Stony Brook University

Identification of L1334, a unique glycopeptidolipid isolated from a *M. smegmatis MSMEG1604* mutant and functional characterization of *Rv3409c* from *M. tuberculosis*

A Dissertation Presented

By

Jin Gao

To

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Jin Gao

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Nicole S. Sampson, Ph. D., Dissertation Advisor

Professor, Department of Chemistry

Daniel P. Raleigh, Ph. D., Chairperson of Defense

Professor, Department of Chemistry

Kathlyn A. Parker, Ph. D., Third Member

Professor, Department of Chemistry

Robert S. Haltiwanger, Ph. D., Outside Member

Professor, Department of Biochemistry and Cell Biology

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Identification of L1334, a unique glycopeptidolipid isolated from a *M. smegmatis MSMEG1604* mutant and functional characterization of *Rv3409c* from *M. tuberculosis*

By
Jin Gao
Doctor of Philosophy
in
Chemistry
Stony Brook University
2010

M. tuberculosis has a cholesterol oxidizing activity that initiates sterol metabolism, which may be related to its virulence. *Rv3409c* of *M. tuberculosis* H37Rv was annotated as a putative cholesterol oxidase, an enzyme that catalyzes cholesterol modification, based on low similarities with established cholesterol oxidase genes. The *Rv3409c* gene was heterologously expressed in *Mycobacterium smegmatis* mc^2 155 and purified. Using established cholesterol oxidase assays, no cholesterol oxidizing activity was detected. The *M. smegmatis* ortholog *MSMEG1604* shares 83% amino acid identity with *Rv3409c* from *M. tuberculosis* H37Rv. The *M. smegmatis MSMEG1604* transposon mutant (Myc11) oxidized and degraded cholesterol. Therefore, we concluded that *Rv3409c/MSMEG1604* does not encode a cholesterol oxidase. The colony morphology of Myc11 is different from wild type as observed by light microscopy. Lipidomic analysis revealed that a single lipid, L1334, was present only in the Myc11 cell envelope. Accumulation of L1334 in the cell wall of Myc11 resulted in a rough cell morphology

phenotype as compared to the smooth wild-type *M. smegmatis* phenotype. Sliding motility was also retarded by disruption of *MSMEG1604*. Complementation of the Myc11 mutant with *MSMEG1604* resulted in loss of L1334 from the cell envelope. L1334 was purified from Myc11 grown to the stationary phase for identification. Exact mass measurements and thin layer chromatography experiments revealed that L1334 was similar, but not identical to known glycopeptidolipids (GPLs), which are the major surface-exposed lipids of *M. smegmatis*. High resolution and multidimensional mass spectrometry and 1D and 2D NMR spectroscopy indicated that L1334 is an unusual GPL that includes the glycopeptidolipid tetrapeptide core structure, two rhamnose residues, one 6-deoxytalose residue and one 3-methoxy C_{24} fatty acyl chain. The 6-deoxytalose moiety in L1334 was not acetylated as is typically seen in GPLs. The GPL biosynthetic genes have all been identified. Therefore, we propose that *MSMEG1604/Rv3409c* is indirectly involved in lipid production. Mutation of *MSMEG1604* may alter metabolite, e.g. sugar concentrations, or disrupt the production of a regulatory molecule.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

ABTS	2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)
BSA	Bovine serum albumin
$CHCl_3$	Chloroform
CID	Collision induced dissociation
C-Rv3409c	Rv3409c with a C-terminal histidine6 tag
DCIP	2,4-dichloroindophenol
DHBA	2, 3-Dihydroxybenzoic acid
ESI	Electrospray ionization
FAD	Flavin adenine dinucleotide
GMC	Glucose-methanol-choline
GPL	Glycopeptidolipids
3βHSD	3β-hydroxysteroid dehydrogenase
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl β-D-thiogalactoside
LB	Luria broth
MALDI	Matrix-assisted laser desorption ionization
MDR-TB	Multi-drug resistant *Mycobacterium tuberculosis*
M. smeg	*Mycobacterium smegmatis*
M. tb	*Mycobacterium tuberculosis*
Myc11	*Mycobacterium smegmatis MSMEG1604* transposon mutant
NAD^+	β-nicotinamide adenine dinucleotide
NADH	β-nicotinamide adenine dinucleotide reduced form

NADP	β-nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
N-Rv3409c	Rv3409c with a N-terminal histidine6 tag
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein data bank
PGL	Phenolic glycolipids
PMS	Phenazine methosulfate
R. equi	*Rhodococcus equi*
Rha	Rhamnose
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
6-dTal	6-deoxytalose
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
UV-vis	Ultraviolet-visible
WT	Wild type
XDR-TB	Extremely-drug resistant *Mycobacterium tuberculosis*

Acknowledgement

I am very grateful to get many people help me finish this dissertation. This is a great opportunity to express my respect and appreciation. With all of you, my graduate studies at Stony brook are rich of fun and memorable.

I would like to sincerely thank my research mentor Prof. Nicole S. Sampson. With her great support, patience and advices, I can explore my interests on the research and get trained greatly for my future career development. Her precise scientific attitude and passion on science make her a great idol for me. I will miss her and this graduate experience under her supervision.

I would like to thank Prof. Daniel P. Raleigh for teaching me two important graduate courses on Chemical Biology and serving as my dissertation committee chairperson and Prof. Kathlyn A. Parker as my dissertation committee third member. I appreciate their valuable suggestions on my research and patience to my questions. I am also grateful for Prof. Haltiwanger serving as my dissertation committee fourth member.

I would like to thank all people who have been working and studying in Sampson's Lab. Thank you for the helpful discussion, and happy time together. Thank many nice faculties and staffs at Stony Brook University. I will cherish the experience with all of you forever. Thank many friends at Stony Brook. Thank my parents for loving me without any condition. I love being your daughter.

Chapter 1

Introduction

I. *Mycobacterium tuberculosis* overview

M. tuberculosis is one of the most successful pathogens causing human infectious disease. The data from World Health Organization (WHO) indicate that tuberculosis causes about two million deaths every year, and five thousand people die from the disease daily. It is estimated that one third of the world's population is currently infected with *M. tuberculosis* and the number is mounting with 8.8 million new cases each year (*1-2*). Tuberculosis is most prevalent in Africa and Asian countries. About 15% of human immunodeficiency virus (HIV)-infected individuals are co-infected with tuberculosis and about one third of HIV-TB co-infected patients die every year (*3*). Only about 10% of healthy individuals infected with *M. tuberculosis* develop active tuberculosis. In the remaining 90%, the bacterium is in a latent stage. The latent *M. tuberculosis* could be reactivated when host immune system gets compromised due to HIV or malnutrition or aging (*4-5*).

Mycobacterium tuberculosis (*M. tuberculosis*) bacilli were first isolated and identified by Dr. Robert Koch in 1882. It is the causative agent of tuberculosis in humans. The mechanism by which *M. tuberculosis* invades the host immune system and survives in the early phagosome remains to be identified. There are several mechanism models suggesting that *M. tuberculosis* could block phagosome-lysome (P-L) fusion (*6-8*), prevent the phagosome maturation (*9-13*) or disrupt the IFN-γ-mediated signaling pathway (*14-15*).

The traditional tuberculosis treatment is less effective on tuberculosis caused by drug-resistant tuberculosis (DR-TB) and multidrug-resistant tuberculosis (MDR-TB) (*16*). Current tuberculosis disease treatment requires isoniazid, rifampicin, pyrazinamide

and ethambutol or streptomycin using a two-stage dosage. Complete treatment takes a long time and is expensive; therefore many tuberculosis-positive patients discontinue the regimen at later stages of the treatment or take lower drug dosages to save the medical expenses. This incomplete/wrong treatment could trigger the production of drug resistant strains of *M. tuberculosis*, which are more difficult to treat than normal tuberculosis. The challenges to study the *M. tuberculosis* virulence mechanism and develop more effective anti-tuberculosis (anti-TB) drugs are very important for human health in the future.

1.1 *M. tuberculosis* cell wall

M. tuberculosis has a very unique waxy lipid-rich cell envelope (Figure 1-1). Besides plasma membrane, *M. tuberculosis* has a cell wall layer composed of covalently linked peptidoglycan, arabinogalactan and mycolic acids with various non-covalently linked lipids, carbohydrates and proteins, covered by other polysaccharides, proteins and lipids as a capsule structure (*17*). In 1998, the entire genome of *M. tuberculosis* H37Rv was deciphered and about 4,000 genes from its genome were annotated (*18*). Genes involved in cell wall biosynthesis and cell processes and lipid metabolism account for about 24.8% of the genome (Table 1-1) (*19*). A large number of enzymes utilized in cell wall biosynthesis and lipid metabolism are essential for its pathogenicity. Targeting cell wall biosynthesis has emerged as an area for designing anti-TB therapeutics. The cell wall lipid biosynthetic pathway provides a great number of potential anti-TB drug targets because genes that are involved in this pathway do not have human gene counterparts. The effective anti-TB drug, isoniazid, targets inhA, an enzyme involved in the mycolic acid biosynthetic pathway. The inhibition of inhA could cause accumulation of

unsaturated fatty acids, which are toxic to *M. tuberculosis* (*20*). The *M. tuberculosis* cell wall provides a strong barrier against harsh environments, toxic compounds and lysosomal enzymes and also is importantly involved during the bacterial pathogenicity. The studies about mycobacterial cell wall stimulate many scientists' interests.

