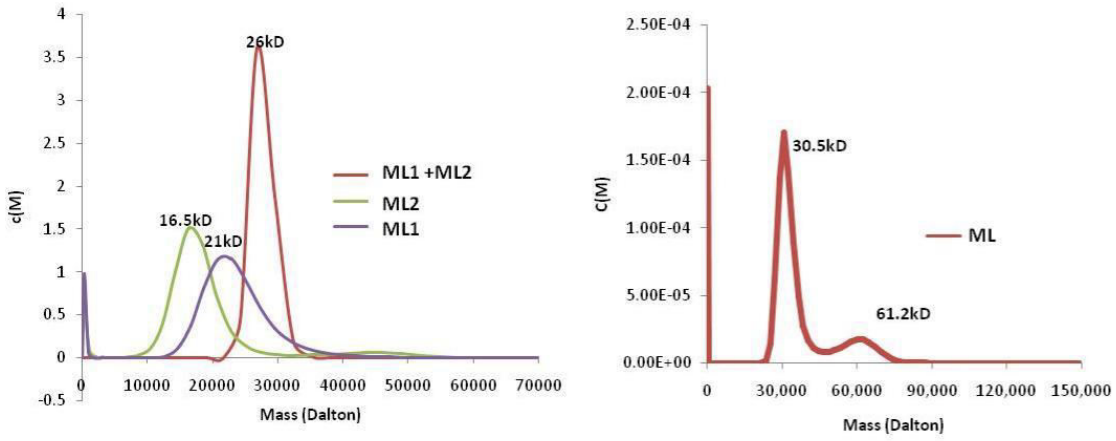


Figure 3-15 Analytical ultra centrifugation of MmpL3 non-TM domain constructs ML1, ML2 and ML. Analytical ultracentrifugation reveals that ML1, ML2 behaves as monomers (left) and ML, mainly exists as monomer but a small fraction behaves as dimer (right).

Table 3-1 Analytical ultra centrifugation data

	ML1	ML2	ML1+ML2	ML
Frictional ratio	1.422982	1.493157	1.390529	1.95
Rmsd	0.015109	0.025853	0.009866	0.0151
Mr (kDa) (from analysis)	21.7	16.66	26.7	30.6
				61.2

3.4.3.4 Protein crystallography

MmpL3 non-TM domain constructs ML, ML1 and ML2 proteins were purified to homogeneity and were used to set up crystal trays (Hanging drop trays) using commercially available crystal screens. After screening several conditions for each protein only ML-C_{HIS} produced crystals in 0.1 M Bis.Tris propane buffer pH 8.5, 0.2 M NaF and 20% w/v PEG3350. The crystal however did not diffract and our attempts to reproduce crystallisation or to optimise conditions to grow fresh crystals were not successful. Repeated efforts to obtain ML1 and ML2 crystals were unsuccessful. Stoichiometric mixtures of ML1 and ML2 were also used for protein crystallisation; however different stoichiometric mixtures did not produce any crystals.

3.4.3.5 Protein-NMR to resolve the structure of ML1 and ML2

Our collaborators Dr. Mark Jeeves (School of Cancer Sciences, University of Birmingham) did initial studies with [^{15}N] labelled ML1 and ML2 using NMR. ML could not be used in this studies due to size limitation (>20kD). Initial studies revealed that the ML1 showed peaks corresponding to random coils and β sheets andML2 showed peaks corresponding to β sheets with very little random coil. This bodes well with our predictions where ML1 at both insertion points shows large unstructured/randomly coiled structures but ML2 is more compact without such random coils. However this requires further NMR studies which could also help us solve the structures of ML1 and ML2.

3.4.3.6 Ligand binding studies

ITF exploits the spectral properties of aromatic amino acid residue tryptophan in proteins. Tryptophan emits with a emission maxima of 340nm when excited at 295nm. Ligand binding affects the fluorescence either by increasing or decreasing the intensity and/or causing shift in emission maxima. Our initial efforts to study substrate binding studies with trehalose, part of

its cognate substrate TMM, gave very little evidence to show substrate binding. Change in fluorescence was equally affected in the presence and absence of the substrate.

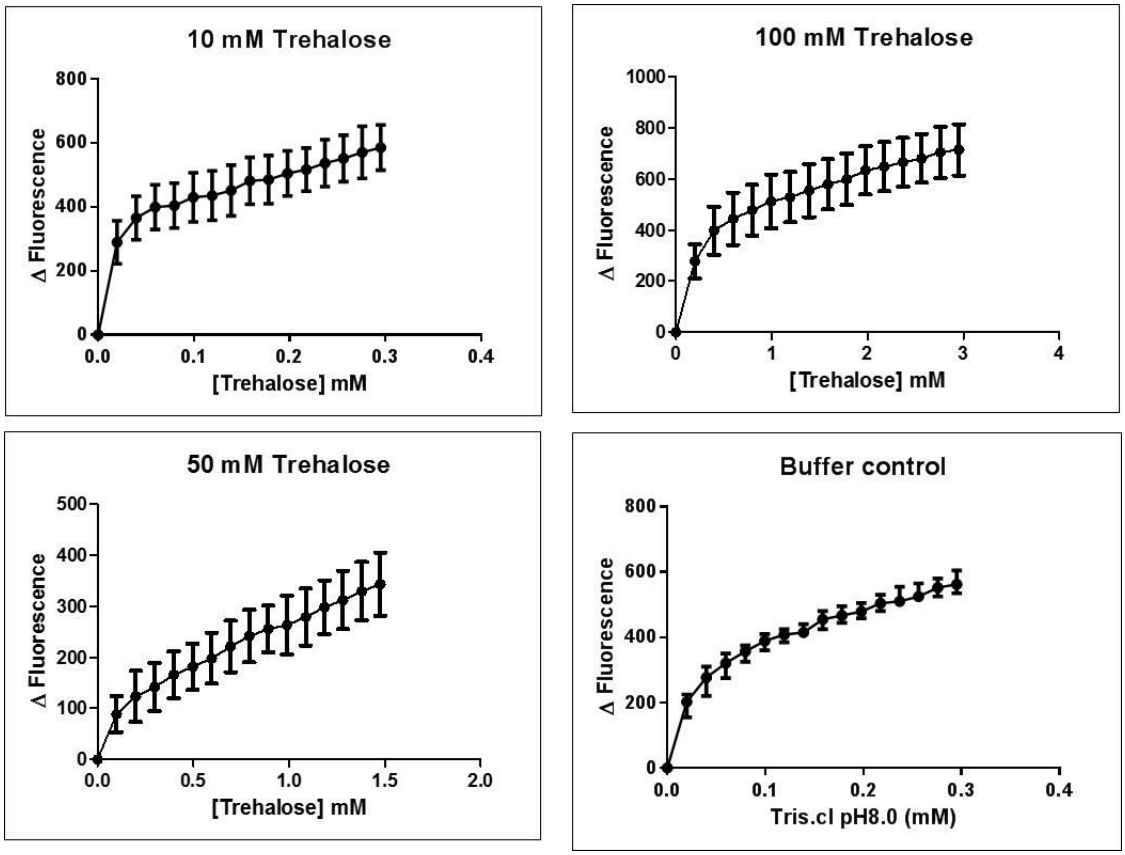


Figure 3-16 Ligand binding assay using ITF. Binding of trehalose to ML protein was evaluated by monitoring intrinsic tryptophan fluorescence. Increasing concentrations of trehalose did not affect the change in fluorescence significantly.

To further evaluate substrate binding we have used fluorophore bound substrate analogue TMP-BODIPY (trehalose monopalmitate-boron dipyrromethane). TMP-BODIPY emits strongly at 515 nm when it is in hydrophobic environment or when bound to protein. Binding of TMP-BDIPY to ML was evaluated by monitoring fluorescence at 515 nm. Inhibitors previously shown to affect mycolic acid transport such as SQ109, BM212, adamantyl urea, Spiro (410A) and THPP (366A) were used to validate the assay and to evaluate the specificity of binding. Binding of TMP-BODIPY was inhibited by only selected compounds (410A and adamantyl urea) suggesting that binding of TMP-BODIPY is specific (Figure 3-16).

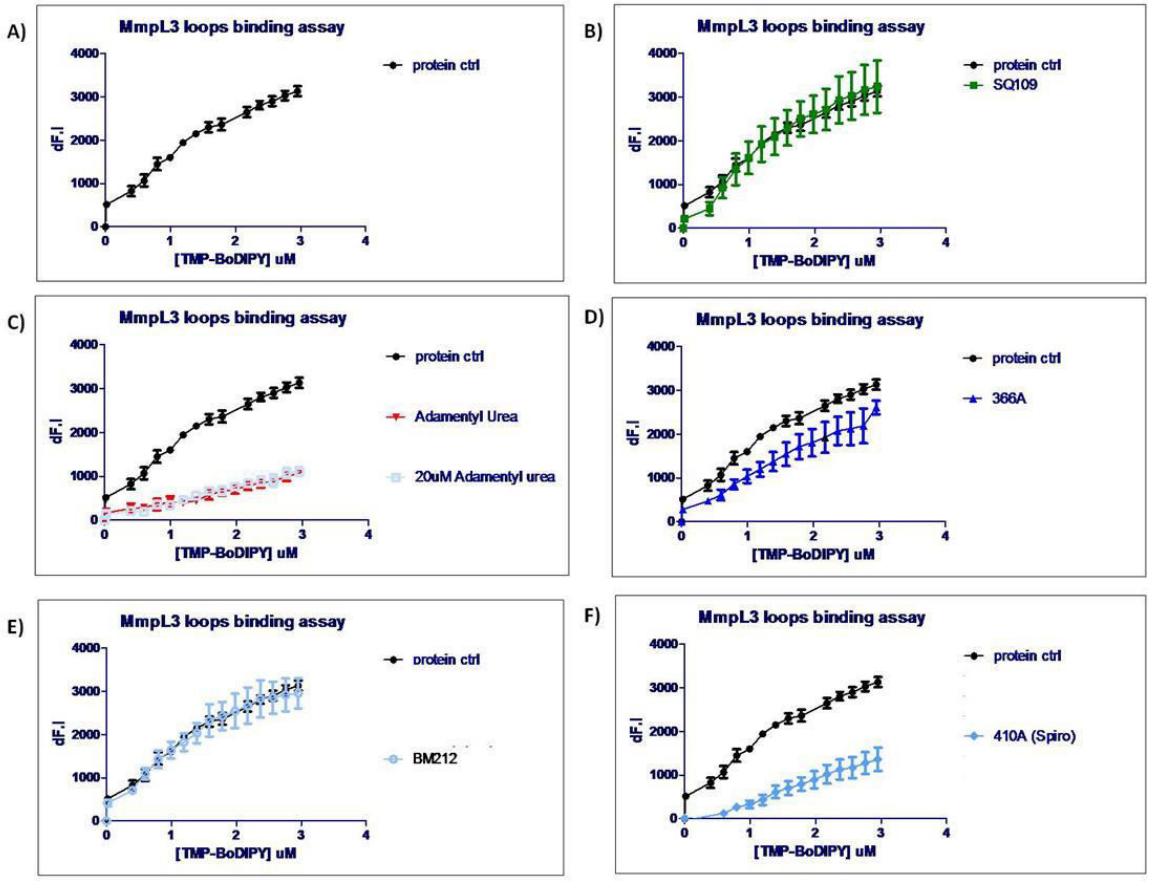


Figure 3-17 Ligand binding evaluation using TMP-BODIPY fluorescence assay. Known inhibitors of mycolic acid transport were used to validate the specificity of the assay. Only adamantyl urea (C) and 410A (F) were shown to affect TMP-BODIPY binding to ML, however other inhibitors SQ109 (B), THPP-366A (D) and BM212 (E) did not affect the binding of TMP-BODIPY to ML.

While characterising the non-TM regions of MmpL3, alternative attempts were made to optimise the expression of full length MmpL3. At this stage decision was made to pursue with the characterisation of full length protein which will be described in the following sections.

3.4.4 Expression and purification of full length MmpL3

For optimising the expression of full length MmpL3 protein, MmpL3 constructs (C_{HIS} construct and N_{HIS} construct) were expressed in *E. coli* expression strains C41 (DE3) and

protein expression was induced with 1 mM IPTG. Western blot analysis revealed that MmpL3-C_{HIS} construct expressed better than MmpL3-N_{HIS} construct, however the level of expression was very low and equally distributed between soluble and insoluble fraction. To optimise the expression levels, protein expression was induced at different temperatures (37°C, 25°C and 16°C) and best expression was seen at 37°C (Figure 3-18). MmpL3-C_{HIS}construct was used for further studies.

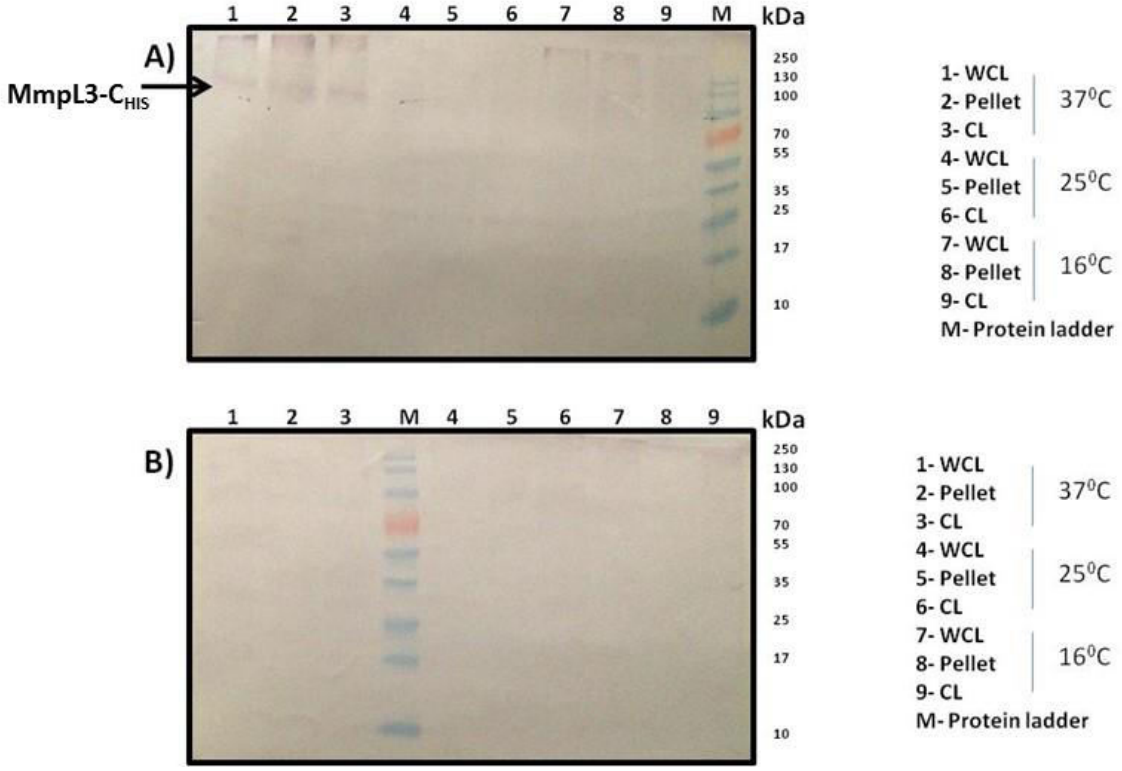


Figure 3-18 Western blot analysis of expression of MmpL3 protein. MmpL3 protein expressed in *E. coli* strain C41 (DE3) at different temperatures 37°C, 25°C and 16°C. MmpL3-C_{HIS} has expressed atlow quantities at 37°C but expression was not detected at lower temperatures (A) but MmpL3-N_{HIS} construct did not express protein at any temperatures tested. Analysed samples include whole cell lysate, pellet and supernatant (CL) from 10,000 x g spin. Protein of interest indicated with an arrow.

The presence of divalent cations in the growth media was shown to improve the soluble expression of proteins in *E. coli* (White S A. School of Biosciences, University of Birmingham, unpublished data). $MgCl_2$ was used at a concentration of 2mM and 10mM as a supplement in the growth media to improve the solubility of protein. A parallel expression

study was done in alternative growth media (Terrific broth) to test for improved protein expression. However, presence of $MgCl_2$ did not have the desired results and MmpL3 solubility was not improved (Figure 3-19). Use of Terrific broth as an alternative growth media did not improve the protein expression significantly either.

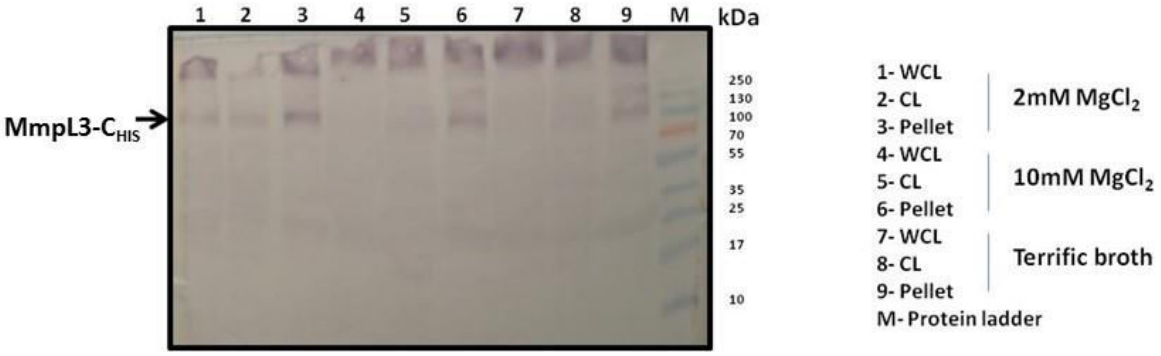


Figure 3-19 Western blot analysis of MmpL3 expression in the presence of cations. MmpL3 protein expressed in *E. coli* strain C41 (DE3) in the presence of $MgCl_2$ (lanes 1-6). MmpL3-C_{HIS} expressed atlow quantities but protein was found in pellet fraction due to lack of soluble expression. Use of different media was not very helpful for soluble expression (lanes 7 to 9).

A codon alteration strategy which optimises the GC content at the 5' end of the gene which helps to ease the secondary structure formed in the transcript and thereby facilitates better expression of the protein was adopted (Usha *et al*. 2006). For this, *mmpL3* gene was PCR amplified using codon altered primers (Figure 3-20) and sub-cloned into pET41c. pET41c: MmpL3 (CB) was transformed into *E. coli* C41 (DE3) cells for expression studies. Protein expression experiment was done under the earlier described conditions. Western blot analysis (Figure 3-21) of the expressed proteins did not help improve the expression of protein significantly.

A) GTGTTCGCCTGGTGGGGTCGAACTGTGTACCGCTACCGGTTCATCGTAATC
B) GTGTTTGCATGGTGGGGTCGTACTGTATATCGATACCGGTTCATCGTAATC

Figure 3-20 Codon alteration strategy used to optimise the expression of MmpL3. Codoon alteration strategy used to optimise GC content at the 5' end of the gene. Native sequence of the *mmpL3* gene (A) and altered gene sequence (B) highlighting altered nucleotide residues in colour, used for priming PCR amplification of *mmpL3*

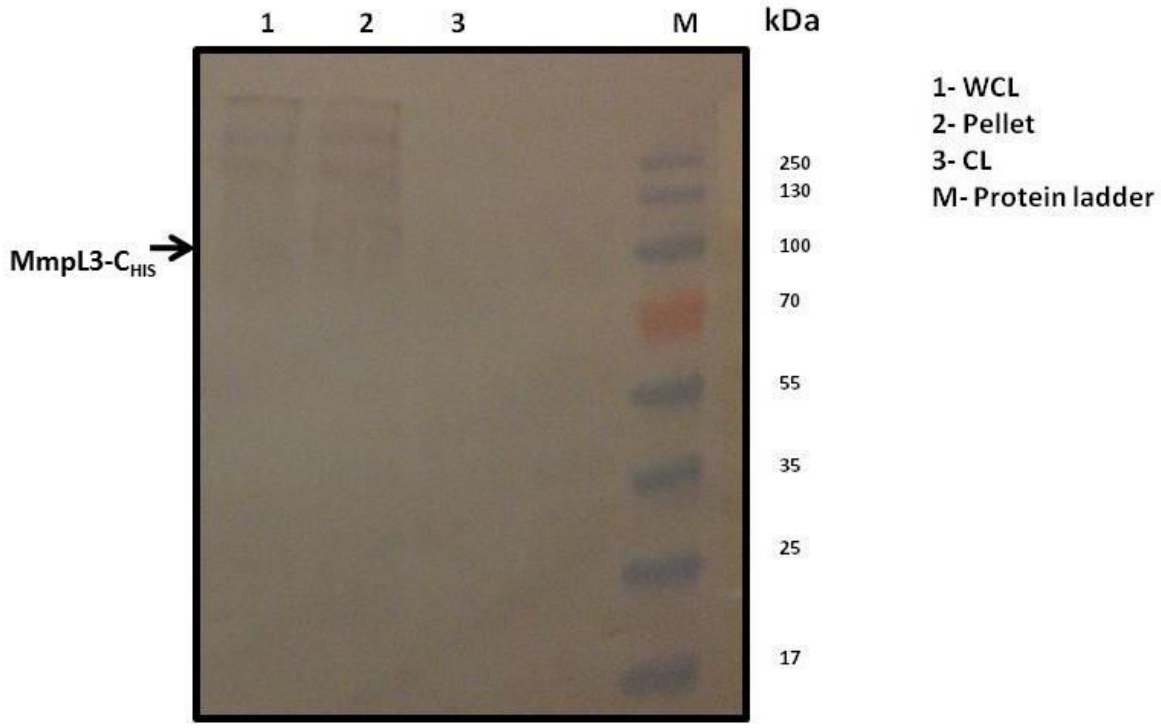


Figure 3-21 Western blot analysis of MmpL3 expression from codon altered construct. MmpL3 protein expressed in *E. coli* strain C41 (DE3) using codon altered construct. Protein expression was not significant and expected band size is indicated with an arrow.

One well known strategy to improve the expression of recombinant protein, is the use of alternate tags that may alter the solubility (Maina *et al*. 1988; Tropea *et al*. 2007; Smith & Johnson 1988). We have used a maltose binding protein and glutathione S-transferase as fusion tags to improve protein expression and solubility. For this we cloned MmpL3 into pMALp2X and pGEX-4T-1 expression vectors for expression in *E.coli* C41 (DE3). Western blot analysis of expression revealed that fusion tag did not express the protein at detectable levels (Appendix-7).

Mycobacterial chaperones Cpn60.1 and Cpn60.2 have previously been shown to improve solubility of proteins when co-expressed in *E. coli*(Henderson *et al*. 2010; Batt *et al*. 2012). We have co-expressed MmpL3 with Cpn60.1 and Cpn60.2 in an effort to enhance solubility and improve protein yields. Co-expression with either Cpn60.1 or Cpn60.2 did not result in detectable level of expression of MmpL3.

Codon optimisation has long been used to improve the expression of recombinant proteins in *E. coli* (Grosjean & Fiers 1982). To investigate whether codon optimisation would improve the expression of MmpL3, the gene was codon optimised for expression in *E. coli* (synthesised from Genscript.Inc) and cloned into the pET28b expression vector with a C-terminal His-tag. Expression studies were done as described earlier using *E. coli* C41 (DE3) strain. SDS-PAGE analysis followed by Western blot analysis using anti-His antibody showed improved expression of MmpL3 which allowed purification of MmpL3.

Most methods of membrane protein purification involve the use of harsh conditions such as the use of detergents. We employed a SMA polymer previously described for the purification of membrane proteins from their native environment (Jamshad *et al*. 2011). The MmpL3 protein was purified using the SMA polymers that encapsulate the membrane protein along with the surrounding membrane. The MmpL3 was purified from solubilised membrane fraction (Section 3.3.5.2) using affinity chromatography. Purified fractions were analysed by SDS-PAGE and Western blot (Figure 3-22). Fractions containing MmpL3 were pooled, concentrated to 1 mg/ml and purified further using a size exclusion column (Figure 3-23). The purified protein was concentrated using Amicon centriprep (GE Amersham) and used for subsequent biophysical characterisation.

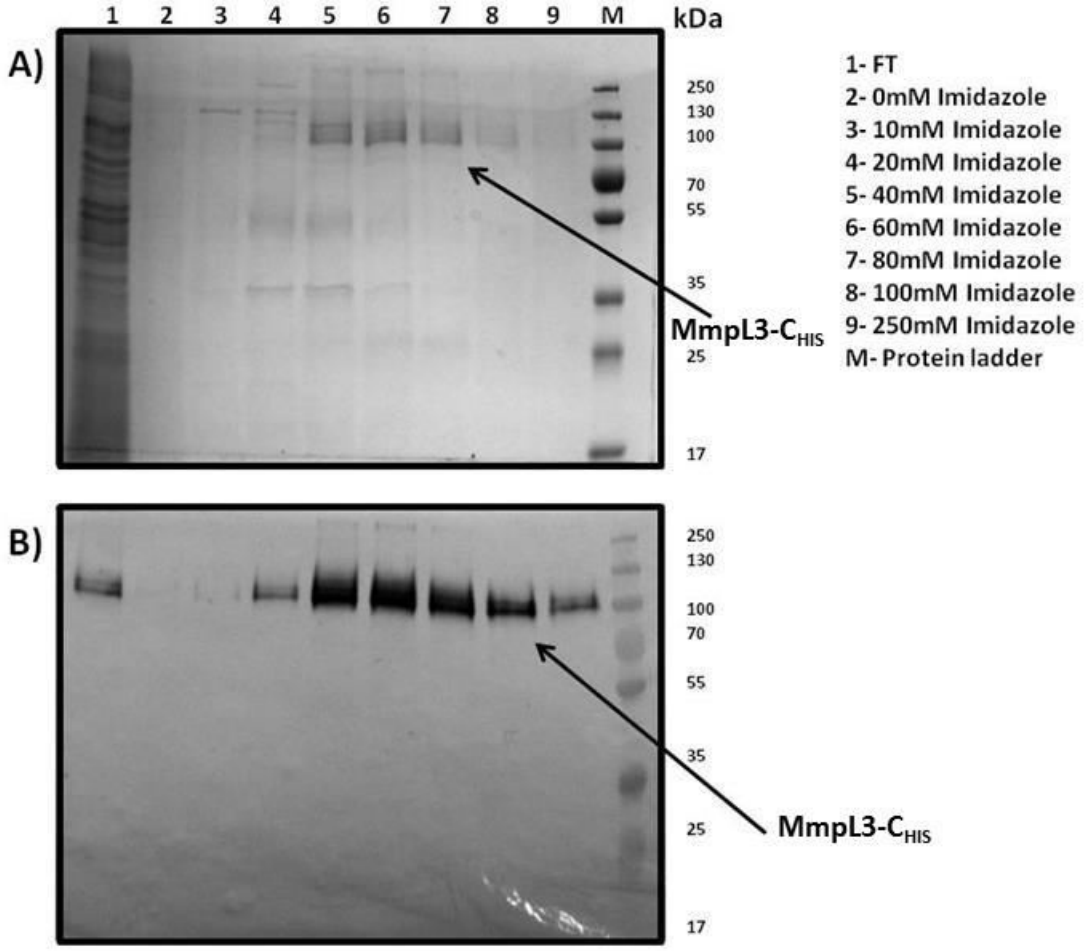


Figure 3-22 Purification of SMA solubilised MmpL3 using affinity chromatography. Codon optimised MmpL3 was expressed in *E. coli* strain C41 (DE3) and purified using affinity chromatography, analysed by SDS-PAGE (A) and confirmed by Western blot analysis of the same showing purified MmpL3-C_{HIS} (B). Protein of interest is shown with an arrow.

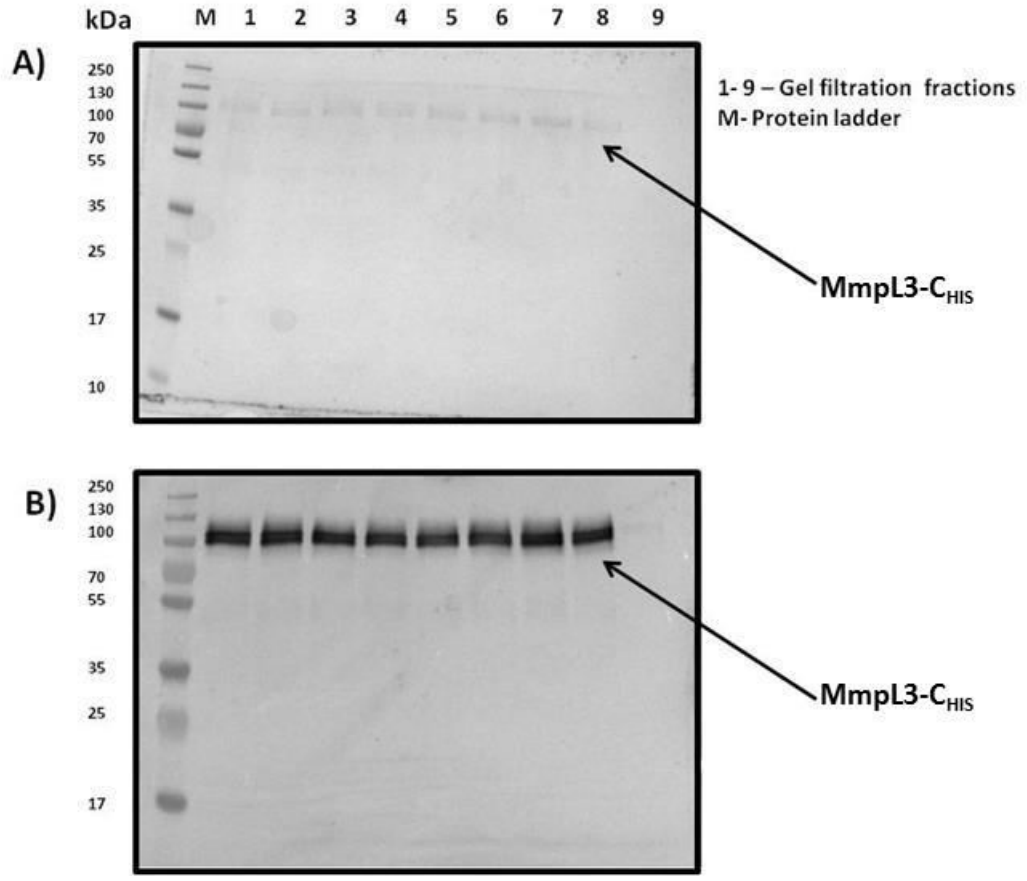


Figure 3-23 Further purification of SMA solubilised MmpL3 using Size exclusion chromatography. MmpL3 partially purified using affinity chromatography was further purified using size exclusion chromatography. Purified fractions were analysed by SDS-PAGE (A) and confirmed by Western blot (B). Protein of interest is shown with an arrow.

3.4.5 Biophysical characterisation of MmpL3 protein

3.4.5.1 Size exclusion chromatography

SMA solubilised MmpL3 purified to homogeneity, was evaluated using size exclusion chromatography, as an analytical tool to evaluate the oligomeric status of MmpL3.The protein distributes into two peaks across 20ml fractions when analysed usingWestern blot analysis (Figure 3-24). This suggests that MmpL3 exists in at least two oligomeric states. These fractions were subjected to AUC to ascertain their oligomeric status.

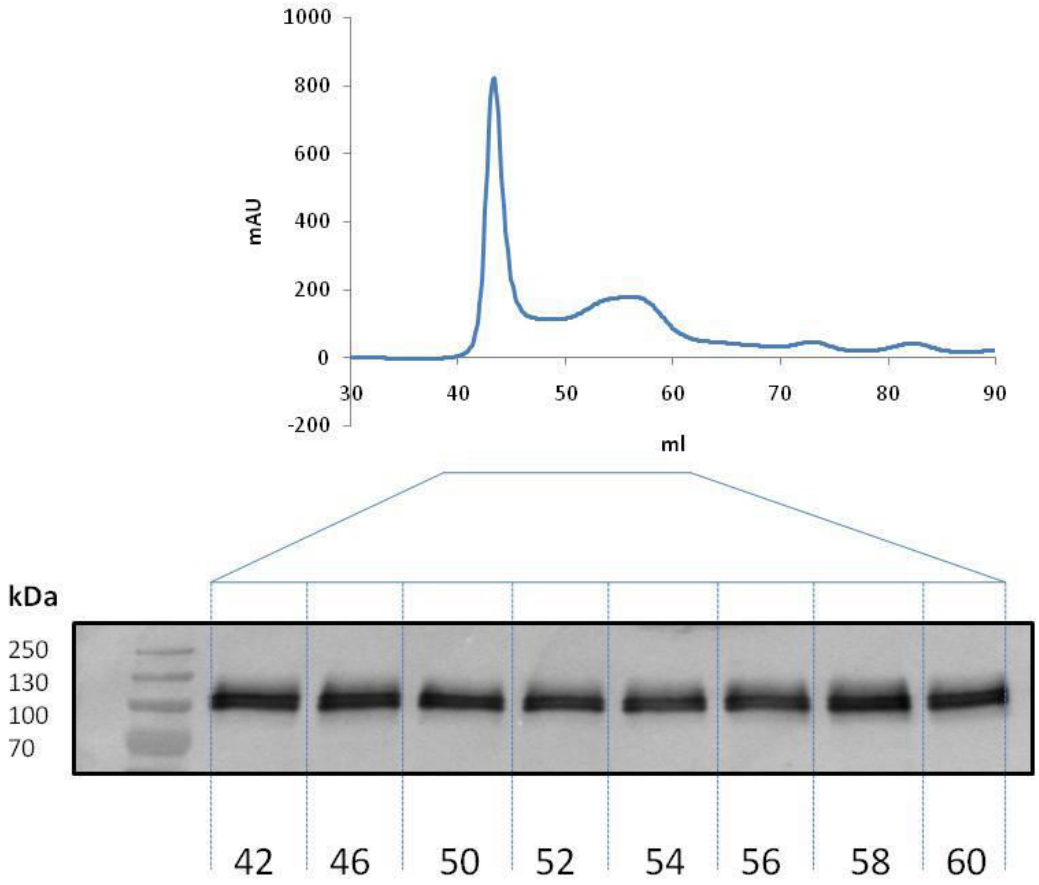


Figure 3-24 Size exclusion chromatogram for MmpL3. MmpL3 chromatogram reveals distribution of two peaks corresponding to different oligomeric states of the protein and Western blot analysis of fractions corresponding to peaks shown below in the insert.

3.4.5.2 AUC

One of the key aims of the study was to evaluate the oligomeric status of the MmpL3 which would help examine the similarities drawn with RND proteins. Other RND proteins exist in higher oligomeric states (Long *et al*. 2010; Murakami *et al*. 2002; Sennhauser *et al*. 2009). To determine the oligomeric status of the protein, purified MmpL3 fractions from size exclusion chromatography were subjected to AUC. Sedimentation velocity experiment data was analysed using SedFit software and revealed that MmpL3 distributes into two distinct species (Figure 3-25) that correspond to a dimer (221kDa) and the other corresponding to a tetramer (395kDa). This result corroborates the earlier results obtained with the MmpL3 non-TM domain fusion protein (ML).

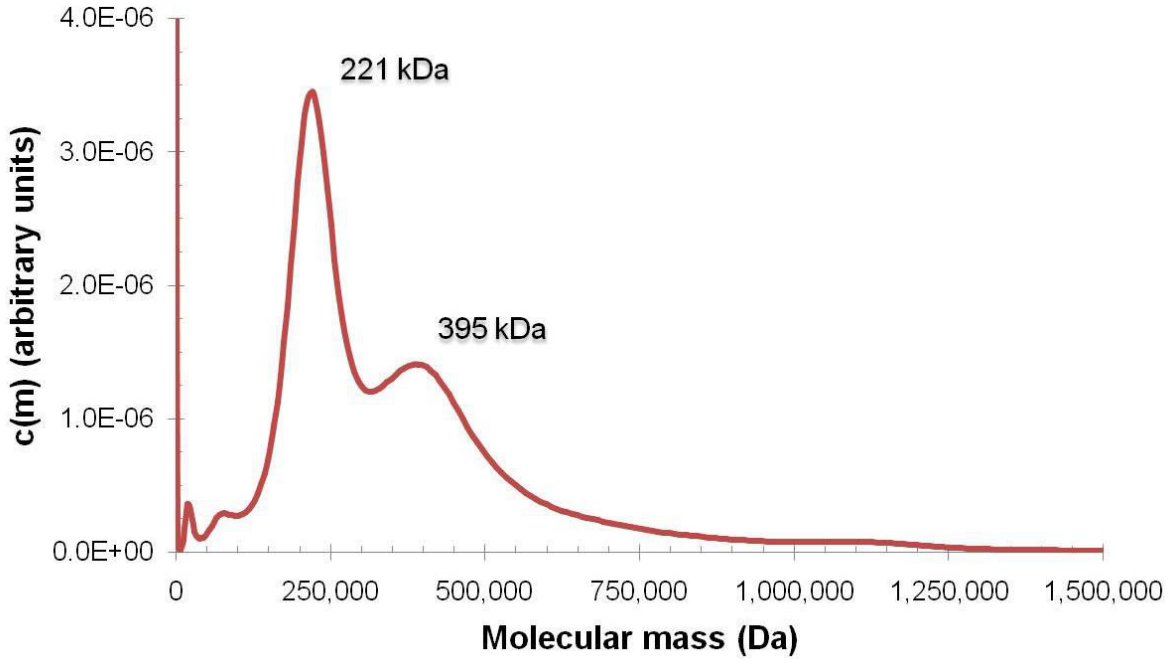


Figure 3-25 Analytical ultra centrifugation of MmpL3. SMA solubilised MmpL3 was purified to homogeneity and subjected to analytical ultracentrifugation which revealed that MmpL3 exists in two distinct oligomeric states that corresponds to dimer (221 kD) and a small fraction exist as tetramer (395 kD).

Table 3-2 Analysis of analytical ultra centrifugation data

	MmpL3 (SMA solublised)
Friction Ratio	1.50
Rmsd	0.007567
Mw	395 kDa
Mw	221 kDa

3.4.5.3 Circular Dichroism

Heterologous and overexpression of proteins often leads to inappropriate folding of proteins, formation of inclusion bodies and protein aggregation (Baneyx & Mujacic 2004). Although, SMA encapsulates membrane proteins from its native environment, it does not give conclusive evidence concerning protein folding. Circular Dichroism (CD) has long been used to study secondary structural features of protein and to get insights into protein folding. UV-CD (ultra violet circular dichroism) exploits the spectral properties of mainly peptide bonds

at near UV regions to get insights into protein folding (Kelly *et al*. 2005). Thus, CD was employed to study folding of SMA solubilised MmpL3. CD spectra (Figure 3-26) revealed that that protein gave negative ellipticity at 222 nm and 208 nm and a positive ellipticity at 193 nm, suggesting that MmpL3 mainly consists of α-helical structures, a characteristic feature of transmembrane proteins (MmpL3 has twelve transmembrane helices) and implicates that the secondary structure of the protein is intact. Signals that correspond to disordered protein (very low ellipticity at 210 nm and negative ellipticity at 193 nm) were not observed suggesting proper folding of the protein (Greenfield 2006).

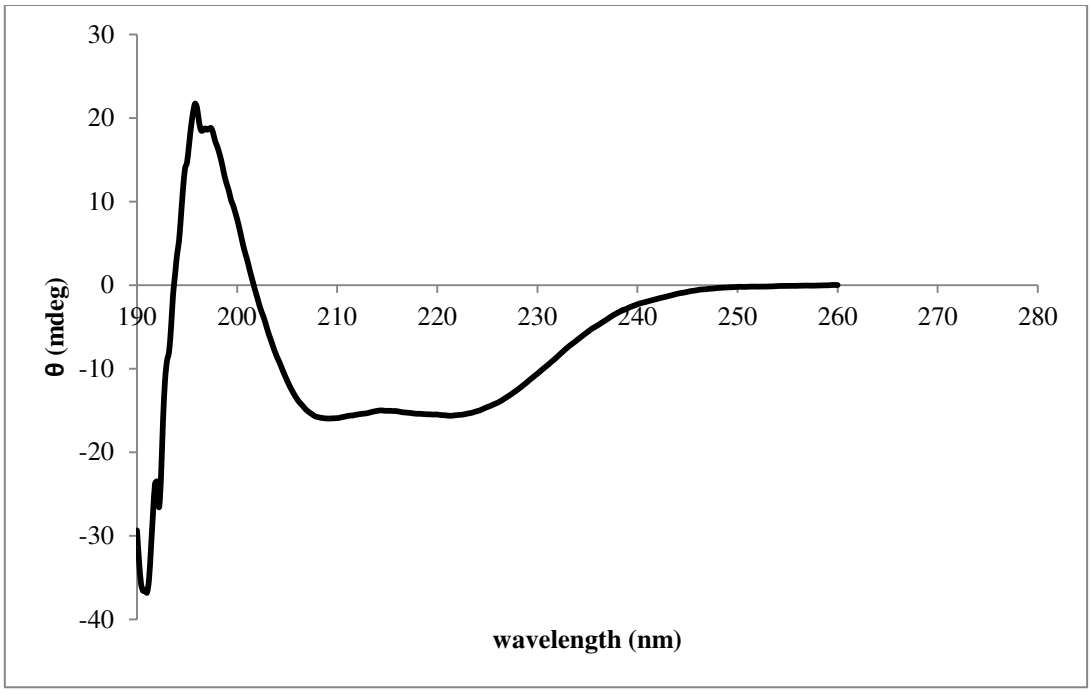


Figure 3-26 CD spectra of MmpL3. SMA solubilised MmpL3 folding was examined by using CD which revealed a negative ellipticity at 222 nm and 208 nm and a positive ellipticity at 193 nm typical of α-helical proteins. MmpL3 which contains twelve transmembrane helices is predicted to be α-helical.

3.4.5.4 Cryo electron microscopy (Cryo-EM)

Another key aspect of this study was to obtain a crystal structure of the protein. However, MmpL3 could not be expressed in sufficient quantities to set up protein crystallisation trays. Thus, alternative tools to get this structural insight were sought. Cryo-EM was previously used to obtain low resolution structural data for proteins which are difficult to crystallise, such as the mycobacterial FAS-I complex (Boehringer *et al*. 2013). Cryo-EM also helps to validate model predictions and thus is a valuable tool for the study of protein structures. Cryo-EM studies were carried out in collaboration with Prof. Tim Dafforn, School of Biosciences, University of Birmingham. Data collection and analysis were performed by Dr. Sarah Lee, School of Biosciences, University of Birmingham and a MmpL3 model was built by Dr. Vassily Bavro and Dr. KlausFutterer , School of Biosciences, University of Birminghm.

Preliminary data analysis (Figure 3-27) indicates that the envelope could accommodate two molecules of predicted MmpL3 model, suggesting a dimeric status of the protein. This conclusively corroborates similar observations made with other experiments. A detailed analysis of the data will help us gain more structural information of the protein.

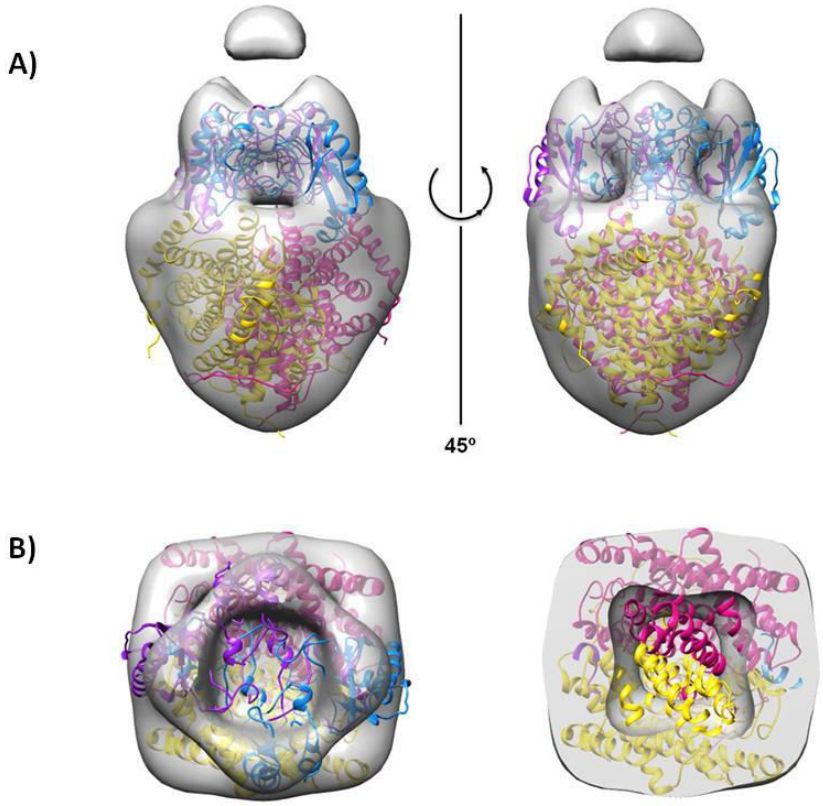


Figure 3-27 The reconstituted particle of MmpL3 from cryo-electron microscopy. MmpL3 model built on CusA template and fitted into the cryo-EM density. Lateral view of the fitted cryo-EM model shows arrangement of transmembrane regions and the two non-TM domain regions (ML1 and ML2) accommodated at the top (A) and view from top (on the left, B) and view from the bottom (on the right, B). Two molecules of MmpL3 are colour coded to highlight the dimeric state.