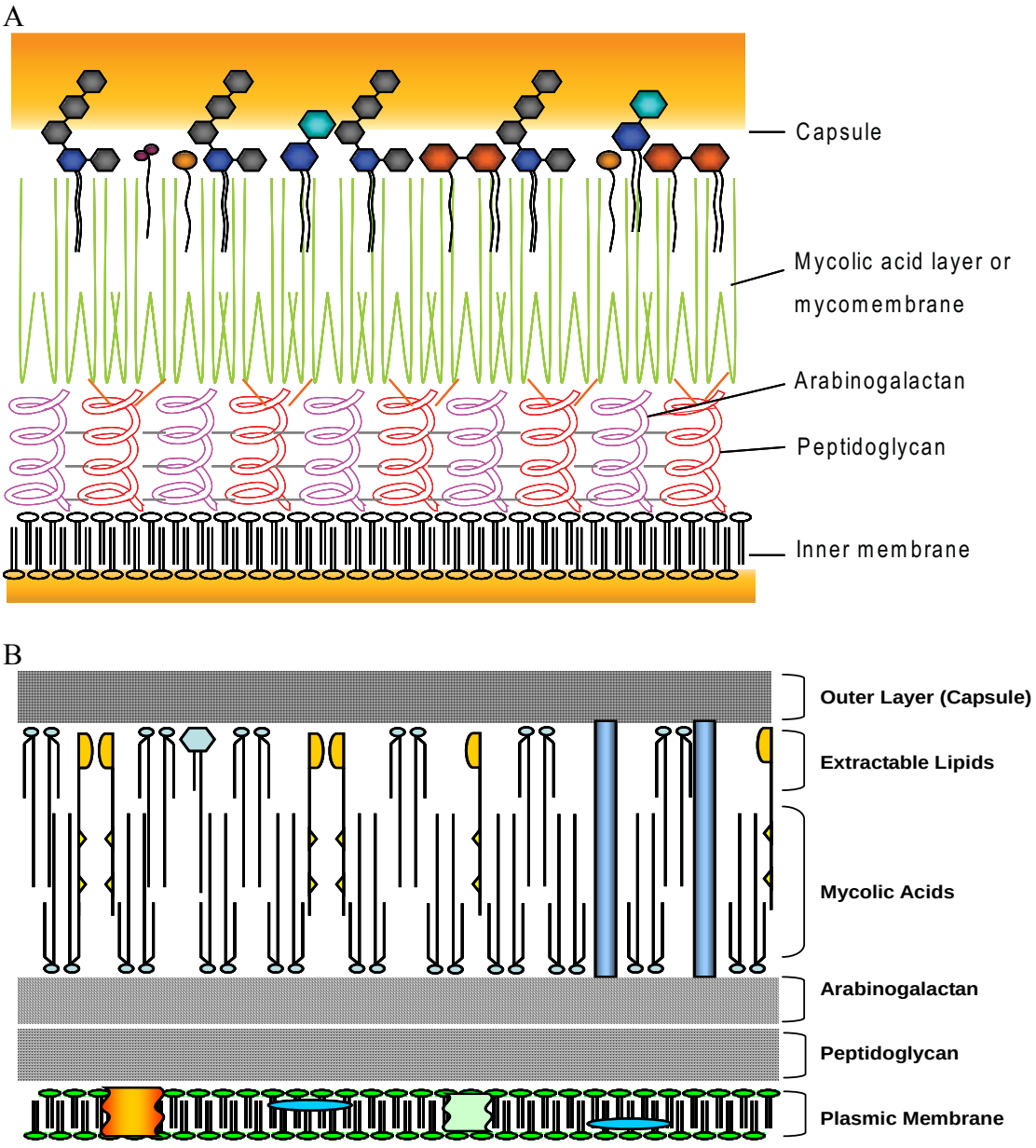


Figure 1-1. A) The *Mycobacterium tuberculosis* cell envelope, modified from reference (*21*) and B) the *Mycobacterium smegmatis* cell envelope (*22*).

Table 1-1. General classification of *M. tuberculosis* genes; cited from reference (*19*).

Function	No. of genes	% of total	% of Total coding capacity
Lipid metabolism	225	5.7	9.3
Lipid metabolism	207	5.2	6.1
Cell wall and cell processes	517	13.0	15.5
Stable RNAs	50	1.3	0.2
IS elements and bacteriophages	137	3.4	2.5
PE and PPE proteins	167	4.2	7.1
Intermediary metabolism and respiration	877	22.0	24.6
Regulatory proteins	188	4.7	4.0
Virulence, detoxification and adaptation	91	2.3	2.4
Conserved hypothetical function	911	22.9	18.4
Proteins of unknown function	607	15.3	9.9

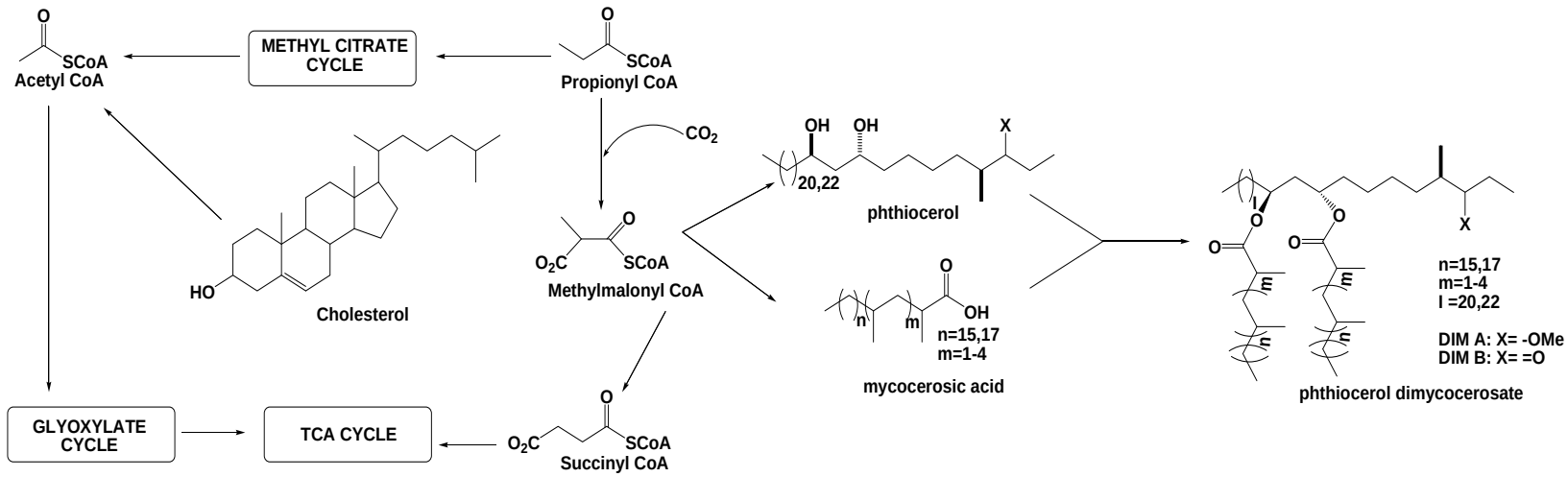
1.2 *M. tuberculosis* and cholesterol metabolism

Mycobacteria do not synthesize cholesterol (*23*) and do not contain cholesterol in their plasma membrane. But many studies indicate cholesterol could be utilized by *M. tuberculosis* for virulence or nutrition purposes. Cholesterol was found to accumulate at the site where *M. bovis* is taken up by macrophages (*24*). Cholesterol rich foamy macrophages also accumulate in lung granulomas in infective mouse (*25*) and human lungs (*26*). *M. tuberculosis* shifts to a lipid-based metabolism in the intracellular host environment (*27-28*). And cholesterol is an abundantly available lipid in the intracellular environment. Isotope labeling of cholesterol indicated that *M. tuberculosis* degrades cholesterol into acetate and propionate, which are further converted into both CO_2 and phthiocerol dimycoserate (PDIM) (*29*). The pathway shown in Scheme 1-1 indicates

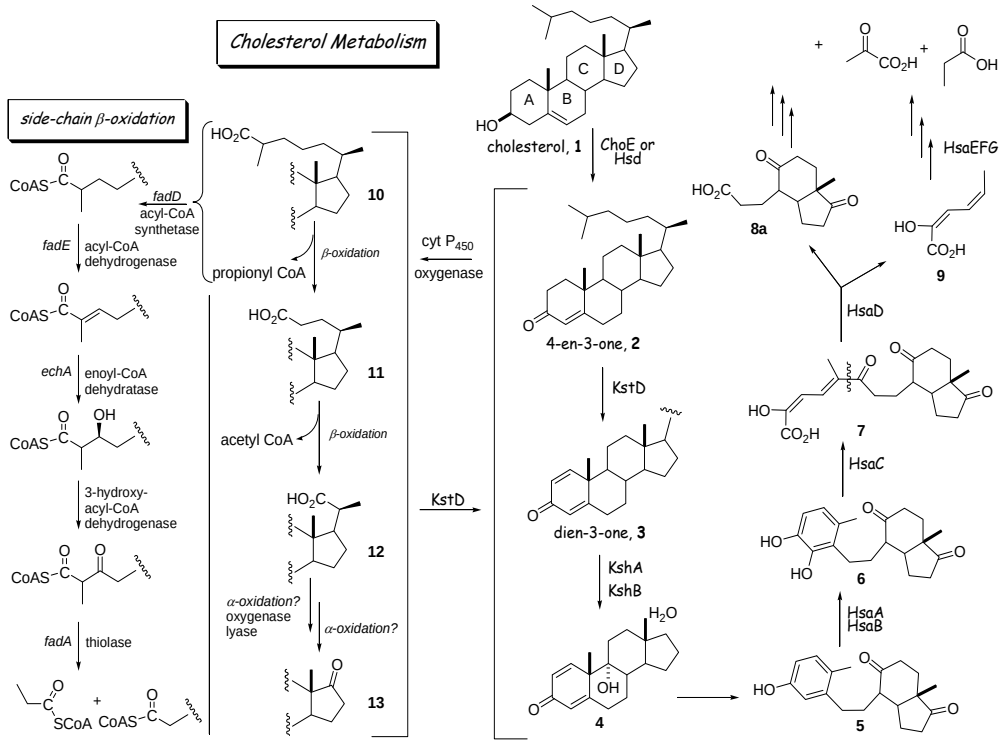
how CO_2 is formed in the citric acid cycle (TCA) cycle and PDIM is biosynthesized from propionyl CoA. PDIM is a surface-exposed lipid of *M. tuberculosis* capsule and plays very important functions in host response regulation (*30*).

In *Rhodococcus* sp RHA1, cholesterol up-regulates a gene cluster of 58 genes (*ro04482-ro04705)*. The conserved genes are also found in *M. tuberculosis* H37Rv (*Rv3492c-Rv3574*), *Mycobacterium bovis* (*M. bovis*) bacillus Calmette-Guein (*Bcg3556c-Bcg3639*) and *Mycobacterium avium* (*M. avium*) *(Map0571-Map0491)* (*31-32*). In the cholesterol metabolism pathway, cholesterol is metabolized through ring degradation and side chain degradation (*33*). Genes from those conserved cholesterol up-regulated gene clusters encode enzymes involved in cholesterol metabolism (Scheme 1-2).

The oxidation of a 3-beta-hydroxyl group to a ketone group is the first step in the cholesterol degradation pathway (*34*). In the cholesterol oxidation reaction, cholesterol is converted to cholest-4-en-3-one in a two-step reaction: oxidation and isomerization. In nature, two enzymes have the cholesterol oxidizing ability: cholesterol oxidase and cholesterol dehydrogenase. In the *M. tuberculosis* genome, there are two genes annotated as a cholesterol oxidase (*Rv3409c*) and cholesterol dehydrogenase (*Rv1106c*) separately (*18*). The enzymatic activity of Rv3409c has not been characterized *in vitro*. Rv1106c has been purified and this enzyme has been confirmed to catalyze the conversion of 3β-hydroxysteroids into a 3-ketosteroids (*35*).



Scheme 1-1. Flux of metabolites from cholesterol catabolism adapted from reference (*36*).



Scheme 1-2. Cholesterol degradation pathway based on cholesterol degradation studies from *Rhodococcus* (*37*), *Comamonas testosteroni* (*38-39*) and fast-growing mycobacteria *(40)*. Modified from reference (*33*) with permission.

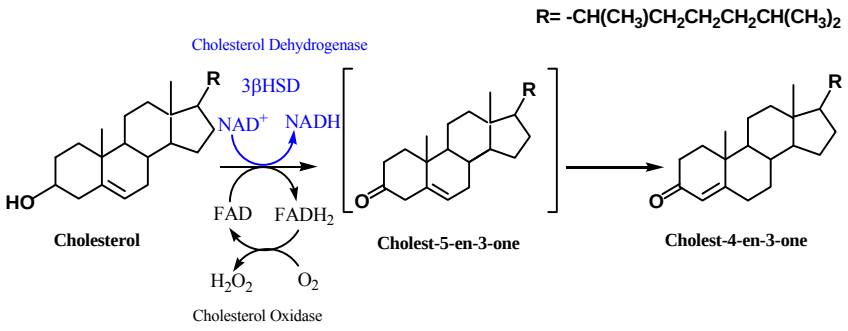
1.3 Cholesterol oxidase and cholesterol dehydrogenase

Cholesterol oxidase could be induced by cholesterol in both pathogenic and non-pathogenic bacteria (*41*). The membrane damaging ability of cholesterol oxidase is considered as a virulence factor for some cholesterol oxidase-producing pathogen bacteria. *R. equi* is a Gram positive coccobacillus bacterium found in soil and most commonly infects foals, and also infects immunocompromised humans (*41-42*). In the *R. equi* genome, there is an identified cholesterol oxidase ChoE and a putative cholesterol oxidase (ChoD). The similarity of ChoD and ChoE is about 25%. ChoE is more close to cholesterol oxidase (ChoA) from *Streptomyces* species, about 65% identity.

Cholesterol oxidases are water-soluble interfacial flavoenzymes with one

molecule of flavin adenine dinucleotide (FAD) cofactor covalently (Type II) or non-covalently (Type I) bound. Although the two types of cholesterol oxidases catalyze the same reaction, the protein structures and sequences are completely different (*34*).

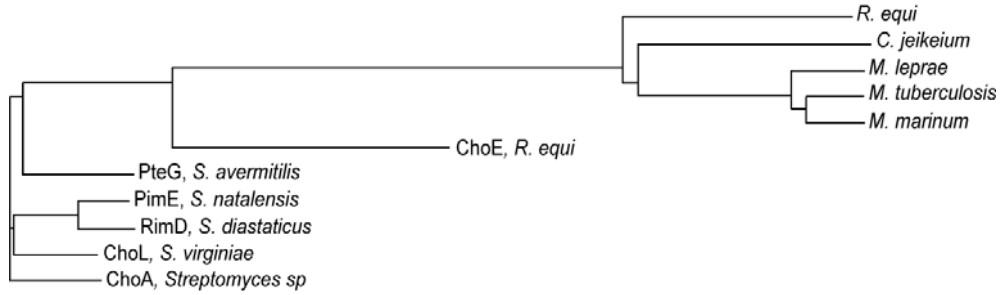
Cholesterol dehydrogenases are found in *Norcardia*. Cholesterol dehydrogenase plays an important role in the virulence of vaccinia virus (*43*). Cholesterol dehydrogenase utilizes an NAD and/or NADP as the cofactor for activity instead of FAD. Different from the cholesterol oxidase reaction, oxygen is not involved in the cholesterol dehydrogenase reaction. Reaction mechanisms of cholesterol oxidase and cholesterol dehydrogenase are shown in Scheme 1-3.



Scheme 1-3. Cholesterol oxidizing reaction catalyzed by cholesterol oxidase with FAD as cofactor or cholesterol dehydrogenase with NAD^+ as cofactor.

Bioinformatic analysis indicates that there are no cholesterol oxidase orthologs of *choA* from Streptomycete or *choE* from *Rhodococcus equi* (*R. equi*) in mycobacteria. On the basis of bioinformatic prediction, mycobacteria like *M. tuberculosis* contain *Rv3409c* annotated as putative cholesterol oxidase and *Rv1106c* annotated as a putative cholesterol dehydrogenase. In other mycobacterial genomes both *Rv3409c* and *Rv1106c* are conserved.

Fusion of the first 95 amino acids from the N-terminus of Rv3409c to green fluorescence protein has been reported to promote secretion of green fluorescence protein (*44*). However Rv3409c does not contain a signal peptide sequence as judged by analysis by sequence analysis and Rv3409c is most likely an intracellular enzyme (*33*).



Scheme 1-4. Phylogenetic tree of functionally characterized and proposed cholesterol oxidases from Streptomyces, Rhodococcus and *Mycobacteria*, cited from reference (*33*).

The *M. tuberculosis* cholesterol oxidizing ability is abrogated when *Rv1106c* is disrupted by insertion of a kanamycin gene cassette into this gene (*35*). The *MSMEG1604* (*Rv3409c* ortholog in *M. smegmatis*) transposon mutant of *M. smegmatis* and *Rv3409c* transposon mutant of *M. tuberculosis* grows on medium containing cholesterol as the only carbon source in our studies.

The phylogenetic tree shows that all mycobacterial homologues of *Rv3409c* are highly similar to *choE* from *Rhodococcus*, but the similarity of *Rv3409c* from mycobacteria is low with other well studied cholesterol oxidases eg. *choA* from Streptomycete (Scheme 1-4). The cholesterol oxidizing ability has been reported in mycobacteria, but so far no cholesterol oxidase has been isolated and identified from mycobacteria. Bioinformatic analysis on protein enzyme function is very limited and could predict a wrong function easily because the information protein data base is

still being updated. That Rv3409c enzyme functions as a cholesterol oxidase is doubtful. Cellular lysates of *M. smegmatis* overexpressing *MSMEG1604* were reported to have cholesterol oxidase activity (*45*), but the Rv3409c enzyme has not been purified and the activity has not been characterized in *in vitro* assays. The cholesterol oxidizing activity could be due to the Rv1106c orthologue that is conserved in *M. smegmatis* mc^2155.

II. *Mycobacterium smegmatis* mc^2155

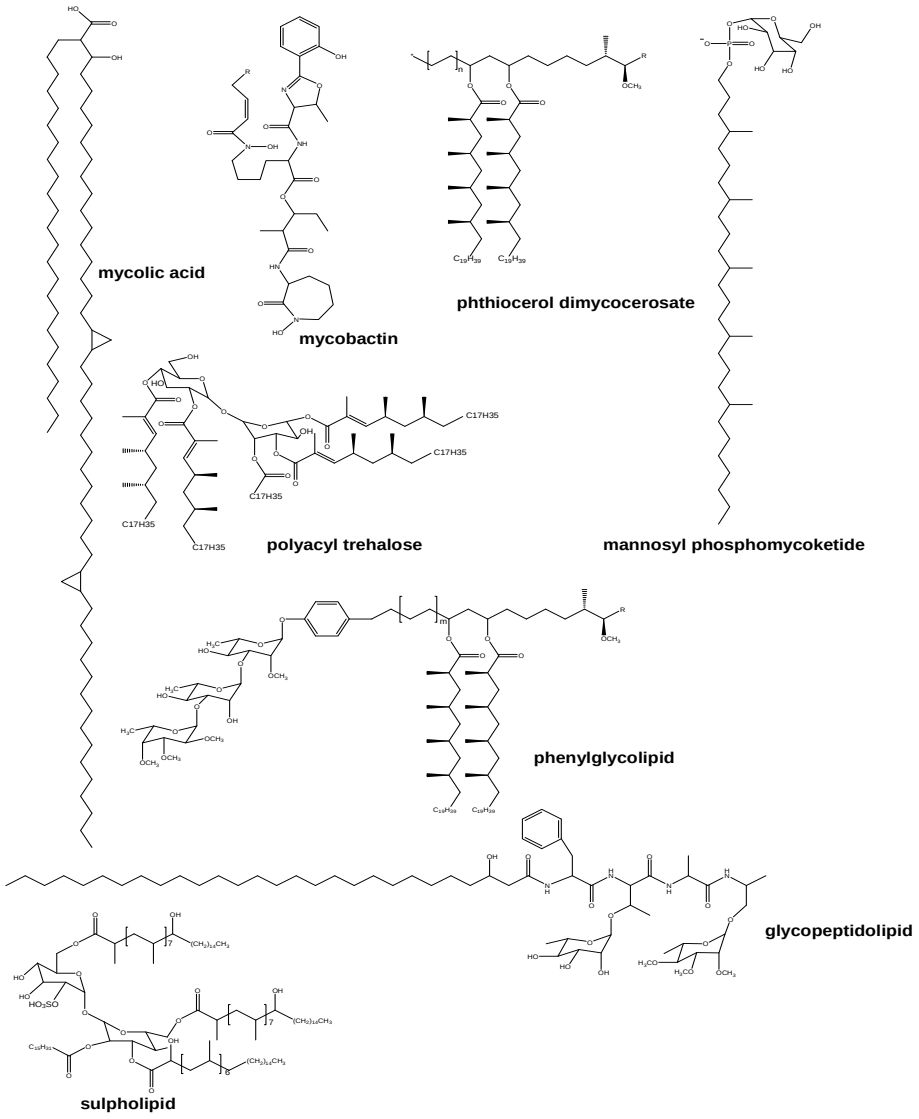
Mycobacterium smegmatis is a saprophytic non-pathogenic fast-growing mycobacterium. The whole genome was sequenced by TIGR (http://www.tigr.org). The cell wall of the mycobacterium contains a lot of unique and structurally-complicated lipid components. *M. smegmatis* mc^2155 cell wall is the simplest among mycobacteria. It is routinely used as the laboratory mycobacterial model to study pathogenic mycobacteria. It is also a good heterologous over-expression system for other genes from pathogenic mycobacteria (*46-47*).