3.4.5.5 Ligand binding studies

ITFassay was employed to test ligand binding. The natural substrate of MmpL3, TMM and known inhibitors of MmpL3-adamantyl ureas, BM212, THPP, Spiro compounds were used as ligands. However, it was difficult to interpret the binding data since the protein encapsulated in SMALP contains hydrophobic membrane components which contributed to the change in fluorescence observed. Thus we could not conclusively differentiate between specific and non-specific binding.

1-Anilinonaphthalene-8-Sulfonic acid (1,8 ANS) was used to design an alternative competition binding assay. ANS is a amphiphilic dye that strongly fluoresces when exposure to water is lowered or in other words when in a hydrophobic environment. The assay was designed to monitor the decrease in fluorescence upon its displacement by ligands. However the hydrophobic environment contributed by membrane components resulted in no change in fluorescence even at high concentrations of ligand.

3.4.6 Optimisation of fermentation conditions for increasing protein yield for structural studies

To carry out structural studies using X-ray crystallography, we needed higher amounts of protein to set up crystal trays and to optimise conditions for crystal growth. We have initially tried to achieve this by bulk culturing (up to 12 L in shake flasks), however the protein induction across several flasks was not consistent along with batch to batch variation. Thus we sought to use fermentation for homogenous induction of protein in large volumes. To achieve this, we first optimised growth conditions in a shake flask experiments and later tried to extrapolate into fermentor system. Cell density, protein expression and plasmid retention were monitored through the optimisation process to estimate the protein yield.

Earlier described conditions were used as a starting point for optimisation for fermentation. As it was previously described, protein expression in *E. coli* C41 (DE3) cells was better when induced with 1mM IPTG. However, plasmid retention was quite low at ~ 15%. Thus even though cell density could reach as high at ~ 4.0 at OD600nm, protein yield was very low. Thus, to increase the plasmid retention a range of IPTG concentrations were tried. C41 (DE3) cells transformed with MmpL3-C_{HIS} expressing plasmid were grown at 37°C to an OD_{600} of 0.5 and induced with 10 µM, 50 µM, 250 µM and 1 mM IPTG. Cells

were harvested at the end of 1 hr, 2 hrs, 4 hrs and overnight induction and analysed for protein expression (Figure 3-28).

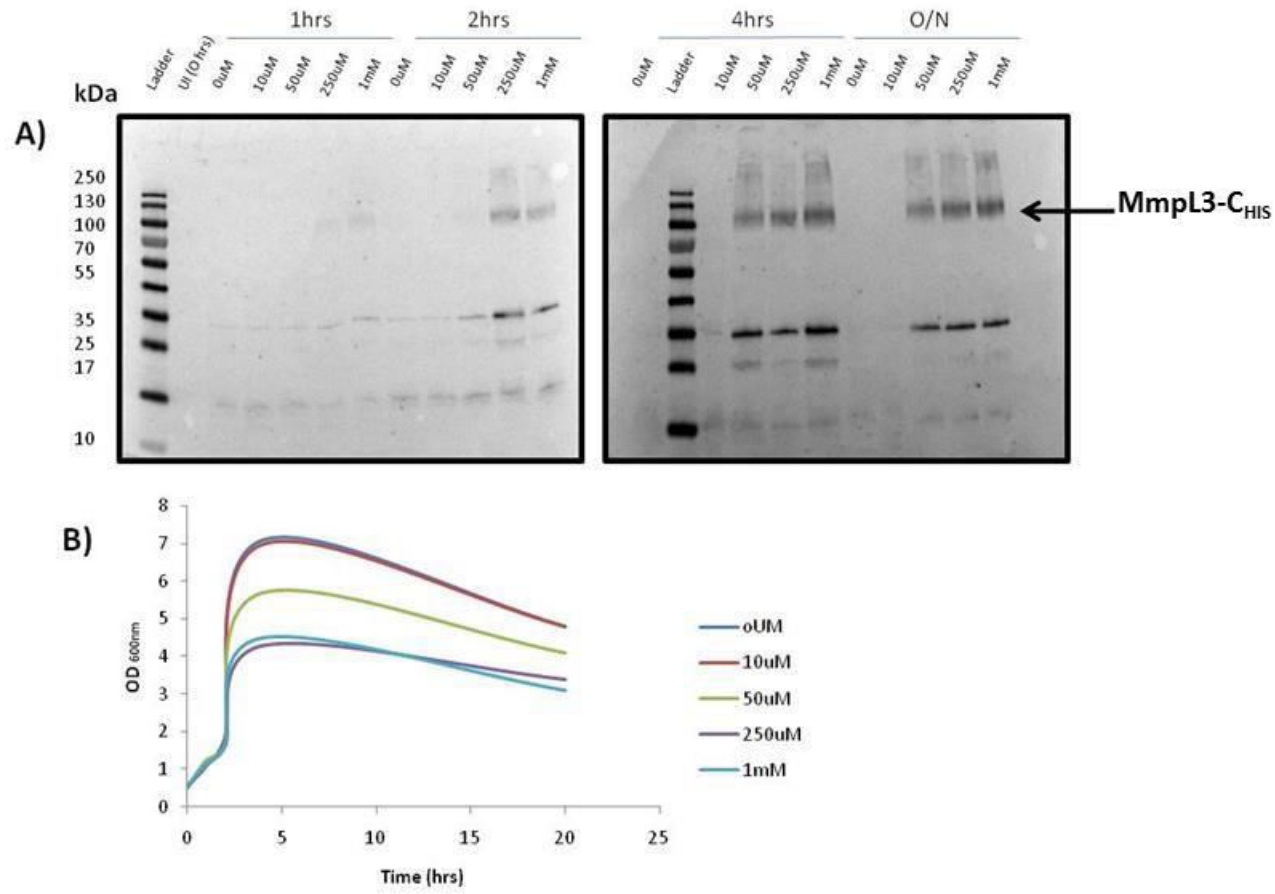


Figure 3-28 Western blot analysis of MmpL3-C$_{HIS}$ over-expression and growth kinetics of MmpL3-C$_{HIS}$ expressing cells. Western blot analysis of MmpL3 expression in *E. coli* C41 (DE3) cells monitored until 20 hrs after induction (A) Growth kinetics of MmpL3 expressing cells compared with WT C41 (DE3) cells (B). Best expression was observed after overnight induction however the plasmid retention was very low coupled with lower overall growth. Protein of interest is indicated with an arrow.

These expression conditions were not optimal for higher protein yield due to very low plasmid retention and low expression levels. C41 (DE3) cells have very tight regulation of T7 promoter which could explain low expression levels. Thus we used the BL21 (DE3) strain, where the T7 promoter is less tightly regulated, which allows better expression of protein. BL21 (DE3) cells transformed with MmpL3-C$_{HIS}$ expressing plasmid were grown at 25°C (to slow the growth and expression of protein) to an OD_{600} of 0.4 and induced with 50 µM, 100 µM, 250 µM and 1 mM IPTG. Cells were harvested at the end of 10 hrs, 12 hrs, 14 hrs, 16

hrs, 18 hrs and 20 hrs of induction and analysed for protein expression and grown kinetics (Figure 3-29).

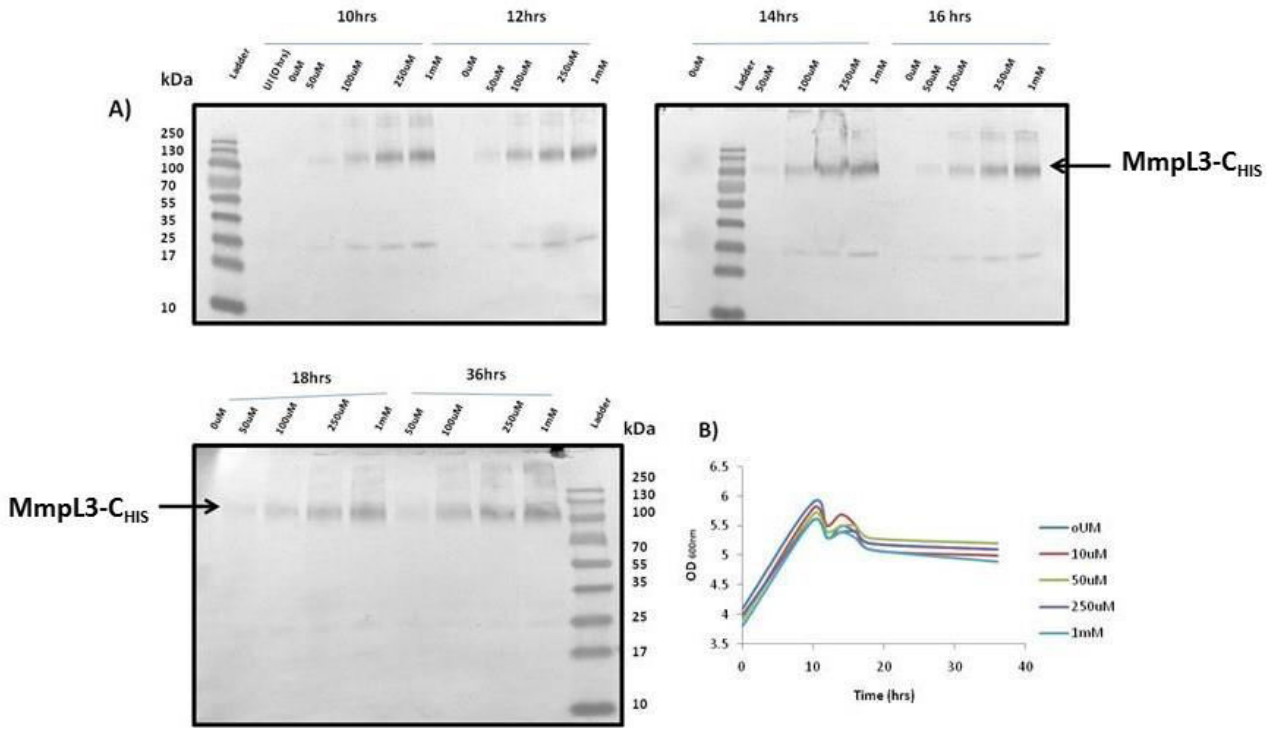


Figure 3-29 Western blot analysis of MmpL3-C$_{HIS}$ over-expression and growth kinetics of MmpL3-C$_{HIS}$ expressing cells. Western blot analysis of MmpL3 expression in *E. coli* BL21 (DE3) cells monitored until 20 hrs after induction (A) Growth kinetics of MmpL3 expressing cells compared with WT BL21 (DE3) cells (B). Best expression was observed after 14-16 hrs induction however the plasmid retention was still low but the overall growth was similar to WT BL21 (DE3) cells. Protein of interest is indicated with an arrow.

MmpL3 was expressed well at 250 µM and at 1 mM it was only marginally better. Cell density was optimum as well however plasmid retention was poor suggesting protein yield would be quite low. Thus we had to further optimise the conditions to improve protein yield (Figure 3-30). We reduced the [IPTG] concentration for induction and started harvesting the cells 2 hrs after induction as opposed to 10hrs after induction. The growth and induction temperatures were not changed.

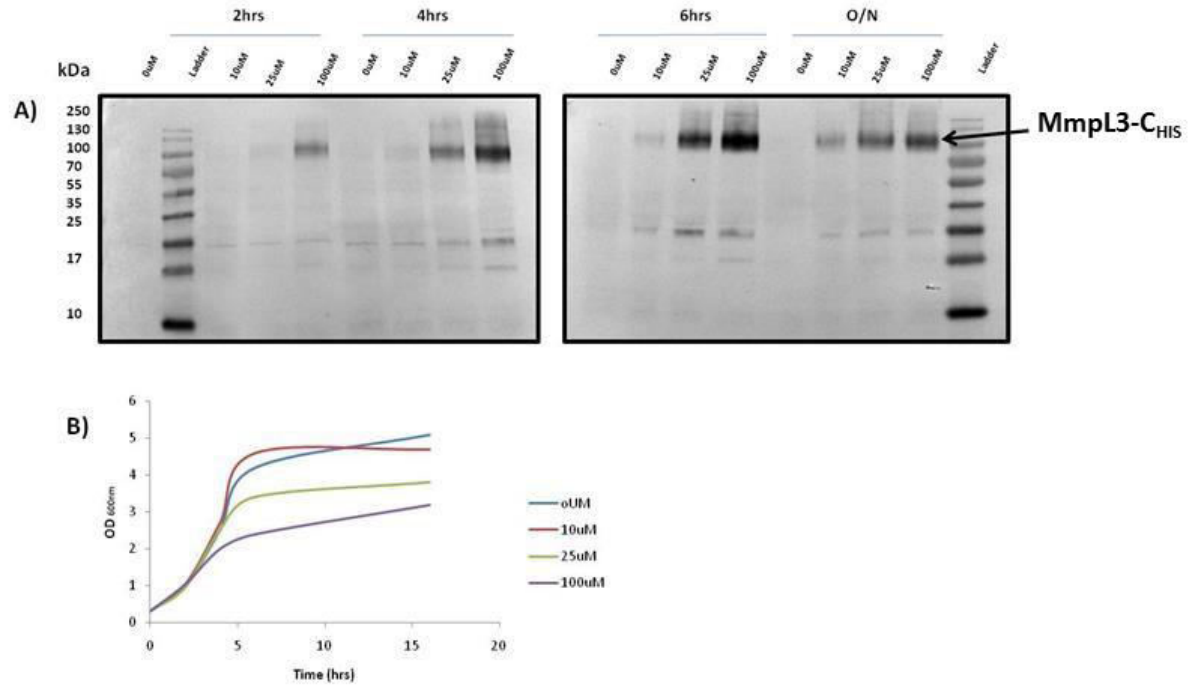


Figure 3-30 Western blot analysis of MmpL3-C$_{HIS}$ over-expression and growth kinetics of MmpL3-C$_{HIS}$ expressing cells. Western blot analysis of MmpL3 expression in *E. coli* BL21 (DE3) cells monitored until 20 hrs after induction (A) Growth kinetics of MmpL3 expressing cells compared with WT BL21 (DE3) cells (B). Best expression was observed after 6 hrs of induction with 100 µM IPTG, also plasmid retention was high. Protein of interest is indicated with an arrow.

Low [IPTG] concentrations improved plasmid retention drastically up to 85-90% and a moderate cell density of ~ 3.5 OD600 (cell mass ~ 0.75 dry weight/L) which resulted in better expression of protein. Since, these conditions resulted in optimum cell density, plasmid retention and protein expression, we have used these conditions for scaling up in a fermentor.

Initial results from fermentation were encouraging in terms of protein expression. However, plasmid retention was very low (25%) even when induced with 25 µM [IPTG]. Further optimisation for plasmid retention was required. We compromised on protein expression level and reduced the [IPTG] concentration to 8 µM, which resulted in better plasmid retention with a slight decrease in expression levels but compensated by better cell density. MmpL3 was purified (Figure 3-32) using SMA solubilisation method described

earlier, from cells harvested from fermentation. Protein yield has increased 4 folds from 0.5 mg/L to 2 mg/L.

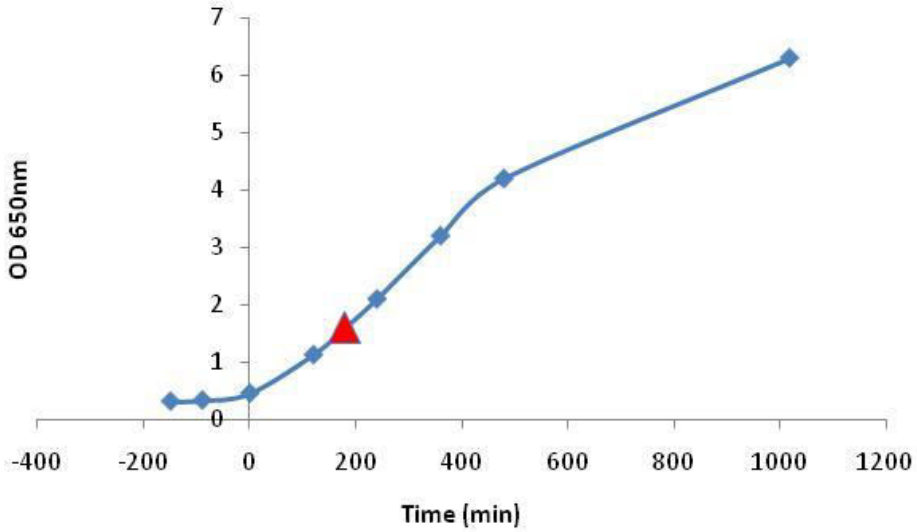


Figure 3-31 Growth kinetics of *E. coli* BL21(DE3) expressing MmpL3 in a fed batch fermentor. Time point 0 represents induction with 8 µM IPTG and red triangle indicating the initiation of feeding.

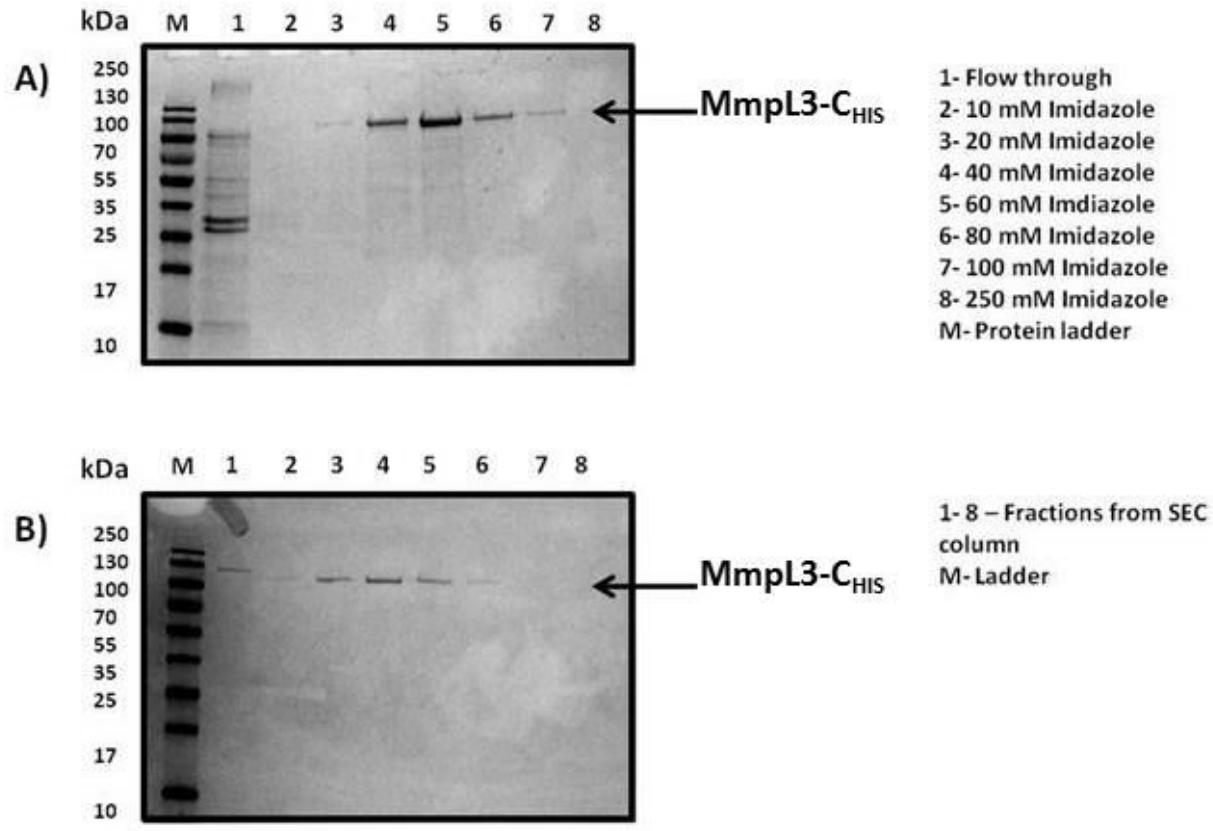


Figure 3-32 SDS-PAGE analysis of purification of MmpL3-C$_{HIS}$ from fed-batch fermentation. MmpL3 was solubilised using SMA and purified using first, affinity chromatography (A) and later using SEC (B). Protein of interest is indicated with an arrow.

3.4.7 Site directed mutagenesis

Our model predictions established the similarities between MmpL3 and other RND transporter proteins. Recent reports suggest that MmpL3, like other RND transporter proteins depend on proton motive force (Yang *et al*. 2014). Amino acid residues that participate in proton relay have been identified in other RND transporters AcrB and SecDF (Tsukazaki *et al*. 2013; Seeger *et al*. 2009). MmpL3 model predictions have helped us identify amino acid residues that are likely to participate in a proton relay or affect proton relay (Figure 3-6).

Aspartatic acid (D408) on TM4 facing lysine (K940) on TM10 forms a proton relay pump in AcrB and similarly aspartic acid (D340) on TM4 facing arginine (R671) on TM11

forms a proton relay pump in SecDF (*Thermus thermophilus*). We have looked at similar regions on MmpL3 to find D-R or D-K pairs. We found D251 on TM4 but positively charged amino acid was not in close proximity to form a proton relay pump. Interestingly, further down in the same transmembrane helix (TM4) we found a glutamate (E262) residue facing an arginine (R653) from the neighbouring transmembrane helix (TM10). These residues were mutated to study structure function relationship. The mutations carried out involved I250T, D251A, Y252A, Y252D, E262A and R623A.

To test the functionality, *mmpL3* was cloned into the mycobacterial vector pMV261 (*apra*) and mutations were carried out using Q5 site directed mutagenesis kit. The mutant constructs were introduced into a*M. smegmatis* conditional mutant Δ*mmpL3* generated previously (Varela et al. 2012). The ability of the mutant constructs to complement MmpL3 function in the absence of acetamide was monitored.

We were successful in generating all six mutants (I250T, Y252A, Y252D, D251A, E262A and R653A), but could only test three mutants (I250T, Y252A and Y252D) in a survival experiment. I250T andY252A mutations did not significantly affect the function of MmpL3, but the Y252D mutation slightly affected the growth rate, although did not abolish the growth. The result suggests that the tested amino acid changes might not play a vital role in generating a proton motive force (Figure 3-33).

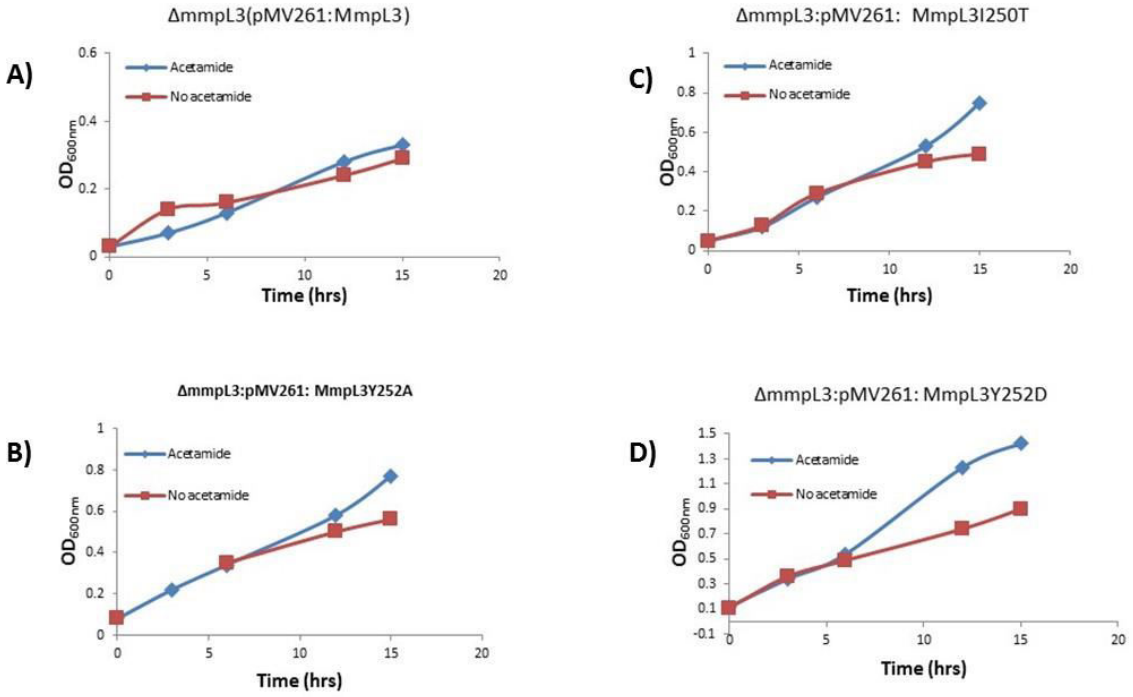


Figure 3-33 Functionality analysis of MmpL3 site directed mutants. Functionality analysis of MmpL3 in Δ*mmpL3* strain through complementation studies using WT MmpL3 (A), MmpL3-I250A (B), MmpL3-Y252A (C) and MmpL3-Y252D (D).

3.5 Discussion

M. tuberculosis consists of family of proteins known as MmpLs that are involved in the transport of complex lipid molecules across the membrane (Varela *et al*. 2012; Reed *et al*. 2004; Cole *et al*. 1998; Converse *et al*. 2003; Sondén *et al*. 2005). These MmpL proteins belong to the RND superfamily of proteins, characterised by twelve transmembrane helices and 2 non-transmembrane domains (Murakami *et al*. 2002; Tsukazaki *et al*. 2013). Twelve such MmpLs are found in *M. tuberculosis* of which only *mmpL3*is shown to be essential for viability (Domenech *et al*. 2005; Varela *et al*. 2012), thus is a target of several anti-TB drugs (Tahlan *et al*. 2012; La Rosa *et al*. 2012; Grzegorzewicz *et al*. 2012; Remuiñán *et al*. 2013). In recent years, MmpL3 has emerged as an important drug target for these reasons and structural studies would help advance MmpL3 as a drug target. No structural information

isavailable for the MmpL family of proteins either from mycobacteria or from related species. In this chapter, some of the structural and mechanistic features have been revealed.

Recently, non-TM domains of these MmpL proteins have been shown to interact with late biosynthetic enzymes to form scaffolds assisting in transport of target lipid molecules suggesting their functional role (Jain & Cox 2005). Previously, non-TM regions from other RND proteins were shown to form independent, properly folded globular domains (Tsukazaki *et al*. 2013). Our model prediction studies revealed that these non-TM domains have a conserved motif across the superfamily, each domain consisting of four β-strands flanked by two α-helices (Figure 3-3 D). The two domains are homologous and come together to form a pseudo-symmetrical 8-stranded anti-parallel β-sheet structure (Tsukazaki *et al*. 2013). This lead us to design the ML constructs (fusion of two non-TM regions with an artificial linker peptide mimicking the native special arrangement) to study structural and biophysical properties. Although we could not obtain the crystal structure, the biophysical properties shed some light on its structural features. Protein cross-linking andAUC suggests a tendency to form dimers and higher oligomers. Initial NMR studies showed that the non-TM regions are mainly composed of β-strands supporting our model predictions. Limited ligand binding studies showed that ML could independently bind the TMM substrate analogue specifically and is displaced by known inhibitors of MmpL3, adding more confidence on specificity. However, more elaborative ligand binding studies could not be carried out since our focus was to solve the crystal structure of the full length protein.

We have expressed MmpL3 in sufficient quantities and purified it to homogeneity. We have used a novel method that employs a SMA co-polymer to encapsulate the transmembrane protein in its native form (Jamshad *et al*. 2011). To our knowledge this is the first successful application of SMA solublisation process to purify a mycobacterial protein.

Biophysical studies of MmpL3 revealed several key aspects of protein assembly in the membrane. CDstudies confirmed the proper folding of the protein inside the SMALP disc. CD studies also revealed that MmpL3 is mainly α-helical, characteristic of transmembrane proteins. Learning about the oligomeric states is key to understand how MmpL3 functions. Previously other RND proteins namely AcrB, MexA, SecDF and CusA were shown to function as either dimer or trimers (Tsukazaki *et al*. 2013; Murakami *et al*. 2002; Long *et al*. 2010; Sennhauser *et al*. 2009). Size exclusion chromatography and AUC data has shed some light on oligomeric states of the MmpL3. Size exclusion chromatography data suggests that MmpL3 exists in more than one oligomeric state and AUC data has revealed that MmpL3 mainly exist a dimer, with a small population of tetramer. We did not observe oligomeric mixtures, composed of at least two oligomeric states, of MmpL3 and one of the reasons could be that protein oligomers are encapsulated in the polymer and could not exist as mixtures. These results align well with the earlier results obtained with ML, which showed a tendency to form higher oligomers.

Further, in collaborations with Prof. Tim Dafforn, SAXS and cryo-EM studies were carried out. Although, good quality data was obtained from SAXS, we were not able to process the data as the model building softwares could not account for the membrane components encapsulated within the SMALPS. Since, the SMA solublised protein purification process still in its infancy, software to process these data are not yet available and software development is in its final stages. Our collaborators are processing the cryo-EM data and the early indications are that the results compare well with AUC data and suggest that MmpL3 exist as dimer.

Ligand binding studies with full length MmpL3 did not yield much information on substrate specificity or binding affinity. This was due to difficulties in interpreting the

binding data, as membrane components within SMALP encapsulation was interfering with ITF. Alternative assays were not very helpful either. Exploring binding assays with radio labelled ligand is a possibility. Another possibility is to use gel mobility shift assays. However, due to lack of time that could not be explored. Establishing these assays will have high impact as it can be used for *in* vitro screening of small molecule inhibitors against MmpL3.

The protein yield of MmpL3 was successfully increased 4- fold (from 0.4 mg/L to 2.0 mg/L) to using small scale fermentation. There is still scope to increase the cell density and there by protein yield. The initial results were promising and with further optimisation protein could be purified in sufficient quantity to set up protein crystallisation trays.

In the future, resolving the MmpL3 structure will give insights into its function and also understand how other MmpLs work. Mycobacteria has 12 MmpLs involved in the transport of different class of lipids (Varela *et al*. 2012; Jain & Cox 2005; Reed *et al*. 2004; Sondén *et al*. 2005). Structural studies will shed light on the specificity of MmpL3 to a particular class of lipid and understanding of the mechanism of action will be made easier. Most of the RND transporters are driven by a proton motive force (Seeger *et al*. 2009; Tsukazaki *et al*. 2013) and in this chapter we have tried to test whether MmpL3, like its counterparts depend on a proton relay for its function. Recent studies with Corynebacterial CmpL3 using efflux pump inhibitors suggested this was the case, however important amino acid residues involved were not reported (Yang *et al*. 2014). Our structural studies helped us predict which amino acids that could be involved in proton relay and were subsequently tested in a functional complementation assay. We could not make a definite conclusion because of two major hurdles. Firstly, complementation in mycobacterial strains was difficult as they could not tolerate a defect in mycolic acid transport, and secondly, all the planned

mutations could not be finished due to time constrains. We have designed an alternative complementation assay system in the surrogate organism *C. glutamicum* and the validation of the assays is currently in process.

Further, structural studies will open the door for structure based drug discovery. This is important as *mmpL3* is an essential gene for mycobacterial survival and is a high value drug target. In this chapter, we have made significant steps towards achieving the goals. We have optimised conditions for production of sufficient quantities of the protein and established its biophysical characteristics. Cryo-EM and SAXS experiments have been performed and these results are key to further establish our model which could serve as template for structure based drug discovery until a crystal structure is made available.

Chapter 4

Characterisation of function of genes present in the extended *mmpL3* cluster

4.1 Introduction

The biophysical characterisation of the mycolic acid transporter protein MmpL3 was discussed in the previous chapter. This chapter describes work aimed at identifying missing links in the late processing of mycolic acids and proteins associated with MmpL3 that may facilitate mycolic acid transport. Towards this, the aim was to investigate the role of genes present in the extended *mmpL3* cluster (Figure 4-1), which is conserved across the mycobacterial species, including *M. leprae* which has an 'essential' minimal genome. Of the eight genes in the cluster in *M. tuberculosis*, seven of them are conserved in *M. leprae*, and one of them (*Rv0205*) is annotated as a pseudogene. The function of two of the genes in the cluster, *mmpL3* and *mmpL11* are implicated with the transport of mycolic acids and related molecules (Varela *et al*. 2012; Pacheco *et al*. 2013). The function of other genes are however still unknown.

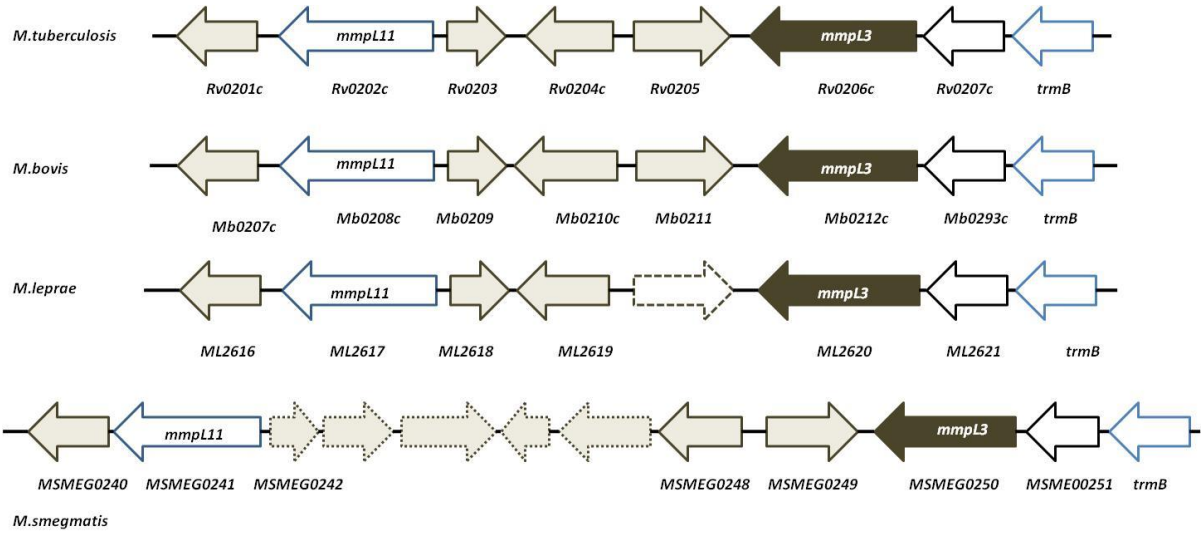


Figure 4-1 Maps of the *mmpL3-mmpL11* region in different mycobacteria. Comparison of the *mmpL3* gene cluster shows *mmpL3* in solid arrow, conserved genes in similar arrows. Pseudogene indicated with a hatched border arrow and genes exclusively found in *M. smegmatis* represented in dotted arrows. Genes of particular interest for this study arehomologues of *Rv0201c* (*MSMEG0240c*), *Rv0204c* (MSMEG0248c) and *Rv0205* (*MSMEG0249*). Although, *Rv0205* lacks *M. leprae* orthologue, it is listed as essential gene in HIMAR-1 based Transposon mutagenesis studies (Sassetti *et al*. 2003).

Analysis of native molecular composition of protein complexes, in *M. bovis* BCG membrane fraction revealed that MmpL11, Rv0201c and Rv0204 (*Rv* homologues) forms protein complexes (Zheng *et al*. 2011). Previous studies with other MmpL proteins also suggest formation of a protein scaffold to achieve the transport of targeted substrates. Jain and Cox (2005) showed that MmpL7, involved in the transport of phthiocerol dimycocerosate (PDIM), interacts with PpsE, a late biosynthetic enzyme involved in the biosynthesis of phenolphthiocerol (Cox *et al*. 1999; Jain & Cox 2005). Also, MmpL8 responsible for the transport of sulfolipid has been shown to form a complex with Pks2, PapA1,FadD23 and a putative transposase (Rv3890c), suggesting possible formation of protein scaffold to achieve transport (Zheng *et al*. 2011; Bhatt *et al*. 2007). This suggests a common mechanism adopted for the transport of targeted molecules by MmpL proteins and we sought to investigate this possibility of other proteins encoded by the genes in the *mmpL3* extended cluster participating in forming a protein complex to achieve the transport of mycolic acids and related products.

Additionally, MmpL3 and MmpL11belong to RND superfamily of proteins (Domenech *et al*. 2005). The RND proteins of other bacterial sources are invariably associated with periplasmic membrane fusion proteins (MFP), which interact cooperatively to enable the transport. In some cases this complex is shown to interact with a third family of channel forming proteins called outer membrane factor (OMF) family of proteins (Paulsen *et al*. 1996). Recently, periplasmic adaptor proteins, MmpS4 and MmpS5 have been identified in mycobacteria and structural studies revealed that MmpS4/S5 function as membrane adaptor proteins and facilitates MmpL4/L5 in the transport of siderophores (Wells *et al*. 2013). The *M. tuberculosis* genome consists of five genes that encode MmpS proteins that co-localise with respective *mmpL* genes, with an exception to *mmpL3*. Additionally, outer membrane factor proteins have not yet been identified in mycobacteria which has led us to

investigate whether the genes that co-localise *mmpL3* and *mmpL11* and annotated as hypothetical membrane proteins fulfil any of these roles.

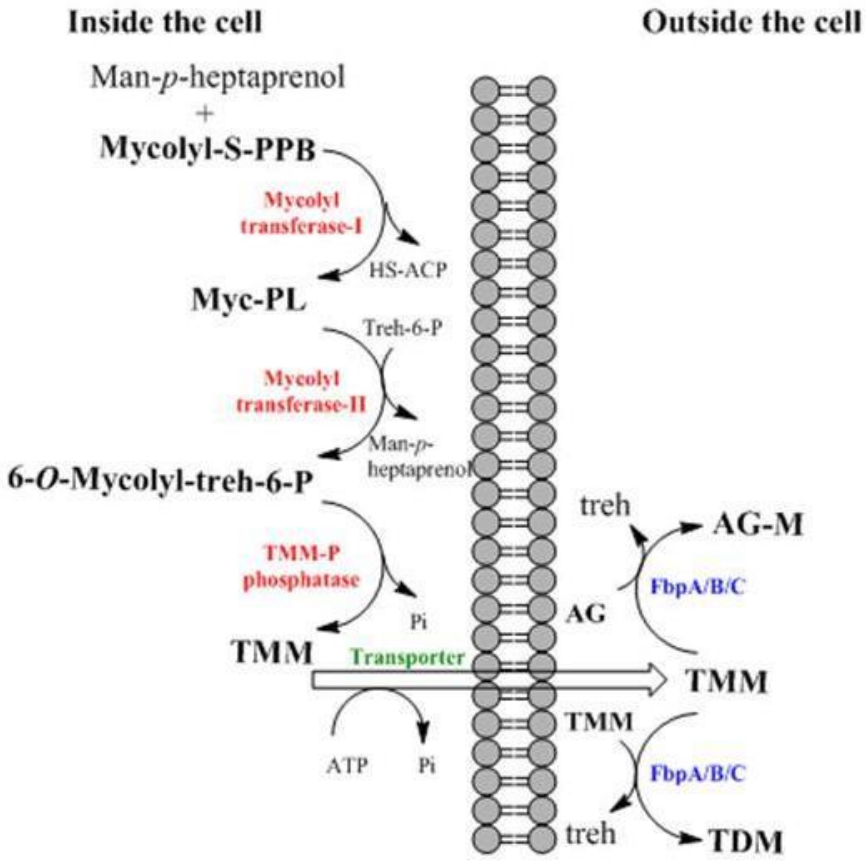


Figure 4-2 Schematic presentation of early hypothesis for transport of nascent mycolic acid for cell wall mycolylation. Nascent mycolic acid transferred from processing enzymes on to prenol carrier which is eventually transferred to trehalose to form TMM. TMM is then transferred across the membrane, where it acts as a substrate for mycolylation of cell wall AG. Man-p-heptaprenol- mannosylphosphorylheptaprenol; treh-trehalose; AG-arabinogalactan; AG-M-arabinogalactan-mycolate; TMM-trehalose monomycolate and TDM-trehalose dimycolate. Mycolyl transferases and the phosphotase involved in the late processing of mycolic acid to convert them into TMM, are yet to be identified, and indicated in red. The transporter is now identified as MmpL3 and is indicated in green. Mycolyl transferases FbpA/B/C involved in the cell wall mycolylation and the synthesis of TDM, are indicated in blue.

Another class of enzymes hypothesised to be involved in the processing of mycolic acids prior to its transport, are yet unidentified are the mycolyltransferases (Figure 4-2). The mycolyltransferases are characterised by a thioesterase activity that cleaves the thioester

linkage between nascent mycolic acid and cystine residue on PKS13. Recently, *Rv3802*, encoding a thioesterase was identified from the mycolic acid biosynthetic cluster, which is conserved across the mycobacterial species including *M. leprae,* and was shown to be essential for mycobacterial growth and survival (Parker *et al*. 2009). Further, structural studies revealed a conserved α/β-hydrolase polypeptide fold and a catalytic triad formed by Serine-Glutamic acid-Histidine, that are characteristic of FbpA/B/C proteins, previously reported as mycolyltransferases (Meniche *et al*. 2009; Parker *et al*. 2009; Kremer *et al*. 2002; Ronning *et al*. 2000; Belisle *et al*. 1997). We thus tried to evaluate the involvement of Rv3802 in mycolic acid processing.

We employed genetic tools to characterise the function of the genes in the extended *mmpL3* cluster using *M. tuberculosis* surrogate organism *M. smegmatis*. Further, we used protein-protein interaction studies to evaluate association of Rv0204, Rv0205 and Rv3802 with MmpL3 and MmpL11 to assess if they play a role in facilitating the transport of mycolic acids or its related molecules.

4.2 Materials and Methods

4.2.1 *In silico* analysis

Prediction of interacting partners of MmpL3 were determined by STRING, a database of predicted and actual protein-protein interaction networks for a set of over 1100 organisms (Franceschini *et al*. 2013). The selection criteria for testing possible interaction included co-localisation with *mmpL3* (*Rv0201* (*MSMEG_0240*), *Rv0204c* (*MSMEG_0248*), *Rv0205* (*MSMEG_0249*)), literature (*Rv0205, Rv3802*), other proteins implicated with similar function (MmpL11). To test the possible interactions and to identify the role of individual genes, the study aims to generate null mutants of *MSMEG_0240, MSMEG_0248* and

MSMEG_0249 and also carry out protein-protein interaction studies using bacterial two hybrid interaction tools. The gene sequences for *M. smegmatis* genes *MSMEG_0240, MSMEG_0248* and *MSMEG_0249* were obtained from NCBI genome database (http://www.ncbi.nlm.nih.gov/genome). The gene sequences for *M. tuberculosis* genes *mmpL3, mmpL11, Rv3802, Rv0204* and *Rv0205* were obtained from the Tuberculist database (http://genolist.pasteur.fr/TubercuList/genome.cgi).

4.2.2 Plasmids, DNA manipulations and bacterial growth conditions

Table 4-1lists the plasmids, bacterial strains and recombinant mycobacteriophages used in this study. *E. coli* strains were routinely cultured in LB broth at 37°C. *M. smegmatis* strain mc^2155 was cultured at 37°C in TSB supplemented with 0.05% Tween 80. The generation and propagation of mycobacteriophages was done on lawn of *M. smegmatis* at 30°C in 7H10 basal plates overlaid with 7H9 soft agar as described in Chapter 7. High titres of mycobacteriophages were generated using protocols described in Chapter 7 (Larsen *et al*. 2007). Antibiotics were added at required concentrations where required. Hygromycin was routinely used at 150 µg/ml for *E. coli* and at 100 µg/ml for *M. smegmatis* and kanamycin was used at 50 µg/ml for *E. coli* and at 25 µg/ml for *M. smegmatis*.