2.1 *M. smegmatis* cell wall

M. smegmatis cell wall is composed of three layers: the plasma membrane, the periplasmic layer and the outer capsule. The extractable lipid components from *M. smegmatis* include: triacylglycerols (TAGs), glycopeptidolipids (GPLs), trehalose dimycolate (TDMs), trehalose monomycolates (TMMs), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylinositol (PI) and phosphatidylinositolmannosides (PIMs) (*48*). The structures of some polyketide lipids are shown in Scheme 1-5.

M. smegmatis mc^2155 colony morphology and correlation with lipids components in the cell wall has been characterized. *M. smegmatis* mc^2155 strains

have been reported with many different morphologies. But generally *M. smegmatis* strains are grouped into two morphology phenotypes: rough morphology and smooth morphology. The cell colony morphology of *M. smegmatis* mc^2155 is correlated to the production and structure of a surface exposed glycolipid–glycopeptidolipid (GPL). In *M. smegmatis*, genes involved in GPL biosynthesis were disrupted to study colony morphology changes. The *M. smegmatis gtf3* (encoding a glycosyltransferase) and *M. smegmatis atf* (encoding an acetyltransferase) and *M. smegmatis gap* (GPL addressing protein) produce different phenotypes from wild type due to changes in the GPLs (*49-50*).

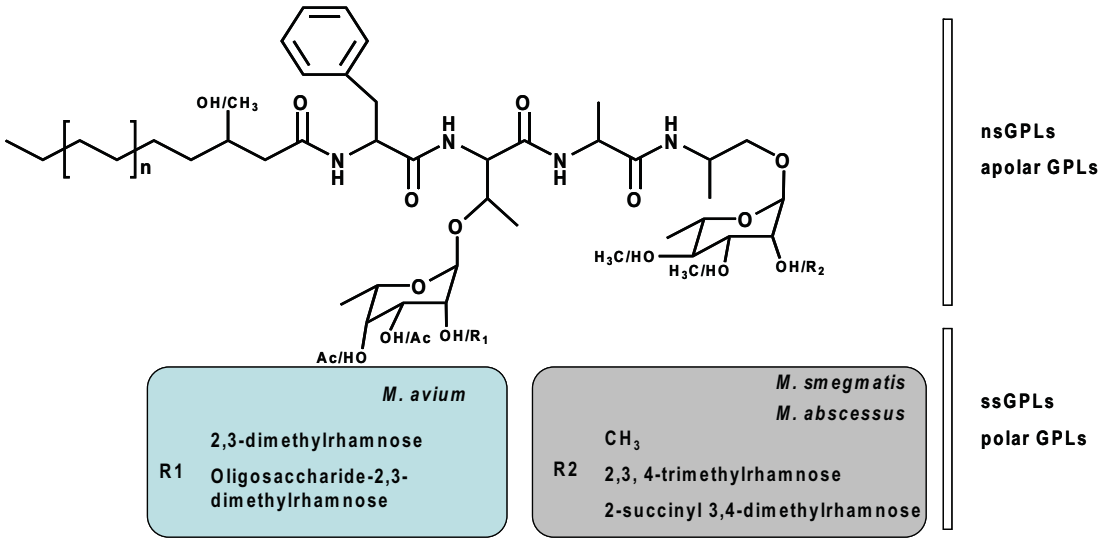


Scheme 1-5. Some polyketide lipids from mycobacteria (*5, 17*).

2.2 Glycopeptidolipids

Glycopeptidolipids (GPLs), also called C-mycosides, are amphiphilic lipids located in the outermost layer of mycobacterial cell walls. The structure of GPL was identified as early as 1960 (*51*). All GPLs share one common tetrapeptide core structure of D-Phe-D-allo-Thr-D-Ala-L-alaninol except for one immunogenic GPL X-I isolated from *Mycobacterium xenopi* (*M. xenopi*) strain NCTC 10042 (*52-53*). GPL production and structures are species-dependent and environment-dependent. There are generally two types of GPLs, non-specific GPLs (nsGPLs) and species-specific GPLs (ssGPLs) (Scheme 1-6). For non-specific GPLs (nsGPLs), the allo-Thr residue is linked with 6-deoxytalose (6-dTal) and the L-alaninol residue is substituted with O-methylated rhamnose. 6- d Tal and rhamnose can be o modified by O-acetylation and O-methylation (*54*). *M. smegmatis* GPLs are not further glycosylated on the 6-dTal residue, while M. avium produces serovar -specific GPLs in which 6-dTal could is further linked to haptenic oligosaccharides (*55-57*).

GPLs are related to mycobacterial colony morphology, sliding motility and biofilm formation (*58*). Besides affecting the cell morphology, GPLs isolated from pathogen mycobacterium *M. avium* and *M. abscessus* also have important biological activities (*59*). *M. avium* which belongs to *Mycobacterium avium* complex (MAC) can cause pulmonary diseases in humans. Isolated GPLs are able to disrupt the macrophage membrane (*60*) and inhibit the human macrophages phagocytosis of mycobacterium (*61*). GPLs also protected MAC by blocking lysosomal enzymes outside the cell (*62*). More recently, GPLs were reported to induce TNF-α synthesis in murine macrophages (*63*). The virulence role played by GPLs could be important for GPL-expressing pathogens.



Scheme 1-6. Common *M. smegmatis* isolated glycopeptidolipids structure adapted from (*64*).

The GPLs biosynthetic gene cluster of *M. smegmatis* was proposed (Scheme 1-7). The GPLs biosynthetic pathway in *M. smegmatis* is shown in Scheme 1-8. There is a 65 kbp size gene locus containing about GPL 25 biosynthesizing genes in the genome (*65*). The gene functions are listed in Table 1-2. Two *mps* genes, *mps1* and *mps2*, are in *M. smegmatis* GPLs locus. *mps* genes encode a peptide synthetase active in transferring amino acids to a peptide acceptor as well as converting the first three amino acid residues to D-isomers. The *pks* gene encoding a polyketide synthase is suggested to be required for synthesis of the monounsaturated β-hydroxyl fatty acid linked to the tetrapeptide core. Cloning experiments have not been accomplished because of *pks* large gene size, 11 kbp. Bioinformatics approaches have revealed that other genes in this GPL locus may be important for fatty acid attachment to the tetrapeptide core. FadD23, the homologue of FadD28 in *M. tuberculosis* catalyzes activation for mycocerosic acid acyl transfer. *papA* genes, which encode a group of acyltransferases, are proposed be involved in polyketide synthesis. For example, during PDIM biosynthesis the *M. tuberculosis papA5* catalyze diesterification. *M.*

smegmatis papA3 might be involved in the polyketide biosynthesis steps for GPL biosynthesis. FadE5, an acyl coA dehydrogenase may catalyze dehydrogenation of the long chain fatty acid unit. Variable modifications to the lipidopeptide core including glycosylation, O-methylation, and O-acetylation provide the different GPLs structures. Glycosyltransferases gtf1, gtf2 and gtf3 are proposed to transfer the rare hexoses moieties (deoxyltalosyl and rhamnosyl) to link to the glycolipidpeptide core units. gtf1 and gtf2 are responsible for transferring deoxytalose and rhamnose respectively while gtf3 is involved in transferring sugar to the triglycosylated GPL form. *rmlA*, encoding a putative glucose-1-phosphate thymidylytransferase and *rmlB*, encoding a putative dTDP-glucose-4,6 dehydrogenase are probably involved in biosynthesis of rhamnose. Four methyltransferases rhamnosyl-3-O-methyltransferase, rhamnosyl-4-O-methyltransferase and rhamnosyl-2-O-methyltransferase are all responsible for O-methylation of rhamnose. fmt, is responsible for methylating the hydroxyl group on the fatty acid unit of GPLs. O-methylation first happens to O3 position followed by O4 and O2. *aftA*, encoding a putative acetyltransferase is responsible for O-acetylation of GPLs.