Table 4-1 Plasmids and strains used in this study.

Plasmids, phages and strains	Description	Reference
Plasmids		
p0004S	Cosmid vector containing Hyg-sacB cassette	Larsen *et al*., 2007
p∆*MSMEG_0240*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MSMEG_0240*	This work
p∆*MSMEG_0248*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MSMEG_0248*	This work
p∆*MSMEG_0249*	Derivative of p0004s_Apramycin cosmid obtained by cloning upstream and downstream flanks of *MSMEG_0249*	This work
pTIC6a:*MSMEG_0248*	Functional copy of *MSMEG_0248* cloned into pTIC6a under the regulation of inducible pTet promoter	This work
Phages		
phAE159	Thermo sensitive, conditionally replicating shuttle phasmid derived fromlytic mycobactiophage TM4	Larsen *et al*., 2007
ph∆*MSMEG_0240*	Derivative of phAE159 obtained by cloning p∆*MSMEG_0240* into its PacI site	This work
ph∆*MSMEG_0248*	Derivative of phAE159 obtained by cloning p∆*MSMEG_0248* into its PacI site	This work
ph∆*MSMEG_0249*	Derivative of phAE159 obtained by cloning p∆*MSMEG_0249* into its PacI site	This work
Bacterial strains		
E. coli Top 10	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen™
E. coli HB101	*E. coli* K-12 _F- mcrB mrr hsdS20(rB- mB-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(SmR) glnV44 λ-	Invitrogen™
M. smegmatis mc^2155	Wild type strain, E*pt*mutant of *M. smegmatis* strain mc^26	Snapper *et al.* 1990
Δ*MSMEG_0240*	*M. smegmatis* strain with *MSMEG_0240* replaced with Hyg-sacB cassette	This work
Δ*MSMEG_0249*	*M. smegmatis* strain with *MSMEG_0249* replaced with Hyg-sacB cassette	This work

4.2.3 Construction of knockout phages

Approximately 1kb regions upstream and downstream of gene were PCR amplified using genomic DNA from *M. smegmatis* mc^2155. The PCR products were purified, digested with respective enzymes and cloned on either side of a hygromycin (*hyg*) cassette in the plasmid p0004S (Stover *et al*. 1991) to generate an allelic exchange substrate plasmids listed in Table

4-1. The allelic exchange plasmids were sequenced to confirm the presence of upstream and downstream regions of the respective genes. The allelic exchange plasmids were then cloned into temperature sensitive phasmid phAE159 at a *PacI* site and packaged into empty λ-phage heads. The packaged phasmids were transduced into *E. coli* HB101cells and selected on hygromycin plates. Positive clones were identified by *PacI* digestion.

One such positive phasmid for each gene was transformed by electroporation into *M. smegmatis* mc^2155 and recovered at 30°C for 3-4 hrs in TSB supplemented with 0.05% Tween 80 (1 ml). The recovered cells were split into 200 µl and 800 µl and mixed with 100 µl of log phase *M. smegmatis* and 5 ml of molten 7H9 soft agar. This mix was overlaid on 7H10 basal plates and incubated at 30°C for 2-3 days for plaques to be formed. The plates were soaked in mycobacteriophage (MP) buffer for 4-6 hrs and syringe filtered and either used straight away or stored at 4°C.

4.2.4 Generation of null mutants

Specialised transduction of *M. smegmatis* mc^2155 was performed as described before (Bardarov *et al*. 2002). *M. smegmatis* mc^2155 cultures were grown in TSB supplemented with 0.05% Tween 80 to an OD_{600nm} of 0.8 and harvested. The cell pellet was washed twice with equal volume of MP buffer. Finally, the cells were suspended in small volume of MP buffer (2 ml). High titre (10^8-10^{10}) phage lysate (1 ml) was mixed with 1 ml of cells for transduction and 1 ml of MP buffer mixed with 1 ml of cells for control reaction and were incubated at 37°C for 3-4 hrs, harvested and recovered in TSB supplemented with 0.05% Tween 80 for at 37°C 4 hrs. The recovered cells were harvested and resuspended in 400-600 µl of TSB and plated on TSA containing 100 µg/ml hygromycin for selection. Hygromycin resistant colonies were inoculated into 10 ml of TSB supplemented with 0.05% Tween 80 and 100

µg/ml of hygromycin for genomic DNA extraction and further Southern blot. Allelic exchange of each gene with *hyg* cassette in transductants was confirmed by Southern blot.

4.2.5 Lipid analysis of *M. smegmatis* strains

M. smegmatis mc^2155, Δ*MSMEG0240* and Δ*MSMEG0249* strains were grown in TSB supplemented with 0.05% Tween to an OD of 0.8 and labelled with [^{14}C] acetate (50 µCi/ml). The cells were harvested, washed and dried, and cell wall associated lipids were extracted using methods described by Dobson *et al*. (1985). Briefly, surface exposed, non-covalently bound lipids were first extracted using petroleum ether (60-80 bp). Cells were resuspended in 2 ml of petroleum ether (60-80 bp) and mixed for 30 minutes. Upper layer containing lipids were removed into fresh tube and dried. This step was repeated twice for efficient extraction. Apolar and polar lipids were then extracted from the remaining cell pellet. Cell pellet was resuspended in methanolic saline (methanol: 0.3% NaCl, 10:1 v/v) and apolar lipids were extracted with 2 ml of petroleum ether (60-80b.p) twice, dried and re-constituted in chloroform: methanol (2:1). Polar lipids containing glycolipids were extracted first with chloroform: methanol: water (9:10:3) and twice with chloroform: methanol: water (5:10:4). The two extractions were pooled and mixed with 1.3 ml of chloroform and 1.3 ml of 0.3% NaCl, mixed and centrifuged. The lower organic layer containing polar lipids is collected, dried and re-constituted in chloroform: methanol (2:1). Apolar lipids were analysed by 2D TLC using SystemC and SystemD for altered mycolic acid biosynthesis/transport and visualised by autoradiographs (Dobson *et al*. 1985).

4.2.6 Fatty acid methyl esters (FAMEs) and mycolic acid methyl esters (MAMEs) extraction from defatted cells and whole cells

Alkaline hydrolysis was performed on de-fatted cells after the extraction of polar lipids as described in Chapter 7. Briefly, 2 ml of 5% aqueous tetrabutylammonium hydroxide (TBAH)

was added to defatted cells and incubated at 100°C overnight. Base hydrolysed fatty acids and mycolic acids were extracted as methyl esters using 4 ml of dichloromethase (CH_2Cl_2), 300 µl of methyl iodide (CH_3I), and 2 ml of water with mixing for 30 min. The lower layer of organic phase containing FAMEs and MAMEs was removed and were washed three times with water and dried, and re-dissolved in diethyl ether. The organic layer was taken into separate tube and dried under air. The samples were dissolved in CH_2Cl_2 (200 µl), analysed by single dimension TLC using petroleum ether: acetone (95:5) and visualised by autoradiograph.

4.2.7 Protein-protein interaction studies

To identify proteins associated with MmpL3, the mycolic acid transporter Bacterial two hybrid system BACTH (Euromedex) (protocol detailed in Chapter 7) was employed (Karimova *et al*. 1998). Briefly, candidate genes were cloned into bait (pKT25) and prey (pUT18) vectors with complementary fragments (T25 and T18) of catalytic domain of adenylatecyclase as fusion tags. The interacting partners bring together the complementary fragments (T25 and T18) in physical proximity required to restore the activity of adenylate cyclase in *cya*$^-$ strain of *E. coli*, resulting in a cascade of reactions activating transcription of resident genes including *lac* and *mal* operons. MmpL3 and cytoplasmic domains Loop1 (hereafter referred to ML1) and Loop2 (hereafter referred to as ML2) and MmpL11were cloned into pKT25 vector and genes for putative interaction partners Rv3802, Rv0204, Rv0205, were cloned into pUT18 vector. The pKT25 and pUT18 constructs were co-transformed into *E. coli cya*$^-$ strain BTH101. The transformants were plated on MacConkey agar containing maltose for screening positive interactions. Maltose utilisation leads to lowering of pH turning the agar pink, and is considered positive for protein-protein interaction.

4.3 Results

4.3.1 *In-silico* analysis

STRING, a database of predicted and actual protein-protein interaction networks for a set of over 1100 organisms, was used to analyse possible proteins interacting with MmpL3 (Franceschini *et al*. 2013). STRING uses several criteria to predict/identify interacting partners which include neighbourhood, gene fusion, cooccurance, coexpression, experiments, database, textmining and homology and ranks the interacting proteins, generating a network of protein-protein interactions. To identify possible interactors of MmpL3, we used sa set of criteria using all of aforementioned filters, with a medium confidence (0.400) and generated network using top 10 ranked interactors which include only the direct intractor proteins. The possible interacting proteins of MmpL3 included other MmpL's, probable conserved transmembrane proteins- MT0214 (Rv0204) and Rv0205, conserved hypothetical protein-MT0217 (Rv0207c) and a putative methyl transferase-TrmB (Figure 4-3).

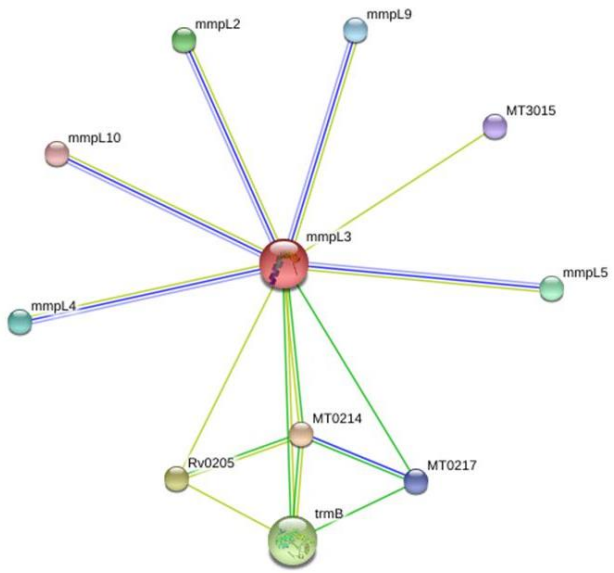


Figure 4-3 Predicted interacting partners of MmpL3 based on database mining by STRING.

The genes *Rv0205* and *trmB* are predicted to be essential (Sassetti *et al*. 2003) and disruption of *Rv0204* was shown to affect virulence (Camacho *et al*. 1999). A detailed STRING analysis which also includes interacting proteins that are indirectly linked showed links between Rv0204 and MmpL11 (Appendix-5). This prediction was supported by blue native 2D SDS-PAGE analysis where MmpL11 was shown to form a complex with Rv0204 and Rv0201(Zheng *et al*. 2011). Further, Varela *et al.*(2012) suggested that Rv0204 and Rv0205 could possibly play a role in mycolic acid transport. Thus, we attempted to probe into the functions of *Rv0201c, Rv0204c* and *Rv0205* genes by generating knockouts of the homologues (*MSMEG0240c, MSMEG0248c* and *MSMEG0249*) in *M. smegmatis.*

4.3.2 Generation of null mutants

Specialised transduction, a highly efficient mycobacterial knock out method, was employed to generate null mutants in *M. smegmatis*. Null mutants *MSMEG0240* and *MSMEG0249* were successfully generated, indicating that they are non-essential for the survival of the organism, contradicting earlier reports, which suggested that homologue of *MSMEG0249* was essential (Sassetti *et al*. 2003). However, repeated attempts to delete *MSMEG_0248* were not successful. *MSMEG0240* and *MSMEG0249* knockouts confirmed by Southern blot (Appendix 10) were used for further characterisation.

4.3.3 Analysis of total lipids from Δ*MSMEG_0240* and Δ*MSMEG_0249*

To assess the loss of *MSMEG_0240* and *MSMEG_0249* function on the cell wall lipids in general and mycolic acid transport in particular, WT and mutant strains were grown and lipids were labelled with [^{14}C] acetate. Petroleum ether extracts containing surface exposed, non-covalently bound lipids and apolar lipids from WT and mutant strains were analysed by 2D-TLC using the solvent systems described by Dobson *et al*. (1985). 2D TLC analysis of using System C (Figure 4-4, Figure 4-6) and System D (Figure 4-5, Figure 4-7) revealed no

significant difference in mycolic acid and related lipids (TMM, TDM, GMM), indicating that *MSMEG0249*and *MSMEG0240*do not affect the biosynthesis and transport of mycolic acids.

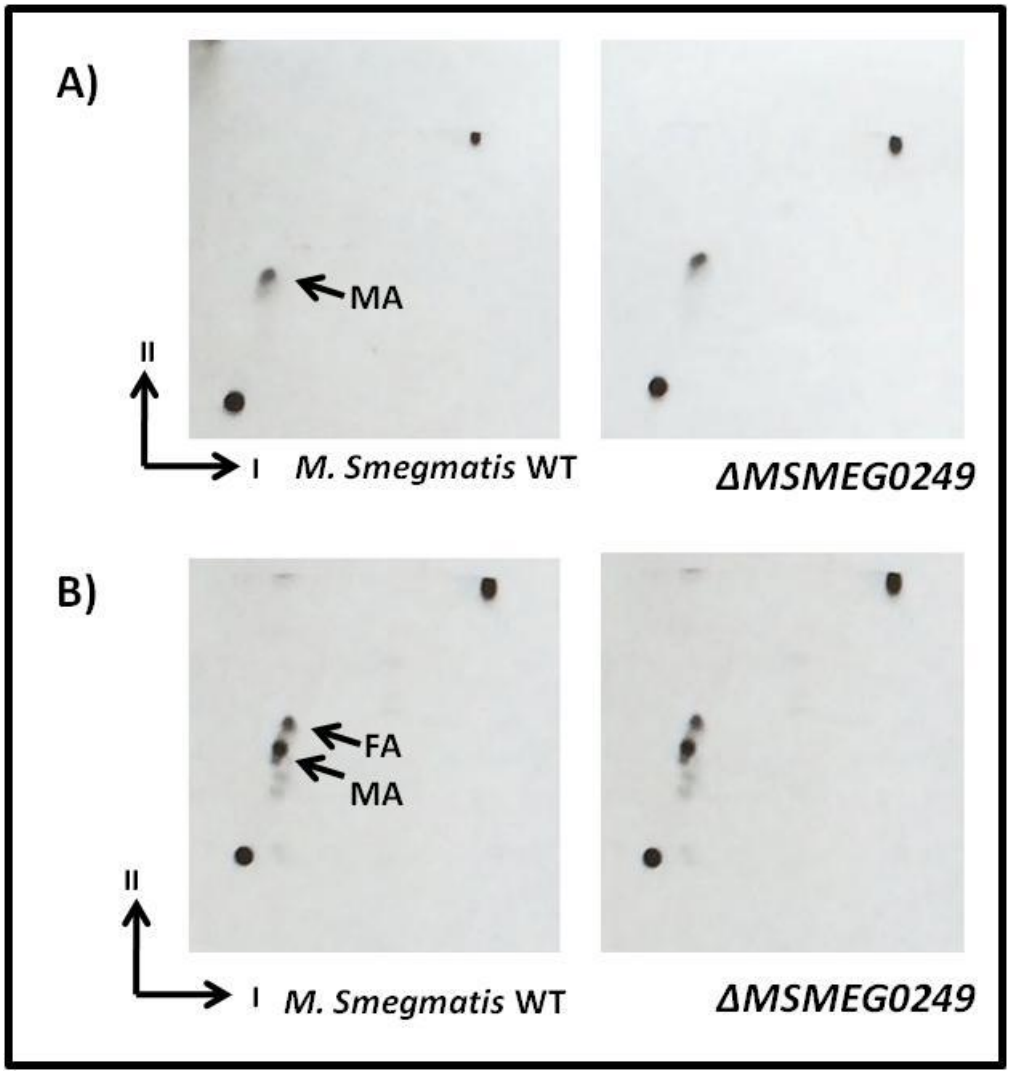


Figure 4-4 Lipid analysis of the Δ*MSMEG0249* mutant.2D TLC analysis of [^{14}C]-labelled lipids from the *M. smegmatis* WT and Δ*MSMEG0249* mutant. Petroleum ether extracts (A) and intracellular apolar lipids (B) were separated using chloroform:Methanol (96:4 v/v) in Direction I and toluene: acetone (80:20 v/v) in Direction II. Positions of mycolic acids (MA) and fatty acids (FA) are indicated with arrows.

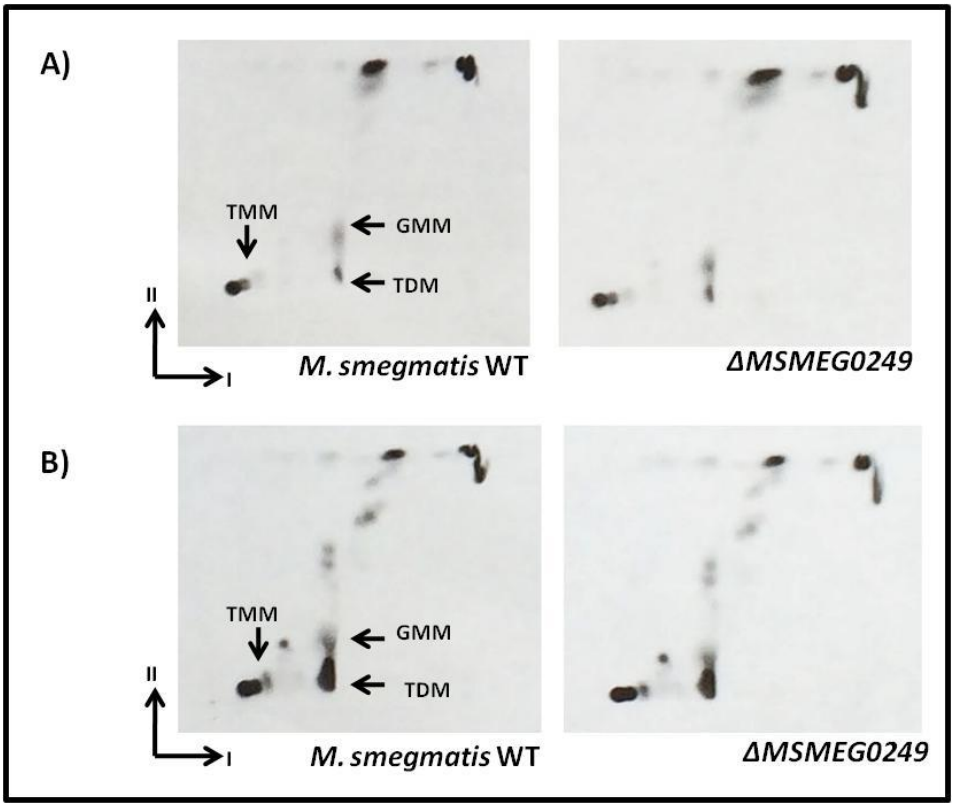


Figure 4-5 Lipid analysis of the Δ*MSMEG0249* mutant. 2D TLC analysis of [^{14}C]-labelled lipids from the *M. smegmatis* WT and Δ*MSMEG0249* mutant. Petroleum ether extracts (A) and intracellular apolar lipids (B) were separated using Chloroform: methanol: water (100:14:0.8 v/v/v) in Direction I and chloroform: acetone: methanol: water (50:60:2.5:3 v/v/v/v) in Direction II. Positions of TMM, TDM and GMM indicated with arrows.

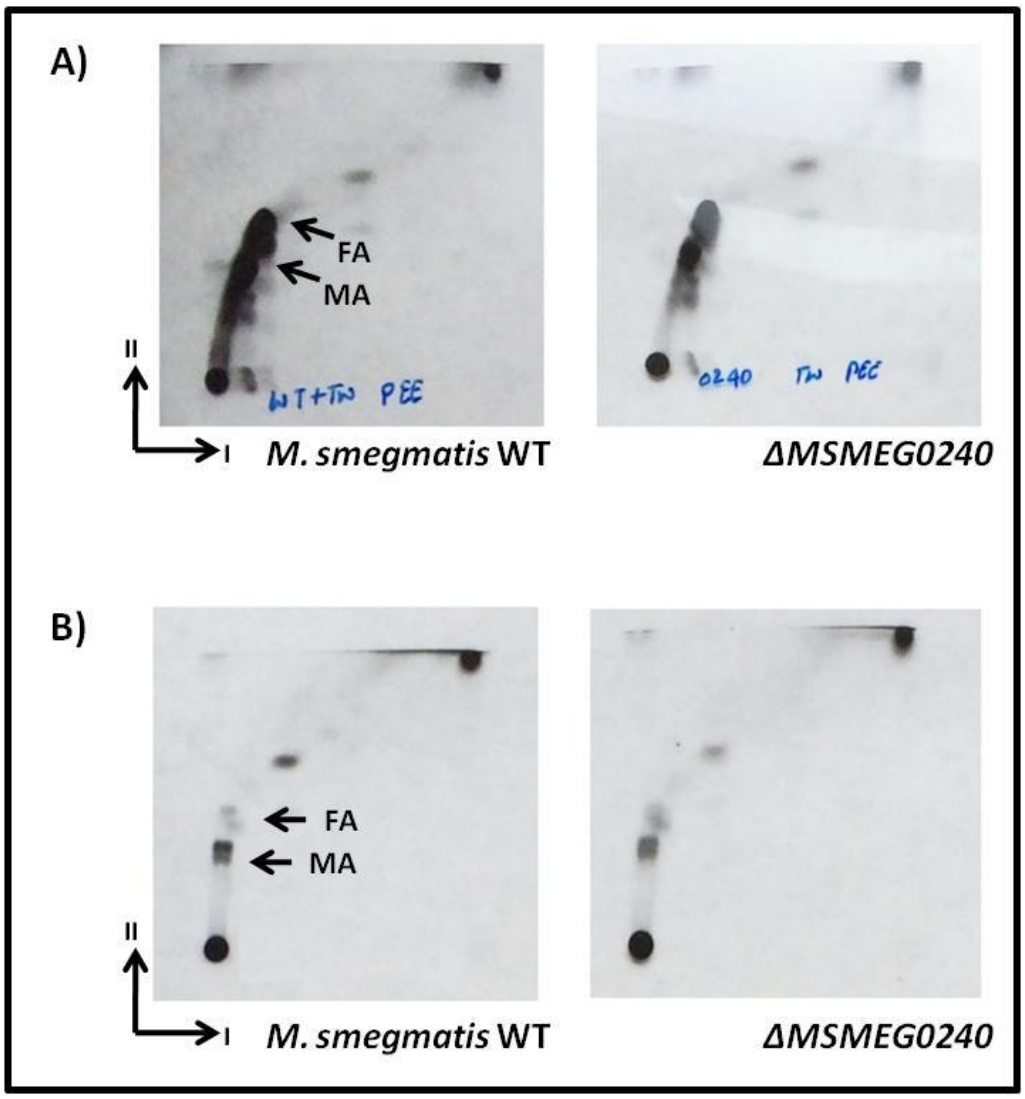


Figure 4-6 Lipid analysis of the Δ*MSMEG0240* mutant.2D TLC analysis of [^{14}C]-labelled lipids from the *M. smegmatis* WT and Δ*MSMEG0240* mutant. Petroleum ether extracts (A) and intracellular apolar lipids (B) were separated using chloroform:Methanol (96:4 v/v) in Direction I and toluene: acetone (80:20 v/v) in Direction II. Positions of mycolic acids (MA) and fatty acids (FA) are indicated with arrows.

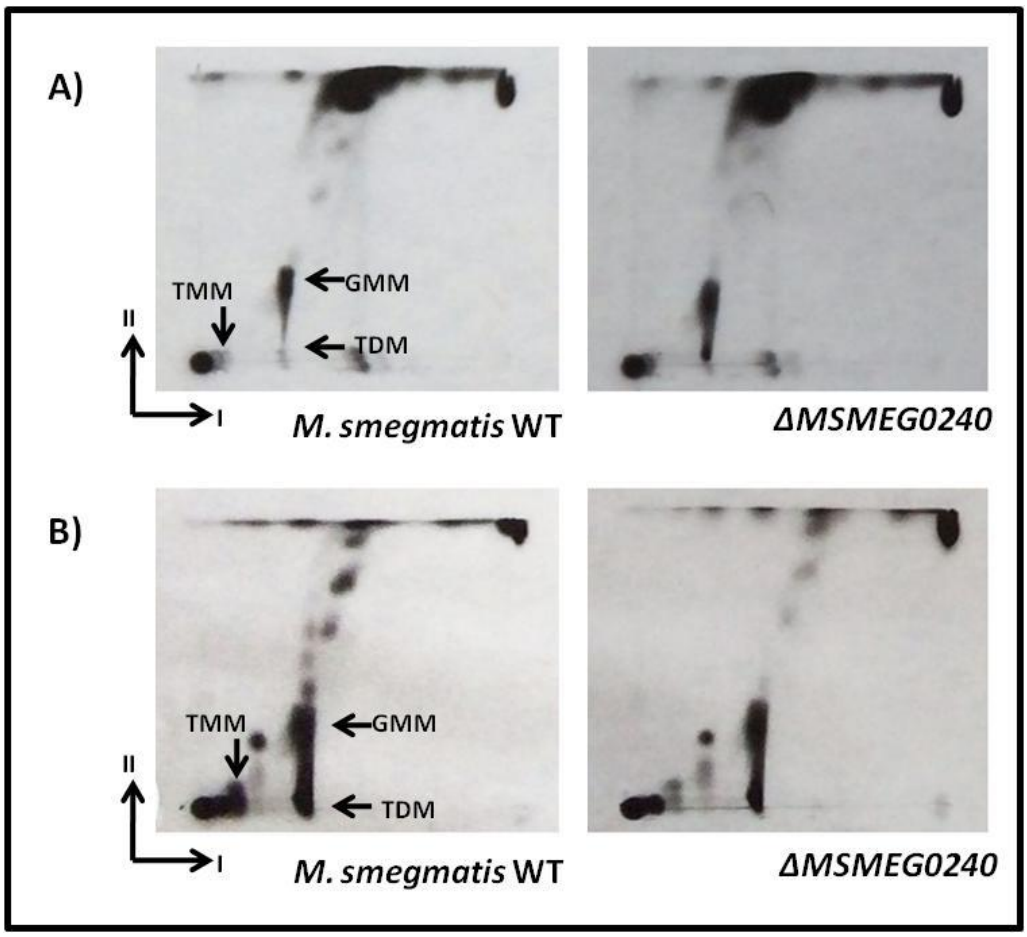


Figure 4-7 Lipid analysis of the Δ*MSMEG0240*mutant.2D TLC analysis of [^{14}C]-labelled lipids from the *M. smegmatis* WT and Δ*MSMEG0240* mutant. Petroleum ether extracts (A) and intracellular apolar lipids (B) were separated using Chloroform: methanol: water (100:14:0.8 v/v/v) in Direction I and chloroform: acetone: methanol: water (50:60:2.5:3 v/v/v/v) in Direction II. Positions of TMM, TDM and GMM indicated with arrows.

Further, cell wall bound mycolates were esterified as methyl esters and analysed using TLC (Figure 4-8), which revealed no significant difference in methyl esters of sub-classes of mycolic acids, suggesting that these genes may not contribute in either biosynthesis or to the transport of these lipids.

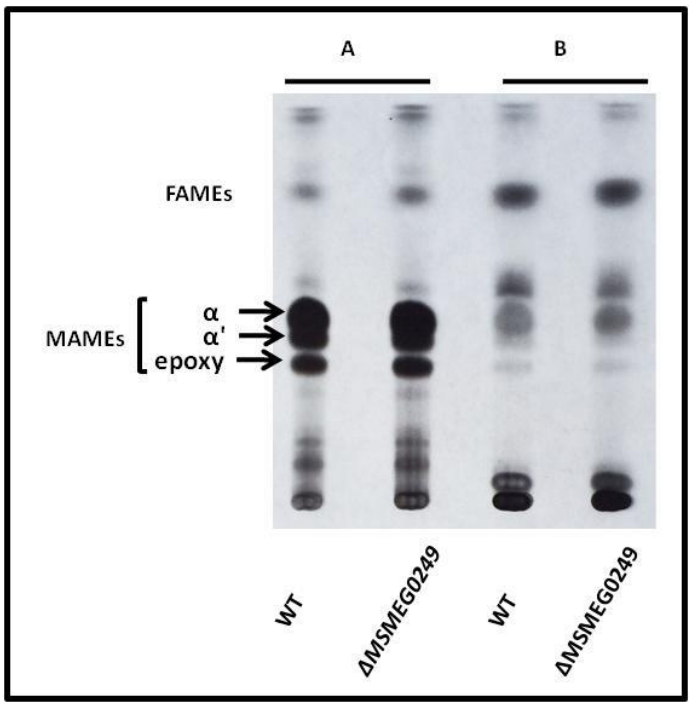


Figure 4-8 TLC analysis of Mycolic acid methyl esters.[^{14}C]- labelled, cell wall bound mycolic acids (MAMEs) from delipidated cells (A) and apolar lipids (B) were separated usingpetroleum ether: acetone (95:5 v/v) as the solvent system. Methyl esters of the different sub-classes of mycolic acids are indicated with arrows.

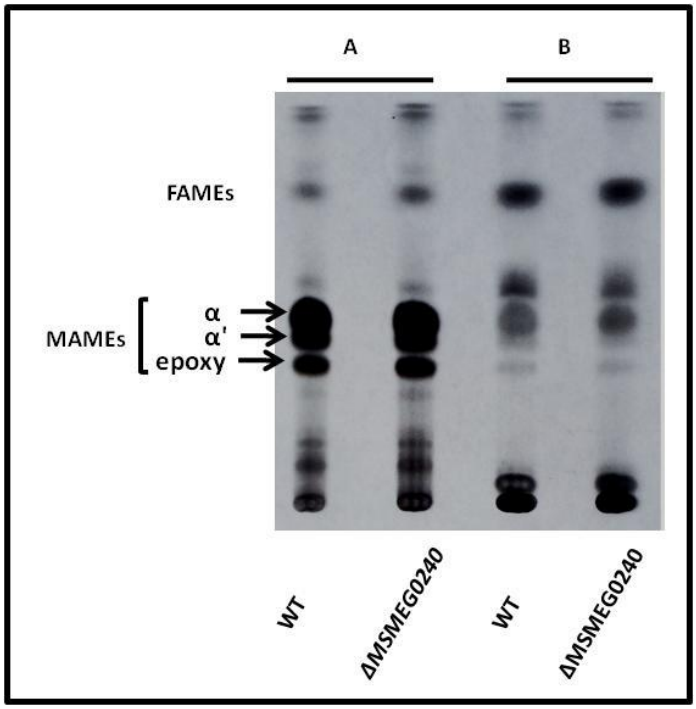


Figure 4-9 TLC analysis of Mycolic acid methyl esters.[^{14}C]- labelled, cell wall bound mycolic acids (MAMEs) from delipidated cells (A) and apolar lipids (B) were separated usingpetroleum ether: acetone (95:5 v/v) as the solvent system. Methyl esters of the different sub-classes of mycolic acids are indicated with arrows.

4.3.4 Screening for protein-protein interactions using bacterial two hybrid system

Parallel to the genetic approach, protein-protein interaction studies were carried out to identify the functions of *Rv0204, Rv0205* and *Rv3802* in the processing and transport of mycolic acids. MmpL3, ML1, ML2 and MmpL11 were used as 'bait' protein, to identify functional partners that facilitate the transport or bridge the biosynthesis and transport and to test whether this is a common model adopted by MmpL family of proteins (Zheng *et al*. 2011; Jain & Cox 2005; Rao & Ranganathan 2004). Table 4-2 lists the combinations of protein-protein interactions tested and summary of results.

Table 4-2 Summary of bacterial two hybrid test from MacConkey agar screen

	Rv3802	Rv0204	Rv0205	pUT-Zip
MmpL3	-	-	-	-
ML1	-	-	-	-
ML2	-	-	-	-
MmpL11	-	-	-	-
pKT-Zip	-	-	-	+

The positive interactions are indicated as (+) and negative interactions as (-)

The candidate genes were cloned in-frame with adenylatecyclase T25 or T18 fragments and expressed in *E. coli cya*$^{-}$ BTH101. The co-transformants containing two plasmids encoding putative interaction partners were plated onto selective MacConkey agar plates supplemented with 1% maltose, 0.5 mM IPTG and appropriate antibiotics. Interacting proteins bring T25 and T18 fragments in physical proximity restoring the catalytic activity and switches on the *lac* and *mal*operon, enabling the *cya*$^{-}$strain BTH101 to use maltose as unique carbon source and consequent red colouration of the colony due to the acidification of the media. T25 and T18 fragments fused to leucine zipper protein was used as a Positive control and T25 and T25 Zip construct co-transformed with counter empty plasmid used as a negative control. Initial screening results from bacterial two hybrid test indicated that the

candidate genes tested did not interact with any of the 'bait' proteins (Table 4-2, Figure 4-10). Rv0204 interaction with MmpL3 was observed but was turned out to be a false positive, since the corresponding control with empty vector also gave positive interaction (Appendix 11).

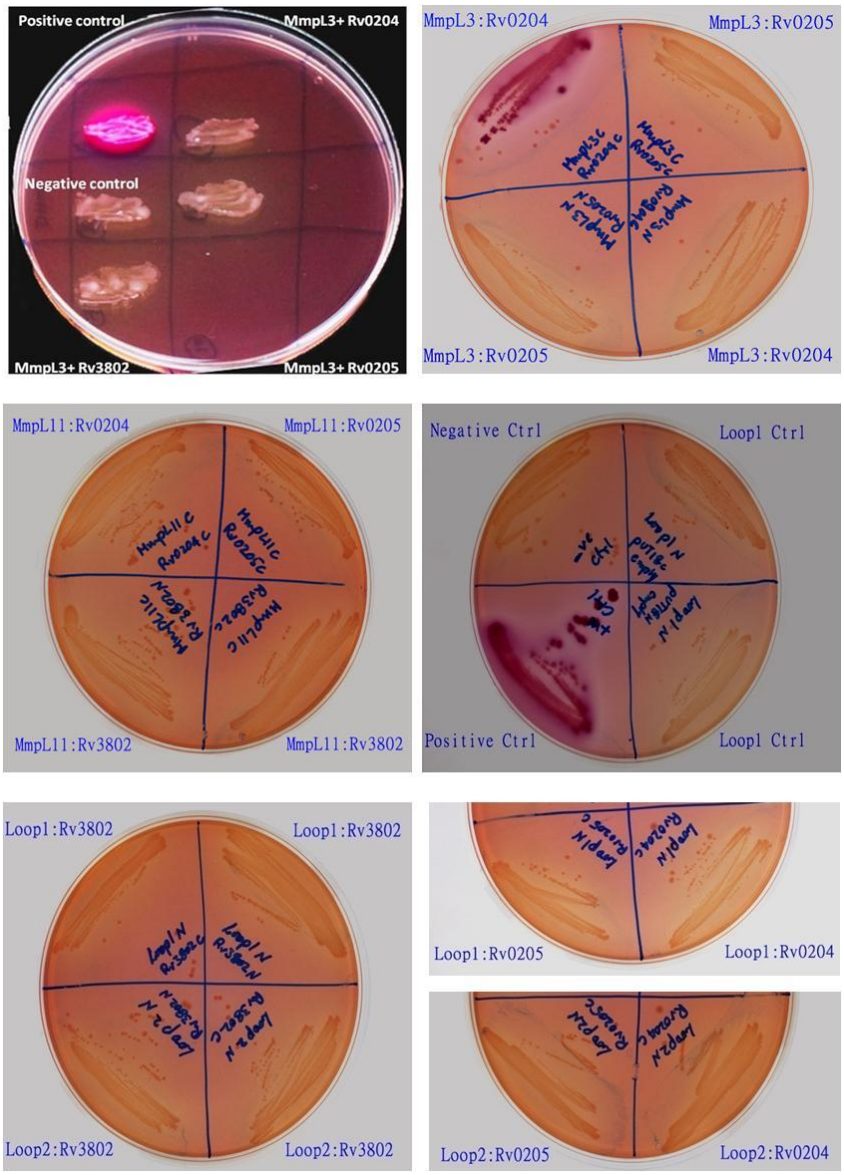


Figure 4-10 Bacterial two hybrid screening testing for interactions between 'bait' and 'prey' proteins on MacConkey agar plates. The candidate genes were cloned in-frame with adenylatecyclase T25 or T18 fragments and expressed in *E. coli cya*$^{-}$ BTH101. The co-transformants containing two plasmids encoding putative interaction partners were plated onto selective MacConkey agar plates supplemented with 1% maltose, 0.5 mM IPTG and appropriate antibiotics. The positive interaction is indicated by the red colouration of agar, on utilisation of maltose by reconstituted adenylatecyclase catalytic domain. T25 and T18 fragments fused to leucine zipper protein was used as a Positive control and T25 and T25 Zip construct co-transformed with counter empty plasmid used as a negative control.

4.4 Discussion

Mycolic acidsare a vital cell wall component essential for mycobacterial survival and pathogenicity (Takayama *et al*. 2005; Dubnau *et al*. 2000; Glickman *et al*. 2000). Mycolic acid biosynthesis has been thoroughly investigated however further studies are required to delineate the processing and transport of mycolic acids. Identification of the mycolic acid transporter MmpL3 was a major step forward in this regard (Varela *et al*. 2012). However, enzymes catalysing the transfer of nascent mycolic acids to trehalose to generate TMM and its subsequent transfer to MmpL3 still need to be clarified.

Jain and Cox (2005) have shown that non-TM domains of MmpL7 interact with polyketide synthase PpsE suggesting a link between biosynthesis and transport. Further, Zheng *et al* (2011) showed that these MmpLs form protein complexes with other proteins suggesting that the transport of different lipid molecules are brought about by protein scaffolds rather than single transporter. Previous studies have shown that mycolic acid biosynthesis is achieved through formation of large protein complexes that involve FabH, FAS-II components and PKS13 (Veyron-Churlet *et al*. 2004; Veyron-Churlet *et al*. 2005). However, beyond PKS13 step, its processing is not well studied. Transport of other complex lipid molecules *viz.* PDIM, sulfolipids, points at possible coupling of biosynthesis and transport (Domenech *et al*. 2004; Jain & Cox 2005).

In this regard we have investigated the function of genes co-localised in the*mmpL3* extended cluster and evaluated the possible formation of protein complexes to achieve transport. Towards this, we have generated null mutants of *MSMEG_0240c* and *MSMEG_0249*. Biochemical characterisation of these mutants revealed that these genes do not play a significant role in mycolic acid processing and transport. Although protein-protein

interaction studies seem to support the biochemical data, lack of interaction observed could be due to several reasons, which will be discussed in the due course.

Rv3802, a thioesterase with a conserved mycolyltransferase domain, is suggested to play a role in mycolic acid processing (Parker *et al*. 2009). The genetic locus of the gene also supports the hypothesis. The present hypothesis for the transfer of mycolic acids to its transporter suggest that the event is membrane associated and is likely to take place in the proximity of the transporter thus we have used MmpL3 as a bait protein in the bacterial two hybrid screens to identify possible proteins contributing to the process (Takayama *et al*. 2005; de Souza *et al*. 2008). Bacterial two hybrid screens did not help us to identify the links between these proteins. The lack of interaction between the combinations tested could be due to one of several reasons. There is a possibility that the interaction may require assistance of third protein or a recruiting factor which could not be accommodated in the bacterial two hybrid screens. Additionally, a lack of interaction could also be contributed by inappropriately expressed proteins or complete lack of either one or both the fusion proteins. Topological orientation of fusion proteins might also keep the complementary adenylatecyclase fragments apart, resulting in lack of interaction. This leaves a possibility that the hybrid proteins might still be interacting *in-vivo* but the interaction is not detectable. This limits us from completely ruling out the role of the tested proteins in facilitating mycolic acid transport. One possibility we have not tried in this chapter was to test whether any of these proteins interact with Pks13, the last known biosynthetic enzyme in the mycolic acid biosynthetic pathway. This could be important experiment since PKS13 was previously shown to form a complex with FAS-II enzymes (Veyron-Churlet *et al*. 2004; Veyron-Churlet *et al*. 2005) Interaction between Pks13 and MmpL3 could not be ruled out either since Pks13 also contains a thioesterase activity required to transfer the mycolic acid to recipient and earlier studies with other MmpLs supports the hypothesis (Jain & Cox 2005).

MmpL proteins belong to RND super family of proteins which are frequently found to be associated with a outer membrane protein and a membrane fusion protein (Paulsen *et al*. 1996). Wells *et al.* (2013) showed that MmpS4/S5 are membrane associated proteins that facilitate the MmpL4/L5 in the transport of siderophores. This class of proteins may be involved in the transport of other lipid molecules in association with respective MmpL partners. Transposon insertion in *mmpS3* gene resulted in slow growth pattern, however a detailed characterisation of the null mutant has not been carried out (Cole *et al*. 1998). The role of MmpS3 in mycolate transport and the link between MmpS3 and MmpL3 needs to be evaluated. Wells *et al* (2013) also showed that MmpS4 and MmpS5 are functionally redundant, which also opens up a possibility of other MmpS proteins associating with MmpL3 to facilitate mycolate transport.

In summary, we show that *MSMEG_0240c* and *MSMEG_0249* are non-essential for the growth and survival of mycobacteria, contradicting an earlier report that has listed *MSMEG0249* homologue as essential. Biochemical characterisation of these mutants revealed that they do not play significant role in the processing and transport of mycolic acids.

Chapter 5

Identification and characterisation of genes involved in LOS biosynthesis in *M. kansasii* and *M. canetti*

5.1 Introduction

The complex mycobacterial cell wall contains several solvent extractable glycolipids, some of which occur in all strains eg. TDM, while some are found only in select species *eg.* sulfolipids, glycopeptidolipids, phenolic glycolipids and lipooligo-saccharides (LOSs) (Minnikin et al. 2002; Indrigo et al. 2003; Hunter et al. 1983; Ortalo-Magné et al. 1996). The LOSs are solvent extractable, highly polar sugar-containing lipids. LOS producers include the opportunistic pathogen *M. kansasii*, the fish pathogen *M. marinum*, and the *M. tuberculosis* complex strain *M. canetti* (Hunter *et al*. 1985; Minnikin *et al*. 1989; Daffe *et al*. 1991). An early understanding of LOS chemistry came from studies carried out in *M. kansasii*, where seven sub-classes of LOSs were identified and characterised (Hunter *et al*. 1983; Hunter *et al*. 1984). The core consists of a polyacylated a trehalose which is further glycosylated. The variable sugars identified were xylose, 3-*O*-methyl-rhamnose, fucose and 4,6–dideoxy-2–*O*-methyl-3-*C*-methyl-4-(2'-methoxy-propionamido)-α-L-mannohexo-pyranose (Hunter *et al*. 1983; Hunter *et al*. 1985).

However, most of the biosynthetic knowledge concerning LOS assembly comes from *M. marinum* (Burguière *et al*. 2005; Ren *et al*. 2007; Sarkar *et al*. 2011). The presence of LOSs in *M. marinum* was first shown by Minikkin *et al*. (1989), and subsequently characterised by Burguière *et al*. (2005). *M. marinum* produces 4 sub-classes of LOSs, each with an identical core consisting of a polyacylated tetra-glucose unit with a varying oligosaccharide (Minnikin *et al*. 1989; Burguière *et al*. 2005). The oligosaccharide was shown to contain 3-*O*-Me-rhamnose, xylose and unique sugar residues, caryophyllose (α-3,6 dideoxy-4-*C*-(D-altro-1,3,4,5-tetrahydroxyhexyl)-D-xylo-hexopyranose) and an *N*-acyl amino sugar (*N*-acylated dideoxygalactose) (Burguière *et al*. 2005; Rombouts *et al*. 2009).

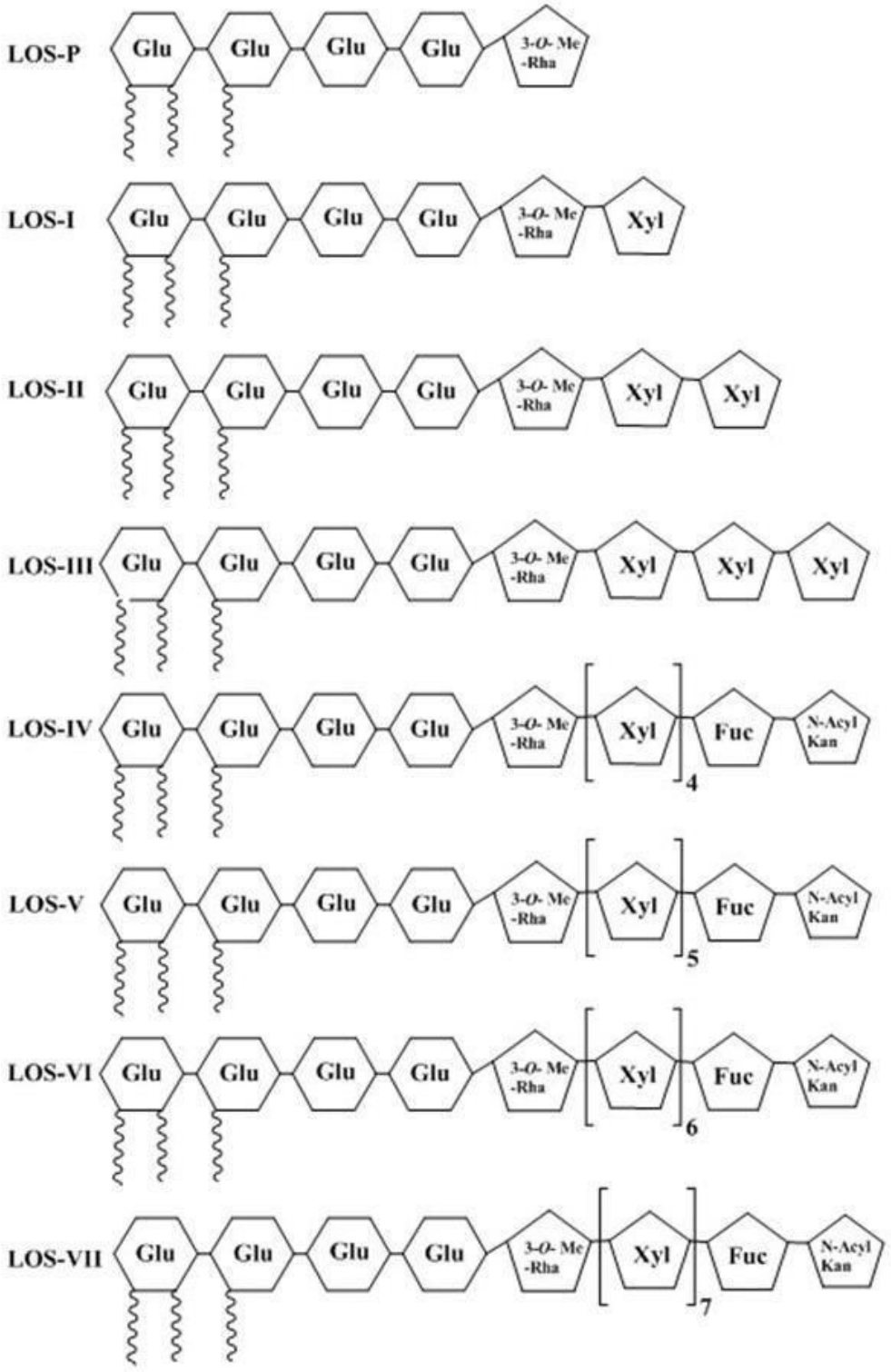


Figure 5-1 Schematic presentation of LOS subclasses from *M.kansasii*. Glu: D-Gluco-pyranose; 3-*O*-Me-Rha: 3-*O*- Methyl-α-L-Rhamno-pyranose; Xyl: β-D-Xylo-pyranose; Fuc: Fuco-pyranose; *N*-acyl-Kan: *N*-acetyl-kansosamine.

While *M. kansasii* has a similar cell wall lipid profile to that of *M. tuberculosis*, it has several antigenic glycolipids that are absent in *M. tuberculosis.* One class of such glycolipids are LOSs which renders the outer membrane of *M. kansasii* highly polar. The seven sub-classes of LOSs are produced by *M. kansasii* have less complex structure than the four sub-classes produced by *M. marinum* (Ortalo-Magné *et al*. 1996; Hunter *et al*. 1983; Hunter *et al*. 1984; Hunter *et al*. 1985; Belisle & Brennan 1989). The LOSs of *M. kansasii* are similar to that of *M. marinum* and consists of 3-*O*-Me-rhamnose, xylose, fucose and a unique disaccharide, *N*-acyl-kansosamine. The first 3 sub-classes of LOSs have a 3-*O*-Me-rhamnose residue with increasing residues of xylose. The sub-classes IV-VII, which are more polar are characterised by the presence of a fucose residue and a novel species-specific *N*-acyl-kanosamine added to LOS-III, with increasing amounts of xylose residues between the fucose and 3-*O*-Me-rhamnose. Acylation occurs at 3, 4 and 6 position of the terminal glucose with 2, 4-dimethyl tetradecanoic acid (Figure 5-1). *M. canetti,* a member of the MTBC complex, produces only one LOS species, in which the trehalose residue was found to be methylated at position 6' and is either 2,3,6- or 3,4,6- tri-*O*-acylated (Daffe *et al*. 1991). The oligosaccharide consists of glucose, 2-*O*-Me and 4-*O*-Me- rhamnose, 2-*O*-Me-fucose, *N*-acyl-4-amino-4, 6-dideoxy-galactose.

The presence of LOSs give the colonies of these strains a smooth texture and are closely linked to virulence. It has been shown that rough variants of *M. kansasii* can establish chronic systemic infections in mice, whereas the smooth variants are quickly cleared from the organs of infected mice (Daffe *et al*. 1991; Collins & Cunningham 1981). This early theory of clearance of smooth variants comes from the theory that the LOSs prevent the exposure of 'virulence factors' thereby making the strains susceptible to host immune responses. This may explain why the rough variants of *M. kansasii* are successful at establishing an infection in mice (Daffe *et al*. 1991). Similarly, the LOS-producing MTBC strain *M. canetti* is an

opportunistic human pathogen, unlike *M. tuberculosis* which is virulent in humans. These findings indicate that a loss of LOSs may be one, if not the only, defining factor in the increased virulence of *M. tuberculosis* in humans. Furthermore, the role of individual sub-classes of LOSs in antigenicity and immunomodulation is poorly understood in the context of a correlation between LOS production and severity of disease. Recently, LOS-IV, but not the other LOS-subclasses from *M. marinum* were shown to inhibit TNF-α release in LPS-activated macrophages (Rombouts *et al*. 2009). Similarly, a LOS-IV-deficient mutantshowed increased virulence in infected Zebrafish embryos (van der Woude *et al*. 2012). Furthermore, Alibaud *et al*. (2014) observed that *M. marinum* strains deficient in LOS production were phagocytosed more efficiently than those accumulating only early LOS intermediates, while wildtype strains and only LOS-IV lacking mutants were least efficient. A fine dissection of the role of different LOS-subclasses and their correlation to the ability to cause disease in models, such as *M. kansasii* and *M. canetti*, can eventually help us better understand how *M. tuberculosis*, which has lost the ability to make LOS's through reductive evolution (Ren *et al*. 2007), became a more virulent human pathogen than other LOS producing members of the MTB complex, like *M. canetti*.

To ascertain a definite role for LOSs, it is imperative to understand the biosynthesis of LOSs through genetic dissection of the biosynthetic pathway. A transposon mutant of *MMAR2313* (*losA*) was identified to be defective in LOS biosynthesis, showing its role as a possible glycosyltransferase. The mutant was defective in producing LOS-IV, indicating that LosA inhibits transfer of *N*-acylated-α-amino-4,6-dideoxy galactose to LOS-III (Burguière *et al*. 2005). Characterisation of *M. marinum* strain Mma7 supports this study. *M. marinum* strain Mma7 which carries a chromosomal deletion that extends from the 3' end of *losA* to *MMAR2318,* accumulates LOS-III and cannot produce LOS-IV suggesting that the genes in this region may be involved in LOS-IV synthesis, supporting earlier studies (Rombouts *et al*.

2009). There were three other genes that have been identified to have a biosynthetic role in LOSs biosynthesis. Transposon mutagenesis studies have shown that a UDP-glucose dehydrogenase (*MMAR2309*) and a carboxylase (*MMAR2332*) are essential for the biosynthesis of LOS, deletion of which led to the accumulation of intermediates of LOS-I and LOS-II. Further, deletion of *MMAR2333*, resulted in accumulation of LOS-II*, another LOS-II intermediate (Ren *et al*. 2007; Sarkar *et al*. 2011). Interestingly, *M. tuberculosis* has also shown to possess a LOS biosynthetic gene cluster, however is missing a number of genes (Table 5-1) (Ren *et al*. 2007). The gene cluster responsible for LOSs biosynthesis has also been identified in *M. kansasii*. This led us to attempt to investigate the function of several glycosyltransferases, polyketide synthases and mycobacterial membrane protein large (MmpL) that are present in this gene cluster. Although, the genetic locus has not been identified and annotated for *M. canetti*, the recent publication of a whole genome sequence of *M. canetti* (Supply *et al*. 2013) has helped us to identify homologues regions for *losA* and *mmpL12* and LOS biosynthetic cluster (Table 5-1). In this study we have attempted to generate LOS deficient null mutants in *M. kansasii* and *M. canetti* and characterise them in an effort to get insight into the biosynthesis of LOSs and their role in virulence and immunomodulation.

Table 5-1 Comparative table of genes involved in LOS biosynthesis. LOS biosynthetic cluster from *M. tuberculosis, M. kansasii* and *M. canetti* was compared with genes involved in LOS biosynthesis in *M. marinum.*

M.tuberculosis	*M.canetti*	*M.kansasii*	*M.marinum*	Function
Rv1491c	MCAN_15091	MKAN27340	MMAR_2301	Conserved membrane protein of unknown function
Rv1492	MCAN_15101	MKAN27345	MMAR_2302	Methylmalonyl-CoA mutase small sub unit, MutA
Rv1493	MCAN_15111	MKAN27350	MMAR_2303	Methylmalonyl-CoA mutase large sub unit, MutB
Rv1494	MCAN_15121	No homology	No homology	Hypothetical conserved protein/Antitoxin
Rv1495	MCAN_15131	No homology	No homology	Hypothetical conserved protein/Toxin
Rv1496	MCAN_15141	MKAN27355	MMAR_2304	Arginine/ornithine transport system ATPase
Rv1497	MCAN_15151	MKAN27360	MMAR_2305	LipL (esterase)
Rv1498c	MCAN_15161	No homology	No homology	Methyl transferase
Rv1498a	MCAN_15171	MKAN27365	MMAR_2306	Conserved hypothetical protein
Rv1499	MCAN_15181	No homology	No homology	Conserved hypothetical protein
No homology	No homology	MKAN27370	MMAR_2307	Hypothetical transmembrane protein
No homology	No homology	MKAN27375	MMAR_2308	Hypothetical membraneprotein
No homology	MCAN_03241	MKAN27380	MMAR_2309	UDP-Glucose/ GDP-Mannose dehydrogenase
No homology	MCAN_36461	MKAN27385	MMAR_2310	UDP-glucose 4- epimerase
No homology	No homology	MKAN27390	MMAR_2311	Glycosyl transferase
No homology	No homology	No homology	MMAR_2312	Nucleoside diphosphate sugar epimerase
Rv1500	MCAN_15191	No homology	MMAR_2313	Glycosyl transferase
Rv1501	MCAN_15201	No homology	MMAR_2314	Hypothetical conserved protein
No homology	No homology	No homology	MMAR_2315	Methyltransferase
No homology	No homology	No homology	MMAR_2316	Transcriptional regulator
No homology	No homology	No homology	MMAR_2317	O-methyltransferase
Rv1502	MCAN_15211	No homology	MMAR_2318	Glycosyl hydrolase (GH superfamily)
No homology	No homology	MKAN27395	MMAR_2802	Conserved membrane protein
No homology	No homology	MKAN27400	MMAR_2307	Hypothetical transmembrane protein
No homology	No homology	MKAN27405	No homology	Hypthetical protein
No homology	No homology	No homology	MMAR_2333	Glycosyl transferase
No homology	No homology	MKAN27410	MMAR_2337	Polysaccharide biosynthesis protein, GtrA

No homology	No homology	MKAN27415	MMAR_2924	Methyl transferase/SAM
No homology	No homology	MKAN27420	No homology	Hypothetical protein
No homology	No homology	MKAN27425	No homology	Glycosyl transferase, GT-A type (DPM/DPG)
No homology	No homology	MKAN27430	No homology	NAD dependent epimerase/dehydratase family
No homology	No homology	MKAN27435	No homology	Glycosyltransferase (GTA)
No homology	No homology	MKAN27440	No homology	Glucose-1-phosphate cytidiltransferase
No homology	No homology	MKAN27445	No homology	NAD dependent epimerase/dehydratase family
No homology	No homology	MKAN27450	No homology	Rhamnose epimerase
No homology	No homology	MKAN27455	No homology	Methyltransferase
No homology	No homology	MKAN27460	No homology	Perosamine synthase
No homology	No homology	MKAN27465	No homology	Carbomoyl phosphate synthase
No homology	No homology	MKAN27470	No homology	Hypothetical protein
No homology	No homology	MKAN27475	No homology	Methyltransferase
No homology	No homology	MKAN27480	MMAR_2339	SAM methyltrasnferase
Rv1527c	MCAN_15481	MKAN27485	MMAR_2340	Polyketide synthase
Rv1521	NC_019950	MKAN27490	MMAR_2341	Fatty acid co-A synthatase
No homology	No homology	MKAN27495	No homology	Hypothetical protein
No homology	No homology	MKAN27500	MMAR_4618	Hypothetical conserved protein
No homology	No homology	MKAN27515	No homology	WhiB4 transcriptional regulator
No homology	No homology	MKAN27525	No homology	Hypthetical conserved protein
Rv1522c	MCAN_15431	MKAN27530	MMAR_2342	MmpL family transport protein-12
Rv1527c	MCAN_15481	MKAN27535	MMAR_2343	Hypthetical conserved protein
No homology	No homology	MKAN27540	MMAR_2344	Polyketide synthase
No homology	No homology	MKAN27545	No homology	Hypothetical protein
No homology	No homology	MKAN27550	MMAR_2345	Decarboxylase
No homology	No homology	MKAN27555	No homology	Thioesterase
No homology	No homology	MKAN27560	MMAR_2346	Monooxygenase
No homology	No homology	MKAN27565	No homology	Transcriptional regulator
No homology	No homology	MKAN27570	No homology	Hypothetical protein
No homology	No homology	MKAN27575	MMAR_2349	Rhamnosyl transferase
No homology	No homology	MKAN27580	MMAR_2351	Glycosyl transferase
No homology	No homology	MKAN27585	No homology	Hypothetical protein
No homology	No homology	MKAN27590	MMAR_2350	Mehtyl transferase
No homology	No homology	MKAN27595	MMAR_2352	Hypthetical protein
No homology	No homology	MKAN27600	MMAR_2353	Glycosyl transferase

No homology	No homology	MKAN27605	MMAR_2354	Hypothetical protein
No homology	No homology	MKAN27610	MMAR_2355	Acyl transferase
No homology	No homology	MKAN27615	No homology	Hypothetical protein
No homology	No homology	MKAN27620	MMAR_2356	Isoleucyl t-RNA synthase
No homology	No homology	MKAN27625	MMAR_1375	Aryl sulfatase
No homology	No homology	MKAN27630	MMAR_2358	Methyl transferase type11
No homology	No homology	MKAN27635	MMAR_2359	DNA polymerase IV
No homology	No homology	MKAN27640	MMAR_2360	Asperginase
No homology	No homology	MKAN27645	MMAR_2361	Lipoprotein peptidase
No homology	No homology	MKAN27650	No homology	RNA pseudo uridine synthase
No homology	No homology	MKAN27655	No homology	RNA polymerase
No homology	No homology	MKAN27660	No homology	RNA polymerase subunit sigma 70
No homology	No homology	MKAN27665	No homology	Hemin receptor
No homology	No homology	MKAN27670	No homology	Hypothetical protein
No homology	No homology	MKAN27675	MMAR_2366	Oxido reductase
No homology	No homology	MKAN27680	MMAR_2367	Keto acyl reductase
No homology	No homology	MKAN27695	No homology	Glycosyl transferase

5.2 Materials and Methods

5.2.1 In silico analysis

The gene sequences for *M. kansasii* genes *MKAN27435* (codes for a glycosyl transferase), *MKAN27485* (codes for a polyketide synthase) and *MKAN27530* (codes for *mmpL12*) were obtained from the *M. kansasii* database (http://www.ncbi.nlm.nih.gov/genome) and for *M. canetti* genes MCAN_15191 (*losA*) and MCAN_15431 (*mmpL12*) were obtained from *M. canetti* database (http://www.ncbi.nlm.nih.gov/genome). The protein sequence alignments were done using blastp, two sequence alignments on (http://blast.ncbi.nlm.nih.gov).

5.2.2 Plasmids, DNA manipulations and bacterial growth conditions

Table 5-2 lists the plasmids, bacterial strains and mycobacteriophages used in this study. *E. coli* strains were routinely cultured in LB broth at 37°C. *M. smegmatis* strain mc^2155 was cultured at 37°C in TSB with 0.05% Tween 80. The generation and propagation of

mycobacteriophages was done on a lawn of *M. smegmatis* at 30°C in 7H10 basal plates overlaid with 7H9 soft agar as described in Chapter 7. High titres of mycobacteriophages were generated using protocols described in Chaper 7 (Larsen *et al*. 2007). *M. kansasii* and *M. canetti* strains were grown in 7H9 broth supplemented with 10% OADC (oleic acid, albumin, dextrose and catalase) and 0.05% Tween 80, at 37°C. For growth on plate, 7H10 media was supplemented with 10% OADC and incubated at 37°C. Antibiotics were added for selection where required. Hygromycin was used at 150 µg/ml and 100 µg/ml for *E. coli* and *M. smegmatis* respectively and at 50 µg/ml for *M. kansasii* and *M. canetti*. Kanamycin was used at 50 µg/ml for *E. coli* and at 25 µg/ml for *M. smegmatis, M. kansasii and M. canetti*.

Table 5-2Plasmids, phages and Bacterial strains used in this study.

Plasmids, phages and strains	Description	Reference
Plasmids		
p0004S	Cosmid vector containing Hyg-sacB cassette	Larsen *et al*., 2007
p∆*MKAN27435*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MKAN27435*	This work
p∆*MKAN27485*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MKAN27485*	This work
p∆*MKAN27530*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MKAN27530*	This work
p∆*MCAN_15191*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MCAN_15191*	This work
p∆*MCAN_15431*	Derivative of p0004s cosmid obtained by cloning upstream and downstream flanks of *MCAN_15431*	This work
pMV261	Kan^R, *E. coli*-mycobacterial shuttle vector (ColE1-MP*hsp60*)	Stover *et al*., 1991
pMV261- *MKAN27435*	Functional copy of *MKAN27435* cloned into pMV261 shuttle vector	This work
Phages		
phAE159	Thermo sensitive, conditionally replicating shuttle phasmid	Larsen *et al*., 2007

	derived fromlytic mycobactiophage TM4	
ph∆*MKAN27435*	Derivative of phAE159 obtained by cloning p∆*MKAN27435* into its *PacI* site	This work
ph∆*MKAN27485*	Derivative of phAE159 obtained by cloning p∆*MKAN27485* into its *PacI* site	This work
ph∆*MKAN27530*	Derivative of phAE159 obtained by cloning p∆*MKAN27530* into its *PacI* site	This work
ph∆*MCAN_15191*	Derivative of phAE159 obtained by cloning p∆*MCAN_15191* into its *PacI* site	This work
ph∆*MCAN_15431*	Derivative of phAE159 obtained by cloning p∆*MCAN_15431* into its *PacI* site	This work
Bacterial strains		
E. coli Top 10	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen
E. coli HB101	*E. coli* K-12 _F- mcrB mrr hsdS20(rB- mB-) recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20(SmR) glnV44 λ-	Invitrogen
M. smegmatis mc^2155	Wild type strain, E*pt*mutant of *M. smegmatis* strain mc^26	Snapper *et al*. 1990
M. kansasii ATCC12478	Wild type strain of *M. kansasii*	Generous gift from DR. Ben Appelmelk, VUMC, Amsterdam
M. canetti	Wild type strain of *M. canetti*	
Δ*MKAN27435*	*M. kansasii* strain with *MKAN27435* replaced with *hyg* cassette	This work
Δ*MKAN27435*-C	Δ*MKAN27435* complemented withpMV261-*MKAN27435*	This work

5.2.3 Construction of knockout phages

Approximately 1kb regions upstream and downstream of gene (s) identified in 5.2.1, were PCR amplified using genomic DNA from *M. kansasii* ATCC-12478 and *M. canetti*. The primer pairs are listed on Appendix-9. The PCR products were purified, digested with respective enzymes and cloned on either side of a hygromycin (*hyg*) cassette in the plasmid p0004S (Stover *et al*. 1991) to generate the allelic exchange substrate plasmid listed in Table

5-2. The allelic exchange plasmids were sequenced to confirm the presence of upstream and downstream regions of the respective genes. The allelic exchange plasmids were then cloned into the temperature sensitive phasmid phAE159 at the *PacI* site and packaged into empty λ-phage heads. The packaged phasmids were transduced into *E. coli* HB101cells.

The positive phasmid for each gene was transformed by electroporation into *M. smegmatis* mc^2155 and recovered at 30°C for 3 to 4 hrs in 1 ml of TSB containing 0.05% Tween 80. The recovered cells were split into 200 µl and 800 µl and mixed with 100 µl of log phase *M. smegmatis* and 5 ml of molten 7H9 soft agar. This mix was overlaid on 7H10 basal plates and incubated at 30°C for 2 to 3 days for plaques to be formed. The plates were soaked in MP buffer for 4 to 6 hours and syringe filtered and either used straight away or stored at 4°C.

5.2.4 Generation of null mutants

Specialised transduction of *M. kansasii* ATCC-12478 and *M. canetti* were performed as described for other mycobacterial species (Bardarov *et al*. 2002). *M. kansasii* and *M. canetti* cultures were grown in 7H9 media containing 10% OADC and 0.05% Tween 80 to an OD_{600nm} of 0.8 and harvested by centrifugation. The cell pellet was washed twice with an equal volume of MP buffer (composition in Chapter 7). Finally, the cells were suspended in small volume of MP buffer (2 ml). High titre (10^8-10^{10}) phage lysate (1ml) was mixed with 1 ml of cells for transduction, and 1 ml of MP buffer mixed with 1 ml of cells for a control reaction and incubated at 37°C for 24 hrs, harvested and recovered on 7H9 media supplemented with 10% OADC and 0.05% Tween 80 for at 37°C. The recovered cells were harvested and resuspended 400-600 µl of 7H9 media and plated on 7H10 agar supplemented with 10% OADC and 50 µg/ml hygromycin for selection. After 3 to 4 weeks of incubation, hygromycin resistant colonies that appear on the plates were inoculated into 10 ml of 7H9

media supplemented with 10% OADC, 0.05% Tween 80 and 50 µg/ml of hygromycin for genomic DNA extraction and further characterisation. Allelic exchange of each gene with *hyg* cassette in transductants was confirmed by Southern blot.

5.2.5 Generation of complemented strains

MKAN27435 was amplified from *M. kansasii* WT genomic DNA using the primer pair MKAN27435_F 5'ATGCGAATTCGTGGATTCGGTCAGCGTTG3' and MKAN27435_R 5'ATGCAAGCTTTCAGTCCGCCGGATTGTCGAAG3'. The PCR product was digested with *EcoRI* and *HindIII* and cloned into mycobacterial replicative plasmid pMV261 (Stover *et al*. 1991). The cloned PCR product was verified by sequencing and subsequently electroporated into the Δ*MKAN27435* strain. The transformants were selected on 7H10 plates containing 100 µg/ml hygromycin and 25 µg/ml kanamycin. One such transformant was called Δ*MKAN27435*-C and used for all subsequent characterisation.

5.2.6 Analysis of cell wall lipids from *M. kansasii* strains

M. kansasii WT, Δ*MKAN27435* and Δ*MKAN27435*-C cultures were grown at 37°C, to mid-log phase in 10 ml of 7H9 medium supplemented with 10% OADC and 0.05% Tween -80. [^{14}C]-acetate, (50 µCi, 57 mCi/mmol, GE Healthcare, Amersham Biosciences) was added to a mid-log phase culture and grown for a further 24 hrs to label lipids. The labelled bacterial cultures were harvested and washed with PBS. The washed pellet was resuspended in 2 ml of methanolic saline (methanol: 0.3% NaCl, 10:1 v/v) and apolar lipids extracted using 2 ml of petroleum ether (60-80 bp). Polar lipids containing glycolipids were extracted first with chloroform: methanol: water (9:10:3) and twice with chloroform: methanol: water (5:110:4). The two extractions were pooled and mixed with 1.3 ml of chloroform and 1.3 ml of 0.3% NaCl, mixed and centrifuged. The lower organic layer containing polar lipids were collected, dried and re-constituted in chloroform: methanol (2:1). The polar lipids were analysed by 2D

TLC visualised by autoradiography by exposing a Kodak MR film to the TLC plates overnight (Dobson *et al*. 1985).

5.2.7 Purification of LOS's from *M. kansasii* WT and Δ*27435* mutant species

180 mg of polar lipids extracted from 20 g of dried cells of *M. kansasii* WT and Δ*MKAN27435* strains. DEAE-cellulose column was equilibrated with chloroform: methanol (2:1) before applying polar lipids. The column was eluted with first, 10 column volumes of chloroform: methanol (2:1) followed by 10 column volumes of 10 mM, 25 mM, 100 mM and 500 mM ammonium acetate in chloroform: methanol (2:1). The fractions were collected and analysed using 2D TLC (Dobson *et al*. 1985) and visualised by spraying the TLC plates with α-naphthol/sulphuric acid.

5.2.8 Mass spectroscopic analysis of LOS subclasses from *M. kansasii* WT and mutant

The LOS subclasses from WT and mutant *M. kansasii* were permethylated using the sodium hydroxide procedure. For each sample, about 5 NaOH pellets were ground to fine powder in a dry mortar with a pestle. About 3 ml of anhydrous DMSO (for each sample) was added to form slurry. About 1 ml of the resulting slurry was added to the sample before 0.5 ml of methyl iodide was added. The reaction mixture was mixed on a vortex and then agitated on an automatic shaker for 30 minutes at room temperature. The reaction was quenched with water, while constantly shaking the tube. Permethylated samples were extracted with 1 ml of chloroform and washed several times with 3 ml of water. The chloroform was dried down under a gentle stream of nitrogen gas before sample purification using a Sep-Pak C18 cartridge. The permethylated LOS were then dissolved in methanol before an aliquot was mixed at a 1:1 ratio (v/v) with 10 mg/ml 3, 4-diaminobenzophenone in 75% acetonitrile. The glycan-matrix mixture was spotted on a target plate and dried. MALDI-TOF MS data were

obtained using a 4800 MALDI-TOF/TOF mass spectrometer (AB Sciex UK Limited) in the positive ion mode. The obtained MS data were viewed and processed using Data Explorer 4.9 (AB Sciex UK Limited).

5.2.9 Macrophage infection studies

The ability of *M. kansasii* and Δ*MKAN27435* strains to survive in macrophages was tested using J774 macrophage cell line. Briefly, J774 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) containing glutamine and 10% foetal bovine albumin, 100 µg/ml penicillin and 100 µg/ml streptomycin (complete DMEM). The J774 macrophage cells were enumerated on a haemocytometer and adjusted to 1 X 10^5cells per ml and adhered to a 24 well plate overnight at 37°C. *M. kansasii* WT, Δ*MKAN27435* and Δ*MKAN27435*-C cultures were grown to late log phase and harvested. The harvested cultures were washed with PBS and resuspended in complete DMEM to give 1 X 10^6 cells per ml. J774 macrophage cells were activated with 15 µg/ml PMA (phorbol 12-myristate 13-acetate) for 1 hour and infected with bacterial cultures resuspended in complete DMEM with a multiplicity of infection (moi) of 10 (i.e. 10 bacteria per macrophage cell) for 4 hrs. The infection was done in triplicates. The infection medium was removed at the end of 4 hrs, washed twice with PBS and once with complete DMEM containing 50 µg/ml gentamycin. Cells were incubated in 1 ml complete DMEM at 37°C. This post-infection media was collected after 24 hrs, 72 hrs and 120 hrs for TNF-α analysis and intracellular bacterial survival is determined by lysing the macrophages using PBS containing 0.01% triton-X 100 and plating the dilutions (10^{-2}-10^{-6}) on 7H10 plates supplemented with 10% OADC.

5.3 Results

5.3.1 In silico analysis

We first compared the gene cluster from *M. kansasii*, *M. canetti* and *M. tuberculosis* with *M. marinum* to identify the LOS biosynthetic cluster (Table 5-1). *MKAN27435* and *MCAN_15191* were both annotated as putative glycosyltransferases (GTFs) in their respective genome databases. *MKAN27485* was annotated as a polyketide synthase and *MKAN27530* and *MCAN_15431* were both annotated as MmpL that are involved in the transport of complex lipids in mycobacteria.

BLASTP analysis using *MKAN27435* revealed matches to LosA (MMAR2313) and WcaA (MMAR2333), but relatively low scores: 26% identity and 42% similarity with LosA and 23% identity and 43% similarity with WcaA. A preliminary amino acid sequence analysis of MKAN27435 revealed a domain characteristic of members of the GT-2 family of glycosyltransferases that are similar to eukaryotic dolichol phosphate mannose (DPM) synthases. These GTFs, that typically contain a GT-A fold, catalyse the transfer of sugarsto dolichol phosphate by using sugar nucleotide as substrates. Topology predictions also indicated the presence of two transmembrane helices, indicating a membrane-associated function. Additionally, the C-terminal region of MKAN27435 contains a domain similar to the solute binding domain of SLC5 proteins which are Na^+/sugar co-transporters with affinity for D-galactose, D-glucose, and D-fucose. Given the presence of fucose in *M. kansasii* LOSs, it seemed likely that *MKAN27435* played a direct or indirect role in the addition of a fucose residue during LOS biosynthesis in *M. kansasii* (Figure 5-2).

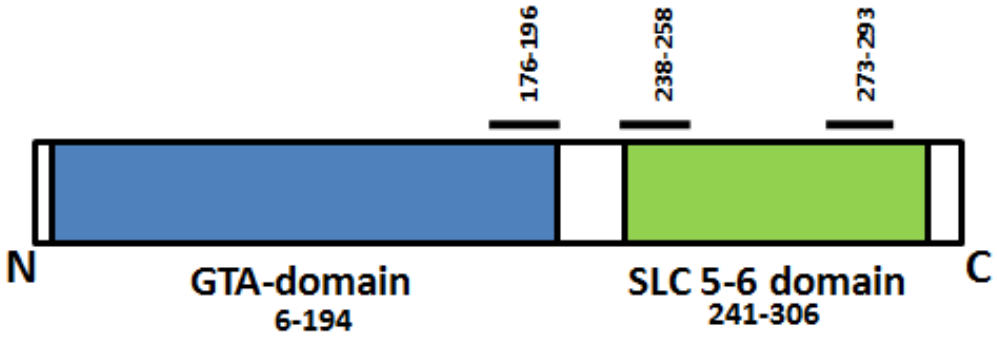


Figure 5-2 Predicted topology of MKAN27435showing domain organisation (shown in colour) and transmembrane domains shown with bars along with spanning regions.

BLASTP analysis of *MCAN_15191* revealed homology to *MMAR_2313* (Figure 5-3), annotated as *losA*, which is involved in the generation of a lipid bound N-acyl 4,6 dideoxygalactose which is subsequently transferred by an extracellular glycosyltransferase to LOS-III to form LOS-IV (Burguière *et al*. 2005). Primary sequence analysis of *MCAN_15191* for a conserved domain also revealed the presence of a DPM-like domain. The genetic locus of *MCAN_15191* near LOS biosynthesis genes and that it encodes for a putative GTF; is likely to be involved in LOS biosynthesis in *M. canetti*.

```
MCAN_15191      1 MRLSIVTTMYMSEPYVLEFYRRARAAADKITPDVEIIFVDDGSPDAALQQAVSLLDSDPC
MMAR_2313       1 MRLSIVTTLYMSEPYVLEFYRRVRAAADKITSDVEMIFVDDGSPDGSLDQAVSLLGKDPS
                  ******** ************* ******** *** *********  * ******  **

MCAN_15191     61 VRVIQLSRNFGHHKAMMTGLAHATGDLVFLIDSDLEEDPALLEPFYEKLISTGADVVFGC
MMAR_2313      61 VRVIQLSRNFGHHKAMMTGLAHATGDLVFLIDSDLEEDPALLEQFYEKLIATDADVVFGC
                  ****************************************** ****** * *******

MCAN_15191    121 HARRPGGWLRNFGPKIHYRASALLCDPPLHENTLTVRLMTADYVRSLVQHQERELSIAGL
MMAR_2313     121 QAQRPGSWWKNFGPKMHYRASAMLCNPPLHENVLTVRLMRADYVRCLVQHQERELSIAGL
                   * *** *  ***** ****** ** ****** ****** ***** **************

MCAN_15191    181 WQITGFYQVPMSVNKAWKGTTTYTFRRKVATLVDNVTSFSNKPLVFIFYLGAAIFIISSS
MMAR_2313     181 WQITGFNQIPMSVHKASKGTTTYTFAHKVKALVDNVTSFSNKPLVFIFYLGAVIFMISSL
                  ****** * **** ** ********  **  ********************** ** ***

MCAN_15191    241 AAGYLIIDRIFFRALQAGWASVIVSIWMLGGVTIFCIGLVGIYVSKVFIETKQRPYTIIR
MMAR_2313     241 AAGYLIIDRLFFRVLLGGWPSLIVSIWMLGGLTIFCLGLIGIYISKIFTETKQRPYTIIR
                  ********* *** *  ** * ********* **** ** *** ** * ***********

MCAN_15191    301 RIYGSDLTTREPSSLKTAFPAAHLSNGKRVTSEPEGLATGNR
MMAR_2313     301 RIYGPNFTSQGPASLEAAFRAMPPASWERFGADPVRSPRGDR
                  ****   *   * **  ** *       *    *     * *
```

Figure 5-3 Alignment of the *MCAN_15191*with *MMAR_2313*, known glycosyl transferase. The conserved amino acids are marked by anasterisk (*) below the alignment.

The gene *MKAN27485* annotated as *pks5*, encodes for a putative polyketide synthase. *MKAN27485* encodes a 2095 amino acid protein that shares 76% identities with *MMAR_2340*, which also encodes for *pks5*. Polyketide synthases are multi-domain enzymatic complexes and consist of an acyltransferase domain, Gro-ES like alcohol dehydrogenase domain, zinc-binding dehydrogenase domain, ketoreductase domain and a phosphpantheteine attachment site towards the C-terminus, that catalyse formation of polyketides by processive condensation reactions (Porter *et al*. 2013).

MKAN27530 and *MCAN_15431* are annotated as MmpL12 in *M. kansasii* and *M. canetti*genome databases, respectively. MmpL's are transmembrane proteins with 12-13 transmembrane helices and are involved in the transport of complex lipid molecules across the membrane for cell wall assembly (Varela *et al*. 2012). Albeit, the role of MmpL's in the transport of LOS are not well established. We wanted to test this hypothesis since these genes are located close to the LOS biosynthetic cluster.

5.3.2 Generation of null mutants

Specialised transduction, a highly efficient mycobacterial knock out method, was employed to generate null mutants in *M. kansasii* and *M. canetti*. This is the first report of the use of phages for delivering allelic exchange substrates for targeted gene knockouts in either species. We could only generate a null mutant of *MKAN27435* out of three genes attempted and transductants were obtained for the two of the *M. canetti* genes. The gene deletion was confirmed by Southern blot (Appendix 12). The knock out strain was complemented by introducing a plasmid borne functional copy of *MKAN27435* and henceforth will be refered to as Δ*MKAN27435*-C.

5.3.3 Effect of deletion of *MKAN27435* on colony morphology

Colony morphology of mycobacteria is affected by changes in the composition of cell wall lipids (Chen *et al*. 2006; Alexander *et al*. 2004; Sarkar *et al*. 2011). Previously, a LOS deficient *M. kansasii* was shown produce rough colonies (Daffe *et al*. 1991).Thus, affecting LOS biosynthesis was expected to alter the colony morphology in the mutant strain. *M. kansasii* WT, mutant (Δ*MKAN27435*) and complemented (Δ*MKAN27435*-C) strains were grown 7H10 agar plates supplemented with 10% OADC. Surprisingly, the mutant strain Δ*MKAN27435* did not show any alterations in colony morphology (Figure 5-4).

Figure 5-4 Colonies of *M. kansasii* WT, Δ*MKAN27435* and Δ*MKAN27435*-C grown on 7H10 agar supplemented with 10% OADC, after incubation at 37°C for 10 days.

5.3.4 Analysis of total lipids from the Δ*MKAN27435* mutant

To assess the effects of loss of Δ*MKAN27435* function on cell wall lipid composition, cultures of *M. kansasii* WT, mutant (Δ*MKAN27435*)and complemented (Δ*MKAN27435*-C) strains were grown in 7H9 medium supplemented with 10% OADC and 0.05% Tween-80 and labelled with [^{14}C]-acetate. Polar lipids were extracted and analysed by 2D TLC using System E, designed to separate LOSs and phospholipids as described by Dobson *et al*. (1985). Autoradiographs of 2D TLCs, revealed an altered pattern of LOS species in the mutant strain; while the WT produced all the subclasses of LOS's (LOS-I, LOS-II, LOS-III, LOS-IV, LOS-V, LOS-VI and LOS-VII) the mutant produced LOS-I, LOS-II and LOS-III but LOS-IV, LOS-V, LOS-VI and LOS-VII were missing in the strain **(Figure 5-5)**.

Furthermore, a new set of less abundant polar lipid species were detected in the mutant strain (**Figure 5-5**). Synthesis of all LOS species was restored in the mutant strain on introduction of a plasmid-encoded copy of *MKAN27435* (**Figure 5-5**), indicating that the observed alteration in LOS patterns in the mutant was solely due to the loss of *MKAN27435* function.

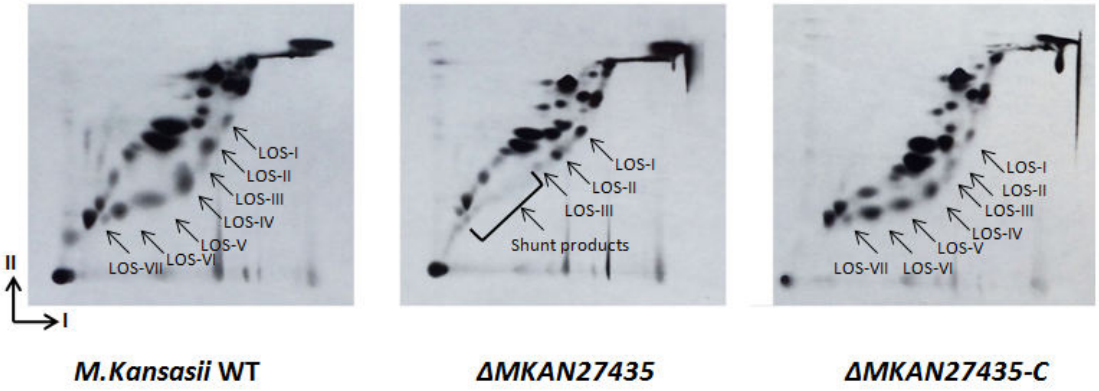


Figure 5-5 Autoradiograph of 2-D TLC analysis of polar lipids extracted from *M. kansasii* WT, Δ*MKAN27435*, Δ*MKAN27435*-C strains. The different LOS species are indicated by arrows. I and II indicate system E dimension 1 and 2 respectively. Dimension I: Chloroform: methanol: water (60:30:6) and dimension II: Chloroform: acetic acid: methanol: water (40:25:3:6).

Further, the polar lipids were analysed by 2D-TLC using system E and stained with phosphate stain. This was done to distinguish between phospholipids and LOSs.

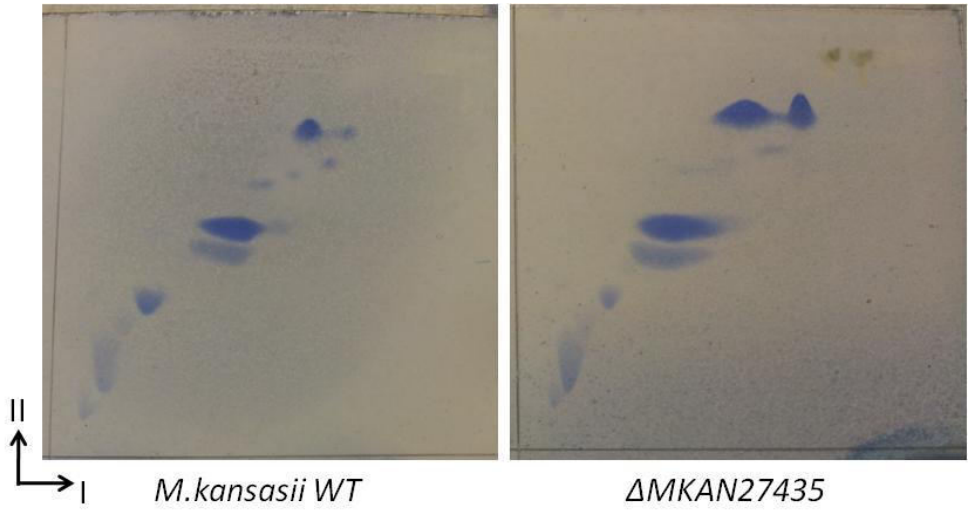


Figure 5-6: 2D-TLC analysis of polar lipids extracted from *M. kansasii* and Δ*MKAN27435* strains and stained with phosphate stain to visualise phospholipids. I and II indicate system E dimension 1 and 2 respectively. Dimension I: Chloroform: methanol: water (60:30:6) and dimension II: Chloroform: acetic acid: methanol: water (40:25:3:6).

5.3.5 Purification of LOS sub-classes from *M. kansasii* WT and Δ*MKAN27435*

Polar lipids extracted from *M. kansasii* WT and Δ*MKAN27435* were separately applied to a DEAE-cellulose column and eluted with chloroform: methanol (2:1). All the subclasses of LOSs were expected not to bind to the column and come through in the flow through (FT), as LOSs are non-ionic polar lipids leaving behind charged phospholipids. The FT fraction was analysed by 2D-TLC using system E and visualised by α-naphthol/ sulphuric acid, which revealed, as expected all the LOS subclasses were present in the FT fraction (**Figure 5-7**).

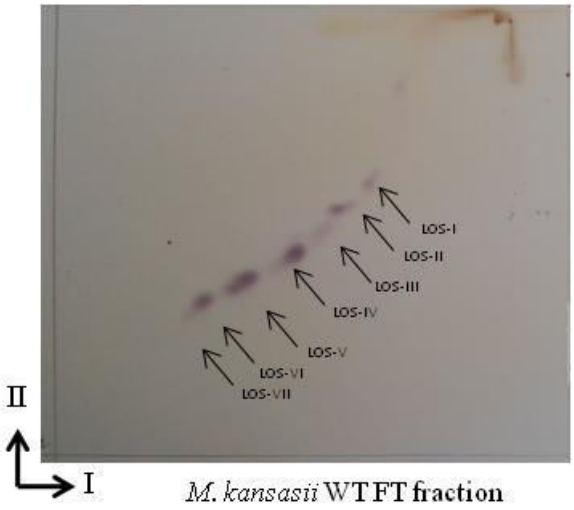


Figure 5-7: 2D-TLC analysis of purified LOSs from *M. kansasii* WT polar lipids. I and II indicate system E dimension 1 and 2 respectively. Dimension I: Chloroform: methanol: water (60:30:6) and dimension II: Chloroform: acetic acid: methanol: water (40:25:3:6).

The LOS subclasses from Δ*MKAN27435* polar lipids were eluted first with chloroform: methanol (2:1) followed by 10mM, 25mM and 100mM Ammonium acetate in chloroform: methanol (2:1). The fractions were analysed by 2D-TLC using system E and visualised by α-naphthol/ sulphuric acid. The 2D-TLC revealed that all the subclasses of LOS are produced by Δ*MKAN27435* does not bind to the column and comes through in the FT fraction (Figure 5-8).