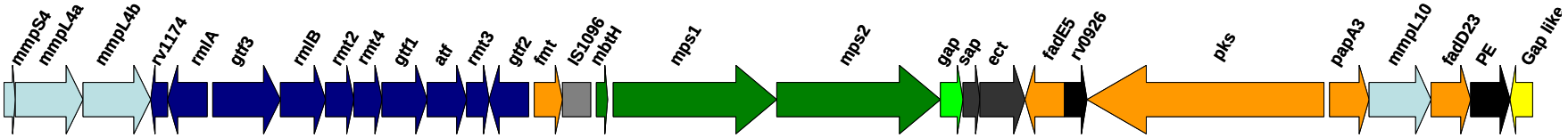
Within the GPL biosynthetic locus, there are four *mmp* genes and one *gap* gene proposed to be involved in GPL transport to the cell envelope surface. The *gap* gene has been characterized by Sondon and coworkers in 2005 in exporting GPL to cell surface (*59*). Three transmembrane proteins *mmpS4*, *mmpL4a* and *mmpL4b* located upstream of the *gap* gene are supposed to be involved in both GPL biosynthesis and transportation. The down stream *mmpl10* function in GPL remains unclear.

2.3 Phenolic glycolipids

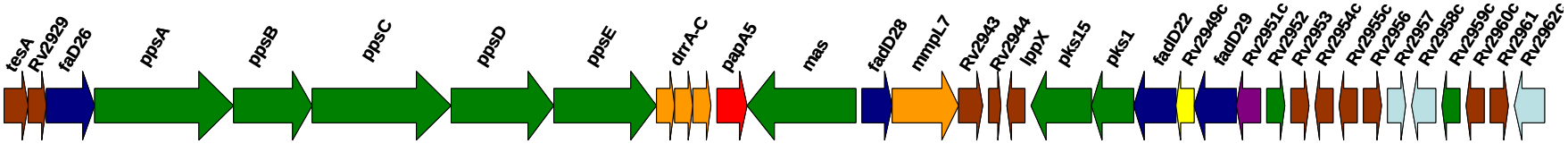
The *M. tuberculosis* cell wall does not contain glycopeptidolipids (GPLs). Phenolic

glycolipids (PGLs) are proposed as functional homologs of GPLs (*66*). PGLs are produced as surface exposed lipids by *M. tuberculosis*, *M. leprae* and other slow-growing mycobacteria, most of which are pathogens (*67-68*). No mycobacterium strain has been found to produce both GPLs and PGLs in its cell wall (*64*). PGLs share a common lipid, phthiocerol diester with phthiocerol dimycocerosates (PDIM). PGLs and PDIMs are mixtures of long chain β-diols esterified by polymethyl-branched fatty acids (*69*). Recently, a new type of lipid compound structurally related to PGLs has been isolated from *M. tuberculosis*, the glycosylated p-hydroxybenzoic acid methyl esters (β-HBAD) (*70*).

M. smegmatis



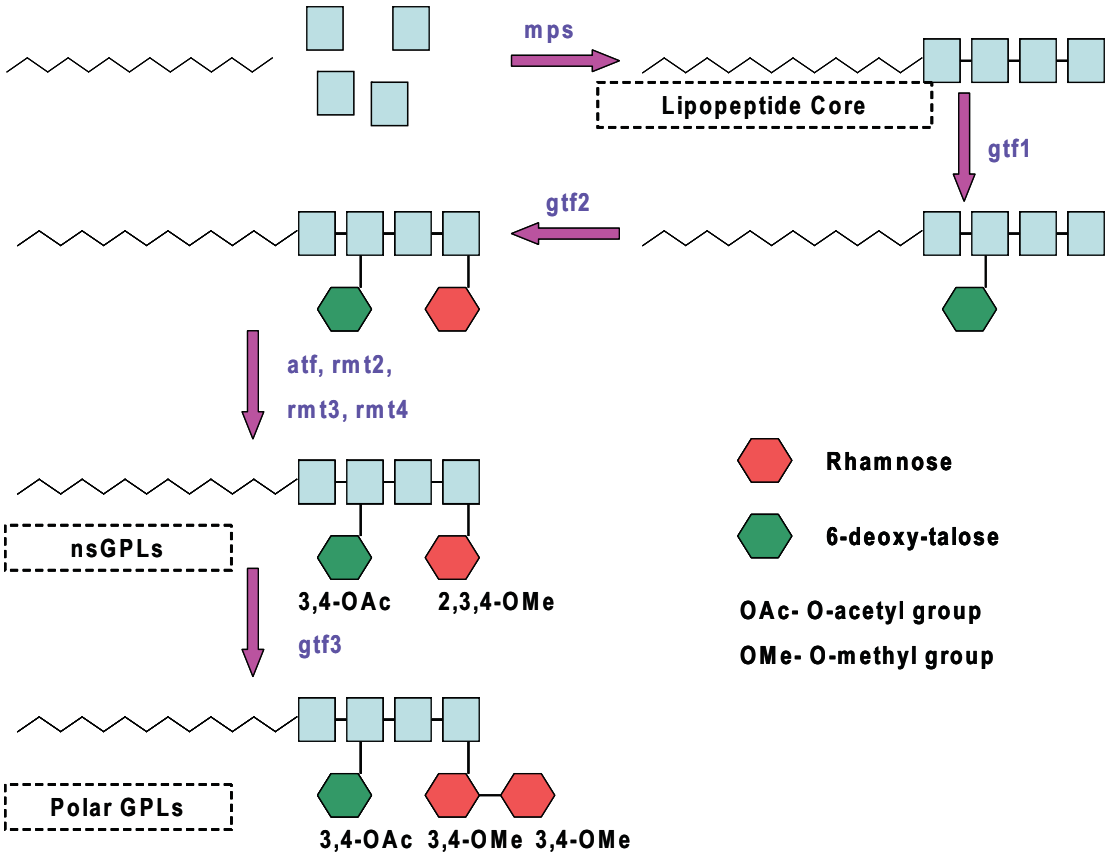
M. tuberculosis



Scheme 1-7. *M. smegmatis* GPL biosynthesis gene cluster (*65*) and *M. tuberculosis* PGL biosynthesis gene cluster (*71*).

Table 1-2. Proposed functions of genes within GPL and PGL biosynthesis cluster (*71-72*).

GPLs biosynthesis (*M. smegmatis*)		PGLs biosynthesis (*M. smegmatis*)	
Gene	**Proposed function**	**Gene**	**Proposed function**
mmpS4	Membrane associated. Interaction with pks	*ppsB*	Linear PKS
mmpL4a	Membrane associated. Interaction with pks	*ppsC*	Linear PKS
mmpL4b	Membrane associated. Interaction with pks	*mas*	Type I iterative PKS
Rv1174	None	*pks7*	Type I iterative PKS
rmlA	Alpha-D-hexose-1-phosphate-thymidylyl-transferase	*pks12*	Type I iterative PKS
gtf3	D-Rhamnose rhamnosyltransferase	*Pks10*	CHS
rmlB	UDP-hexose 4-epimerase	*pks11*	CHS
rmt2	Rhamnose 2-O-methyltransferase	*pks15/1*	Type I iterative PKS
rmt4	Rhamnose 4-O-methyltransferase	*papA5*	Acyltransferase
gtf1	D-allo-Threonine 6-deoxytaloseyltransferase	*drrB*	ABC transporter
atf	Integral membrane protein. 6-deoxytalose 3,4-O-acetltransferase	*drrC*	ABC transporter
		mmpL7	-
rmt3	Rhamnose 3-O-methyltransferase	*fadD26*	-
gtf	L-alaninol rhamnosyltransferase	*fadD28*	-
fmt	Fatty acid O-methyltransferase	*tesA*	Type II thioesterase
mps1	Non-ribosomal protein synthase. Synthesis of the dipeptide	*Rv2951*	Oxidoreductase
mps2	Non-ribosomal protein synthase. Synthesis of the amino acid alcohol	*Rv2952*	Methyltransferase
		Rv2958c	Glycosyltransferase
gap	Integral membrane protein. Required for GPL export	*Rv2957*	Glycosyltransferase
ecf	Sigma factor of the ECF family. Required for Regulation.	*Rv2959*	Methyltransferase
fadE5	Fatty acid desaturase.	*Rv2962c*	Glycosyltransferase
pks	Fatty acid Synthesis and activation	*nat*	Arylamine
papA3	Transfer of the Pks-bound fatty acid to the pseudotetrapeptide		N-acyltransferase
		Mb0100	Oxidoreductase
mmpL10	Membrane associated. Interaction with pks	*(Rv0097)*	
fadD23	Long chain fatty acyl-AMP ligase		
gap-like	Integral memrane protein. Role in the transport of GPLs.		



Scheme 1-8. *M. smegmatis* GPL biosynthetic pathway modified from reference (*73*).

PGLs may have an important role in pathognesis during *M. tuberculosis* virulence. Earlier studies showed that the PGL-type I and de-acylated PGL-type I from *M. leprae*, protect *Staphylococcus aureus* (*S. aureus*) against oxidative stress from molecules like hydrogen peroxide (H_2O_2) and OH radical *in vitro*, and dPGL-I help *S. aureus* survive inside of the human monocyte-derived macrophages (*74*). Reed and co-workers found that PGLs from *M. tuberculosis* suppress the pro-inflammatory TNF-alpha, interleukin (IL)-6 and IL-2 and monocyte chemotactic protein-1 (MCP-1) production (*30*). The clinical isolate HN878 *pks15/1* mutant, which was not able to produce PGLs shows attenuated infection of the mouse and rabbit (*30, 75*). There was also a dose-dependent inhibition activity of the pro-inflammatory response when treating mouse bone-marrow macrophages with

isolated PGLs (*30*). Lipids extracted from CDC1551 lacking PGLs stimulate phagocyte activation by inducing IL-12 and other molecules. The lipid extracts containing PGLs will inactivate the phagocyte by inducing IL-4 and IL-13 (*76*). Although direct evidence for PGLs in *M. tuberculosis* virulence has not been identified, it is believed that PGLs from *M. tuberculosis* (PGL-tb) contribute by modulating the immune response in the host and delaying or disrupting immunity (*66*).

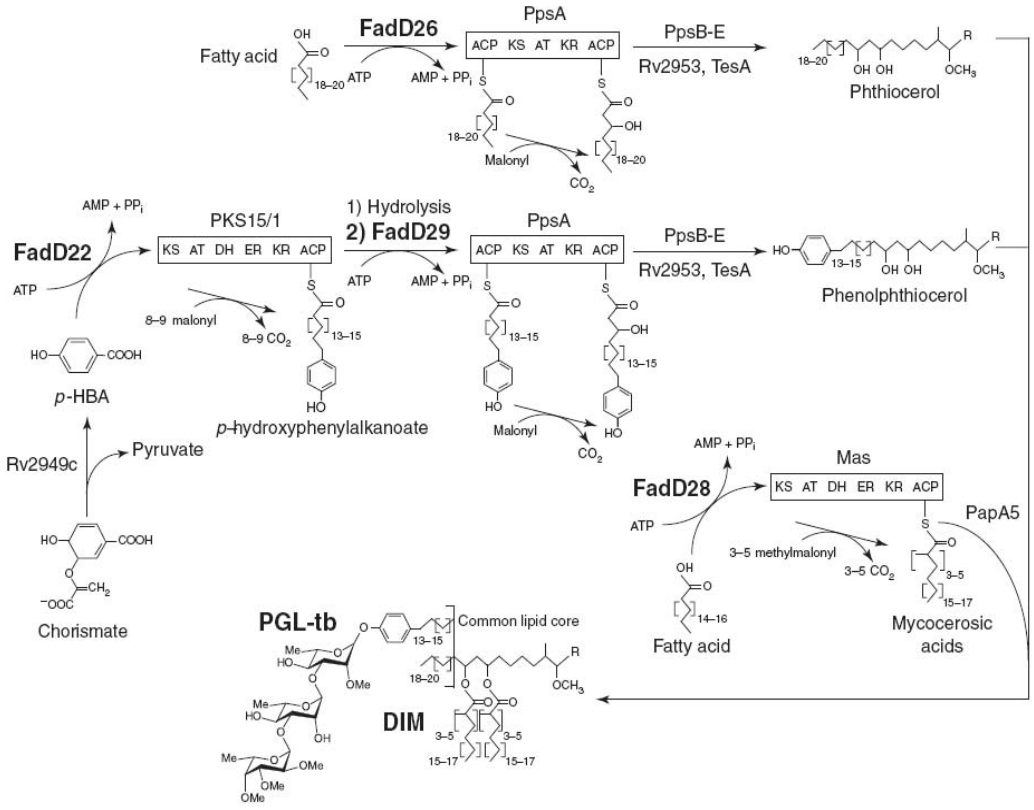
About 20 genes for DIM, PGL and *p*-HBAD biosynthesis and translocation are clustered within a 73 kbp region of the *M. tuberculosis* chromosome (*18, 71*) (Scheme 1-7). The PGL biosynthetic pathway is shown in Scheme 1-8. The PGL biosynthetic gene cluster contains proteins responsible for phthiocerol and phenolphthiocerol chain synthesis (*Rv2929c, fadD26 and ppsA-E*), mycocerosate chain synthesis and transfer (*mas, fadD28*), phenolphthiocerol chain synthesis (*pks1/15*), *p*-HBAD synthesis (Rv2949), PGL and *p*-HBAD glycosidic domains formation (*Rv2954c -Rv2959c*) and DIM, PGL translocation and lipidic domain modification including methylation (*Rv2952*) and glycosylation (*Rv2957, Rv2958c and Rv2962*) (*77*). The pathway and the gene cluster are shown in Scheme 1-9. The functions of most of the genes are listed in Table 1-2.

III. Summary

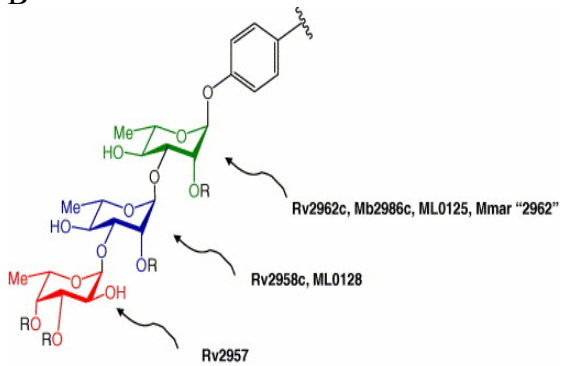
M. tuberculosis cell wall biosynthesis provides a lot of genes as potential anti-TB drug targets that do not have mammalian cell counterparts. This thesis work is identification of the true function of one important *M. tuberculosis* enzyme (Rv3409c). Rv3409c was annotated as a putative cholesterol oxidase, for the first step of cholesterol degradation in *M. tuberculosis*. However, our studies disproved that

hypothesis. Current studies on its enzymatic function in combination with lipidomics approaches were undertaken to determine Rv3409c's real function.

A



B



Scheme 1-9. A) Biosynthesis of PGL and DIM of *M. tuberculosis*; B) PGL glycosyltransferases. The unknown methyltransferases may be Rv2952, Rv2959, and Rv2966 (66).

Chapter 2

Colony Morphology Characterization of *M. smegmatis* mc^{2}155 *MSMEG1604* transposon mutant

I. Introduction

The waxy and thick mycobacterial cell envelope is very important for protecting mycobacteria from harsh environments and the host immune system. The mycobacterium cell wall contains many unique mycobacterial lipids like mycolic acid, "cord factor", trehalose dimycolate (TDM), phthiocerol dimycocerosate (PDIM), and some surface-exposed species and type-specific glycolipids like phenolic glycolipids (PGLs) and glycopeptidolipids (GPLs). Mycobacterial cell envelope lipid components account for as much as 60% of dry weight of the whole bacteria, which is also unique for mycobacteria (*78*). A large fraction of genes are employed by many mycobacteria for lipid metabolism and about 250 genes are involved in lipid metabolism in *M. tuberculosis* genome (*18-19*). The outer most cell envelope of mycobacterium cells interact directly with the host immune system. Those cell surface-exposed lipids could play an important role in the mycobacterium pathogen virulence.

Mycobacterium tuberculosis (*M. tuberculosis*) has infected one third of world's population and is one of the most successful pathogens in history. The cell wall of *Mycobacterium tuberculosis* was thought to play an important role in *M. tuberculosis* virulence. But as an infectious pathogen, and slow-growing mycobacterium, *M. tuberculosis* cell wall studies take more time and money. *Mycobacterium smegmatis* serves as a good lab strain model for *M. tuberculosis (79)*. It is a fast-growing, non-pathogen micro-organism. *Mycobacterium smegmatis* mc^2155 was isolated in 1990 as a transformable mutant of *Mycobacterium smegmatis* ATCC 607 strain (*46*). *M. smegmatis* serves well as an over-expression system for some GC rich mycobacterial genes that are difficult to express in *E. coli*. *M. smegmatis* cell envelope is considered as one of the simplest in all the mycobacterium species.

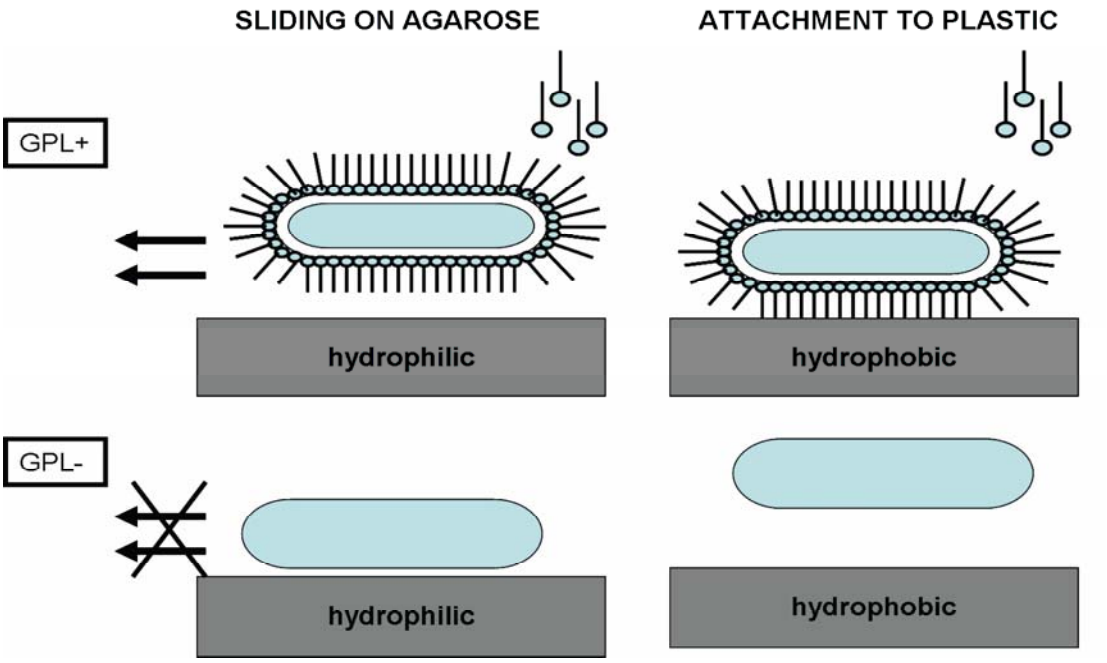


Figure 2-1. The sliding motility model and attachment to surface proposed by Retch et al. (*80*).

Cell morphology is a basic cell characterization for identifying bacteria. The most typical types of cell morphology /bacterial cell shape include spherical (coccus), rod (bacillus) and spiral (spirillum) (http://en.wikipedia.org/wiki/Bacterial_cell_structure). Cell morphology can vary according to growth conditions and environment. For example, bacteria will have different colony morphologies when cultured in liquid medium vs. agar plates. Mutants will be different from wild-type mycobacterium in the cell morphology phenotype (*59, 81*). A number of studies indicate that mycobacterium cell morphology is determined by surface-exposed lipid components. The cell morphology changes indicate that cell wall lipid components change. *M. smegmatis* cell morphology is greatly affected by an important surface-exposed type of lipid - glycopeptidolipid. In general, *M. smegmatis* has the smooth and rough colony phenotypes. The smooth phenotype has more polar GPLs (pGPLs) on the outer layer of the cell envelope. GPL-deficient mutants are usually the rough phenotypes.