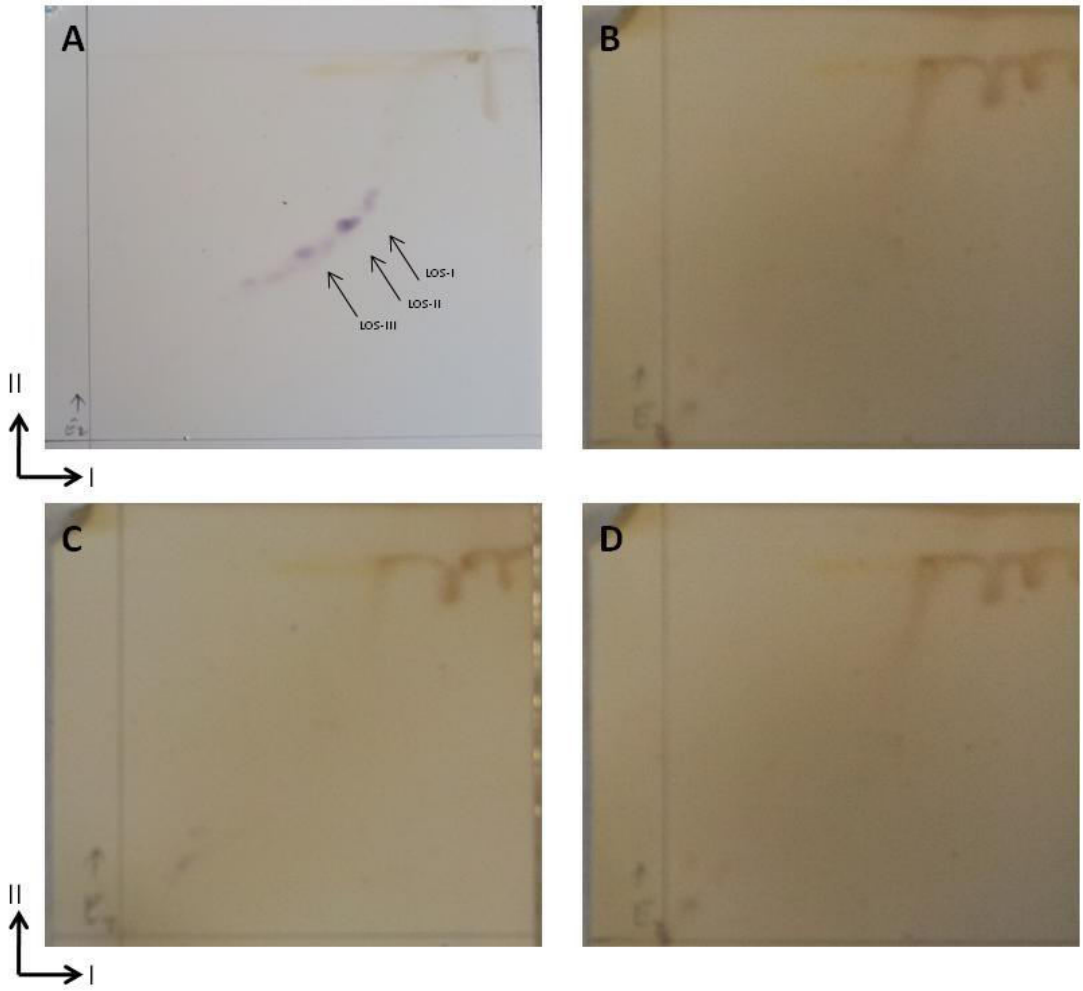


Figure 5-8: 2D-TLC analysis of purified LOS fractions from Δ*MKAN27435*. (A) FT fraction containing all the subclasses of LOSs (B) 10 mM ammonium acetate fraction (C) 25 mM ammonium acetate fraction and (D) 100mM ammonium acetate fraction

5.3.6 Characterisation of LOSs

Purified LOSs from *M. kansasii* WT and Δ*MKAN27435* strains were subjected to mass spectroscopic analysis. For this, LOS's were subjected to either per-*O*-methylation and then subjected to MALDI-TOF. All the subclasses of LOSs except LOS-VII were identified from *M. kansasii* WT purified LOSs while mutant showed the accumulation of LOS-P, precursor molecule for LOS biosynthesis and the presence of LOS-I, LOS-II and LOS-III (Figure 5-9). Higher subclasses of LOS's namely LOS-IV, LOS-V, LOS-VI and LOS-VII were missing. Mass spectroscopic analysis of the LOSs from mutant Δ*MKAN27435*, also showed the

presence of shunt product in relatively low abundance (Figure 5-9) (a detailed mass spectroscopic analysis provided in Appendix-13). The shunt products were characterised by incremental molecular weight corresponding to a pentose. The mass spectroscopic analysis is summarised in Table 5-4.

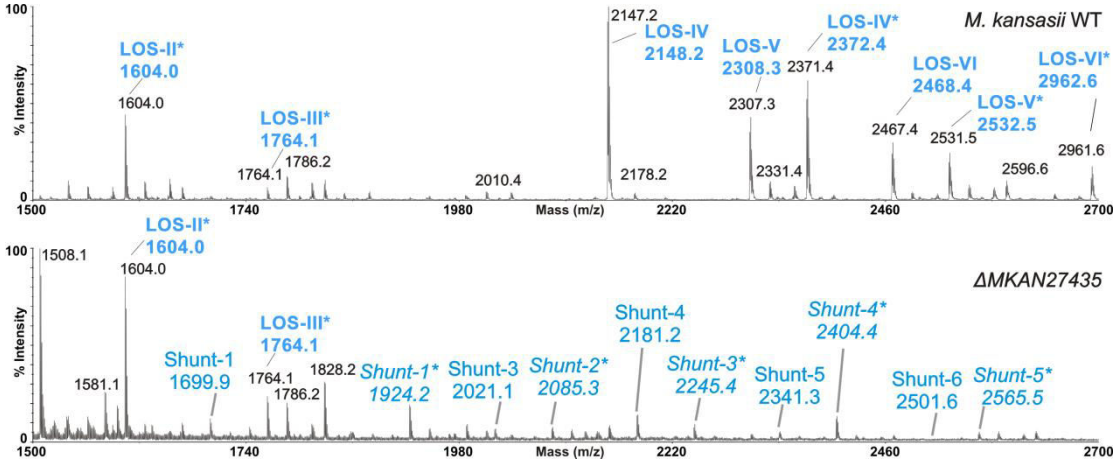


Figure 5-9: Mass spectroscopic analysis of purified LOSs from WT and Δ*MKAN27435* strains. LOS sub-classes are indicated with corresponding *m/z* values. Species indicated with an asterisk (*) correspond to *m/z* values with an intact acyl group are indicated

LOS-III is converted to LOS-IV by the addition of fucose through glycosidic linkage to xylose in the growing oligosaccharide chain. Further addition of *N*-acyl kansosamine to fucose yields LOS-IV. Similarly, LOS-V, LOS-VI and LOS-VII are formed by the addition of fucose and *N*-acyl kansosamine residues to LOS-III core with an additional xylose residue acting as precursor molecule. Mass spectroscopic analysis reveals that the mutant strain is defective in adding fucose residue to the growing oligosaccharide chain thus accumulating shunt product lacking fucose and *N*-acyl kansosamine. Given the similarity of *MKAN27435* to DPM like glycosyl transferases, the findings suggest that *MKAN27435* is involved in the transfer of nucleotide bound fucose residue to a polyprenol based lipid carrier. Polyprenol bound fucose could then be 'flipped' across the membrane and subsequently utilised as a substrate by an extracellular fucoosyl transferase to add the fucose to LOS-III.

Table 5-3: MALDI-TOF mass spectroscopy analysis tabulated to list LOS species with corresponding *m/z* values in *M. kansasii* and Δ*MKAN27435* strains. Asterisk (*) indicates species with an intact acyl group. Presence (+) or absence (-) of each species indicated for each species, *m/z* values correspond to [M+Na^+].

m/z	LOS sub-classes	*M.kansasii* WT	Δ*MKAN27435*
1059.6	LOS-P	+	+
1283.8	LOS-P*	+	+
1219.7	LOS-I	+	+
1443.91	LOS-I*	+	+
1379.8	LOS-II	+	+
1604	LOS-II*	+	+
1539.9	LOS-III	+	+
1764.1	LOS-III*	+	+
2148.2	LOS-IV	+	-
2372.4	LOS-IV*	+	-
2308.3	LOS-V	+	-
2532.5	LOS-V*	+	-
2468.4	LOS-VI	+	-
2692.6	LOS-VI*	+	-
2628.5	LOS-VII	-	-
2852.7	LOS-VII*	-	-

Table 5-4 MALDI-TOF mass spectroscopic analysis tabulated to list shunt products and corresponding *m/z* values in Δ*MKAN27435* strain; shunt products with an asterisk (*) correspond to *m/z* values with an intact acyl group. *m/z* values correspond to [M+Na^+].

m/z	Shunt product	Δ*MKAN27435*
1699.9	Shunt product-1	LOS-III + pentose
1860.1	Shunt product-2	LOS-III + 2 (pentose)
2021.1	Shunt product-3	LOS-III + 3 (pentose)
2181.2	Shunt product-4	LOS-III + 4 (pentose)
2341.3	Shunt product-5	LOS-III + 5 (pentose)
2501.6	Shunt product-6	LOS-III + 6 (pentose)
1924.2	Shunt product-1*	LOS-III + pentose
2085.3	Shunt product-2*	LOS-III + 2 (pentose)
2245.3	Shunt product-3*	LOS-III + 3 (pentose)
2404.4	Shunt product-4*	LOS-III + 4 (pentose)
2565.5	Shunt product-5*	LOS-III + 5 (pentose)

5.3.7 Intracellular survival of the *MKAN27435* null mutant in macrophages

Previous studies with LOS deficient strains of *M. kansasii* has shown that they could persist longer in mice (Collins & Cunningham 1981) and consistent with this observation, LOS deficient *M. marinum* strains were shown to be effectively phagocytosed (Alibaud *et al*. 2014). But on the contrary, a *M. marinum* strain defective in LOS-IV production was inefficient in entering murine macrophages (Burguière *et al*. 2005; Alibaud *et al*. 2014). However, both a LOS-IV deficient *M. marinum* mutant and LOS-II* accumulating *M. marinum* mutant had elicited increased levels of cytokine release (Burguière *et al*. 2005; Sarkar *et al*. 2011). The *M. kansasii* mutant strain Δ*MKAN27435* was missing higher LOS species namely LOS-IV, LOS-V, LOS-VI and LOS-VII. In order to assess the role of *MKAN27435* deficient strain in virulence, an intracellular survival assay was performed using the J774 macrophage cell line. Macrophages were infected with Δ*MKAN27435* to determine the ability to survive in the macrophages and alter immune response. The number of intracellular bacteria (colony forming units) was enumerated by lysing the macrophages and plating the lysate dilutions on 7H10 agar plates. The infection experiments were done in triplicate.

Disruption of *MKAN27435* did not affect the ability of the mutant strain to survive with in J774 macrophage cell line (Figure 5-10). *M. marinum* strains with gene disruptions leading to defects in LOS biosynthesis had also showed similar results (Burguière *et al*. 2005; Sarkar *et al*. 2011), however it led to alteration in immune response. In order to assess this, TNF-α level in the cell culture supernatant was estimated. To detect the TNF-α levels ELISA was performed, which revealed that the mutant strain Δ*MKAN27435* did not affect the TNF-α production significantly (Figure 5-11).

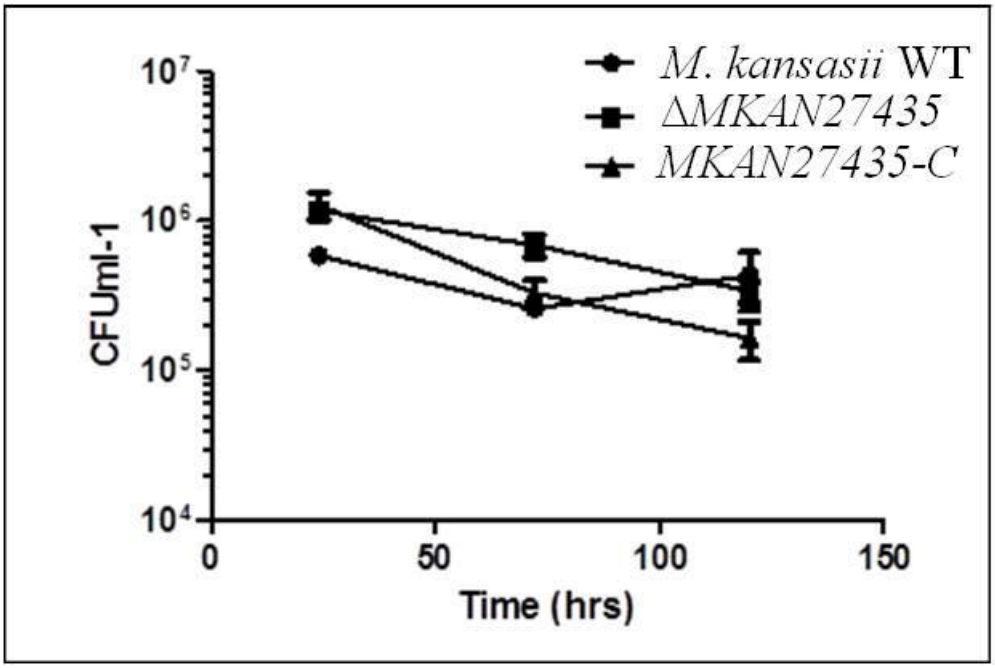


Figure 5-10: Survival of *M. kansasii* strains in infected J774 macrophage cell line. Survival of the WT, mutant and complemented strains were measured by plotting CFU/ml on log scale as a function of time. CFU counting was done by lysing the infected macrophages and plating the dilutions of lysate on 7H10 plates.

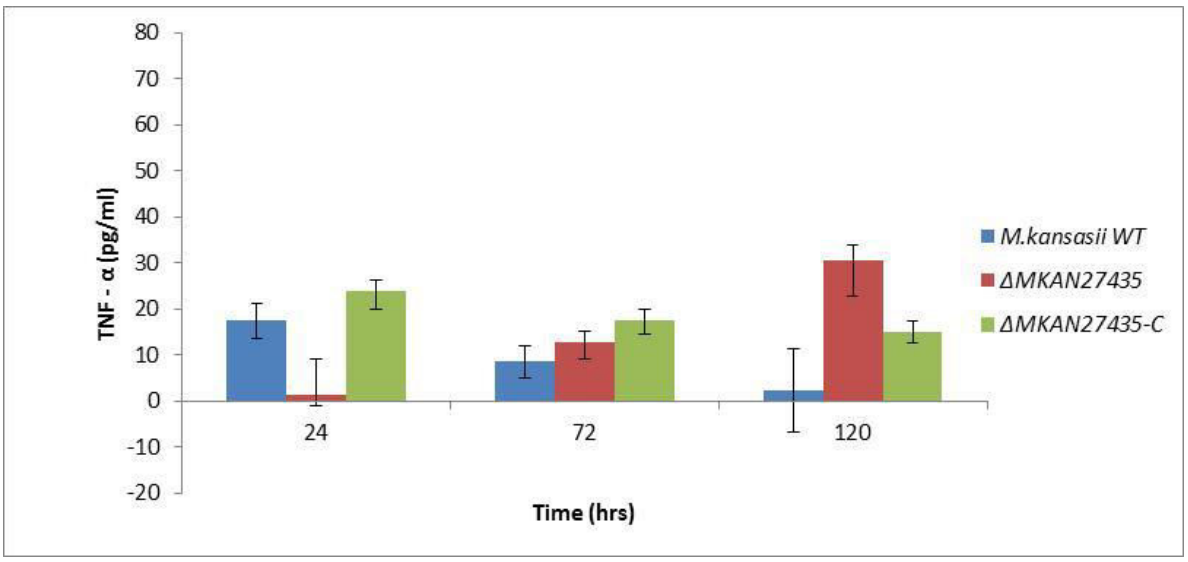


Figure 5-11: TNF-α production by J774 macrophages post infection with *M. kansasii* strains. Cytokine released in the spent media from the infected J774 cell line was measured using mouse TNF-α kit at different time points.

5.4 Discussion

M. kansasii causes chronic pulmonary disease in immune-compromised individuals and is the second most prevalent of non-tuberculosis mycobacterial (NTM) infection (Han *et al*. 2010; Anon 1997). *M. kansasii* is genetically close to *M. tuberculosis* and can be handled without BSL-3 facilities. These factors make *M. kansasii* an attractive model to study mycobacterial pathogenesis.

Mycobacterial cell wall associated lipids play a vital role in virulence and antigenicity (Daffe & Draper 1998; Domenech *et al*. 2004). Lipooligosaccharides are one such antigenic surface glycolipid. Interestingly, the *M. tuberculosis* complex organism *M. canetti* is an opportunistic pathogen produces LOSs (Daffe *et al*. 1991), but the highly virulent *M. tuberculosis* does not produce LOSs. Recent studies have shown that approximately one third of the LOS gene cluster is present in *M. tuberculosis* (Ren *et al*. 2007). LOSs have been implicated to play a role in sliding motility, biofilm formation and eliciting an immune response (Ren *et al*. 2007; Burguière *et al*. 2005; Sarkar *et al*. 2011). The role of LOS's in virulence has been shown in mice studies, wherein LOS deficient strains persisted longer in mice lungs as opposed to LOS producing strains (Belisle & Brennan 1989). While the structure of the LOS's has been largely understood, there's been little research on their biosynthesis (Minnikin *et al*. 1989; Picardeau *et al*. 1997). The role of individual LOS sub-classes in antigenicity and its immunomodulation can be made possible by generating deletion mutants that are defective in making particular LOS subspecies.

In this study, our aim was to dissect the biosynthesis of LOS's by generating deletion mutants of genes involved in the LOS biosynthesis. In this regard, we have generated a deletion mutant of *MKAN27435* that is defective in making four of the seven subclasses of LOS's. This study also demonstrates the utility of specialised transduction to generate null

mutants of *M. kansasii*. This is the first report of the use of phages for delivering the allelic exchange substrate for targeted gene knockouts in this species. Bioinformatics analysis of *MKAN27435* revealed that it encodes a putative glycosyltransferase. A preliminary amino acid sequence analysis of MKAN27435 revealed a domain characteristic of members of the GT-2family of glycosyltransferases that are similar to eukaryotic dolichol phosphatemannose (DPM) synthases and a solute binding domain of SLC5 proteins which are Na^+/sugar co-transporters with affinity for D-galactose, D-glucose, and D-fucose. Given the presence of fucose in *M. kansasii* LOSs, it seemed likely that *MKAN27435* plays a direct or indirect role in the addition of a fucose residue during LOS biosynthesis in *M. kansasii*. Consistent with the bioinformatics analysis, biochemical results implicated that *MKAN27435* intercepts conversion of LOS-III to LOS-IV. Mass spectroscopic analysis and 2D TLC analysis revealed that mutant strain lacks the ability to produce higher LOS sub-classes *viz.* LOS-IV, LOS-V, LOS-VI and LOS-VII. Interestingly, LOS-VII was not identified in the mass spectroscopic analysis of purified LOSs from both *M. kansasii* WT and mutant strain. One possible explanation is that LOS profile is affected by growth phase and media conditions (Burguière *et al*. 2005).

The resemblance of *MKAN27435* to bacterial DPM-like synthases suggests that it is unlikely that the glycosyltransferase (GTF) is involved directly in the formation of LOS-IV or higher subclasses. It is very likely that it is involved in the generation of a sugar donor in the form of polyprenol phosphate or a similar lipid carrier. This carrier could then be flipped across the membrane by either the same protein or by yet unidentified 'flippase' where it is used as a substrate by an extracellular fucosyltransferase to add the fucose to LOS-III. *In* vitro confirmation of these activities could not be accomplished due to limitations such as, availability of substrates. This is the first insight into LOS biosynthesis in *M. kansasii*. The LOS biosynthetic gene cluster consists of few more DPM-like GTFs and further studies of

these genes will confirm this hypothesis. That should also enable us to build a model for LOS biosynthesis and its subsequent transport across the cell membrane. In this study we had also attempted to generate a null mutant of *MKAN27530* which encodes a transporter protein MmpL12, which could have shed light into the transport of LOSs.

Alteration in the colony morphology was not observed on inactivation of *MKAN27435.* Disruption in the LOS biosynthesis was expected to lead to rough colony morphology, however the Δ*MKAN27435* colony had a smooth/glassy appearance, which may be due to LOS-I, LOS-II and LOS-III still being formed in considerable amounts. Intervention of biosynthesis further upstream in the pathway may lead to abolition of complete biosynthesis and may lead to rough variant. Generating null mutants of other annotated glycosyltransferases identified in the gene cluster, such as *MKAN27425* and *MKAN27580* will add to our current understanding.

Disruption of LOS biosynthesis by inactivating *MKAN27435* in *M. kansasii* did not affect its ability to survive in macrophages. The three subclasses of LOSs are still made in considerable amounts in Δ*MKAN27435* and could explain unaltered pathogenicity. Abolishing LOS biosynthesis completely or further upstream of the pathway may be more useful in understanding the role of LOSs in virulence.

Further, efforts to understand the role of LOS in mycobacterial virulence were attempted by generating null mutants in *M. canetti.* However, due to time constrains we could not characterise the transductants.

Chapter 6

General Discussion

Over the last seven decades or so our understanding of mycobacterial cell wall has made a lot of progress taking giant leaps in the last two decades. The complex nature of the cell wall has come under scrutiny and intense investigation by the international research community. Number of components that make up the complex cell wall has been shown to be indispensable for the survival of mycobacteria and the importance of mycolic acid for the integrity of the cell wall (Brennan 2003; Brennan & Nikaido 1995; Barry *et al*. 1998). Biosynthesis of mycolic acid is thoroughly studied and the enzymes involved are biochemically characterised (Bhatt *et al*. 2005; A. Bhatt, Molle, *et al*. 2007; Jackson *et al*. 1999; Marrakchi *et al*. 2002; Quémard *et al*. 1995; Smith *et al*. 2003; Choi *et al*. 2000). Some of the key aspects of mycolic acids biosynthesis are however still remain as important research questions. One of the key steps in the mycolic acid biosynthesis is bridging FAS-I and FAS-II, which is believed to be carried out by a type-III β-ketoacyl ACP synthase, FabH. *In-vitro* enzyme assays and structural studies supported this belief, however *in-vivo* data is still elusive (Scarsdale *et al*. 2001; Choi *et al*. 2000). This criticality of this reaction stems from inability for *de-novo* synthesis of fatty acids by FAS-II and dependence on short chain acyl co-A (C_{20-26}) substrates ligated to ACP for its initiation (Takayama *et al*. 2005; Choi *et al*. 2000a). Although, *fabH* is listed as non-essential (Sassetti *et al*. 2003), no reports of successful generation of *fabH* KO mutant points at possible essentiality. This led to biochemical and structural characterisation of the protein and pursuing it as an anti-tubercular drug target (Choi *et al*. 2000a; Scarsdale *et al*. 2001; Lv *et al*. 2009; Liu *et al*. 2012). The essentiality of both *fabH* and the reaction it catalyses was thus probed in this thesis.

The genetic studies was carried out in *M. bovis,* routinely used as a surrogate organism to study biochemical processes of *M. tuberculosis*; to generate deletion mutant of *fabH*. A successful generation of *fabH* deletion mutant indicated its non-essentiality. Further, deletion of *fabH* did not affect the biosynthesis of mycolic acids. These results provided

direct evidence contradicting general belief and supporting TraSH data. A careful bioinformatics analysis and literature survey were carried out to identify redundant β-ketoacyl synthases present in *M. tuberculosis* since previously functional redundancies have been reported both in mycobacteria and its surrogates (Puech *et al*. 2002; Brand *et al*. 2003; Sousa-D'Auria *et al*. 2003; Meniche *et al*. 2009). Also, recently functional redundancy was also reported for FabH function in pseudomonas (Yuan, Leeds, *et al*. 2012). Bio-informatics analysis and literature reports pointed at Pks18 as a potential substitute for FabH (Saxena *et al*. 2003). *pks18*, like *fabH* is also non-essential. Multiple attempts at generating *fabH: pks18* double KO was not successful indicating at requirement of at least one of the genes for its survival. We are now using a genetic tool termed CESTET, a tool designed to test the gene essentiality in mycoabacterial species, to generate the double KO (Bhatt *et al*. 2005). CESTET method involves generating a merodeploid by introducing a functional copy of the gene to be tested under the control of inducible promoter and then substituting the native copy of the gene with a selection marker. Generation of double KO will help us provide crucial evidence for functional redundancy in mycobacterial FabH function.

Another important question we sought to answer in this study is to get insights into transport of mycolic acids. Although mycolic acid biosynthesis was studied for years, it's processing and transport is still in its rudimentary stage. The initial suggestions on how mycolic acids processing and transport could occur was outlined by Takayama and later by de Souza (Takayama *et al*. 2005; de Souza *et al*. 2008). In the recent years several class of inhibitors that affect mycolic acid biosynthesis were shown to generate mutations in MmpL3, indicating its possible role in mycolic acids biosynthesis and transport (Tahlan *et al*. 2012; La Rosa *et al*. 2012). MmpL3 was later shown to play a central role in mycolate transport (Varela *et al*. 2012). However, the mechanism of transport and the proteins associated with MmpL3 still remain a research question. Since MmpL3 is a potential high value drug target,

delineating its mechanism could help in accelerating the drug discovery apart from understanding the underlying biochemistry.

Towards this, studies aimed to characterise MmpL3 using biophysical techniques and resolve its structure by combining molecular model predictions with low resolution structure determination using Cryo-EM. However, MmpL3 is a transmembrane protein with twelve transmembrane helices and heterologous expression of such huge protein itself is difficult that hinders downstream characterisation. Thus the issue was approached by first, trying to characterise soluble cytoplasmic domains (Loop1 (ML1) and Loop2 (ML2)) with parallel efforts to express full length protein in sufficient amounts for biophysical and structural studies. We were successful in expressing these domains in sufficient quantity and purifying them to homogeneity. We showed that these domains when fused with an artificial linker peptide, behaves as a dimer, albeit in a minor population. Attempts at crystallising these domains were futile. Initial studies using NMR spectroscopy suggested that these domains fold properly but requires detailed study for resolving the structure. We are currently establishing collaborations for detailed NMR studies.

The conditions were optimised for the expression of sufficient amounts of full length protein and purified to homogeneity. A styrene based co-polymer was used to isolate membrane protein from its native environment to purify the protein (Jamshad *et al*. 2011). Purified protein was confirmed using western blot and mass spectrometric analysis of trypsin digested fragments. Circular Dichroism study indicated proper folding of protein and thus, was used for downstream biophysical studies. A large population of MmpL3 behaves as a dimer and this observation was consistent across different biophysical experiments carried out. Further this was supported by earlier observations made with cytoplasmic domains, which exhibited a tendency to form dimers.

MmpL3 belong to a class of RND proteins with 12 transmembrane helices and two cytoplasmic domains. The general architecture of the protein is depicted in Chapter 3. The model was built (in collaboration with Dr. Vassily Bavro) using other RND proteins (AcrB, CusA and SecDF), crystal structures of which are available (Tsukazaki *et al*. 2013; Murakami *et al*. 2002; Long *et al*. 2010). Protein yield for MmpL3 was not sufficient to set up protein crystallisation. Previously, Cryo-EM was reported to be useful in solving low resolution structure of large mycobacterial proteins or protein complexes that are difficult to crystallise and to validate the model prediction of the proteins (Boehringer *et al*. 2013). We used Cryo-EM and SAXS to analyse the structure of MmpL3 and to validate our model predictions. Analysis of the SAXS data was made difficult due the presence of membrane components and the polymer used for purification. Our collaborators are optimising the model building tool to accommodate polymer which should allow us to get refined low resolution structure of MmpL3. Analysis of Preliminary Cryo-EM data (done by our collaborators Dr. Sarah Lee, University of Birmingham) revealed that MmpL3 is a dimer which is in line with our earlier observations and the predicted model fit in the electron density envelope generated by Cryo-EM data analysis. More data is currently being collected to get a higher resolution images. Model predictions also suggested that the mycolate transport could be possibly driven by proton motive force (Seeger *et al*. 2009; Long *et al*. 2010; Tsukazaki *et al*. 2013). Results obtained so far could not conclusively prove this hypothesis and requires more experiments.

The conditions are now optimised for the sufficient protein yield of MmpL3 using fermentation process. Fermentation has helped increase the yield nearly 4 fold from 0.5 mg/L to 2 mg/L, and bulk cultures should give enough protein to carry out protein crystallography. However, an alternative or a traditional purification steps needs to be optimised since no crystallisation success has been reported with use of SMA solubilised protein yet which could be due to presence of polymer that encapsulates individual protein molecule, which might

interfere with crystallisation. The protein thus purified might also find its usefulness in ligand binding assays due to the inherent fluorescence of the polymer. Alternatively radiolabelled ligands could be used in place of fluorescent ligands to overcome the interference of polymer in the binding assays. With protein purified in sufficient quantity also allows us to develop functional assays which will be helpful in deciphering function of the protein and also mechanism of action. The functionality assay will also provide a platform to test site directed mutants generated in this study to understand the role of individual aminoacids in transport mechanism.

Previous studies with other MmpL proteins show that these proteins are likely to function as protein scaffolds. This allows coupling of biosynthesis and transport of these complex lipid molecules. MmpS proteins association with some of the MmpLs have also been shown indicating that transport is facilitated by associated proteins (Jain & Cox 2005; Zheng *et al*. 2011; Wells *et al*. 2013). Studies of protein complexes in mycobacteria and the STRING database indicated that genes co-localised in *mmpL3* cluster are the most likely to assist in mycolate transport (Zheng *et al*. 2011; Franceschini *et al*. 2013). Pacheco *et al*. provided evidence to support these prediction by showing MmpL11 role in transporting mycolate containing lipids (Pacheco *et al*. 2013). The function of *MSMEG0240* (orthologues of *Rv0201*), *MSMEG0248* (*Rv0204)* and *MSMEG0249* (*Rv0205*) were probed and evaluated its association with MmpL3 using protein-protein interaction studies using comprehensive bacterial two hybrid screens. Multiple attempts to delete *MSMEG0248* failed suggesting its possible essentiality.

Bacterial two hybrid screens have proven helpful in deciphering protein interactions and understanding the networking in accomplishing biological processes in mycobacteria (Jankute *et al*. 2014). A comprehensive bacterial two hybrid screen was carried out using

MmpL3 as a bait protein. However, the screen could not identify the interacting partner which could be due to one of the (a) the prey proteins tested did not interact with MmpL3 and are not associated with mycolate transport, (b) the prey proteins investigated are associated with mycolate transport however requires presence of third protein to interact with MmpL3 or (c) either or both bait and prey proteins are not expressed/ folded properly to give positive interaction. Bacterial two hybrid system has its limitations in that it may not be best suited to express membrane proteins. The fact that we had difficulties in expressing MmpL3 in *E. coli* suggests (c) would be the likely scenario. Using yeast two hybrid system in the place of bacterial two hybrid system may help to overcome these difficulties and yield conclusive results. Biophysical techniques such as AUC, size exclusion chromatography and protein cross linking studies have proven useful in studying protein-protein interactions require large amounts of purified protein (Roy *et al*. 2013; Tang & Bruce 2009). Availability purified MmpL3 in sufficient quantity will now help explore these avenues. Further, purified protein may also help generating protein specific antibodies that could be used in co-immunoprecipitation experiments.

This study has also attempted to understand the biosynthesis and transport of other important components of mycobacterial cell wall, lipooligosaccharides (LOSs). Interestingly virulent strains of mycobacteria *viz. M. tuberculosis* has selectively lost the ability to produce LOSs (Ren *et al*. 2007; Burguière *et al*. 2005) and also LOS deficient strains of opportunistic pathogen *M. kansasii* are more virulent (Collins & Cunningham 1981). These data indicates that LOSs behaves as surface antigens and plays a role in immunomodulation. Dissecting biosynthesis of LOSs through defined null mutants have proved useful in understanding specific role of LOSs in general (Ren *et al*. 2007; Rombouts *et al*. 2009; Rombouts *et al*. 2010 Sarkar *et al*. 2011; Alibaud *et al*. 2014). Most of our current biosynthetic knowledge stems from genetic studies carried out in *M. marinum*. In this study we have attempted to

study biosynthesis and transport of LOSs from two other mycobacterial strains that are evolutionary closely related to *M. tuberculosis; M. kansasii* and *M. canetti* (Veyrier *et al*. 2009; Veyrier *et al*. 2011). LOS biosynthetic gene clusters were identified in *M. kansasii* and *M. canetti* by comparing it with *M. marinum* and representative genes from three classes of enzymes involved in LOS biosynthesis and transport were chosen for study. The three classes included glycosyltransferases, polyketide synthases and transporter proteins (MmpLs). Using a Specialised transduction, a null mutant of *MKAN27435,* a glycosyltransferase involved in LOS biosynthesis in *M. kansasii,* was generated (Bardarov *et al*. 1997). This is first report of targeted gene knock out using Specialised transduction in *M. kansasii*. Characterisation of the mutant revealed that the gene is involved in the conversion of LOS-III to LOS-IV and intercepts before the addition of fucose moiety. Deletion of *MKAN27435* resulted in loss of higher subspecies of LOSs and accumulation of shunt products in low abundance. Loss of LOS subspecies did not affect TNF-α release from activated human macrophages nor survival.

To date there is no clear idea about the location of the different stages of biosynthesis of LOSs. Sarkar *et al*. (2011) have suggested that early stages of biosynthesis occur in cytoplasm and the later stages are carried out by extra-cytoplsmic glycosyltransferases. Like in *M. marinum*, DPM like glycosyltransferases also found present in *M. kansasii* which are involved in generating polyprenol bound sugar substrates which are ultimately utilised by extracytoplasmic glycosyltransferases to generate higher subspecies of LOSs. In *M. marinum* the xylose is predicted to be added from a polyprenol donor by extracytoplasmic glycosyltransferase with polyprenol bound sugar substrate generated by a DPM like glycosyl transferase. Thus, one could suggest, in *M. kansasii*, LOS-I precursor synthesised in the cytoplasm is flipped across the membrane by a transporter protein (MmpL12) encoded by *MKAN27530*. *M. kansasii* LOS biosynthetic cluster consists of three DPM like

glycosyltransferases encoded by *MKAN27390, MKAN27425* and *MKAN27695* and a DPM like active site containing glycosyltransferase encoded by *MKAN27435*, which are likely be involved in polyprenol bound sugar donors and a separate set of glycosyltransferases encoded by *MKAN57580* and *MKAN57600* which might be involved in elongation of oligosaccharide chain of LOSs to generate different subspecies of LOSs. However, generating null mutants of individual genes might shed more light into the biosynthesis of LOSs and the null mutant of MmpL12 would not only show its role in transport but also would help in understanding the location of different stages of LOS biosynthesis.

Deletion of *MKAN27435* did not affect the ability of the mutant to survive in activated macrophages nor did it induce or suppress the release of TNF-α. Earlier studies with purified LOS-IV showed suppression of TNF-α release from LPS activated macrophages (Rombouts *et al*. 2009). In a recent study an inverse relationship was established between phagocytosis and the LOS production in *M. marinum* (Alibaud *et al*. 2014). It would be interesting to test phagocytosis of the mutant generated in this study. Also intercepting LOS production at earlier stages would be interesting in terms of understanding its role in antigenicity, since more drastic changes might needed to see these effects.

In conclusion, the study establishes the non-essentiality of *fabH* and suggests possible functional substitute in the form of *pks18*. Biophysical characterisation of MmpL3 revealed its oligomeric status and preliminary Cryo-EM data validates model predictions. The conditions for production of sufficient quantities of protein are now optimised, which was a bottleneck in the structural studies. The study also shows a clear role for MKAN27435, a glycosyl transferase in the biosynthesis of LOSs in *M. kansasii*.

Chapter 7

General materials and methods

7.1 Media preparations

7.1.1 Luria-Bertani (LB) broth

LB broth was routinely prepared by dissolving 25 g of LB (Tryptone-Yeast extract-NaCl 2:1:2 w/w/w) (Sigma-Aldrich) in 1 litre of distilled water and sterilised by autoclaving.

7.1.2 LB agar

LB agar was routinely prepared by dissolving 37 g of LB agar mix (Tryptone-Yeast extract-NaCl-Agar 2:1:2:3 w/w/w/w) (Sigma-Aldrich) was dissolved in 1 L of distilled water and autoclaved. Alternatively, 15 g of bacto agar (BD) was added to 1 L LB broth and autoclaved. Antibiotics were added where required, to molten agar after cooling it down to approximately 55°C, mixed thoroughly before pouring into petri dishes (25 ml) aseptically.

7.1.3 Tryptic Soy Broth (TSB)

Tryptic soy broth was prepared by dissolving 30 g of TSB mix (Pancreatic casein digest-Papaic soybean digest-dextrose-NaCl-K_2HPO_4 34:6:5:10:5 w/w/w/w/w) (Sigma Aldrich) in 1 L of distilled water and sterilised by autoclaving.

7.1.4 Tryptic Soy Agar (TSA)

TSA was prepared by adding 15 g of agar in 1 L of TSB and sterilised by autoclaving. Antibiotics were added where required, to molten agar after cooling it down to approximately 55°C, mixed thoroughly before pouring into petri dishes (25 ml) aseptically.

7.1.5 Middlebrooks 7H9 broth

7H9 broth was prepared by dissolving 4.7 g of the Middlebrook 7H9 media (BD) in 900 ml of distilled water. 2 ml of glycerol was added to this and was sterilised by either filter sterilisation or autoclaving. 7H9 broth was supplemented with 10% ADC or OADC (BD) as required.

7.1.6 Middlebrooks 7H10 agar

7H10 agar was prepared by dissolving 19 g of the 7H10 Middlebrook media (BD) in 900 ml distilled water containing 5 ml glycerol and was sterilised by autoclaving. 100 ml of OADC (BD) and antibiotics were added to molten agar after cooling it down to approximately 55°C, mixed thoroughly before pouring into petri dishes (25 ml) aseptically.

7.1.7 Middlebrooks 7H11 agar

7H11 agar was prepared by dissolving 4.7 g of 7H11 Middlebrook media (BD) in 900 ml distilled water containing 5 ml glycerol and was sterilised by autoclaving. 100 ml of OADC (BD) and antibiotics were added to molten agar after cooling it down to approximately 55°C, mixed thoroughly before pouring into petri dishes (25 ml) aseptically.

7.1.8 7H9 soft agar

7H9 soft agar or top agar was prepared by dissolving 4.7 g of 7H9 Middlebrook media in 1 L of distilled water containing 2 ml glycerol. Bacto agar (BD) was added to a final concentration of 0.75% (7.5 g/L) and wassterilised by autoclaving.

7.1.9 MacConkey agar

MacConkey agar was prepared by dissolving 40 g of MacConkey agar base media (Difco) in 1 L distilled water and autoclaved. Glucose free Maltose (10%) was filter sterilised and added to the cooled agar to a final concentration of 1% along with other antibiotics and IPTG.

7.1.10 Antibiotics and supplements

Table 7-1: List of antibiotics and supplements

Additives	Stock concentration	Storage
Hygromycin B	50 mg/ml	4°C, protected from light
Kanamycin	50 mg/ml	-20°C, filter sterilised
Ampicillin	100 mg/ml	-20°C, filter sterilised
Apramycin	50 mg/ml	-20°C, filter sterilised
Oleic acid-albumin-dextrose-catalase (OADC)	10%	4°C,
Albumin-dextrose-catalase (ADC)	10%	4°C,
Tween80	10%	Room temperature, protect from light
IPTG	1M	-20°C, filter sterilised
X-gal	40mg/ml	-20°C, filter sterilised
Maltose	20%	Room temperature
Acetamide	20%	Room temperature

7.2 Molecular biology techniques

7.2.1 DNA electrophoresis

Agarose gel electrophoresis was routinely used for analysis of PCR products, restriction digestion of plasmid constructs and genomic DNA. Different percentage of agarose gels (0.7% to 1.0%) were prepared by dissolving/melting molecular biology grade agarose (Bioline) in 1X-Tris acetate EDTA (TAE) buffer. DNA samples were separated under a horizontal electric field (100-140 V, 400mA). DNA was visualised by staining the gel with ethidium bromide and viewing under UV light (Bio-Rad Gel Doc systems).

7.2.2 Polymerase chain Reaction (PCR)

Table 7-2: PCR mix using Phusion polymerase

Component	Volume (µl)	Stock concentration	Final concentration
Forward Primer	1	10 µM	0.5 µM
Reverse Primer	1	10 µM	0.5 µM
Template DNA	1	10 ng/µl	10 ng
dNTP	0.8	2 mM	200 µM
Polymerase	1	100Units	1 Unit
GC buffer	20	5X	1X
DMSO	5	100%	5%
$MgCl_2$	2	50 mM	1 mM
Milli Q water	Made up to total of 100 µl		

7.2.3 Restriction digestion of DNA

20 µl restriction digestion reactions were set up for either single (using one restriction enzyme) or double (using more than one restriction enzyme, usually two) digestion. 5 µl (~ 1 µg) of purified DNA (plasmid DNA/genomic DNA/PCR product) was mixed with 2 µl of 10X digest buffer, 1 µl (~5-10 units) of restriction enzyme and made up to 20 µl with nuclease free water. Reaction mix was supplemented with 2 µl of 10X BSA (Bovine Serum Albumin) wherever required. Reaction mix was incubated at 37°C (or as required for specific enzymes) for 30-60 min. Digested DNA was purified using Qiagen PCR purification kit and used for ligation.

7.2.4 Ligation

Digested DNA fragments with compatible ends were ligated using T4 DNA ligase (NEB). 10 µl reaction mix containing vector and insert fragments at required concentration, 1 µl (5-10 units) of T4 DNA ligase enzyme, 2 µl of 10X T4 DNA ligase reaction buffer and nuclease free water to make up the volume, were incubated at room temparature for 30 min or 4°C overnight.

7.2.5 Preparation of chemically competent *E. coli* cells

Required strain of *E. coli* was streaked onto LB agar to get isolated colonies. 5 ml of LB media was inoculated with a single colony of *E. coli* to prepare overnight cultures. The overnight culture was then used to inoculate 100 ml LB medium containing 20 mM $MgSO_4$. Cells were grown to an $O.D_{600nm}$ of 0.4-0.6. The cells are then harvested by centrifugation at 4500xg for 10 min at 4°C. Cell pellets were gently washed with 25 ml ice cold TFB1. The resuspended cells were incubated for 15 min on ice and centrifuged at 4500xg for 10min at 4°C. The cells were then resuspended in 8 ml ice cold TFB2. Cells were incubated on ice 15-60 min and then aliquoted and flash frozen. The aliquots are stored at -70°C for further use.

- **TFB1**

  30 mM Potassium acetate

  10 mM $CaCl_2$

  50 mM$MnCl_2$

  100 mM $RbCl_2$

  15% glycerol

  Filter sterilised and stored at 4°C

- **TFB2**

  10 mM MOPS or PIPES pH 6.5

  75 mM $CaCl_2$

  10 mM $RbCl_2$

  15% glycerol

  Filter sterilised and stored at 4°C

Table 7-3: Strains of *E. coli* used for different purpose

Strain	Purpose
Top10	Cloning
C41(DE3)	Protein expression
BL21(DE3)	Protein expression
BL26 (DE3) (Tuner)	Protein expression

7.2.6 Transformation of *E. coli* competent cells

Competent cells of *E. coli* strain to be transformed were thawed on ice. The ligation mix (5 µl) or plasmid (100 ng) was added to the cells, gently mixed and incubated on ice for 30 min and transformed by heat shock at 42°C for 45 seconds. Transformed cells are cooled on ice before adding 1 ml of LB broth for recovery. Transformation mix was incubated at 37°C for

about 45 min with shaking/ standing. At the end of 45 min recovery period cells were plated onto LB agar plates containing appropriate selection markers.

7.2.7 Plasmid extraction

Plasmid constructs were routinely extracted using Qiagen plasmid purification kits according to manufacturer's recommendation.

7.3 Generation of knock out phage for null mutant creation using Specialised Transduction

Primers were designed to amplify approximately 1kb sequences of the upstream and downstream of the target gene (gene*X*). Using this set of primers, the upstream and downstream regions of the gene was amplified from *M. smegmatis, M. kansasii, M. canetti or M. bovis* BCG. The PCR products were purified using Quiagen pcr clean up kit and the primer incorporated *Van91I/DraIII/BstAPI/AlwNI* sites were digested with right choice of enzyme. Digested PCR fragments were cloned into *Van91I*-digested (all enzymes are mutually compatible for ligation) p0004s (or p0004S-Apra version) to generate allelic exchange plasmid pΔgene*X*. The allelic exchange plasmid pΔgene*X* was sequenced to confirm the presence of right DNA fragments.

pΔgene*X* and phAE159 DNA were digested with the *PacI* and ligated. Ligation mix was then packaged into empty λ-phage heads and transduced into *E. coli* HB101 competent cells. Cells containing phasmid DNA were selected for on LB agar containing hygromycin or apramycin at 37°C. Presence of allelic exchange plasmid on phage DNA was confirmed by *PacI* digestion. The positive phasmids were electroporated into *M. smegmatis* mc^2155 electro-competent cells at 1800V and recovered at 30°C for ~ 4 hrs in TSB. The recovered cells were then harvested by centrifuging at 10000xg for 10 min and resuspended in 200 ml of MP buffer. This was mixed with 200 µl of freshly growing *M. smegmatis* mc^2155 cells and

5 ml molten 7H9 soft agar (50°C) and overlaid onto 7H10 agar plates. The plates were incubated at 30°C for 2-3 days to allow formation of plaques. The plates were soaked in minimum amount (2.5 ml to 3 ml) of MP buffer for 5-6 hrs to extract phages into solution and the solution containing phages was filtered and stored at 4°C.

Specialised transduction of *M. kansasii, M. canetti, M. smegmatis* and *M. bovis* BCG was performed as described previously (Bardarov et al. 2002). The cultures were grown in 50 ml of 7H9 + 10% OADC (or TSB for *M. smegmatis*) with 0.05% Tween 80 to an OD_{600} of ~0.8 and harvested by centrifugation at 4500g for 10 min. The cell pellet was washed twice with 50 ml of MP buffer and resuspended in 2 ml of MP buffer and high titre (10^{10} pfu/ml) phage lysate (2 ml) was mixed with the cells. A control was set up where 0.5 ml of resuspended cells was mixed with 0.5 ml of MP buffer. The mix was incubated overnight at 37°C followed by harvesting and recovered with 10 ml of 7H9 + 10% OADC with Tween-50 at 37°C overnight (4-5 hrs for *M. smegmatis*). This was plated onto 7H10 + 10% OADC-agar plates with hygromycin (or apramycin as required) and plates were incubated at 37°C for 1-4 weeks depending on the strain of mycobacteria. Hygromycin resistant colonies obtained after transduction were inoculated in 10 ml 7H9 + 10% OADC- Tween 80 with hygromycin for genomic DNA extraction and further characterisation. Allelic exchange of gene*X* with a hygromycin resistance cassette in hygromycin resistant transductants was confirmed by Southern blot. One such transductant was chosen for subsequent experiments.

7.3.1 Genomic DNA extraction

The cells were grown by inoculating a single colony in 10 ml of appropriate broth and growing them to an OD_{600nm} of ~ 1.0. The cells were harvested by centrifugation at 4500g for 10 min at room temperature. The cells are resuspended in 500 µl GTE buffer containing 10 mg/ml lysozyme and incubated at 37°C overnight for complete lysis of the cells. 100 µl of

10% SDS and 50 µl of 10 mg/ml Proteinase K is added to the mix and incubated at 55°C for 3-4 hrs. While SDS aids in efficient lysis of cells, Proteinase K, a broad spectrum serine protease digests the proteins. To this mix, 200 µl of 5 M NaCl is added. The protein and cell debris is then removed by using Chloroform and Chloroform: Isoamyl alcohol (24:1) mix. The DNA is precipitated using Isopropyl alcohol and washed with 70% ethanol. DNA is air dried to get rid of organic solvents and dissolved in distilled water.

7.3.2 Southern blot

The genomic DNA is digested with appropriate restriction enzyme such that wild type and the mutant would give different restriction patterns. The digested DNA was run on 0.7% agarose gel. The gel was then depurinated using 0.25 M HCl. The depurinated gel is then treated with alkali solution (1.5 M NaCl, 0.5 M NaOH) for denaturation. The alkali is then neutralised with 0.5 M Tris-HCl pH 7.2 buffer containing 1 M NaCl. The DNA is then transferred onto a nylon membrane (Nylon membrane, positively charged, No. 11209299001, Roche) by capillary transfer using high salt solution (20X SSC, 3 M NaCl, 0.3 M Sodium citrate, pH 7.0), overnight. The transferred DNA was then cross linked to membrane under UV light. The cross linked membrane was further used for hybridisation, labeling and detection which was performed as described in DIG High Prime Labeling and Detection Starter Kit II (cat. No.11555614910, Roche). The kit uses digoxigenin, a steroid to label DNA probes by random priming. The hybridised probes are then immunodetected using an antibody against digoxigenin which is tagged with Alkaline phosphatase that aids in visualisation by chemiluminiscence.

7.3.3 Preparation of mycobacterial electrocompetent cells

Mycobacterial strain was grown to an OD_{600} of 0.6 and harvested by centrifugation at 4500g for 20 min. The cells were washed twice with sterile 10% ice cold glycerol by centrifugation

at 4500g for 20 min. Washed cells were resuspended in 1/10th volume of sterile 10% ice cold glycerol and aliquoted. The Aliquots were frozen at -70°C.

7.3.4 Electroporation of mycobacteria

Electro-competent cells were thawed on ice and ~1 µg (max. 10 µl) of plasmid was added. The mix was placed into a 1mm electroporation cuvette and incubated on ice for further 15 min. Electroporation was done using Eppendorf Electroporator (model No. 2510) at 1800V. The electroporated cells were recovered in 1 ml 7H9 (TSB for *M. smegmatis*) containing 0.05% Tween 80 at 37°C for one generation time (4 hrs for *M. smegmatis,* 24 hrs for *M. kansasii, M. canetti* and *M.bovis* BCG). At the end of the recovery period, cells were harvested and resuspended in 200 µl of recovery media and plated on to TSA (*M. smegmatis*) or 7H10 or 7H11 (for *M. kansasii, M. canetti* and *M. bovis* BCG) with appropriate selection marker.

7.3.5 Radioactive labelling of lipids

10 ml of mycobacterial culture was grown to mid log phase and was then labelled with 50 µCi/ml of [^{14}C]-acetate (50 mCi/ml, Amersham Pharmacia Biotech). The labelled culture was incubated for 12 hrs (*M. smegmatis*) or 24 hrs (*M. kansasii, M. canetti*and *M. bovis* BCG). The labelled cells were harvested by centrifugation at 4500g for 10 -20 min at RT and washed with PBS before freeze drying.

7.4 Lipid extraction

Polar and apolar lipids were extracted from harvested cells (cold lipids) or cells labelled with [^{14}C]-acetate, using methods described by Dobson et al., (1985). Cells/ labelled cells were resuspended in 2 ml of CH_3OH/0.3% NaCl (100:10, v/v). Apolar lipids were extracted twice with 2 ml of petroleum ether (b.p. 60-80^{0}C). 2 ml of petroleum ether was added the cells resuspended in mathanolic saline and mixed on a rotator for 30 min. The mixture was

centrifuged at 4000g for 5 min to for two layers. The upper layer was removed and stored in a separate tube labelled apolar lipid. 2 ml of petroleum ether was added to lower fraction for complete extraction of apolar lipids. The mixture was allowed to mix on a rotator for 30 min before it was centrifuged again to collect the non polar lipids from upper layer. The combined upper layers' containing apolar lipids was dried under liquid nitrogen and was used for analysis (Besra 1998).

Polar lipids were extracted from the lower fraction by adding 2.3 ml of $CHCl_3/CH_3OH$/0.3% NaCl (90:100:30, v/v/v). The mixture was mixed on a rotator for 1 hr and centrifuged at 4000g for 5 min. The supernatant was removed and stored in a fresh tube. The remaining pellet was extracted twice with 0.75 ml of $CHCl_3/CH_3OH$/0.3% NaCl (50:100:40, v/v/v) and mixing for 30 min each time before centrifugation and removal of supernatant. 1.3 ml of $CHCl_3$ and 1.3 ml of 0.3% NaCl was added to the pooled fraction and mixed for 10 min, centrifuged and the lower layer recovered and dried under Nitrogen for polar lipids. The dried apolar and polar lipids were dissolved in 200 µl of $CHCl_3/CH_3OH$ (2:1, v/v); and 5 µl dried in a scintillation vial and then mixed with 5 ml scintillation fluid and radioactivity incorporated was measured in terms of counts per minutes. The defatted cells were used for further analysis (Besra 1998).

7.4.1 Fatty acid methyl esters (FAMEs) and Mycolic acid methyl esters (MAMEs) extraction from defatted cells and whole cells

Alkaline hydrolysis was performed on de-fatted cells using 2 ml of 5% aqueous tetrabutylammonium hydroxide at 100°C overnight. CH_2Cl_2 (4ml), CH_3I (300 µl) and water (2 ml) was added to the reaction mixture and was mixed on a rotator for 30 min. The mixture was then centrifuged at 4000g for 5 min and the upper aqueous layer was discarded. The lower organic layer was washed 3 times with water and evaporated to dryness. MAMEs were

re-dissolved in diethyl ether, sonicated and vortexed and centrifuged. The organic layer was taken into separate tube and dried under air. The samples were dissolved in either, CH_3Cl: CH_3OH, 2:1 v/v or CH_2Cl_2 (200 µl) and 5 µl was used for counting of labelled samples.

7.4.2 Thin layer Chromatography (TLC) analysis of lipids

Equivalent amounts of cold or labelled lipids (10 µg or 25,000 to 50,000 cpm) of each sample were spotted on TLC plates (5554 silica gel 60F524; Merck). TLC plates were developed using one of the following Systems mentioned below. Cold lipids were visualised by spraying MPA spray, α-naphthol spray or phosphate stain. [^{14}C]- labelled lipids were revealed by 24-72 hrs exposure to Kodak X-Omat AR film.

7.4.2.1 Solvent systems for 2D TLC analysis

Apolar lipids were separated by system A, B, C and D. Polar lipids were separated using systems D and E as described by Dobson *et al*. (1985). TLC plates should be air dried well before each solvent run and after spotting the samples.

- **System A**

  Direction 1

  Thrice with petroleum ether (b.p.60-80^{0}C)/ethyl acetate - 95:2 (v/v)

  Direction 2

  Once with petroleum ether (b.p.60-80^{0}C)/acetone - 92:8 (v/v)

- **System B**

  Direction 1

  Thrice with petroleum ether (b.p.60-80^{0}C)/ acetone – 92:8 (v/v)

  Direction 2

  Once with toluene/acetone – 95:5 (v/v)

- **System C**

  Direction 1

  Once with Chloroform/methanol –96:4 (v/v)

  Direction 2

  Once with toluene/acetone – 80:20 (v/v)

- **System D**

  Direction 1

  Once with Chloroform/methanol/water – 100:14:0.8 (v/v/v)

  Direction 2

  Once with Chloroform/acetone/ethanol/water – 50:60:2.5:3 (v/v/v/v)

- **System E**

  Direction 1

  Once with Chloroform/methanol/water – 60:30:6 (v/v/v)

  Direction 2

  Once with chloroform/acetic acid/methanol/water – 40:25:3:6 (v/v/v/v)

Note: TLC plates were activated by baking at 100°C for 1 hr before application of samples for running in system E. One hour drying between the runs is also required for system E.

7.4.2.2 TLC analysis for Fatty Acid Methyl Esters (FAMEs) and Mycolic Acid Methyl Esters (MAMEs)

The FAMEs/MAMEs mixture (~ 25,000 cpm) were analysed by developing the TLC using petroleum ether (b.p.60-80°C): acetone – 95:5 (v/v). [^{14}C]-labelled FAMEs/ MAMEs were revealed by overnight exposure to Kodak X-Omat AR film.

7.5 Techniques in protein chemistry

7.5.1 Checking the expression of protein expressed in *E. coli*

Cells of *E. coli* expression strain were transformed with plasmid harbouring gene of interest. Overnight cultures of transformants were prepared by inoculating into 5 ml LB broth containing appropriate selection marker with transformants. Overnight cultures were used to inoculate 10 ml of LB broth containing appropriate selection marker and incubated at 37°C (or 30°C or 25°C or 16°C where required) with shaking. The cultures were induced for protein expression with IPTG (at required concentration) at OD_{600} ~0.5 and continued to incubated for required time (3hrs to O/N as required). Cells were harvested at the end of induction period by centrifugation at 4500g for 10-20 min. The cells were washed with an equal volume of PBS. The cells were then resuspended in 300-400 µl of ice cold protein extraction buffer (50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.6% SDS, 10 mM Sodium phosphate, protease inhibitor cocktail tablet 1 in 50 ml) and mixed with equal volume of 2X SDS gel loading dye. The mix was boiled for 10 min to break open the cells and release the proteins and spun at high speed on a microfuge to pellet down the cell debris. The samples are loaded on to SDS-PAGE gel to visualise the bands.

7.5.2 Scaling up protein expression

Overnight cultures of required strain harbouring expression construct was used to inoculate up to 12 L of LB broth to an initial OD_{600} of 0.1 and incubated at 37°C (or as required for the experiment) with vigorous shaking till OD_{600} reached 0.5. Protein expression was induced using 1 mM [IPTG] (or as required by the experiment) and further incubated for another 3 hrs (or O/N at lower temperatures). Cells were harvested at the end of induction period and washed with PBS. Cell pellet was either used immediately or frozen at -70°C for later use.

7.5.3 SDS PAGE

Precast protein gels (any Kda, Biorad) or 8-12% SDS-PAGE gels were used for the analysis of proteins. SDS-gel sample buffer was added to samples and boiled for 10 min. The samples were then spun for 5 min at maximum speed in a microfuge and loaded on to gel. The gel is run using Tris-glycine-SDS buffer with a current of 35- 45 mA and potential difference 200V. The gel is stained using either instant blue or coomassie brilliant blue stain.

7.5.4 Affinity chromatography

His trap HP column (Amersham Biosciences) was used for the purification of His-tagged proteins. The column has highly cross linked agarose beads on which a chelating group is immobilised. The Ni^{2+} is coupled to the chelating matrix. The column is washed with 5-10 column volume of distilled water and equilibrated with 5 column volumes of 50 mM Potassium Phosphate buffer pH 7.9, containing 500 mM NaCl. The cells are lysed using a sonicator (30s on/30s off, 10 cycles). The sonicated lysate is then spun at 15,000g for 45 min. The supernatant is loaded on to the column. The column is washed with different concentration of imidazole (10 mM to 1 M) to wash the contaminant proteins and elute the protein of interest. The fractions are collected and analysed on SDS-PAGE gel.

7.5.5 Ion exchange chromatography

QHP ion exchange column (Amersham biosciences) was used for purification of proteins. The column is washed with 5-10 column volume of distilled water. The column is then equilibrated with 20 mM Tris-HCl, pH 8.0. Cell lysate or partially purified protein from affinity chromatography is then loaded on to the column. The column is washed with increasing concentration of NaCl (20 mM to 1 M) to wash contaminant proteins and elute the protein of interest. The fractions are collected and analysed on SDS-PAGE gel.

7.5.6 Gel filtration chromatography

Gel filtration chromatography was used as a preparative tool to further purify proteins and analytical tool to get insight into oligomeric status of the protein. Partially purified protein from earlier purification methods was concentrated to ~ 1 mg/ml and loaded to Superdex-200 10/300GL (GE healthcare) column equilibrated with 50 mM Tris.Cl, 150 mM NaCl. Protein was eluted by washing the column with elution buffer (50 mM Tris.Cl, 150 mM NaCl). 0.5 ml (or 2 ml) fractions were collected and analysed by SDS-PAGE analysis.

7.5.7 Purification of membrane protein using SMALP

Cells expressing membrane protein was harvested and washed with PBS and resuspended in Buffer-1. Cell lysis was done using French press (10,000psi 4 cycles). Cell debris was removed by centrifugation at 10,000g for 30min. The membrane fraction was then separated from cell lysate by centrifugation at 100,000g for 45min. Membrane thus separated was weighed and resuspended in Buffer-2 to a final concentration of 40 mg/ml and homogenised. The homogenised membrane was then solubilised using SMA [poly (styrene-*co*- maleic acid)] (SMA2000-P). The mix was incubated at room temparature for two hours on a rocker to facilitate solubilisation and to generate SMA-solubilised protein (SMA-lipid particle encapsulating protein). SMA-solubilised membrane fraction was centrifuged at 100,000g for 45 min to remove insoluble membrane particles. The supernatant consisting of SMA-solubilised protein was then incubated with affinity matrix (Ni^{2+}-NTA) at 4°C for O/N to facilitate binding and further purified as described earlier but with buffers containing 10% glycerol.

Composition of buffers:

- **Buffer 1:**

  50 mM Tris.Cl pH 8.0

  2 mM EDTA

  5% Glycerol

  1 mM PMSF

  Protease inhibitor cocktail 1 tablet/50 ml

- **Buffer 2**

  50 mM Tris.Cl pH8.0

  500 mM NaCl

  10% Glycerol

  1 mM PMSF

  Protease inhibitor cocktail 1 tablet/50 ml

7.5.8 Western blot analysis

Protein samples run on SDS-PAGE were transferred onto PVDF membrane (Pierce, Thermo scientific) using transfer buffer (25 mM Tris, 190 mM Glycine and 20% methanol) (30V and 300 mA, 3 hrs) at cold. Membrane was incubated in blocking solution (1% BSA or 3% Milk powder) to prevent non-specific binding of antibodies to the membrane, for 1 hr at RT or O/N at 4°C. The membrane was thoroughly washed with TBS containing 0.1% Tween 20 three times and incubated with required dilution of primary antibody (anti-His or anti-MBP). After one hour of incubation, membrane was again washed with TBS containing Tween 20.

Membrane blot was later incubated with alkaline phosphatase coupled secondary antibody for one hour. Membrane blot was washed before developing signal to visualise the proteins using chromogenic substrate BCIP/NBT (Sigma Aldrich) solution.

7.5.9 Analytical ultracentrifugation

Analytical ultracentrifugation was routinely used to evaluate the oligomeric status of the protein. Protein samples were prepared by adjusting the concentration of the purified protein to ~ 0.5 to 1.0 mg/ml (either by concentrating or diluting). Ultracentrifugation was performed using sedimentation velocity protocol at 20000g (or 30000g where required) for 16 hrs at 20°C in a Beckman Coulter XL ultracentrifuge. Absorbance data collected was analysed using Sedfit program. Parameters required for data analysis was deduced from protein sequence and buffer composition using Sednterp program.

7.5.10 Protein-protein interaction studies

Genes of interest were cloned into pKT25 (T25 fusion tag) or pUT18 (T18 fusion tag) using standard DNA manipulation protocols (Karimova et al. 1998). Two plasmids, expressing bait and prey proteins as fusion tags of T25 fragment and T18 fragment respectively were co-transformed into *E. coli cya*$^{-}$ strain BTH101. Cells were spread on LB plates containing streptomycin (100 µg/ml), ampicillin (100 µg/ml) and kanamycin (50 µg/ml) and incubated at 30°C for 48 hrs. Overnight cultures were prepared by inoculating several colonies into 5 ml LB broth containing above mentioned antibiotics and 0.5 mM IPTG to induce protein expression and incubated at 30°C with shaking. The cultures were then washed three times in minimal M63 media and either spotted or spread on MacConkey agar plates supplemented with above mentioned antibiotics and nutrients.

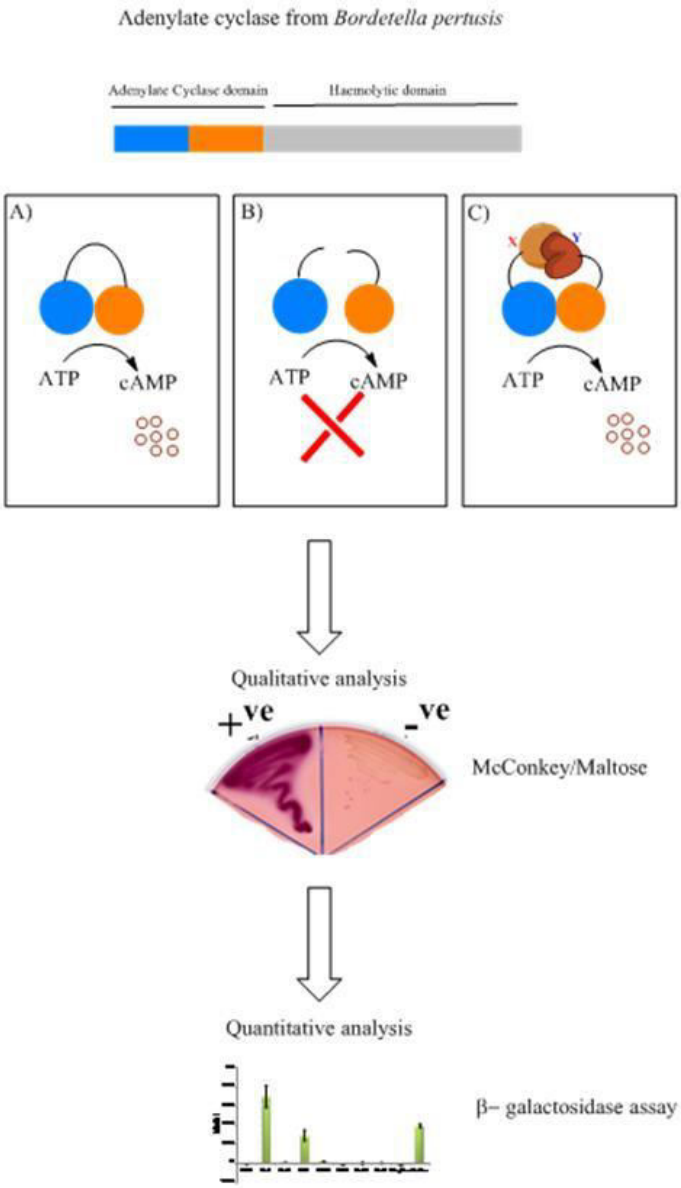


Figure 7-1 Schematic presentation of flow chart of steps involved in protein-protein interaction studies. Domain organisation of *Bordetella pertusis* adenylate cyclase showing T25 fragment (in blue) and T18 fragment (in orange). (A) Expression of chimeric adenylate cyclase alone in *E. coli* cya$^-$ strain BTH101 leads to residual cAMP synthesis (B) when the individual fragments are physically separated its activity is lost leading to cessation of cAMP synthesis (C) Individual fragments brought in physical proximity by interacting protein restores cAMP synthesis. Bait and prey proteins are cloned with T25 and T18 fragments of adenylate cyclase domain of *Bordetella pertusis* adenylate cyclase. Plasmids harbouring genes for expression of bait and prey proteins were co-transformed into *E. coli* cya$^-$ strain BTH101. Screening for interactions was done by plating the co-transformants on McConkey agar containing kanamycin and Ampicillin. Further, quantitative β-galactosidase assay is used to characterise the extent of protein-protein interaction.

References

Acharya, P.V.N. & Goldman, D.S., 1970. Chemical composition of the cell wall of the H37Ra strain of Mycobacterium tuberculosis. *J Bacteriol*, 102(3), pp.733–39.

Albouy, R. et al., 1955. First clinical results obtained with cycloserine in the treatment of human tuberculosis. *Antibiot annu*, 3, pp.148–52.

Alderwick, L.J. et al., 2006. Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis. *J Biol Chem*, 281(23), pp.15653–61.

Alderwick, L.J. et al., 2007. Structure, function and biosynthesis of the Mycobacterium tuberculosis cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets. *Biochem Soc Trans*, 35(Pt 5), pp.1325–8.

Alexander, D.C. et al., 2004. PimF, a mannosyltransferase of mycobacteria, is involved in the biosynthesis of phosphatidylinositol mannosides and lipoarabinomannan. *J Biol Chem*, 279(18), pp.18824–33.

Alibaud, L. et al., 2014. Increased phagocytosis of Mycobacterium marinum mutants defective in lipooligosaccharide production: a structure-activity relationship study. *J Biol Chem*, 289(1), pp.215–28.

Andries, K. et al., 2005. A Diarylquinoline Drug Active on the ATP Synthase of Mycobacterium tuberculosis. *Science*, 307(5707), pp.223–7.

Anon, 1997. Diagnosis and treatment of disease caused by nontuberculous mycobacteria. This official statement of the American Thoracic Society was approved by the Board of Directors, March 1997. *Am J Respir Crit Care Med*, 156(2Pt2), pp.S1–25.

Asselineau, J. & Laneelle, G., 1998. Mycobacterial lipids: a historical perspective. *Front Biosci*, 3, pp.164–74.

Asselineau, J. & Lederer, E., 1950. Structure of mycolic acids from Mycobacteria. *Nature*, 166(4227), pp.782–3.

Austin, M.B. & Noel, J.P., 2003. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep*, 20(1), pp.79–110.

Azuma, I. et al., 1970. Occurrence of N-glycolylmuramic acid in bacterial cell walls. A preliminary survey. *Biochim Biophys Acta*, 208(3), pp.444–51.

Azuma, I., Yamamura, Y. & Fukushi, K., 1968. Fractionation of mycobacterial cell wall. Isolation of arabinose mycolate and arabinogalactan from cell wall fraction of Mycobacterium tuberculosis strain Aoyama B. *J Bacteriol*, 96(5), pp.1885–87.

Ballou, C.E., Vilkas, E. & Lederer, E., 1963. Structural Studies on the Myo-inositol Structural Studies on the Myo-inositol Phospholipids of Mycobacterium tuberculosis (var. bovis, strain BCG). *J Biol Chem*, 238, pp.69–76.

Banerjee, A. et al., 1994. inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science.* 263(5144), pp.227–30.

Banerjee, A. et al., 1998. The mabA gene from the inhA operon of Mycobacterium tuberculosis encodes a 3-ketoacyl reductase that fails to confer isoniazid resistance. *Microbiology,* 144(Pt 10), pp.2697–704.

Baneyx, F. & Mujacic, M., 2004. Recombinant protein folding and misfolding in Escherichia coli. *Nature biotechnol*, 22(11), pp.1399–408.

Bardarov, S. et al., 1997. Conditionally replicating mycobacteriophages: a system for transposon delivery to Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, 94(20), pp.10961–6.

Bardarov, S. et al., 2002. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and *M. smegmatis*. *Microbiology*, 148(Pt 10), pp.3007–17.

Barry, C.E. et al., 1998. Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res*, 37(2-3), pp.143–79.

Batt, S.M. et al., 2012. Structural basis of inhibition of Mycobacterium tuberculosis DprE1 by benzothiazinone inhibitors. *Proc Natl Acad Sci U S A*, 109(28), pp.11354–9.

Baulard, a R. et al., 2000. Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem*, 275(36), pp.28326–31.

Belanger, a E. et al., 1996. The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc Natl Acad Sci U S A*, 93(21), pp.11919–24.

Belánová, M. et al., 2008. Galactosyl transferases in mycobacterial cell wall synthesis. *J Bacteriol*, 190(3), pp.1141–5.

Belisle, J.T. et al., 1997. Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. *Science.* 276(5317), pp.1420–2.

Belisle, J.T. & Brennan, P.J., 1989. Chemical basis of rough and smooth variation in mycobacteria. *J Bacteriol*, 171(6), pp.3465–70.

Bernheim, F., 1941. The effect of various substances on the oxygen uptake of the tubercle bacillus. *J Bacteriol*, 41(3), pp.387–395.

Besra, G.S. et al., 1995. A new interpretation of the structure of the mycolyl-arabinogalactan complex of Mycobacterium tuberculosis as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and 1H nuclear magnetic resonance . *Biochemistry*, 34(13), pp.4257–66.

Besra, G.S. et al., 1994. Identification of the apparent carrier in mycolic acid synthesis. *Proc Natl Acad Sci U S A*, 91(26), pp.12735–12739.

Besra, G.S., 1998. Preparation of cell-wall fractions from mycobacteria. *Methods Mol Biol,* 101, pp.91–107.

Besra, G.S. et al., 1993. Trehalose-containing lipooligosaccharides of Mycobacterium gordonae: presence of a mono-O-methyltetra-O-acyltrehalose "core" and branching in the oligosaccharide backbone. *Biochemistry*, 32(47), pp.12705–14.

Besra, G.S. & Brennan, P.J., 1997. The mycobacterial cell wall: biosynthesis of arabinogalactan and lipoarabinomannan. *Biochem Soc Trans*, 25(3), pp.845–50.

Bhatt, A. et al., 2005. Conditional Depletion of KasA , a Key Enzyme of Mycolic Acid Biosynthesis , Leads to Mycobacterial Cell Lysis. *J Bacteriol*, 187(22), pp.7596–606.

Bhatt, A., Fujiwara, N., et al., 2007. Deletion of kasB in Mycobacterium tuberculosis causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci U S A*, 104(12), pp.5157–62.

Bhatt, A. et al., 2008. Loss of a mycobacterial gene encoding a reductase leads to an altered cell wall containing beta-oxo-mycolic acid analogs and accumulation of ketones. *Chem Biol*, 15(9), pp.930–9.

Bhatt, A., Molle, V., et al., 2007. The Mycobacterium tuberculosis FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol Microbiol*, 64(6), pp.1442–54.

Bhatt, K. et al., 2007. Two polyketide-synthase-associated acyltransferases are required for sulfolipid biosynthesis in Mycobacterium tuberculosis. *Microbiology*, 153(Pt 2), pp.513–20.

Birch, H.L. et al., 2008. Biosynthesis of mycobacterial arabinogalactan: identification of a novel alpha(1-->3) arabinofuranosyltransferase. *Mol Microbiol*, 69(5), pp.1191–206.

Bloch, K. & Vance, D., 1977. Control mechanisms in the synthesis of saturated fatty acids. *Annu Rev Biochem*, 46, pp.263–98.

Boehringer, D., Ban, N. & Leibundgut, M., 2013. 7.5-Å Cryo-Em Structure of the Mycobacterial Fatty Acid Synthase. *J Mol Biol*, 425(5), pp.841–9.

Brand, S. et al., 2003. Identification and functional analysis of six mycolyltransferase genes of Corynebacterium glutamicum ATCC 13032: the genes cop1, cmt1, and cmt2 can replace each other in the synthesis of trehalose dicorynomycolate, a component of the mycolic acid layer of the cell envelope. *Arch Microbiol*, 180(1), pp.33–44.

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

Brennan, P.J. & Ballou, C.E., 1967. Biosynthesis of mannophosphoinositides by Mycobacterium phlei. The family of dimannophosphoinositides. *J Biol Chem*, 242(13), pp.3046–56.

Brennan, P.J. & D.C. Crick, 2007. The cell-wall core of Mycobacterium tuberculosis in the context of drug discovery. *Curr Top Med Chem*, 7(5), pp.475–88.

Brennan, P.J. & Nikaido, H., 1995. The envelope of mycobacteria. *Ann Rev Biochem*, 64, pp.29–63.

Briken, V. et al., 2004. Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Mol Microbiol*, 53(2), pp.391–403.

Broussy, S. et al., 2003. 1H and 13C NMR characterization of hemiamidal isoniazid-NAD(H) adducts as possible inhibitors of InhA reductase of Mycobacterium tuberculosis. *Chemistry*, 9(9), pp.2034–8.

Brown, A.K. et al., 2005. Probing the mechanism of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III mtFabH: factors influencing catalysis and substrate specificity. *J Biol Chem*, 280(37), pp.32539–47.

Burguière, A. et al., 2005. LosA, a key glycosyltransferase involved in the biosynthesis of a novel family of glycosylated acyltrehalose lipooligosaccharides from Mycobacterium marinum. *J Biol Chem*, 280(51), pp.42124–33.

Burman, W.J., 2010. Rip Van Winkle wakes up: development of tuberculosis treatment in the 21st century. *Clin infect dis,* 50 Suppl 3, pp.S165–72.

C. Muschenheim et al., 1954. Pyrazinamide-isoniazid in tuberculosis. I. Results in 58 patients with pulmonary lesions one year after the start of therapy. *Am Rev Tuberc*, 70(4), pp.743–7.

Camacho, L.R. et al., 1999. Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. *Mol Microbiol*, 34(2), pp.257–67.

Chatterjee, D. et al., 1993. Structural definition of the non-reducing termini of mannose-capped LAM from Mycobacterium tuberculosis through selective enzymatic degradation and fast atom bombardment-mass spectrometry. *Glycobiology*, 3(5), pp.497–506.

Chatterjee, D. et al., 1991. Structural features of the arabinan component of the lipoarabinomannan of Mycobacterium tuberculosis. *J Biol Chem*, 266(15), pp.9652–60.

Chatterjee, D. & Khoo, K.H., 1998. Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects. *Glycobiology*, 8(2), pp.113–20.

Chen, J.M. et al., 2006. Roles of Lsr2 in Colony Morphology and Biofilm Formation of Mycobacterium smegmatis. *J Bacteriol*, 188(2), pp.633–41.

Chesne-Seck, M.L. et al., 2008. A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in Mycobacterium tuberculosis H37Ra. *J Bacteriol*, 190(4), pp.1329–34.

Choi, K.H. et al., 2000. Identification and substrate specificity of beta -ketoacyl (acyl carrier protein) synthase III (mtFabH) from Mycobacterium tuberculosis. *J Biol Chem*, 275(36), pp.28201–7.

Cole, S.T., Eiglmeier, K. & Parkhill, J., 2001. Massive gene decay in the leprosy bacillus. *Nature*, 409(6823), pp.1007–11.

Cole, S.T. et al., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393, pp.537–544.

Collins, F.M. & Cunningham, D.S., 1981. Systemic Mycobacterium kansasii infection and regulation of the alloantigenic response. *Infect Immun*, 32(2), pp.614–24.

Converse, S.E. et al., 2003. MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. *Proc Natl Acad Sci U S A*, 100(10), pp.6121–6.

Corbett, E.L. et al., 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med*, 163(9), pp.1009–21.

Corrales, R.M. et al., 2012. Phosphorylation of mycobacterial PcaA inhibits mycolic acid cyclopropanation: consequences for intracellular survival and for phagosome maturation block. *J Biol Chem*, 287(31), pp.26187–99.

Cox, G.L., 1923. Sanatorium treatment contrasted with home treatment. After-histories of 4,067 cases. *Br J Tubec*, 17, pp.27–30.

Cox, H. et al., 2006. Tuberculosis recurrence and mortality after successful treatment: impact of drug resistance. *PLoS Med*, 3(10), p.e384.

Cox, J.S. et al., 1999. Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature*, 402(6757), pp.79–83.

Crubézy, E. et al., 1998. Identification of Mycobacferium DNA in an Egyptian Pott ' s disease of 5400 years old. *C R Acad Sci III,* 321(11), pp.941–951.

Daffe, M. et al., 1991. Novel type-specific lipooligosaccharides from Mycobacterium tuberculosis. *Biochemistry*, 30(2), pp.378–88.

Daffe, M. & Draper, P., 1998. The envelope layers of mycobacteria with reference to their pathogenicity. *Adv.Microb.Physiol*, 39, pp.131–203.

Daffé, M. & Etienne, G., 1999. The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. *Tuber Lung Dis*, 79(3), pp.153–69.

Daffé, M. & Lanée, M.A., 2001. Analysis of the Capsule of Mycobacterium tuberculosis. *methods Mol Med*, 54, pp.217–27.

Daffe, M. & Laneelle, M.A., 1988. Distribution of Phthiocerol Diester , Phenolic Mycosides and Related Compounds in Mycobacteria. *J Gen Microbiol*, 134(7), pp.2049–2055.

Dannenberg, A.M.J. & Rook, G.A.W., 1994. Pathogenesis of pulmonary tuberculosis: an interplay of tissue-damaging and macrophage-activating immune responses—Dual mechanisms that control bacillary multiplication. *In Bloom B (ed), Tuberculosis, ASM press, Washington DC.*Ch27.

Dao, D., Kremer, L. & Guérardel, Y., 2004. Mycobacterium tuberculosis lipomannan induces apoptosis and interleukin-12 production in macrophages. *Infect Immun*, 72(4), pp.2067–74.

Dao, D.N. et al., 2008. Mycolic acid modification by the mmaA4 gene of M. tuberculosis modulates IL-12 production. *PLoS pathog*, 4(6), p.e1000081.

Davidson, L. Draper, P. & Minnikin, D.E., 1982. Studies on the mycolic acids from the walls of Mycobacterium microti. *J Gen Microbiol*, 128(4), pp.823–8.

Davies, G.E. & Stark, G.R., 1970. Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. *Proc Natl Acad Sci USA*, 66(3), pp.651–6.

Deidda, D. et al., 1998. Bactericidal Activities of the Pyrrole Derivative BM212 against Multidrug-Resistant and Intramacrophagic Mycobacterium tuberculosis Strains. *Antimicrob Agents Chemother*, 42(11), pp.3035–037.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

Dobson, G. et al., 1985. Systematic analysis of complex mycobacterial lipids. In M. Goodfellow & D. . Minnikin, eds. *In Chemical methods in bacterial systematics*. London Academic Press, pp. 237–65.

Domenech, P. et al., 2004. The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. *The J Biol Chem*, 279(20), pp.21257–65.

Domenech, P., Reed M.B. & Barry, C.E 3rd, 2005. Contribution of the Mycobacterium tuberculosis MmpL Protein Family to Virulence and Drug Resistance. *Infect Immun*, 73(6), pp.3492–501.

Donoghue, H.D., 2009. Human tuberculosis--an ancient disease, as elucidated by ancient microbial biomolecules. *Microbes and infection / Institut Pasteur*, 11(14-15), pp.1156–62.

Dover, L.G. et al., 2004. Comparative cell wall core biosynthesis in the mycolated pathogens, Mycobacterium tuberculosis and Corynebacterium diphtheriae. *FEMS Microbiol Rev*, 28(2), pp.225–50.

Dover, L.G. et al., 2004. Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. *J Mol Biol*, 340(5), pp.1095–105.

Dubnau, E. et al., 2000. Oxygenated mycolic acids are necessary for virulence of Mycobacterium tuberculosis in mice. *Mol Microbiol*, 36(3), pp.630–7.

Elkins, C.A. & Nikaido, H., 2002. Substrate Specificity of the RND-Type Multidrug Efflux Pumps AcrB and AcrD of Escherichia coli Is Determined Predominately by Two Large Periplasmic Loops. *J Bacteriol*, 184(23), pp.6490–8.

Emsley, P. et al., 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), pp.486–501.

Ernst, J.D., 1998. Macrophage Receptors for Mycobacterium tuberculosis. *Infect Immun*, 66(4), pp.1277–81.

Escuyer, V.E. et al., 2001. The role of the embA and embB gene products in the biosynthesis of the terminal hexaarabinofuranosyl motif of Mycobacterium smegmatis arabinogalactan. *J Bio Chem*, 276(52), pp.48854–62.

Espinal, M. a et al., 2000. Standard short-course chemotherapy for drug-resistant tuberculosis: treatment outcomes in 6 countries. *JAMA*, 283(19), pp.2537–45.

Fattorini, L. et al., 1999. Activity of 16 antimicrobial agents against drug-resistant strains of Mycobacterium tuberculosis. *Microb Drug Resis*, 5(4), pp.265–70.

Forbes, M., Kuck, N.A. & Peets, E.A., 1962. Mode of action of Ethambutol. *J Bacteriol*, 84, pp.1099–103.

Franceschini, A. et al., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*, 41(Database issue), pp.D808–15.

Frothingham, R. et al., 1996. Identification, cloning, and expression of the Escherichia coli pyrazinamidase and nicotinamidase gene, pncA. *Antimicrob Agents Chemother*, 40(6), pp.1426–31.

Gao, L.Y. et al., 2003. Requirement for kasB in Mycobacterium mycolic acid biosynthesis, cell wall impermeability and intracellular survival: implications for therapy. *Mol Microbiol*, 49(6), pp.1547–63.

Ghadbane, H. et al., 2007. Structure of Mycobacterium tuberculosis mtFabD, a malonyl-CoA:acyl carrier protein transacylase (MCAT). *Acta crystallogr Sect F Struct Biol Cryst Commun*, 63(Pt 10), pp.831–5.

Ghuysen, J.M. et al., 1968. Structure of the cell walls of Micrococcus lysodeikticus. 3. Isolation of a new peptide dimer, N-alpha-[L-alanyl-gamma-(alpha-D-glutamylglycine)]-L-lysyl-D-alanyl-N-alpha-[L-alanyl-gamma-(alpha-D-glutamylglycine)]-L-lysyl-D-alanine. *Biochemistry*, 7(4), pp.1450–60.

Gilleronl, M., Vercauterent, J., and Puzo, G. 1993. Lipooligosaccharidic Antigen Containing Dideoxy-a-hexopyranose Typifies Mycobacterium gastri ". *J Biol Chem,* 268(5), pp.3168-79.

Glickman, M.S., Cox, J.S. & Jacobs, W.R., 2000. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. *Molecular cell*, 5(4), pp.717–27.

Gokulan, K. et al., 2013. Crystal structure of Mycobacterium tuberculosis polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *J Biol Chem*, 288(23), pp.16484–94.

Gonzalo Asensio, J. et al., 2006. The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in Mycobacterium tuberculosis. *J Biol Chem*, 281(3), pp.1313–6.

Gopal, P. & Dick, T., 2014. Reactive dirty fragments: implications for tuberculosis drug discovery. *Curr Opin Microbiol*, 21C, pp.7–12.

Goren, M.B., Brokl, O. & Das, B.C., 1976. Sulfatides of Mycobacterium tuberculosis: the structure of the principal sulfatide (SL-I). *Biochemistry*, 15(13), pp.2728–35.

Goren, M.B., Cernich, M. & Brokl, O., 1978. Some observations of mycobacterial acid-fastness. *Am Rev Respir Dis*, 118(1), pp.151–4.

Gouet, P. et al., 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*, 31(13), pp.3320–3323.

Graessle, O.E. & Pietrowski, J.J., 1949. The in vitro effect of para-aminosalicylic acid (PAS) in preventing acquired resistance to streptomycin by Mycobacterium tuberculosis. *J Bacteriol*, 57(4), pp.459–64.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*, 1(6), pp.2876–90.

Grosjean, H. & Fiers, W., 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*, 18(3), pp.199–209.

Grover, S. et al., 2014. Benzothiazinones mediate killing of Corynebacterineae by blocking decaprenyl phosphate recycling involved in cell wall biosynthesis. *J Biol Chem*, 289(9), pp.6177–87.

Grzegorzewicz, A.E. et al., 2012. Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. *Nat Chem Biol*, 8(4), pp.334–341.

Guerardel, Y. et al., 2002. Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains. *J Biol Chem*, 277(34), pp.30635–48.

Han, S.H. et al., 2010. Disseminated Mycobacterium kansasii infection associated with skin lesions: a case report and comprehensive review of the literature. *J Korean med Sci*, 25(2), pp.304–8.

Henderson, B., Lund, P.A. & Coates, A.R., 2010. Multiple moonlighting functions of mycobacterial molecular chaperones. *Tuberculosis (Edinb)*, 90(2), pp.119–24.

Hershkovitz, I. et al., 2008. Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PloS one*, 3(10), p.e3426.

Huet, G. et al., 2009. A lipid profile typifies the Beijing strains of Mycobacterium tuberculosis: identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J Biol Chem*, 284(40), pp.27101–13.

Hunter, R.L. et al., 2006. Review : Multiple Roles of Cord Factor in the Pathogenesis of Primary , Secondary , and Cavitary Tuberculosis , Including a Revised Description of the Pathology of Secondary Disease. *Ann Clin Lab Sci*, 36(4), pp.371–86.

Hunter, S.W. et al., 1984. N-acylkanosamine. A novel N-acylamino sugar from the trehalose-containing lipooligosaccharide antigens of Mycobacterium kansasii. *J Biol Chem*, 259, pp.9729–9734.

Hunter, S.W. et al., 1988. Trehalose-containing lipooligosaccharide antigens of Mycobacterium sp.: presence of a mono-O-methyltri-O-acyltrehalose "core". *Biochemistry*, 27(5), pp.1549–56.

Hunter, S.W. et al., 1983. Trehalose-containing Lipooligosaccharides. *J Biol Chem*, 258(17), pp.10481–10487.

Hunter, S.W. et al., 1985. Trehalose-containing lipooligosaccharides from mycobacteria: structures of the oligosaccharide segments and recognition of a unique N-acylkanosamine-containing epitope. *Biochemistry*, 24(11), pp.2798–805.

Hunter, S.W. & J.Brennan, P., 1990. Evidence for the presence of a phosphatidylinositol anchor on the lipoarabinomannan and lipomannan of Mycobacterium tuberculosis. *J Biol Chem*, 265(16), pp.9272–9.

Indrigo, J., Hunter, R.L.J. & Actor, J.K., 2003. Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. *Microbiology*, 149(8), pp.2049–59.