M. smegmatis cell structure does not have Fimbrae or pili functional structures but it has the sliding ability to move on specially prepared agar medium (*81*). It is a very important translocation/colonization ability for wild-type *M. smegmatis*. The sliding mechanism is proposed to be related to cell surface-exposed lipid-GPLs. *M. smegmatis* mutants which lack sliding motility are all GPL-deficient phenotypes (*80, 82*). The mechanism for GPLs to affect the sliding of mycobacteria was proposed by Dr. Retch. In their sliding and attachment model, GPLs are oriented with sugar decorated hydrophilic groups inside and the hydrophobic fatty acyl chain exposed outside (Figure 2-1) (*80*). The friction reduction between hydrophilic nutrient agar plate surface and hydrophobic cell surface is the key for mycobacterial sliding. Increasing hydrophobic components in nutrition agar plates will cause wild-type to slide. The interaction between cells and PVC hydrophobic surfaces is valid under this GPL orientation model (*83*). Dr. Jean-Marc Reyrat proposed another orientation model in which GPLs bury hydrophobic fatty acyl chains inside and expose the polysaccharide groups on the tetrapeptide core to the outside (*64*). The hydrophobic/hydrophilic properties of cell surface still play an important role in sliding motility. And the changes on sugar groups of GPLs could alter the interaction mechanism and cause mycobacteia to lose sliding motility. The sugar residues and their modification including O-methylation, O-acetylation and O-succinylation will alter the hydrophobic/hydrophilic properties and further affect the sliding motility. This model was also supported by isolating mutants producing deacetylated GPLs and triglycosylated GPLs (*49*). *atf* mutation which produces GPLs without acetylation disrupts sliding motility (*50*). Increasing the polarity of the polysaccharide groups of the GPLs abrogated the strains sliding ability.

Cell wall lipids changes could affect the cell morphology. Disrupting any gene involved in cell wall lipid metabolism could produce a new phenotype. Gopalaswamy and coworkers found the *M. smegmatis* the serine/threonine protein kinase (STPK) *PknF* merodiploid strain cell morphology, biofilm formation and sliding motility are altered (*84*). In *M. avium* 104, a rough phenotype was caused by deletion of genes for GPLs biosynthesis, a 10kbp size region (*85*). Cell morphology characterization of a mycobacteria strain like sliding motility, biofilm formation, and Congo red fixation could detect lipid changes in the cell wall (*59, 80, 86*).

Overall, glycopeptidolipids play a critical role related to mycobacterium colony morphology and sliding motility. In our studies, *Mycobacterium smegmatis MSMEG1604* (Myc11) colony morphology, sliding motility and GPLs components were studied. *M. tuberculosis Rv3409c* is proposed to encode a cholesterol oxidase. But the cell morphology and lipid component changes in cell wall of *M. smegmatis MSMEG1604* transposon mutant strongly suggest that Rv3409c's biofunction could be involved in cell wall lipid biosynthesis/assembly in a certain way.

II. Experimental procedures

Bacterial strains, medium and growth. M. smegmatis strain mc^2155 and *MSMEG1604* transposon mutant (Myc11) (*59*) were grown in 7H9 Middlebrook broth medium supplemented with 0.2% (v/v) glycerol, 0.2% glucose and 0.05% Tween 80. Antibiotics were added as needed at the following concentrations: kanamycin 30 μg/mL; ampicillin 200 μg/mL; hydromycin 50 μg/mL; and cyclohexide 10 μg/mL. ^{15}N labeled medium was prepared from M9 minimal medium: 6.6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1.0 g/L $^{15}NH_4Cl$, 4 g/L glucose, 0.1 mM $CaCl_2$, 0.2 mM $MgSO_4$, 10 mL MEM vitamin solution (100×) (Thermo Hyclone) in DDI H_2O.

Different nutrition media agar for growth rate comparison. All different media were supplemented with 7.5% agar for plates' solidification. Medium 1: 1.9% Middlebrook 7H10 agar with 5 mL/L glycerol; Medium 2: 1.9% Middlebrook 7H10 agar with 5 g/L glucose; Medium 3: 1.9% Middlebrook 7H10 agar with 15 mL/L iso-propanol; Medium 4: 1.9% Middlebrook 7H10 agar with 6.25 mL/L tween80; Medium 5: 1.9% Middlebrook 7H10 agar with 6.25 mL/L Triton X-100; Medium 6: 1.9% Middlebrook 7H10 agar with 2 mM cholesterol and 6.25 mL/L tween80; Medium 7: 1.9% Middlebrook 7H10 agar with 5 mL glycerol and 2 mM cholesterol and 6.35 mL/L tween80; Medium 8: 1.9% Middlebrook 7H10 agar with 2 mM cholesterol and 6.25 mL/L Triton X-100; Medium 9: 1.9% Middlebrook 7H10 agar with 5 mL glycerol and 2 mM cholesterol and 6.25 mL/L Triton X-100; Medium 10 (Minimal Medium) (*32*): 0.5 g/L asparagine, 1 g/L KH_2PO_4, 2.5 g/L Na_2HPO_4, 10 mg/L $MgSO_4$, 0.5 mg/L $CaCl_2$, 0.1 mg/L $ZnSO_4$, 50 mg/L ferric ammonium citrate and 0.6% agar. Same amount of culture were inoculated on the agar plates and the colony size and morphology was scored using microscope after a 2-day incubation at 37 °C.

Congo red assays. All cultures were collected when OD_{600}=0.6. 10 μL was added to a plate. 7H9 MiddleBrook broth medium supplemented with 1.5% agar, 100 μg/mL Congo red (the sodium salt of benzidinediazo-bis-1-naphthylamine-4-sulfonic acid), 0.02% glucose and 30 μg/mL kanamycin. The colony morphology and Congo red staining were measured after two days incubation at 37 °C.

Sliding motility assays. 7H9 Middlebrook broth medium supplemented with 0.3% agar and proper antibiotics was used for preparing sliding motility agar plates. Bacteria cultures were grown to stationary phase with OD_{600}=0.6. Inoculation was conducted by poking a sterile toothpick carrying liquid culture at the center of sliding

agar plates. The inoculated plates were parafilmed to maintain humidity in the plates during colony sliding. The spreading abilities of strains were measured after two weeks at 37 °C in a humidified incubator.

Construction of MSMEG1604 complementary plasmids. 200 μL *M. smegmatis* mc^2155 in 80% glycerol for glycerol-stab was boiled for 2 min at 100 °C. Clear supernatant was acquired by centrifugation and 2 μL of the supernatant was added into PCR reaction mixture as the genome template. For the complementary plasmid pNIP/40b_Msmeg300MSMEG1604 (also named pMs100), 1.25 μL 2 μM forward primer JGP-20 (5'- GGACTA GTGGGCAACGGCTGGTTCTGGCC -3') and 1.25 μL 2 μM backward primer JGP-21 (5'- GGACTAGTCTACCCCGCCGACGACACGGG -3'), 19.5 μL H_2O and 1 μL dimethyl sulfoxide (DMSO) were added into one PCR tube from GE Healthcare PuReTaq Ready-To-Go™ PCR Beads. For the complementary plasmid pMV306.hygro_Msmeg300MSMEG1604 9(also named pMs101), 1.25 μL 2 μM forward primer JGP-22 (5'- CCGATATCGGGCAACGGCTGGTTCTGGCC -3') and 1.25 μL 2 μM backward primer JGP-23 (5'- CGGAAGCTTCTACCCCGCCGACGACACGGG -3') were applied and other conditions are same with pMs100's. The amplification program is:

Segment	Cycles	Temperature	Time
1	1	95 °C	5 min
2	30	95 °C	45 sec
		69 °C	1 min
		72 °C	2.5 min
3	1	72 °C	10 min

The third *MSMEG1604* complementary plasmid named pMV306.hygro_Mtb999Rv3409c (also named pMtb100) was constructed with *M.*

tuberculosis Rv3409c gene with the 999 bp upstream nucleotide region. The forward primer JGP-1 (5'- GGGGGTCATACTGCAGCGATGAAGCCGG -3') and the backward primer JGP-2 (5'- GCCGAAGCTTCTAGCCCGCGTTGC -3') were applied to clone the *Rv3409c* gene with pMCO-100M as a template. the 50 μL PCR reaction mixture include 5 μL dNTP's (2.5 mM each), 10 μL 5× iProof GC buffer, 5 μL 10 μM primer JGP-1 and 5 μL 10 μM primer JGP-2, 200 ng pMCO-100M, 0.5 μL 2 U/μl iProof High-Fidelity DNA Polymerase and proper amount of DDI water. PCR reaction was conducted with the program below:

Segment	**Cycles**	**Temperature**	**Time**
1	1	95 °C	1 min
2	34	95 °C	45 sec
		72 °C	4 min
3	1	72 °C	14 min

The forward primer JGP-3 (5'- GCCCGGATATCCGATCGAGCTCG -3') and backward primer JGP-4 (5'- CCCGATCCAATAATCAGGACGTCGTAATCCGGC TTCATCGCTGCAGTATGACCC -3') were applied to clone the Rv3409c 999 bp upstream region. the 50 μL PCR reaction mixture include 5 μL dNTP's (2.5 mM each), 10 μL 5× iProof GC buffer, 5 μL 10 μM primer JGP-3 and 5 μL 10 μM primer JGP-4, 10 ng *Rv3409c* 999 bp upstream nucleotide sequence, 1.5 μL 100%DMSO, 0.5 μL 2 U/μl iProof High-Fidelity DNA Polymerase and proper amount of DDI water. The amplification program is:

Segment	Cycles	Temperature	Time
1	1	98 °C	5 min
2	30	98 °C	30 sec
		58 °C	30 sec
		72 °C	1 min
3	1	72 °C	8 min