J.R.Brown et al., 2011. The structure-activity relationship of urea derivatives as anti-tuberculosis agents. *Bioorg Med Chem*, 19(18), pp.5585–95.

Jackson, M. et al., 1999. Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the Mycobacterium tuberculosis cell envelope. *Mol Microbiol*, 31(5), pp.1573–87.

Jackson, M., Crick, D.C. & Brennan, P.J., 2000. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem*, 275(39), pp.30092–9.

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

Jamshad, M. et al., 2011. Surfactant-free purification of membrane proteins with intact native membrane environment. *Biochem Soc Trans*, 39(3), pp.813–8.

Janin, Y.L., 2007. Antituberculosis drugs: ten years of research. *Bioorg Med Chem*, 15(7), pp.2479–513.

Jankute, M. et al., 2012. Arabinogalactan and lipoarabinomannan biosynthesis: structure, biogenesis and their potential as drug targets. *Future Microbiol*, 7(1), pp.129–47.

Jankute, M. et al., 2014. Elucidation of a protein-protein interaction network involved in Corynebacterium glutamicum cell wall biosynthesis as determined by bacterial two-hybrid analysis. *Glycoconj*, 13(6-7), pp.475-83.

Jones, D. et al., 1944. Control of Gram-negative bacteria in experimental animals by streptomycin. *Science.*, 100(2588), pp.103–5.

Källberg, M. et al., 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, 7, pp.1511–22.

Kanetsuna, F., Imaaeda, T. & Cunto, G., 1969. On the linkage between mycolic acid and arabinogalactan in phenol-treated myobacterial cell walls. *Biochim Biophys Acta*, 173(2), pp.341–4.

Kanetsuna, F., Imaeda, T. & San Blas, F., 1968. Chemical analyses of the cell wall of the murine leprosy bacillus. *J Bacteriol*, 96(3), pp.860–1.

Kapopoulouemail, A., Lewemail, J.M. & Cole, S.T., 2011. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*, 9(1), pp.8–13.

Karakousis, P.C., Bishai, W.R. & Dorman, S.E., 2004. Mycobacterium tuberculosis cell envelope lipids and the host immune response. *Cell Microbiol*, 6(2), pp.105–116.

Karimova, G. et al., 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci USA*, 95(10), pp.5752–56.

Karplus, K., 2009. SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Res*, 37(Web Server issue), pp.W492–7.

Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30(4), pp.772–80.

Kaufmann, S.H.E., 1993. Immunity to intracellular bcteria. *Annu. Rev. Immunol*, 11, pp.129–63.

Kelly, S.M., Jess, T.J. & Price, N.C., 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta*, 1751(2), pp.119–39.

Khasnobis, S., Escuyer, V.E. & Chatterjee, D., 2002. Emerging therapeutic targets in tuberculosis: post-genomic era. *Expert Opin Ther Targets*, 6(1), pp.21–40.

Khoo, K.H. et al., 1995. Inositol phosphate capping of the nonreducing termini of lipoarabinomannan from rapidly growing strains of Mycobacterium. *J Biol Chem*, 270(21), pp.12380–9.

Khoo, K.H., Tang, J.B. & Chatterjee, D., 2001. Variation in mannose-capped terminal arabinan motifs of lipoarabinomannans from clinical isolates of Mycobacterium tuberculosis and Mycobacterium avium complex. *J Biol Chem*, 276(6), pp.3863–71.

Kilburn, J.O. & Takayama, K., 1981. Effects of ethambutol on accumulation and secretion of trehalose mycolates and free mycolic acid in Mycobacterium smegmatis. *Antimicrob Agents Chemother*, 20(3), pp.401–404.

Koch, R., 1882. Die Atiologic der Tuberkulose. *Berliner Klinische Wochenschrift*, 15(221-30).

Kolattukudy, P.E. et al., 1997. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol*, 24(2), pp.263–70.

Korduláková, J. et al., 2002. Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. *J Biol Chem*, 277(35), pp.31335–44.

Korduláková, J. et al., 2003. Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species. *J Biol Chem*, 278(38), pp.36285–95.

Kremer, L. et al., 2003. Inhibition of InhA activity, but not KasA activity, induces formation of a KasA-containing complex in mycobacteria. *J Biol Chem*, 278(23), pp.20547–54.

Kremer, L. et al., 2002. Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from Mycobacterium tuberculosis. *Biochem J*, 364(Pt2), pp.423–430.

Kremer, L. et al., 2002. The M. tuberculosis antigen 85 complex and mycolyltransferase activity. *Lett Appl Microbiol*, 34(4), pp.233–237.

Kremer, L. et al., 2000. Thiolactomycin and related analogues as novel anti-mycobacterial agents targeting KasA and KasB condensing enzymes in Mycobacterium tuberculosis. *J Biol Chem*, 275(22), pp.16857–64.

Lai, C.-Y. & Cronan, J.E., 2003. Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J Biol Chem*, 278(51), pp.51494–503.

Lannelle, G., 1963. Nature of mycolic acids from Mycobacterium paratuberculosis; application of thin layer chromatography to their fractionation. *C R Hebd Seances Acad Sci*, 257, pp.781–3.

Larsen, M.H. et al., 2007. Genetic Manipulation of Mycobacterium tuberculosis. In *Curr Protoc Microbiol.* pp. A2.1–A2.21.

Lederer, E. et al., 1975. Cell walls of Mycobacteria and related organisms; chemistry and immunostimulant properties. *Mol Cell Biochem*, 7(2), pp.87–104.

Lee, O.Y.C. et al., 2012. Mycobacterium tuberculosis complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, Bison antiquus. *PloS one*, 7(7), p.e41923.

Lee, Y.C. & Ballou, C.E., 1964. Structural studies on the Myo-inositol mannosides from the glycolipids of Mycobacterium tuberculosis and Mycobacterium phlei. *J Biol Chem*, 239, pp.1316–27.

Lima, R.I.A.M.F. et al., 2001. Role of Trehalose Dimycolate in Recruitment of Cells and Modulation of Production of Cytokines and NO in Tuberculosis. *Infect Immun*, 69(9), pp.5305–12.

Liu, Y. et al., 2012. Synthesis of potent inhibitors of β-ketoacyl-acyl carrier protein synthase III as potential antimicrobial agents. *Molecules (Basel, Switzerland)*, 17(5), pp.4770–81.

Long, F. et al., 2010. Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature*, 467(7314), pp.484–488.

Lv, P.C. et al., 2009. Design, synthesis and biological evaluation of novel thiazole derivatives as potent FabH inhibitors. *Bioorg Med Chem Lett*, 19(23), pp.6750–4.

M.G.Szczepina et al., 2010. STD-NMR studies of two acceptor substrates of GlfT2, a galactofuranosyltransferase from Mycobacterium tuberculosis: epitope mapping studies. *Bioorg Med Chem*, 18(14), pp.5123–28.

Maina, C. et al., 1988. An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene*, 74(2), pp.365–73.

Makarov, V. et al., 2011. Benzothiazinones Kill Mycobacterium tuberculosis by Blocking Arabinan Synthesis. *Science*, 324(5928), pp.801–4.

Marrakchi, H. et al., 2002. MabA (FabG1), a Mycobacterium tuberculosis protein involved in the long-chain fatty acid elongation system FAS-II. *Microbiology (Reading, England)*, 148(Pt 4), pp.951–60.

Matsumoto, M. et al., 2006. OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. *PLoS medicine*, 3(11), p.e466.

McDermott, W. et al., 1954. Pyrazinamide-isoniazid in tuberculosis. *Am Rev Tuberc*, 69(3), pp.319–33.

McNeil, M. et al., 1987a. Definition of the surface antigens of Mycobacterium malmoense and use in studying the etiology of a form of mycobacteriosis. *Journal of bacteriology*, 169(7), pp.3312–20.

McNeil, M. et al., 1987b. Demonstration that the galactosyl and arabinosyl residues in the cell-wall arabinogalactan of Mycobacterium leprae and Mycobacterium tuberculosis are furanoid. *Carbohydr Res*, 166(2), pp.299–308.

McNeil, M., Daffe, M. & Brennan, P.J., 1990. Evidence for the Nature of the Link between Peptidoglycan of Mycobacterial Cell Walls. *J Biol Chem*, 265(30), pp.18200–6.

McNeil, M., Daffes, M. & Brennang, P.J., 1991. Location of the Mycolyl Ester Substituents in the Cell Walls of Mycobacteria. *J Biol Chem*, 266(20), pp.13217–23.

McNeil, M.R. et al., 1994a. Enzymatic evidence for the presence of a critical terminal hexa-arabinoside in the cell walls of Mycobacterium tuberculosis. *Glycobiology*, 4(2), pp.165–73.

McNeil, M.R. & Brennan, P.J., 1991. Structure, function and biogenesis of the cell envelope of mycobacteria in relation to bacterial physiology, pathogenesis and drug resistance; some thoughts and possibilities arising from recent structural information. *Res Microbiol*, 142(4), pp.451–63.

Mdluli, K. et al., 1998. Inhibition of a Mycobacterium tuberculosis beta-ketoacyl ACP synthase by isoniazid. *Science.*, 280(5369), pp.1607–10.

Meniche, X. et al., 2009. Partial redundancy in the synthesis of the D-arabinose incorporated in the cell wall arabinan of Corynebacterineae. *Microbiology*, 154(Pt 8), pp.2315–26.

Middlebrook, G., Coleman, C.M. & Schaefer, W.B., 1959. Sulfolipid from virulent tubercle bacilli. *Proc Natl Acad Sci U S A*, 45(1931), pp.1801–04.

Mikusová, K. et al., 1996. Biosynthesis of the Linkage Region of the Mycobacterial Cell Wall. *J Biol Chem*, 271(13), pp.7820–28.

Mills, J. a et al., 2004. Inactivation of the mycobacterial rhamnosyltransferase, which is needed for the formation of the arabinogalactan-peptidoglycan linker, leads to irreversible loss of viability. *J Biol Chem*, 279(42), pp.43540–6.

Minnikin, D.E. et al., 1989. Comparative studies of antigenic glycolipids of mycobacteria related to the leprosy bacillus. *Acta Leprol*, 7(1), pp.51–4.

Minnikin, D.E. et al., 1985a. Distribution of some mycobacterial waxes based on the phthiocerol family. *J Gen Microbiol*, 131(6), pp.1375–81.

Minnikin, D.E. et al., 1985b. Mycolipenates and mycolipanolates of trehalose from mycobacterium tuberculosis. *J Gen Microbiol*, 131(6), pp.1369–74.

Minnikin, D.E. et al., 2002. The methyl-branched fortifications of Mycobacterium tuberculosis. *Chemistry & biology*, 9(5), pp.545–53.

Minnikin, D.E. et al., 1982. The mycolic Acids of Mycobacterium chelonei. *J Gen Microbiol*, 128(4), pp.817–22.

Morlock, G.P. et al., 2003. ethA, inhA, and katG loci of ethionamide-resistant clinical Mycobacterium tuberculosis isolates. *Antimicrob Agents Chemother*, 47(12), pp.3799–805.

Muñoz, M. et al., 1997. Occurrence of an antigenic triacyl trehalose in clinical isolates and reference strains of Mycobacterium tuberculosis. *FEMS Microbiol Lett*, 157(2), pp.251–9.

Murakami, S. et al., 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, 419(6907), pp.587–93.

Obermeyer, Z., Abbott-Klafter, J. & Murray, C.J.L., 2008. Has the DOTS strategy improved case finding or treatment success? An empirical assessment. *PloS one*, 3(3), p.e1721.

Ortalo-Magné, a et al., 1996. Identification of the surface-exposed lipids on the cell envelopes of Mycobacterium tuberculosis and other mycobacterial species. *J Bacteriol*, 178(2), pp.456–61.

Pacheco, S. et al., 2013. MmpL11 transports mycolic acid-containing lipids to the mycobacterial cell wall and contributes to biofilm formation in M. smegmatis. *J Biol Chem,* 288(33), pp.24213-22.

Parish, T. et al., 2007. Functional complementation of the essential gene fabG1 of Mycobacterium tuberculosis by Mycobacterium smegmatis fabG but not Escherichia coli fabG. *J Bacteri*, 189(10), pp.3721–8.

Parker, S.K. et al., 2009. Mycobacterium tuberculosis Rv3802c encodes a phospholipase/thioesterase and is inhibited by the antimycobacterial agent tetrahydrolipstatin. *PloS one*, 4(1), p.e4281.

Paulsen, I.T., Brown, M.H. & Skurray, R. A, 1996. Proton-dependent multidrug efflux systems. *Microbiol Rev*, 60(4), pp.575–608.

Picardeau, M. et al., 1997. Genotypic characterization of five subspecies of Mycobacterium kansasii. *J Clin Microbiol*, 35(1), pp.25–32.

Podinovskaia, M. et al., 2013. Infection of macrophages with Mycobacterium tuberculosis induces global modifications to phagosomal function. *Cell Microbiol*, 15(6), pp.843–859.

Porter, J.L. et al., 2013. The Cell Wall-Associated Mycolactone Polyketide Synthases Are Necessary but Not Sufficient for Mycolactone Biosynthesis. PLoS One, 8(7), pp.1–12.

Portevin, D. et al., 2004. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci U S A*, 101(1), pp.314–9.

Protopopova, M. et al., 2005. Identification of a new antitubercular drug candidate, SQ109, from a combinatorial library of 1,2-ethylenediamines. *J Antimicrob Chemother*, 56(5), pp.968–74.

Puech, V. et al., 2002. Evidence for a partial redundancy of the fibronectin-binding proteins for the transfer of mycoloyl residues onto the cell wall arabinogalactan termini of Mycobacterium tuberculosis. *Mol Microbiol*, 44(4), pp.1109–22.

Quémard, A. et al., 1995. Enzymatic characterization of the target for isoniazid in Mycobacterium tuberculosis. *Biochemistry*, 34(26), pp.8235–41.

Ramaswamy, S.V. et al., 2004. Genotypic analysis of multidrug-resistant Mycobacterium tuberculosis isolates from Monterrey, Mexico. *J Med Microbiol*, 53(2), pp.107–13.

Rao, A. & Ranganathan, A., 2004. Interaction studies on proteins encoded by the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis. *Mol Genet Genomics*, 272(5), pp.571–9.

Rao, V. et al., 2005. Mycobacterium tuberculosis controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule. *J Exp Med*, 201(4), pp.535–43.

Rao, V. et al., 2006. Trans -cyclopropanation of mycolic acids on trehalose dimycolate suppresses Mycobacterium tuberculosis – induced inflammation and virulence. *J Clin Invest*, 116(6), pp.1660-7.

Raviglione, M., 2006. XDR-TB: entering the post-antibiotic era? *Int J Tuberc Lung Dis*, 10(11), pp.1185–7.

Raymond, J.B. et al., 2005. Identification of the namH gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *J Biol Chem*, 280(1), pp.326–33.

Recht, J. & Kolter, R., 2001. Glycopeptidolipid acetylation affects sliding motility and biofilm formation in Mycobacterium smegmatis. *J Bacteriol*, 183(19), pp.5718–24.

Reed, M.B. et al., 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature*, 431(7004), pp.84–7.

Remuiñán, M.J. et al., 2013. Tetrahydropyrazolo[1,5-a]pyrimidine-3-carboxamide and N-benzyl-6',7'-dihydrospiro[piperidine-4,4'-thieno[3,2-c]pyran] analogues with bactericidal efficacy against Mycobacterium tuberculosis targeting MmpL3. *PloS one*, 8(4), p.e60933.

Ren, H. et al., 2007. Identification of the lipooligosaccharide biosynthetic gene cluster from Mycobacterium marinum. *Mol Microbiol*, 63(5), pp.1345–59.

Robinson, N. et al., 2007. A mycobacterial gene involved in synthesis of an outer cell envelope lipid is a key factor in prevention of phagosome maturation. *Infect Immun*, 75(2), pp.581–91.

Rombouts Y, Elass E, Biot C, Maes E, Coddeville B, Burguière A, Tokarski C, Buisine E, Trivelli X, Kremer L, G.Y., 2010. Structural analysis of an unusual bioactive N-acylated lipo-oligosaccharide LOS-IV in Mycobacterium marinum. *J Am Chem Soc.*, 132(45), pp.16073–84.

Rombouts, Y. et al., 2009. Mycobacterium marinum lipooligosaccharides are unique caryophyllose-containing cell wall glycolipids that inhibit tumor necrosis factor-alpha secretion in macrophages. *J Biol Chem*, 284(31), pp.20975–88.

Ronning, D.R. et al., 2000. Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nat Struct Biol*, 7(2), pp.141–46.

La Rosa, V. et al., 2012. MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. *Antimicrob Agents Chemother*, 56(1), pp.324–31.

Rothschild, B.M. et al., 2001. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis*, 33(3), pp.305–11.

Rousseau, C. et al., 2004. Production of phthiocerol dimycocerosates protects Mycobacterium tuberculosis from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. *Cell Microbiol*, 6(3), pp.277–87.

Roy, A., Kucukural, A. & Zhang, Y., 2011. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5(4), pp.725–38.

Roy, R. et al., 2013. Synthesis of α-Glucan in Mycobacteria Involves a Hetero-octameric Complex of Trehalose Synthase TreS and Maltokinase Pep2. *ACS Chem Biol*, 8(10), pp.2245–55.

Rozwarski, D.A. et al., 1998. Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. *Science*, 279(5347), pp.98–102.

Rukmini, R. et al., 2004. Crystallization and preliminary X-ray crystallographic investigations of an unusual type III polyketide synthase PKS18 from Mycobacterium tuberculosis. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 4), pp.749–51.

Russell, D.G., 2007. Who puts the tubercle in tuberculosis? *Nat Rev Microbiol*, 5(1), pp.39–47.

S.T.Cole, 1994. Mycobacterium tuberculosis: drug-resistance mechanisms. *Trends Microbiol*, 2(10), pp.411–15.

Salo, W.L. et al., 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *ProcNatl Acad Sci U S A*, 91(6), pp.2091–4.

Sarkar, D. et al., 2011. Identification of a glycosyltransferase from Mycobacterium marinum involved in addition of a caryophyllose moiety in lipooligosaccharides. *J Bacteriol*, 193(9), pp.2336–40.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J., 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A*, 98(22), pp.12712–7.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, 48(1), pp.77–84.

Saxena, P. et al., 2003. A new family of type III polyketide synthases in Mycobacterium tuberculosis. *J Biol Chem*, 278(45), pp.44780–90.

Scarsdale, J.N. et al., 2001. Crystal structure of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III. *J Biol Chem*, 276(23), pp.20516–22.

Schleifer, K.H. & Kandler, O., 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev*, 36(4), pp.407–77.

Scorpio, A. et al., 1997. Characterization of pncA mutations of pyrazinamide-resistant Mycobacterium tuberculosis in Turkey. N*ew Microbiol*, 41(3), pp.540–543.

Seeger, M.A. et al., 2009. Crucial role of Asp408 in the proton translocation pathway of multidrug transporter AcrB: evidence from site-directed mutagenesis and carbodiimide labeling. *Biochemistry*, 48(25), pp.5801–12.

Sennhauser, G. et al., 2009. Crystal structure of the multidrug exporter MexB from Pseudomonas aeruginosa. *J Mol Biol*, 389(1), pp.134–45.

Sensi, P., 1983. History of the development of rifampin. *Rev Infect Dis*, 5(3), pp.S402–6.

Shin, S.S. et al., 2006. Treatment outcomes in an integrated civilian and prison MDR-TB treatment program in Russia., *Int J Tuberc Lung Dis,* (4), pp.402–8.

Silva, M.T. & Macedo, P.M., 1983. Electron microscopic study of Mycobacterium leprae membrane. *Int J Lepr Other Mycobact Dis.*, 51(2), pp.219–24.

Sloan, D.J., Davies, G.R. & Khoo, S.H., 2013. Recent advances in tuberculosis : New drugs and treatment regimens. *Curr Respir Med Rev*, 9(3), pp.200–10.

Smith, D.B. & Johnson, K.., 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, 67(1), pp.31–40.

Smith, S., Witkowski, A. & Joshi, A.K., 2003. Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res*, 42(4), pp.289–317.

Sondén, B. et al., 2005. Gap, a mycobacterial specific integral membrane protein, is required for glycolipid transport to the cell surface. *Mol Microbiol*, 58(2), pp.426–40.

Sousa-D'Auria, C. De et al., 2003. New insights into the biogenesis of the cell envelope of corynebacteria: identification and functional characterization of five new mycoloyltransferase genes in Corynebacterium glutamicum. *FEMS Microbiol Lett*, 224(1), pp.35–44.

De Souza, M.V.N. et al., 2008. Synthesis and biological aspects of mycolic acids: an important target against Mycobacterium tuberculosis. *ScientificWorldJournal*, 8, pp.720–51.

Sriram, D. et al., 2006. Gatifloxacin derivatives: synthesis, antimycobacterial activities, and inhibition of Mycobacterium tuberculosis DNA gyrase. *Bioorg Med Chem Lett*, 16(11), pp.2982–5.

Steenken, W. (Jr). W. & Wolinsky, E., 1952. Isoniazid in experimental tuberculosis. *Trans Annu Meet Natl Tuberc Assoc*, 48, pp.425–30.

Stover, C.K. et al., 1991. New use of BCG for recombinant vaccines. *Nature*, 351, pp.456–460.

Su, C. et al., 2012. NIH Public Access. , 470(7335), pp.558–562.

Supply, P. et al., 2013. Genome analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of the etiologic agent of tuberculosis. *Nat Genet*, 45(2), pp.172–9.

Sweet, L. et al., 2010. Mannose receptor-dependent delay in phagosome maturation by Mycobacterium avium glycopeptidolipids. *Infection and immunity*, 78(1), pp.518–26.

Tahlan, K. et al., 2012. SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. *Antimicrob Agents Chemother*, 56(4), pp.1797–809.

Takayama, K. & Armstrong, E.L., 1976. Isolation, characterization, and function of 6-mycolyl-6'-acetyltrehalose in the H37Ra strain of Mycobacterium tuberculosis. *Biochemistry*, 15(2), pp.441–7.

Takayama, K. & Armstrong, E.L., 1977. Metabolic role of free mycolic acids in Mycobacterium tuberculosis. *J Bacteriol*, 130(1), pp.569–570.

Takayama, K. & Goldman, D.S., 1970. Enzymatic synthesis of mannosyl-1-phosphoryl-decaprenol by a cell-free system of Mycobacterium tuberculosis. *J Biol Chem*, 245(23), pp.6251–7.

Takayama, K., Wang, C. & Besra, G.S., 2005. Pathway to Synthesis and Processing of Mycolic Acids in Mycobacterium tuberculosis Genetic analysis of syntheis and processing of mycolic acid. *Clin Microbiol Rev*, 18(1), pp.81–101.

Taly, J.-F. et al., 2011. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat Protoc*, 6, pp.1669–1682.

Tang, X. & Bruce, J.E., 2009. Chemical cross-linking for protein-protein interaction studies. *Methods Mol Biol*, 492, pp.283–93.

Taniguchi, H. et al., 1996. Rifampicin resistance and mutation of the rpoB gene in Mycobacterium tuberculosis. *FEMS Microbiol Lett*, 144(1), pp.103–8.

Telenti, A. et al., 1997. The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol. *Nat Med*, 3(5), pp.567–70.

Thomas, J.P. et al., 1961. A new synthetic compound with antituberculous activity in mice: ethambutol (dextro-2,2'-(ethylenediimino)-di-l-butanol). *Am Rev Respir Dis*, 83, pp.891–3.

Tomioka, H., Saito, H. & Sato, K., 1993. Comparative Antimycobacterial Activities of the Newly Synthesized Quinolone AM-1155, Sparfloxacin, and Ofloxacin. *Antimicrob Agents Chemother*, 37(6), pp.1259–63.

Tompsett, R. et al., 1954. The influence of pyrazinamide-isoniazid on M. tuberculosis in animals and man. *Trans Annu Meet Natl Tuberc Assoc*, 67, pp.224–31.

Tropea, J.E. et al., 2007. A generic method for the production of recombinant proteins in Escherichia coli using a dual hexahistidine-maltose-binding protein affinity tag. *methods Mol Med*, 363, pp.1–19.

Tsukazaki, T. et al., 2011. Structure and function of a membrane component SecDF that enhances protein export. *Nature*, 474(7350), pp.235–238.

Udwadia, Z.F. et al., 2012. Totally Drug-Resistant Tuberculosis in India. , 54, pp.579–581.

Usha, V. et al., 2012. Structure and function of Mycobacterium tuberculosis meso-diaminopimelic acid (DAP) biosynthetic enzymes. *FEMS Microbiol Lett*, 330(1), pp.10–6.

Usha, V. et al., 2006. Use of a codon alteration strategy in a novel approach to cloning the Mycobacterium tuberculosis diaminopimelic acid epimerase. *FEMS Microbiol Lett*, 262(1), pp.39–47.

Varela, C. et al., 2012. MmpL Genes Are Associated with Mycolic Acid Metabolism in Mycobacteria and Corynebacteria. *Chem Biol*, 19(4), pp.498–506.

Veyrier, F. et al., 2009. Phylogenetic detection of horizontal gene transfer during the step-wise genesis of Mycobacterium tuberculosis. *BMC Evol Biol*, 9, p.196.

Veyrier, F.J., Dufort, A. & Behr, M.A., 2011. The rise and fall of the Mycobacterium tuberculosis genome. *Trends Microbiol*, 19(4), pp.156–61.

Veyron-Churlet, R. et al., 2004. Protein-protein interactions within the Fatty Acid Synthase-II system of Mycobacterium tuberculosis are essential for mycobacterial viability. *Mol Microbiol*, 54(5), pp.1161–72.

Veyron-Churlet, R. et al., 2005. The biosynthesis of mycolic acids in Mycobacterium tuberculosis relies on multiple specialized elongation complexes interconnected by specific protein-protein interactions. *J Mol Biol*, 353(4), pp.847–58.

Vignal, C. et al., 2003. Lipomannans, But Not Lipoarabinomannans, Purified from Mycobacterium chelonae and Mycobacterium kansasii Induce TNF- and IL-8 Secretion by a CD14-Toll-Like Receptor 2-Dependent Mechanism. *J Immunol*, 171(4), pp.2014–23.

Vilchèze, C. et al., 2005. Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrob Agents Chemother*, 49(2), pp.708–20.

Vilchèze, C. et al., 2011. Coresistance to isoniazid and ethionamide maps to mycothiol biosynthetic genes in Mycobacterium bovis. *Antimicrob Agents Chemother*, 55(9), pp.4422–3.

Vilchèze, C. et al., 2000. Inactivation of the inhA-encoded fatty acid synthase II (FASII) enoyl-acyl carrier protein reductase induces accumulation of the FASI end products and cell lysis of Mycobacterium smegmatis. *J Bacteriol*, 182(14), pp.4059–67.

Vilchèze, C. et al., 2008. Mycothiol biosynthesis is essential for ethionamide susceptibility in Mycobacterium tuberculosis. *Mol Microbiol*, 69(5), pp.1316–29.

Villeneuve, C. et al., 2003. Surface-exposed glycopeptidolipids of Mycobacterium smegmatis specifically inhibit the phagocytosis of mycobacteria by human macrophages. Identification of a novel family of glycopeptidolipids. *J Biol Chem*, 278(51), pp.51291–300.

Walburger, A. et al., 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science.*, 304(5678), pp.1800–4.

Welin, A. et al., 2008. Incorporation of Mycobacterium tuberculosis lipoarabinomannan into macrophage membrane rafts is a prerequisite for the phagosomal maturation block. *Infect Immun*, 76(7), pp.2882–7.

Wells, R.M. et al., 2013. Discovery of a siderophore export system essential for virulence of Mycobacterium tuberculosis. *PLoS pathogens*, 9(1), p.e1003120.

WHO, 2013. Global tuberculosis report 2013.

WHO, 2012, 2012. WHO global report, 2012. *WHO*.

Wietzerbin, J. et al., 1974. Occurance of D-Alanyl-(D)-meso-diaminopimely-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of Mycobacteria. *Biochemistry*, 13(17), pp.3471-6.

Winder, F.G. & Collins, P.B., 1970. Inhibition by Isoniazid of Synthesis of Mycolic Acids in Mycobacterium tuberculosis. *J Gen Microbiol*, 63(1), pp.41–8.

Wolucka, B. a et al., 1994. Recognition of the lipid intermediate for arabinogalactan/arabinomannan biosynthesis and its relation to the mode of action of ethambutol on mycobacteria. *J Biol Chem*, 269(37), pp.23328–35.

Wong, H.C. et al., 2002. The solution structure of acyl carrier protein from Mycobacterium tuberculosis. *J Biol Chem*, 277(18), pp.15874–80.

Van der Woude, A.D. et al., 2012. Unexpected link between lipooligosaccharide biosynthesis and surface protein release in Mycobacterium marinum. *J Biol Chem*, 287(24), pp.20417–29.

Yang, L. et al., 2014. RND transporters protect Corynebacterium glutamicum from antibiotics by assembling the outer membrane. *Microbiologyopen*, 3(4), pp.484–96.

Yeager, R.L., Munroe, W.G. & Dessau, F.I., 1952. Pyrazinamide (aldinamide*) in the treatment of pulmonary tuberculosis. *Trans Annu Meet Natl Tuberc Assoc*, 48, pp.178–201.

Yegian, D. & Vanderlinde, R. j, 1948. A Quantitative analysis of the resistance of mycobacteria to streptomycin. *J Bacteriol,* 56(2), pp.177-86.

Yu, E.W. et al., 2003. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science*, 300(5621), pp.976–80.

Yuan, Y., Sachdeva, M., et al., 2012. Fatty acid biosynthesis in Pseudomonas aeruginosa is initiated by the FabY class of β-ketoacyl acyl carrier protein synthases. *J Bacteriol*, 194(19), pp.5171–84.

Yuan, Y. & Barry, C.E., 1996. A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, 93(23), pp.12828–33.

Yuan, Y., Leeds, J. a & Meredith, T.C., 2012. Pseudomonas aeruginosa directly shunts β-oxidation degradation intermediates into de novo fatty acid biosynthesis. *J Bacteriol*, 194(19), pp.5185–96.

Zhang, Y. et al., 2003. Mode of action of pyrazinamide: disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid. *J antimicrob Chemother*, 52(5), pp.790–5.

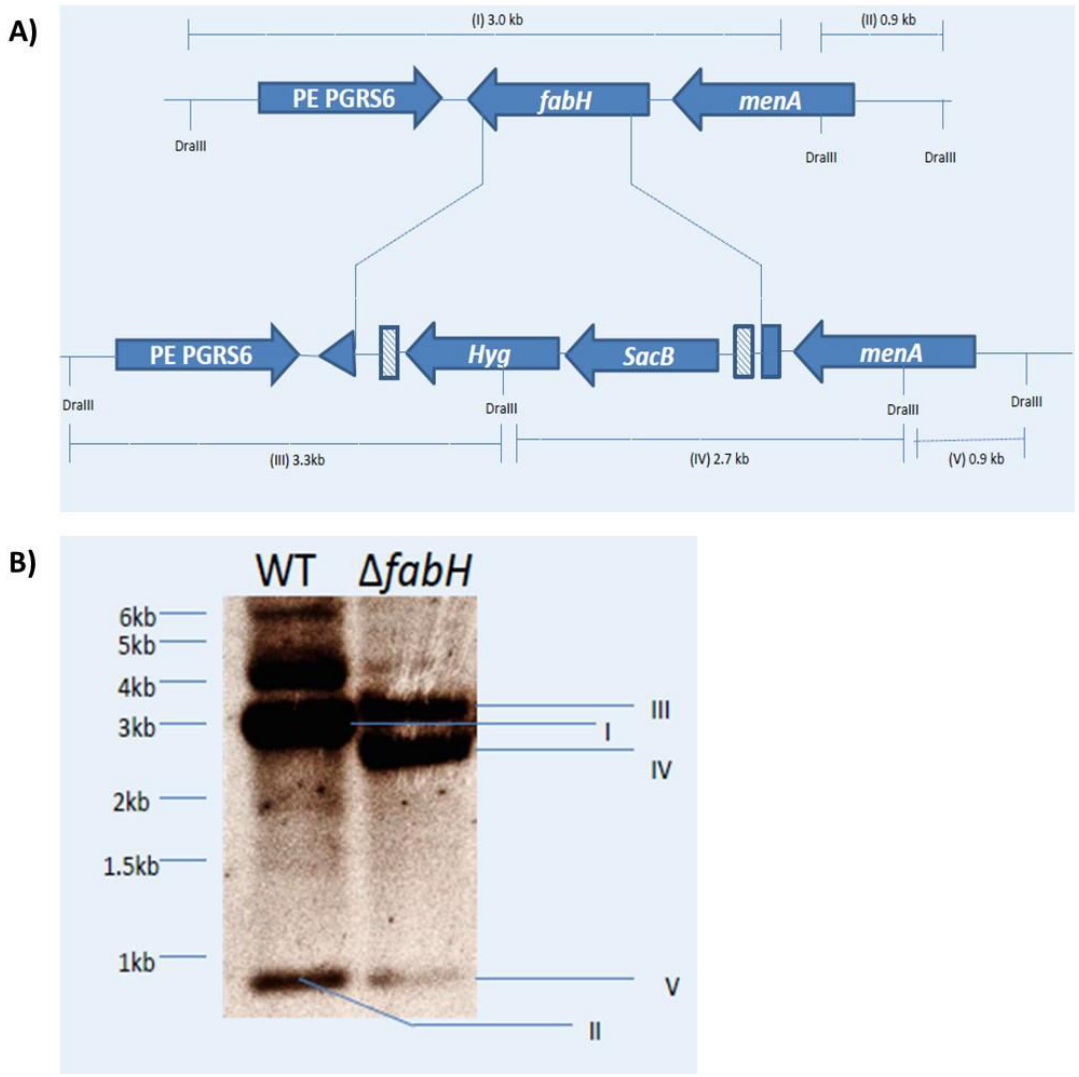
Zhang, Y. et al., 1992. The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. *Nature,* 358(6387), pp.591-3.

Zheng, J. et al., 2011. Combining blue native polyacrylamide gel electrophoresis with liquid chromatography tandem mass spectrometry as an effective strategy for analyzing potential membrane protein complexes of Mycobacterium bovis bacillus Calmette-Guérin. *BMC genomics*, 12(1), p.40.

Zimhony, O., Vilcheze, C. & Jacobs, W.R.J., 2004. Characterization of Mycobacterium smegmatis Expressing the Mycobacterium tuberculosis Fatty Acid Synthase I (fas1) Gene, 186(13), pp.4051–55.

Appendix

Appendix 1



Southern blot analysis of *fabH* mutant. (A) Map of *fabH* region in the parental *M. bovis BCG* and its corresponding region in the Δ*fabH* mutant. *res*, γδ resolvase site; *hyg*, hygromycin resistance gene from *Streptomycin hygroscopicus*; *sacB*, sucrose counter selectable gene from *Bacillus subtilis*. Digoxigenin-labeled probes were derived from ~1 kb upstream and downstream flanking sequences that were used to construct the knockout plasmid. DraIII-digested bands expected in a Southern blot are indicated in roman numerals with sizes. (B) The gel below shows the Southern blot of DraIII-digested genomic DNA from the two strains with expected bands indicated by numbers.

Appendix 2

Plasmids	Description	Reference
pET28a	*E. coli* expression vector with histidine affinity tag	Novagen
pET28b	*E. coli* expression vector with histidine affinity tag	Novagen
pET41c	*E. coli* expression vector with histidine affinity tag	Novagen
pGEX-4T-1	*E. coli* expression vector with glutathion affinity tag	GE Lifesciences
pMAL-p2X	*E. coli* expression vector with maltose binding protein affinity tag	New England Biolabs
pSD26	Mycobacterial expression vector with histidine affinity tag	Daugelat *et al*., 2003
pVV16	Mycobacterial expression vector with histidine affinity tag	BEI resources, NIAID, NIH
pET28a:ML1	MmpL3 non-transmembrane region 1 cloned into pET28a vector in frame with N-term His tag	This work
pET28a:ML2	MmpL3 non-transmembrane region 2 cloned into pET28a vector in frame with N-term His tag	This work
pET28b:ML1	MmpL3 non-transmembrane region 1 cloned into pET28a vector in frame with C-term His tag	This work
pET28b:ML2	MmpL3 non-transmembrane region 2 cloned into pET28a vector in frame with C-term His tag	This work
pSD26:ML1	MmpL3 non-transmembrane region 1 cloned into mycobacterial expression vector pSD26 vector in frame with C-term His tag	This work
pSD26:ML2	MmpL3 non-transmembrane region 2 cloned into mycobacterial expression vector pSD26 vector in frame with C-term His tag	This work
pET28b:ML	Non-transmembrane region 1 and 2 fused with artificial linker peptide (ML) cloned into pET28b expression vector in frame with C-term His tag	This work
pET41c:MmpL3	MmpL3 cloned into expression vector pET41c in frame with C-term His tag	This work
pET28a:MmpL3	MmpL3 cloned into expression vector pET28a in frame with N-term His tag	This work
pGEX:MmpL3	MmpL3 cloned into expression vector pGEX-4T-1 vectorin frame with N-term GST tag	This work

pMAL:MmpL3	MmpL3 cloned into expression vector pMALp2X in frame with N-term MBP tag	This work
pSD26:MmpL3	MmpL3 cloned into mycobacterial expression vector pSD26 vector in frame with C-term His tag	This work
pVV16:MmpL3	MmpL3 cloned into mycobacterial expression vector pVV16 vector in frame with C-term His tag	This work
pET41C:MmpL3-CB	MmpL3 cloned into expression vector pET41c in frame with C-term His tag	This work
pET28b:MmpL3	MmpL3 gene codon optimised for expression in *E. coli* cloned into expression vector pET28bin frame with C-term His tag	This work
pMV261(apra): MmpL3	pMV261 vector with Apramycin selection marker harboring full length *mmpL3* gene for constitutive expression in *M. smegmatis*	This work
Bacterial strains		
E. coli BL21 (DE3)	F^- ompT gal dcmlonhsd$S_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Invitrogen™
E. coli C41 (DE3)	F^- ompT gal dcmhsd$S_B(r_B^- m_B^-)$(DE3)	Invitrogen™
E. coli Tuner	F^- lacYompT gal dcmlonhsd$S_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Invitrogen™
M. smegmatis mc^2155	Wild type strain,E*pt*mutant of *M. smegmatis* strain mc^26	Snapper *et al*. 1990

Appendix 3

Primer name	Primer Sequence
MmpL3_CB_41c F	C GTA CAT ATG GTG TTT GCA TGG TGG GGT CGT ACT GTA TAT CGA TAC CGG TTC ATC GTA ATC
MmpL3_41c_F	C GTA CAT ATG GTG TTC GCC TGG TGG GGT C
MmpL3_41c_R	C GTA AAG CTT AAG GCG TCC TTC GCG GC
MmpL3_pGEX_F	C GTA AGA TCT GTG TTC GCC TGG TGG GGT C
MmpL3_pGEX_R	C GTA GAA TTC TTA AAG GCG TCC TTC GCG GC
MmpL3_pSd26_F	GTC AGA TCT ATG GTG TTC GCC TGG TGG GGT C
MmpL3_pSd_R	C GTA AAG CTT CTA GTG GTGGTGGTGGTGGTGGTGG AAG GCG TCC TTC GCG GC
MmpL3_CB_28a F	C GTA GAA TTC GTG TTT GCA TGG TGG GGT CGT ACT GTA TAT CGA TAC CGG TTC ATC GTA ATC
MmpL3_28a_F	C GTA GAA TTC GTG TTC GCC TGG TGG GGT C
MmpL3_28a_R	C GTA AAG CTT TTA AAG GCG TCC TTC GCG GC
MmpL3_pVV16_R	C GTA AAG CTTC AAG GCG TCC TTC GCG GC
ML1 41cF	ATC CAT ATG CTG AGC CTC GGC AAG CAC
ML1 41cR	ATC AAG CTTCAC CGC CAC CAA CGG CAG
ML2 41cF	ATC CAT ATG AAGCTGTCATTGGGCGGG
ML2 41cR	ATC AAG CTT CAC CAT CAG CGG CAT C
mmpL3_pMV261F	GT CAA GCT TGGGTG TTC GCC TGG TGG GGT C
mmpL3_pMV261R	C GTA ATC GAT CTA AAG GCG TCC TTC GCG GC
ML-N_{HIS}-F	CAC GAA TTC CTG TCT CTG GGC AAA CAC
ML-N_{HIS}-R	ATC CTC GAG TCA TTC CAG TGC CGG CGT GCC
ML1-N_{HIS}-R	AGT CTC GAG TCA TTC CAT GCG ACG CTG ATC
ML1-C_{HIS}-F	AGT ACC ATG GGC CTG TCT CTG GGC AAA CAC
ML1-C_{HIS}-R	GTA CTC GAG TTC CAT GCG ACG CTG ATC
ML2-N_{HIS}-F	CAC GAA TTC GGT ATC TCT GAA AAA TAT CTG
ML2-C_{HIS}-F	AGT ACC ATG GGC GGT ATC TCT GAA AAA TAT CTG
ML2-C_{HIS}-R	GTA CTC GAG TTC CAG TGC CGG CGT GCC

SDM primers for Q5 SDM kit:

I250T (72C) ATC-ACC

I250T F: GGGGATCGCCaccGACTACGGGT

I250T R: AGACCGATCAGCGACACCACG

Y252A (65C) TAC-GCC

Y252A F: CGCCATCGACgccGGGTTGTTCATC

Y252A R: ATCCCCAGACCGATCAGC

Y252D (67C) TAC-GAC

Y252D F: CGCCATCGACgacGGGTTGTTCA

Y252D R: ATCCCCAGACCGATCAGC

W731G (71C)

W731G F: CGATGACTGCgggTGGGCACCGC

W731G R: CCGAGCAGCTTCATCACCGATG

E262A (71C) GAA-GCA

E262A F: CCGGTTCCGCgcaGAGATCGCCG

E262A R: CTCACGATGAACAACCCGTAGTCG

R653A (65C) CGA-GCA

R653A F: GGTCGAGGCGgcaGAGCGCGGCA

R653A R: ATCCGGGACACCAAGAAC

D251A (69C) GAC-GCC

D251A F: GATCGCCATCgccTACGGGTTGT

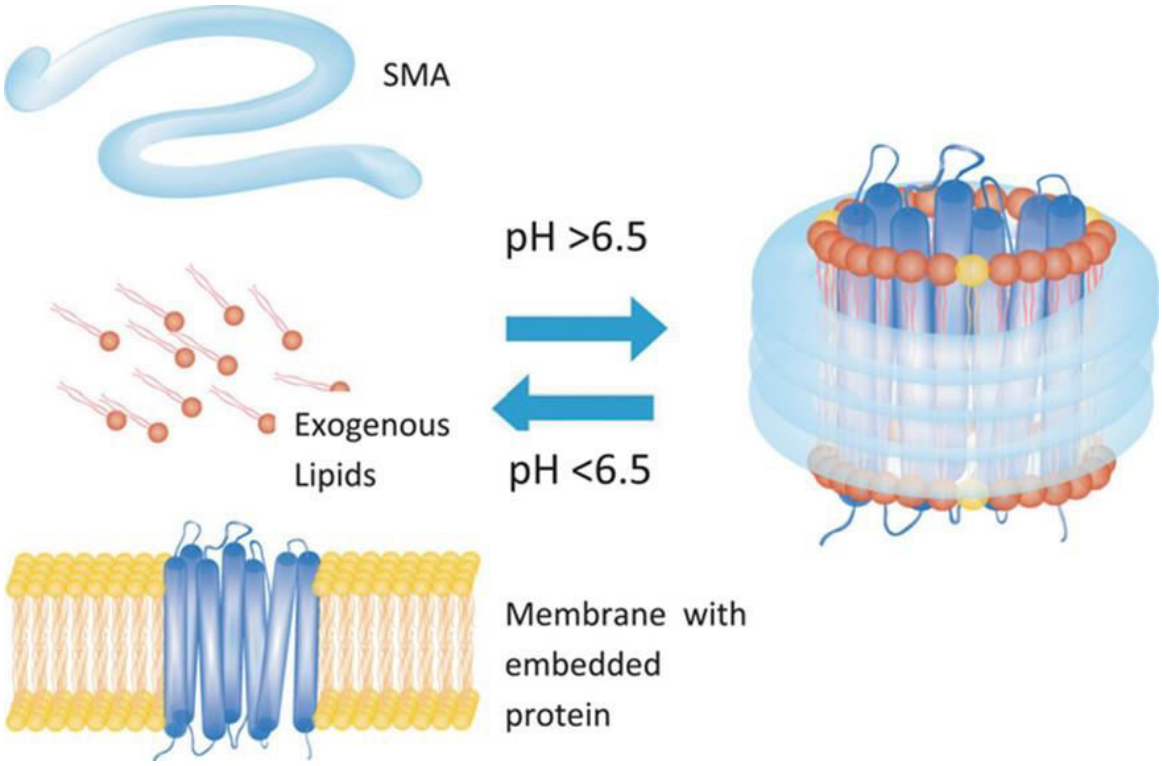
D251A R: CCCAGACCGATCAGCGAC

Appendix 4

Construct	Protein sequence	Vector	Expected Mol.wt (kDa)
MmpL3-C_{His}	MGVFAWWGRTVYRYRFIVIGVMVALCLGGGVFGLSLGKHVTQSGFYDD GSQSVQASVLGDQVYGRDRSGHIVAIFQAPAGKTVDDPAWSKKVVDEL NRFQQDHPDQVLGWAGYLRASQATGMATADKKYTFVSIPLKGDDDDTI LNNYKAIAPDLQRLDGGTVKLAGLQPVAEALTGTIATDQRRMEVLALPL VAVVLFFVFGGVIAAGLPVMVGGLCIAGALGIMRFLAIFGPVHYFAQPVV SLIGLGIAIDYGLFIVSRFREEIAEGYDTETAVRRTVITAGRTVTFSAVLIVA SAIGLLLFPQGFLKSLTYATIASVMLSAILSITVLPACLGILGKHVDALGVR TLFRVPFLANWKISAAYLNWLADRLQRTKTREEVEAGFWGKLVNRVMK RPVLFAAPIVIIMILLIIPVGKLSLGGISEKYLPPTNSVRQAQEEFDKLFPGY RTNPLTLVIQTSNHQPVTDAQIADIRSKAMAIGGFIEPDNDPANMWQERA YAVGASKDPSVRVLQNGLINPADASKKLTELRAITPPKGITVLVGGTPAL ELDSIHGLFAKMPLMVVILLTTTIVLMFLAFGSVVLPIKATLMSALTLGST MGILTWIFVDGHFSKWLNFTPTPLTAPVIGLIIALVFGLSTDYEVFLVSRM VEARERGMSTQEAIRIGTAATGRIITAAALIVAVVAGAFVFSDLVMMKYL AFGLMAALLLDATVVRMFLVPSVMKLLGDDCWWAPRWARRLQTRIGL GEIHLPDERKRPVSNGRPARPPVTAGLVAARAAGDPRPPHDPTHPLAESP RPARSSPASSPELTPALEATAAPAAPSGASTTRMQIGSSTEPPTTRLAAAG RSVQSPASTPPPTPTPPSAPSAGQTRAMPLAANRSTDAAGDPAEPTAALPII RSDGDDSEAATEQLNARGTSDKTRQRRRGGGALSAQDLLRREGRLHHH HHHHH	pET28 b	102.40
ML-C_{His}	MGGLSLGKHTQSGFYDDGSQSVQASVLGDQVYGRDRSGHIVAIFQAPAG KTVDDPAWSKKVVDELNRFQQDHPDQVLGWAGYLRASQATGMATAD KKYTFVSIPLKGDDDDTILNNYKAIAPDLQRLDGGTVKLAGLQPVAEALT GTIATDQRRMEGAEAAAKPGAAHYEAAAKAGGISEKYLPPTNSVRQAQE EFDKLFPGYRTNPLTLVIQTSNHQPVTDAQIADIRSKAMAIGGFIEPDNDP ANMWQERAYAVGASKDPSVRVLQNGLINPADASKKLTELRAITPPKGIT VLVGGTPALEHHHHHHHH	pET28 b	33.750
ML-N_{HIS}	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFGLSLGKHTQSGF YDDGSQSVQASVLGDQVYGRDRSGHIVAIFQAPAGKTVDDPAWSKKVV DELNRFQQDHPDQVLGWAGYLRASQATGMATADKKYTFVSIPLKGDDD DTILNNYKAIAPDLQRLDGGTVKLAGLQPVAEALTGTIATDQRRMEGAE AAAKPGAAHYEAAAKAGGISEKYLPPTNSVRQAQEEFDKLFPGYRTNPL TLVIQTSNHQPVTDAQIADIRSKAMAIGGFIEPDNDPANMWQERAYAVG ASKDPSVRVLQNGLINPADASKKLTELRAITPPKGITVLVGGTPALE	pET28 a	35.75
ML1-N_{HIS}	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFGLSLGKHTQSGF YDDGSQSVQASVLGDQVYGRDRSGHIVAIFQAPAGKTVDDPAWSKKVV DELNRFQQDHPDQVLGWAGYLRASQATGMATADKKYTFVSIPLKGDDD DTILNNYKAIAPDLQRLDGGTVKLAGLQPVAEALTGTIATDQRRME	pET28 a	20.47
ML1-C_{HIS}	MGGLSLGKHTQSGFYDDGSQSVQASVLGDQVYGRDRSGHIVAIFQAPAG KTVDDPAWSKKVVDELNRFQQDHPDQVLGWAGYLRASQATGMATAD KKYTFVSIPLKGDDDDTILNNYKAIAPDLQRLDGGTVKLAGLQPVAEALT GTIATDQRRMELEHHHHHHHHH	pET28 b	18.13
ML2-N_{His}	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFGISEKYLPPTNSV RQAQEEFDKLFPGYRTNPLTLVIQTSNHQPVTDAQIADIRSKAMAIGGFIE PDNDPANMWQERAYAVGASKDPSVRVLQNGLINPADASKKLTELRAITP PKGITVLVGGTPALE	pET28 a	17.65
ML2-C_{His}	MGGISEKYLPPTNSVRQAQEEFDKLFPGYRTNPLTLVIQTSNHQPVTDAQI ADIRSKAMAIGGFIEPDNDPANMWQERAYAVGASKDPSVRVLQNGLINP ADASKKLTELRAITPPKGITVLVGGTPALELEHHHHHHHHH	pET28 b	15.36
Clone	Primer set	Enzym es	Vector
MmpL3-C_{His}	Genscript construct	NcoI/X hoI	pET28b
ML-C_{His}	Genscript construct	NcoI/X hoI	pET28b
ML-N_{His}	ML-N-F/ML-N-R	EcoRI/	pET28a

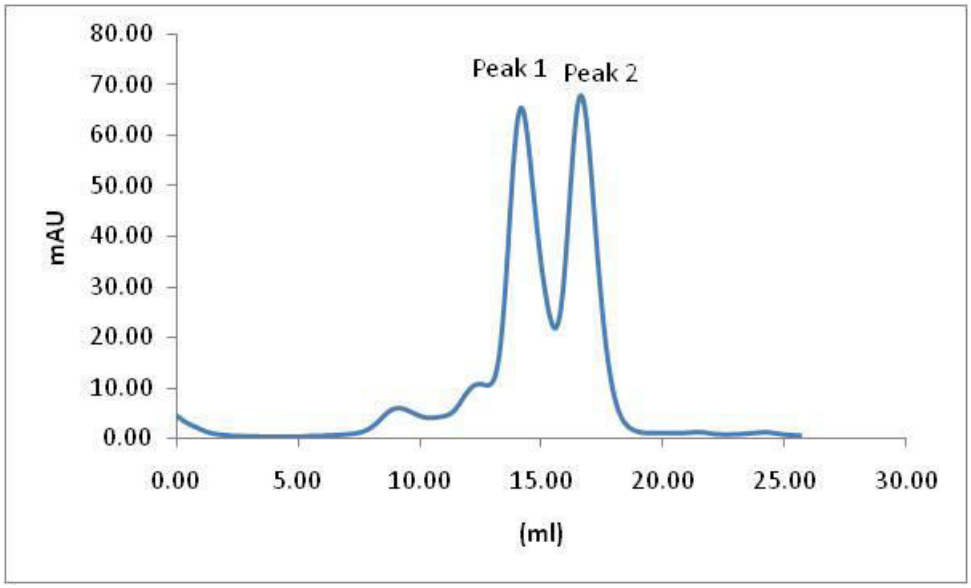
		XhoI	
ML1-N_{His}	ML-N-F/ML1-N-R	EcoRI/ XhoI	pET28a
ML1-C_{His}	ML1-C-F/ML1-C-R	NcoI/X hoI	pET28b
ML2-N_{His}	ML2-N-F/ML-N-R	EcoRI/ XhoI	pET28a
ML2-C_{His}	ML2-C-F/ML2-C-R	NcoI/X hoI	pET28b

Appendix 5



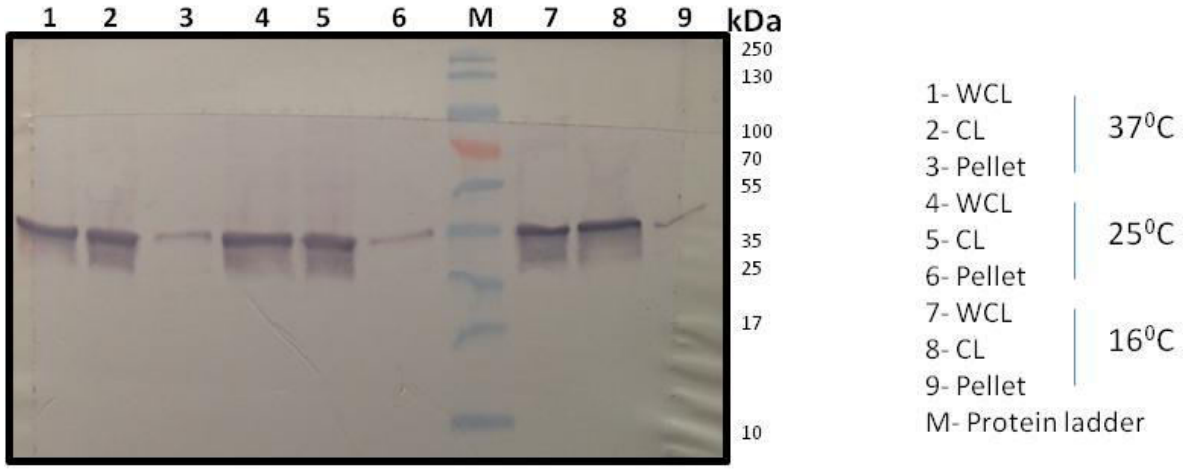
Schematic presentation of extraction of membrane protein using SMA. SMA polymer which is amphipathic in nature, encapsulates membrane protein above pH 6.5 and disassembles below pH 6.5 to give membrane protein in its native form.

Appendix 6

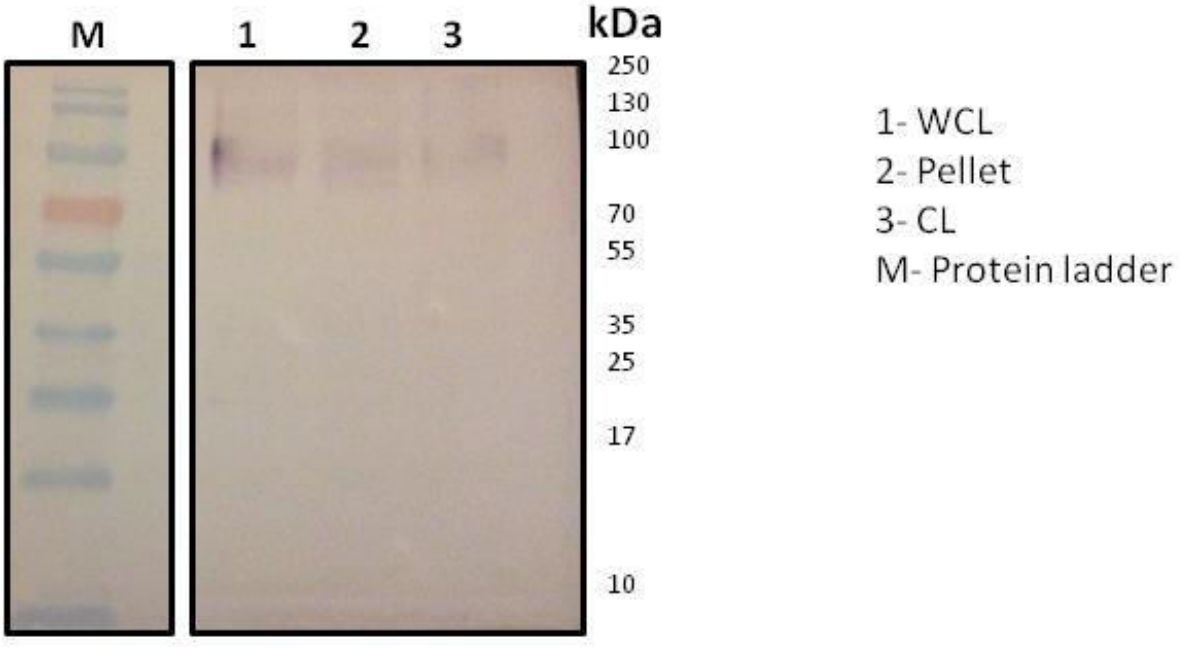


Size exclusion chromatogram of Gel filtration Standards. Peak 1 corresponds to Carbonic anhydrase (28 kD) and peak 2 corresponds to Bovine Serum Albumin (66 kD).

Appendix 7

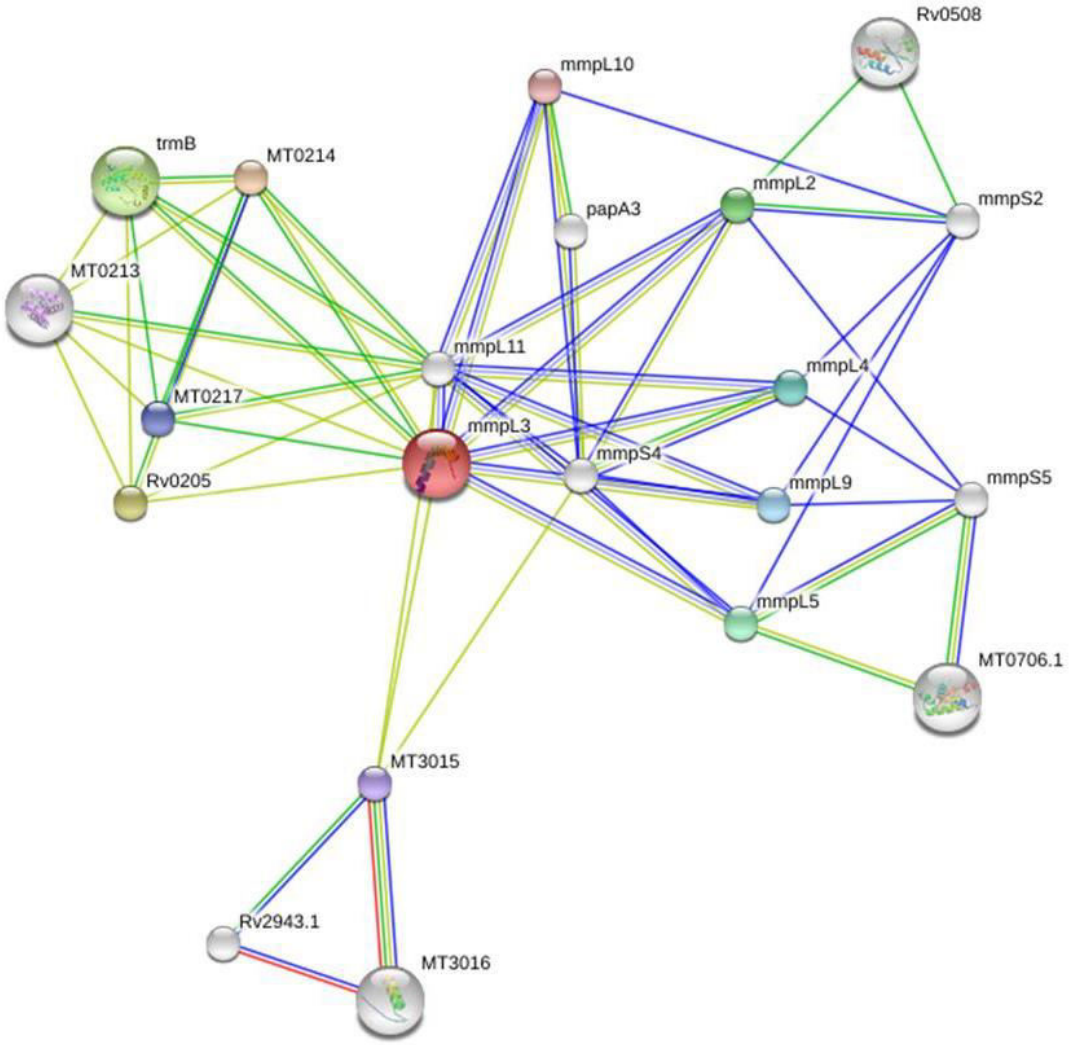


Western blot analysis of MmpL3-MBP fusion protein expression



Western blot analysis of MmpL3-GST fusion protein expression

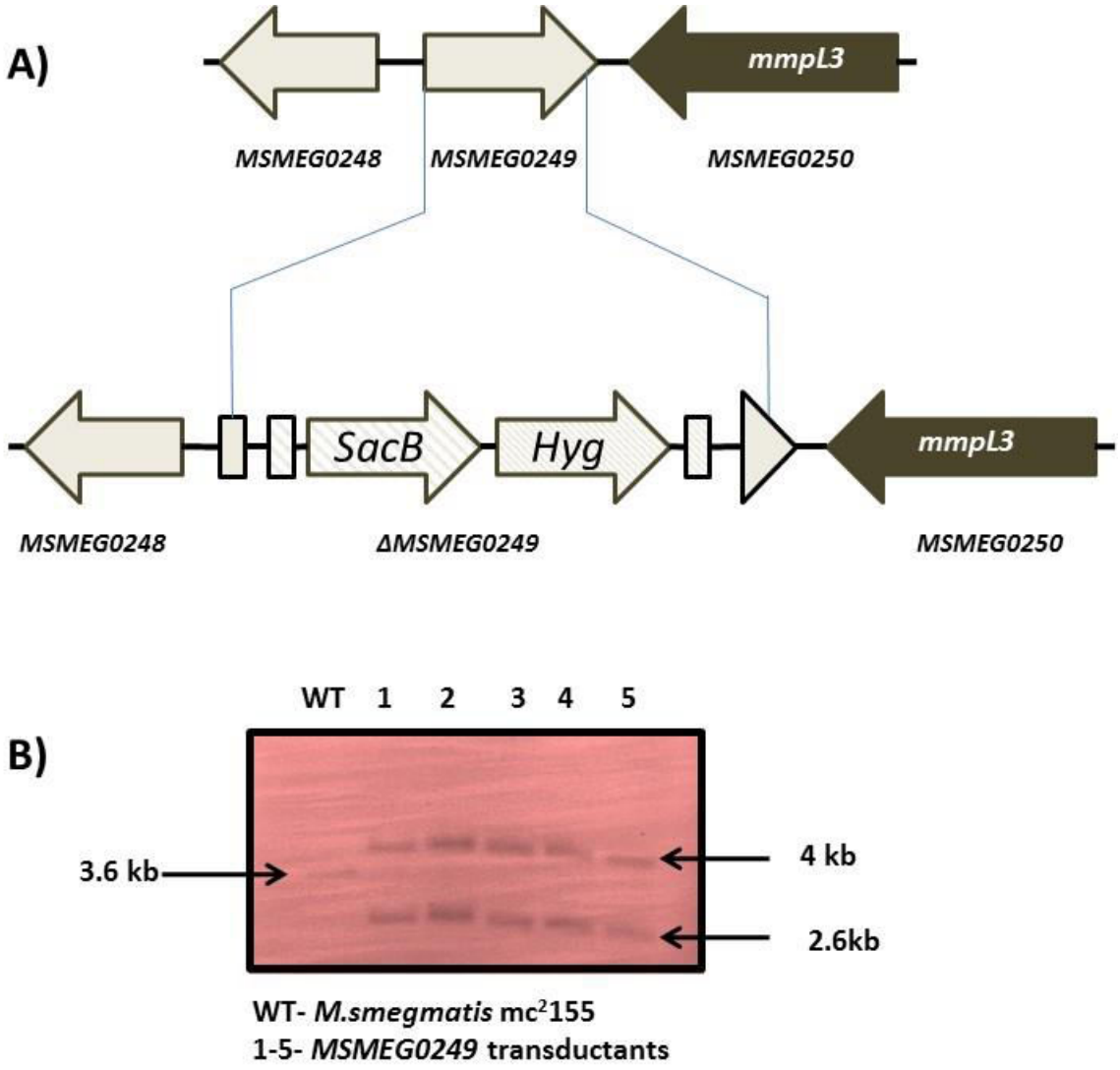
Appendix 8


Rv0508
mmpL10
trmB
MT0214
mmpL2
mmpS2
papA3
MT0213
mmpL11
mmpL4
MT0217
mmpL3
mmpS4
Rv0205
mmpL9
mmpS5
mmpL5
MT0706.1
MT3015
Rv2943.1
MT3016

Appendix 9

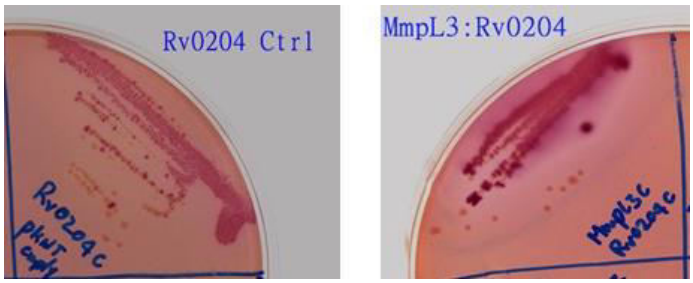
Genes	Primer sequence
MKAN27435_LL	TTTTTTTTCAC**AAA**GTGTTTCGTCGCCAGTGTCAACC
MKAN27435_LR	TTTTTTTTCAC**TTC**GTGAAGTGTGGCCGGATGTTGTC
MKAN27435_RL	TTTTTTTTCAC**AGA**GTGCGCTGCGTTTCGTGTCAATG
MKAN27435_RR	TTTTTTTTCAC**CTT**GTGTCAGGTCCCGAACGTGATAG
MKAN27485_LL	TTTTTTTTCAC**AAA**GTGGTCTGCGGCGCAAGTATCAC
MKAN27485_LR	TTTTTTTTCAC**TTC**GTGCCCATGCGGTTTCCAGAAGC
MKAN27485_RL	TTTTTTTTCAC**AGA**GTGAGTTCGCCGAGGCATTCAAG
MKAN27485_RR	TTTTTTTTCAC**CTT**GTGCGATGAACAGCGCACCAAAG
MKAN27530c_LL	TTTTTTTTCCAT**AAA**TTGGCCGTGTTCGCGCCACATATC
MKAN27530c_LR	TTTTTTTTCCAT**TTC**TTGGCTAAGTCGGCGGGACTAAGC
MKAN27530c_RL	TTTTTTTTCCAT**AGA**TTGGGCACGCTCTTCCGCTATTCG
MKAN27530c_RR	TTTTTTTTCCAT**CTT**TTGGGTCGGTGCAACGCTAGAGTC
MCAN_15191_LL	TTTTTTTTCCAT**AAA**TTGGAGCGGCACACCAGAATAGTC
MCAN_15191_LR	TTTTTTTTCCAT**TTC**TTGG TGCGTACATAGTCGGCTGTC
MCAN_15191_RL	TTTTTTTTCCAT**AGA**TTGG GCCCATGTCCGTAAACAAG
MCAN_15191_RR	TTTTTTTTCCAT**CTT**TTGG ATGGTCCGGTTTCTCTATGC
MCAN_15431_LL	TTTTTTTTCCAT**AAA**TTGGCTAGGTTGCGGGCGATGTAG
MCAN_15431_LR	TTTTTTTTCCAT**TTC**TTGGGTGATGGGCAGGTTCCAAG
MCAN_15431_RL	TTTTTTTTGCAT**AGA**TTGCCGATGATTGGCCGGGCAAAC
MCAN_15431_RR	TTTTTTTTGCAT**CTT**TTGCCGACCGAACCGACCTTTAGC

Appendix 10



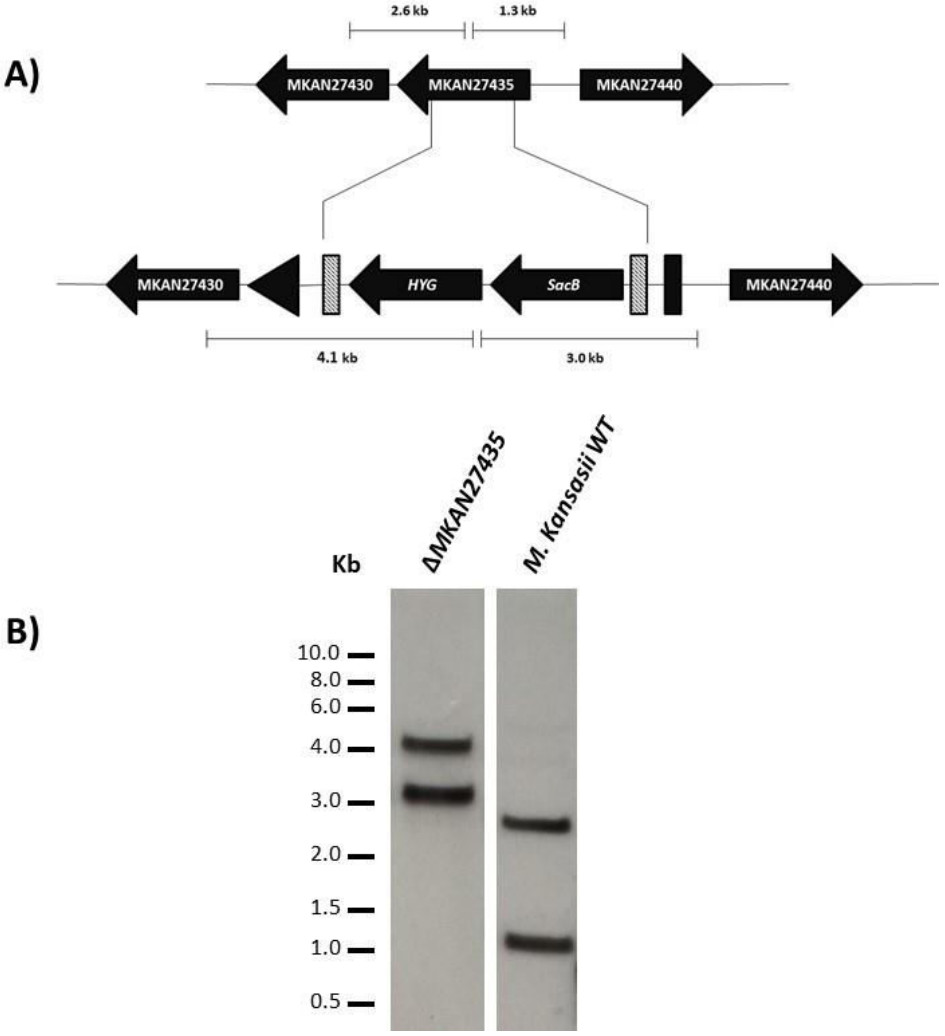
Southern blot analysis of *MSMEG0249* mutant. (A) Map of *MSMEG0249* region in the parental *M. smegmatis* mc^{2}155 and its corresponding region in the Δ*MSMEG0249* mutant. *res*, γδ resolvase site; *hyg*, hygromycin resistance gene from *Streptomycin hygroscopicus*; *sacB*, sucrose counter selectable gene from *Bacillus subtilis*. Digoxigenin-labeled probes were derived from ~1 kb upstream and downstream flanking sequences that were used to construct the knockout plasmid. (B) The gel below shows the Southern blot of KpnI-digested genomic DNA from the two strains with expected bands indicated by arrows.

Appendix 11



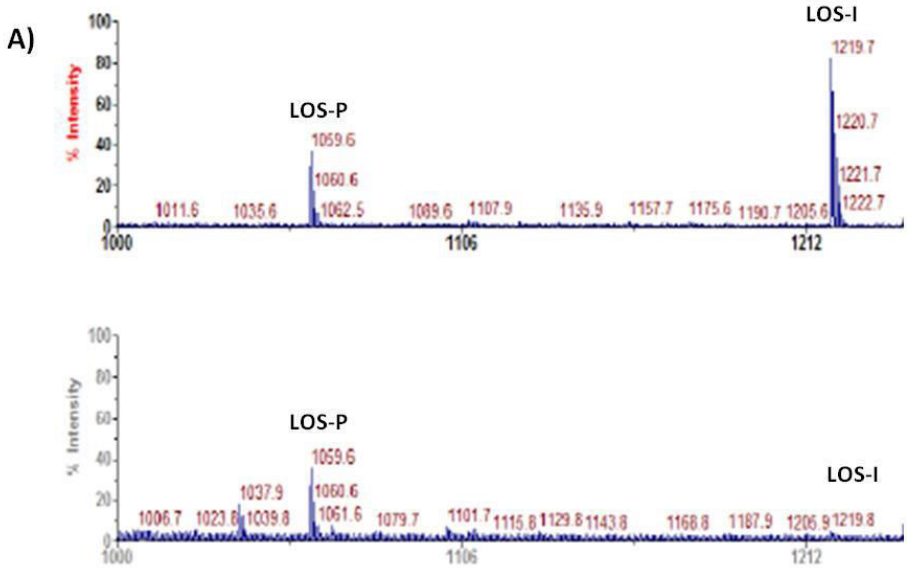
Bacterial two hybrid screening testing for interactions between 'bait' and 'prey' proteins on MacConkey agar plates. The positive interaction observed with both Rv0204 control and its interaction with MmpL3 suggests that the interaction is a false positive.

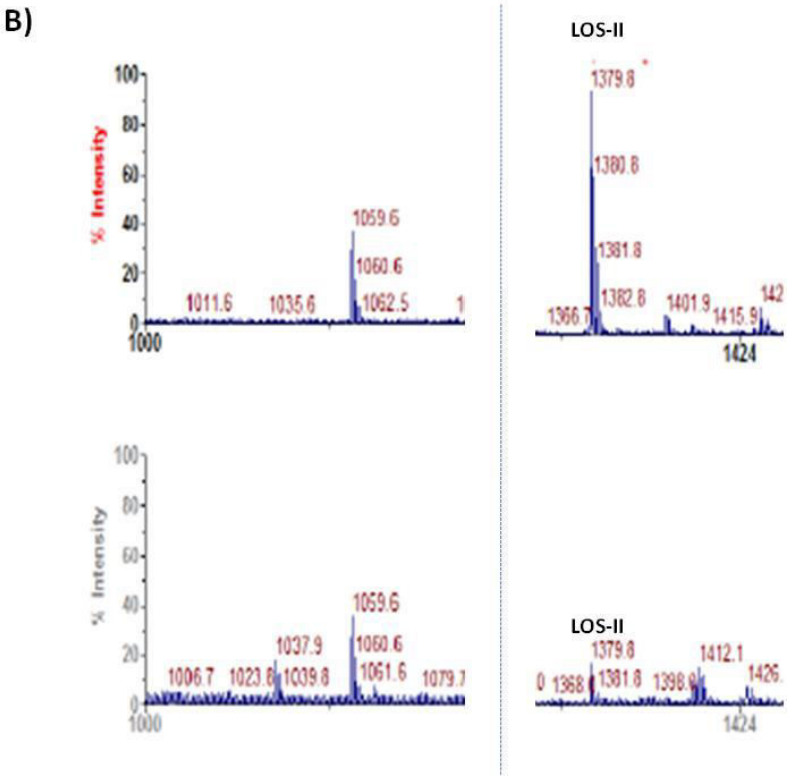
Appendix 12



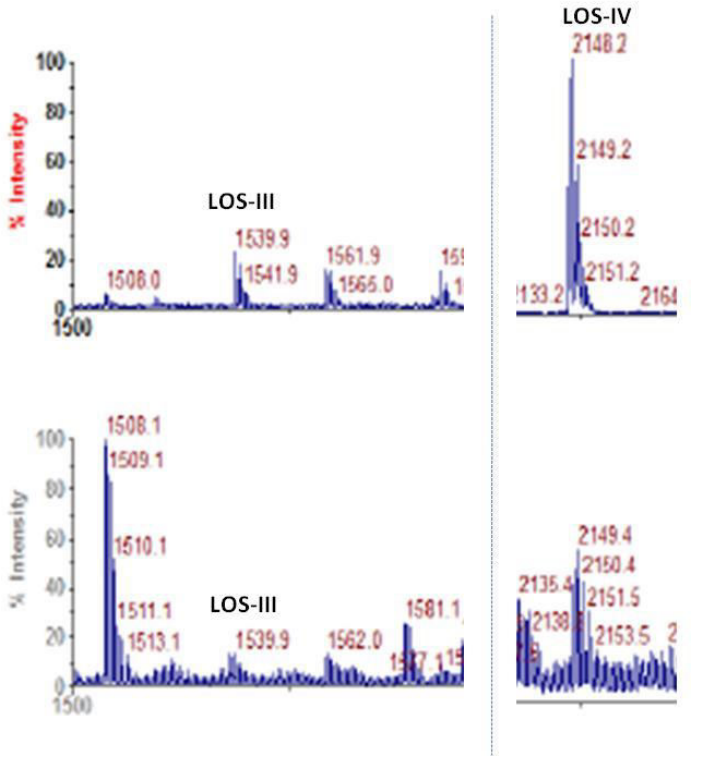
Southern blot analysis of *Mkan27435* mutant. **(A)** Map of *MKAN27435* region in the *M. kansasii* WT and its corresponding region in the mutant; *hyg,* hygromycin resistance gene from *Streptomyces hygroscopicus, SacB,* sucrose counter selectable gene from *Bacillus subtilis.* **(B)** Lanes represents *KpnI* digested genomic DNA from *M. kansasii* WT strain and mutant. Expected restriction digestion pattern: WT- bands corresponding to 2.6 kb and 1.3 kb and Δ*MKAN27435*- bands correspond to 4.1 kb and 3.0 kb; positive control- two bands corresponding to 0.9 kb and 0.7 kb.

Appendix 13

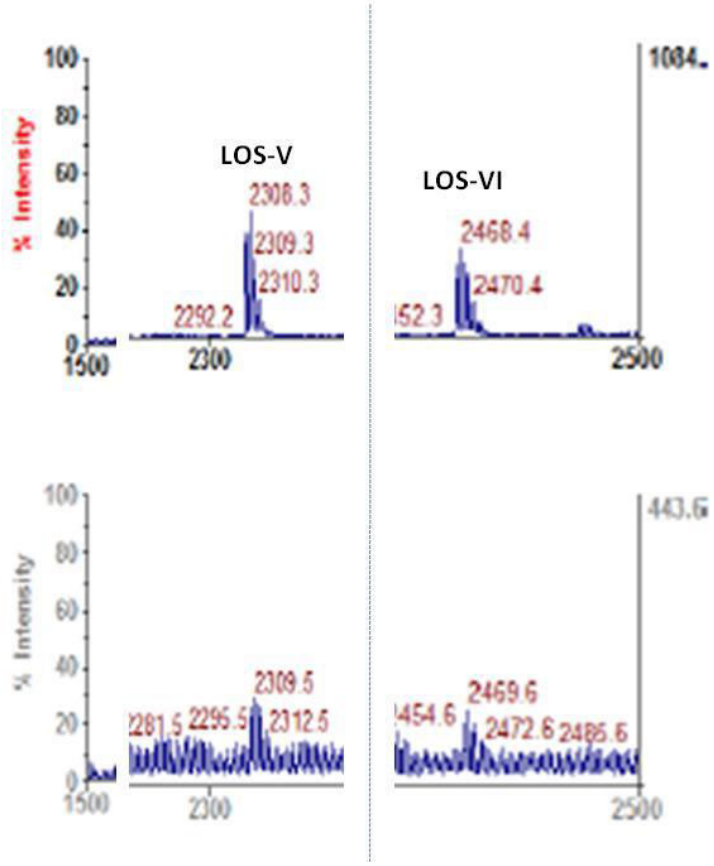

A)
% Intensity
LOS-I
1219.7
LOS-P
1059.6
1220.7
1060.6
1221.7
1062.5
1222.7
1011.6
1035.6
1089.6
1107.9
1135.9
1157.7
1175.6
1190.7
1205.6
1000
1106
1212
% Intensity
LOS-P
1059.6
1037.9
1060.6
1006.7
1023.8
1039.8
1061.6
1079.7
1101.7
1115.8
1129.8
1143.8
1168.8
1187.9
1205.9
1219.8
LOS-I
1000
1106
1212


B)
% Intensity
1059.6
1060.6
1011.6
1035.6
1062.5
1000
LOS-II
1379.8
1380.8
1381.8
1382.8
1401.9
1415.9
1366.7
1424
% Intensity
1059.6
1037.9
1060.6
1006.7
1023.8
1039.8
1061.6
1079.7
1000
LOS-II
1379.8
1412.1
1381.8
1398.9
1426.
1368.6
1424

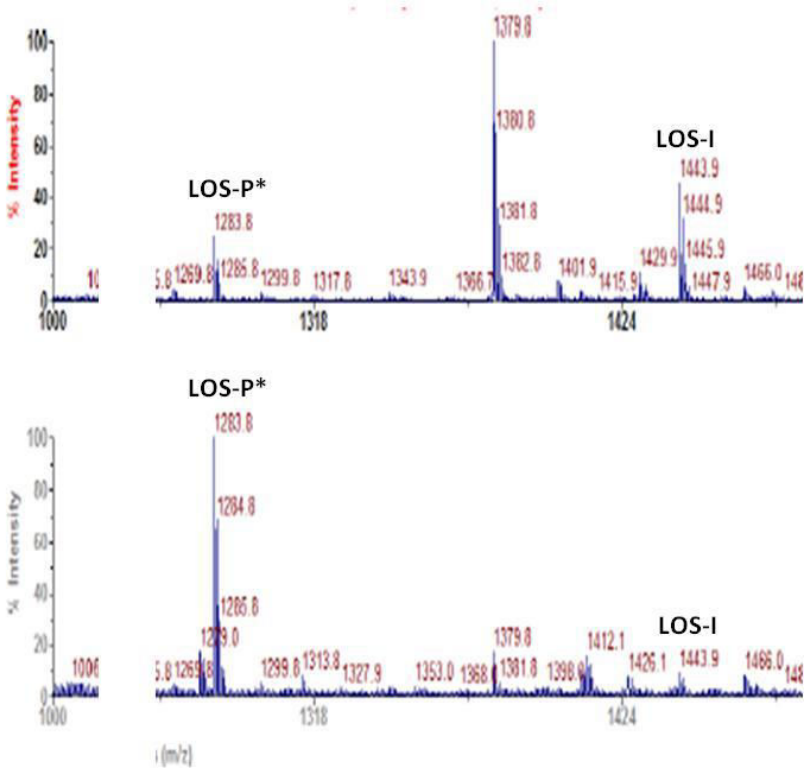
C)


LOS-IV
2148.2
2149.2
2150.2
2151.2
LOS-III
1539.9
1541.9
1561.9
1565.0
1508.0
% Intensity
1508.1
1509.1
1510.1
1511.1
1513.1
LOS-III
1539.9
1562.0
1581.1
2149.4
2150.4
2151.5
2153.5
2135.4

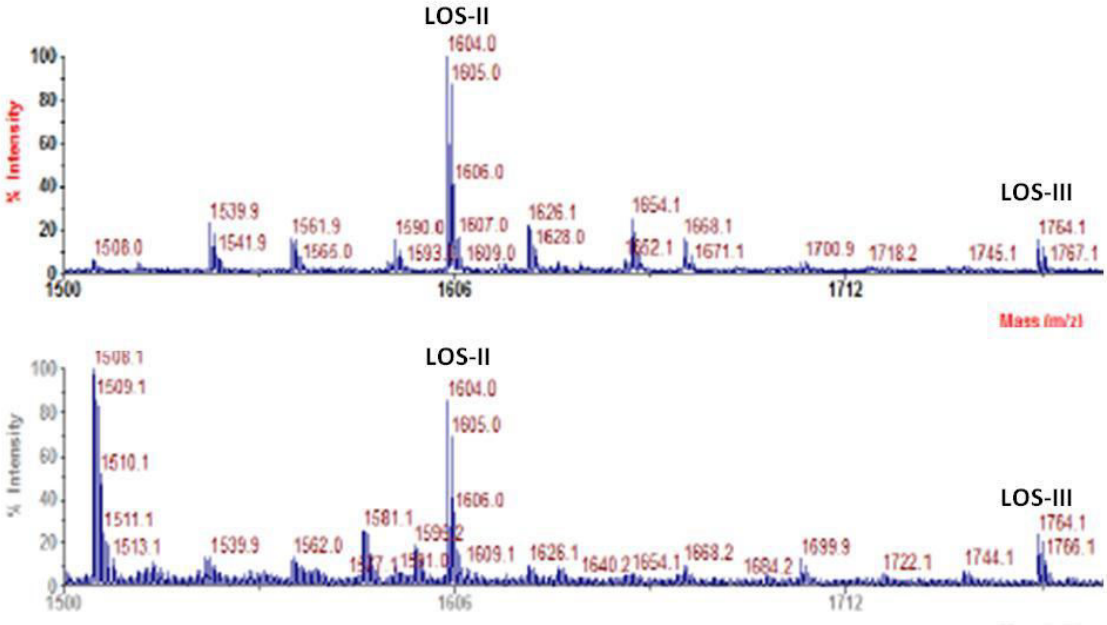
D)


LOS-V
2308.3
2309.3
2310.3
2292.2
LOS-VI
2468.4
2470.4
% Intensity
2309.5
2312.5
2295.5
2469.6
2472.6
2486.6

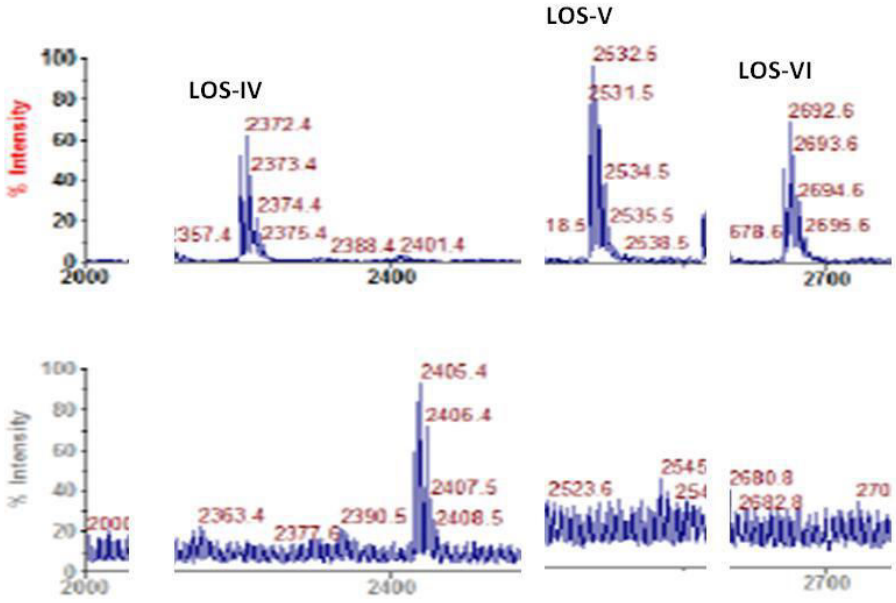
E)


% Intensity
LOS-P*
LOS-I
1379.8
1380.8
1283.8
1443.9
1444.9
1381.8
1445.9
1382.8
1284.8
1286.8
1379.8
1412.1
1443.9
1000
1318
1424
m/z

F)


% Intensity
LOS-II
LOS-III
1604.0
1605.0
1606.0
1539.9
1541.9
1561.9
1626.1
1654.1
1668.1
1764.1
1767.1
1508.0
1607.0
1628.0
1671.1
1700.9
1718.2
1745.1
1500
1606
1712
Mass (m/z)
1508.1
1509.1
1510.1
1511.1
1513.1
1581.1
1609.1
1640.2
1684.2
1699.9
1722.1
1744.1
1766.1

G)



Mass spec analysis of Per-*O*-methylated LOSs from WT and mutant strains

MALDI-TOF analysis of per-*O*- methylated LOSs from WT (top panel) and mutant (lower panel). Fig. A-D per-*O*-methylated LOSs without acyl and Fig. G per-*O*-methylated LOSs with acyl intact, m/z (mol.wt $+Na^+$) corresponding to each LOSs products indicated above the peak