To ligate the Rv3409c and its 999 bp upstream nucleotide sequence, one more PCR reaction was performed. The forwarding primer JGP-3 (5'-GCCCGGATATCCGATCGAGCTCG -3') and the backward primer JGP-2 (5'-GCCGAAGCTTCTAGCCCGCGTTGC -3') were applied. the 50 μL PCR reaction mixture include 4 μL dNTP's (2.5 mM each), 5 μL 10× Herculase reaction buffer, 1 μL 10 μM primer JGP-2 and 1 μL 10 μM primer JGP-3, 20 ng purified Rv3409c, 60 ng purified *Rv3409c* 999 bp upstream nucleotide sequence, 2.5 μL 100% DMSO, 0.5 μL 5 U/μL Herculase Enhanced DNA Polymerase and proper amount of DDI water. The PCR program is:

Segment	Cycles	Temperature	Time
1	1	98 °C	3 min
2	10	98 °C	40 sec
		65 °C	30 sec
		72 °C	3 min
3	20	98 °C	40 sec
		68 °C	30 sec
		72 °C	3 min
4	1	72 °C	10 min

The pMV306.hygro vector and the purified *Rv3409c* gene with its 999 bp upstream nucleotide sequence were treated with HindIII and EcoRV restriction enzymes HindIII and EcoRV followed by ligation experiment. Ligation experiment contains 1 μL T4 ligase enzyme and 2 μL 10 times concentrated T4 ligase reaction buffer and open pMV306.hygro vector and *Rv3409c* gene with its 999 bp upstream nucleotide sequence. The total ligation reaction volume is 20 μL. The ligation was conducted at 4 °C overnight and then was transformed into XL1Blue cells for amplification and isolation.

MSMEG1604 complementary strains preparation. M. smegmatis MSMEG1604 transposon mutant (Myc11) electrocompetent cells were prepared as described (*87*). The complementation plasmids were transformed into competent Myc11 using an electro cell Manipulator 600 (BTX Electroporation System) and 1 mm gap cuvette. 25 μL Myc11 electrocompetent cells with 10 ng of complemention plasmid were gently mixed and aliquoted in the pre-chilled gap cuvettes. The parameters are mode: HV; capacitance: 50 μF; resistance: R5; charging voltage: 1.7 kV; desired filed strength: 17 kV/cm and desired pulse length: 4-5 mSec. After electrocoporation step, 2mL LB medium were added to the mixture and it was cultured for 3 h at 150 rpm, 37 °C. 100 uL of the culture was inoculated on 7H10 agar plates to obtain a single colony.

Extraction and purification of total lipids from Mycobacterium smegmatis. Myc11 or wild type was cultured in 7H9 Middlebrook broth supplemented with 10% ADS (8% sodium chloride, 10% bovine serum albumin fraction V, and 4% glucose) for 3 days. Wet cells were collected by centrifugation at 3,500 rpm and then resuspended in phosphate buffered saline (PBS) and incubated at 60 °C for 90 min. The pellets were collected by centrifugation at 3,500 rpm, removal of the supernatant. Choloroform:methanol (v:v, 2:1) was added to extract lipids (3 mL organic solvent:1

gram wet cells) from pellets at 55 °C for 15 min. The extraction was repeated and the organic phases were combined. Two washes with DDI water (1 mL DDI water:1 gram wet cells) were applied to remove the hydrophilic compounds from the extracts. The organic phase was then evaporated to obtain crude total lipids and stored at -20°C.

Matrix-assisted laser desorption ionization time-of flight mass spectrometry. 10 mg/mL 2, 5-dihydroxybenzoic acid (DHBA) was prepared in $CHCl_3$:CH_3OH (v:v, 1:1) as a matrix solution. Lipid samples were dissolved in chloroform to make 1 mg/mL lipid solution. 1 μL sample was mixed with 1 μL matrix solution and then loaded onto the target plate for measurement in positive ion mode by MALDI-TOF/MS on a Bruker Ultraflex TOF/TOF.

Electrocompetent wild-type M. smegmatis and Myc11 strain preparation. Wild-type *M. smegmatis* or Myc11 cultures were inoculated in 7H9 Middlebrook Broth Medium and harvested when OD_{600}=0.6. The cell pellets were cooled to 4 °C and washed 5 times with sterile 10% glycerol in DDI H_2O. The washing volume was decreasing from 50 mL the first time to 10 ml the final time. After centrifugation, pelleted cells were re-suspended in 1-3 mL 10% glycerol and 50 μL aliquots for electroporation experiments or stored at -80 °C.

III. Results and Discussion

Mycobacterium smegmatis MSMEG1604 is able to grow on cholesterol. MSMEG1604 was predicted to encode a cholesterol oxidase. In *M. smegmatis* genome, there are *M. tuberculosis Rv3409c* (a putative cholesterol oxidase) and *Rv1106c* (a cholesterol dehydrogenase) orthologs. The identity between *MSMEG1604* and *Rv3409c* is 83% (by T-COFFEE). *M. smegmatis MSMEG1604* transposon mutant (Myc11) was constructed by Sonden and co-workers and it showed a RR (rough-red)

colony phenotype on their Middlebrook 7H10 or 7H11 agar plates supplemented with BBL Middlebrook OADC enrichment (*59*). In former studies, *M. smegmatis* could also utilize cholesterol as the only carbon source (*40*). The first step of cholesterol degradation is performed by the MSMEG1604, the ability of Myc11 to grow on cholesterol should be abrogated by disruption of *MSMEG1604*.

Figure 2-2. Growth rate of Myc11 and wild-type *M. smegmatis* on different nutrition agar plates; all plates contain 7H9 MiddleBrook broth except the minimal medium.

Myc11 and wild-type *M. smegmatis* growth on different nutrition agar plates was compared. Besides the Middlebrook 7H10 agar plate supplemented only with cholesterol, other nutrition plates were also tested as controls. The same amount of culture sample was inoculated on each agar plates. Growth rates of both WT and Myc11 on special prepared medium agar plates were carefully monitored by observing the colony diameter and thickness (Figure 2-2). Generally, Myc11 do not show preference on carbon sources different from wild-type. On nutrition agar plates, Myc11 do not grow significant slowly or faster than wild type do. Myc11 is able to grow on cholesterol as the only carbon source. With 100 times magnification, the colony surface of Myc11 does not show any difference while growing on cholesterol as the only carbon source vs. normal carbon nutrition. As in Sonden et al's work, the wild-type colony surface is smooth while Myc11 had more "dendric" structures on the colony surface (Figure 2-3). Neither wild type nor Myc11 could grow on agar

supplemented with Triton X-100. Glucose and glycerol are nice carbon sources, while iso-propanol and Tween 80 are not.

Myc11 still retains the ability to grow on cholesterol as the only carbon source. This finding is consistent with *Rv3409c* studies in *M. tuberculosis*. The cell lysate supernatant of *M. tuberculosis Rv3409c* transposon mutant supernatant still has cholesterol oxidizing ability. But after *Rv1106* is mutated, the cholesterol oxidizing ability is totally abrogated (Dr. Eugenie Dubnau, personal communication, Figure 2-4). Those studies about *M. tuberculosis Rv3409c* and its ortholog in *M. smegmatis* showed that Rv3409c enzyme activity could not be related to cholesterol degradation especially cholesterol oxidation.

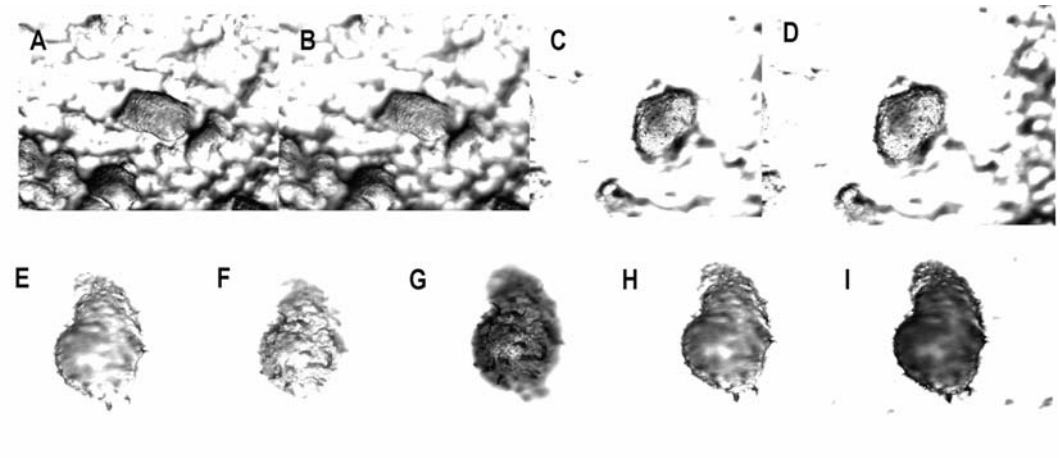


Figure 2-3. Microscopic pictures of colony phenotypes of wild type (A-D) and Myc11 (E-I) on 7H10 agar plates.

Myc11 complemented strain construction. Myc11 complemented strains were constructed. Both *M. tuberculosis Rv3409c* and *M. smegmatis MSMEG1604* were cloned and inserted into pMV306.hygro and pNIP/40b *M. smegmatis* vectors for complementation. Each complementation construct included 300 bp upstream of the cloned gene in order to include the native promoter. In the *Rv3409c* construct, 999 bp was used. All complementation construct are summarized in Table 2-1.

Table 2-1. Summary of Myc11 complementation constructs. (* pNIP40/b is an integrative vector gifted by Dr. Jean-Marc (*88*) and ¶ pMV306.hygro is a L5-based integrating plasmid with the attP site (*89*).)

Plasmid name	Cloned region	Gene	Plasmid used	Resistance marker
pMs100	MSMEG1604 with 300 bp upstream region	*MSMEG1604*	pNIP40/b*	Hygromycin
pMs101	MSMEG1604 with 300 bp upstream region	*MSMEG1604*	pMV306.hygro¶	Hygromycin
pMtb100	Rv3409c with 999 bp upstream region	*Rv3409c*	pMV306.hygro¶	Hygromycin

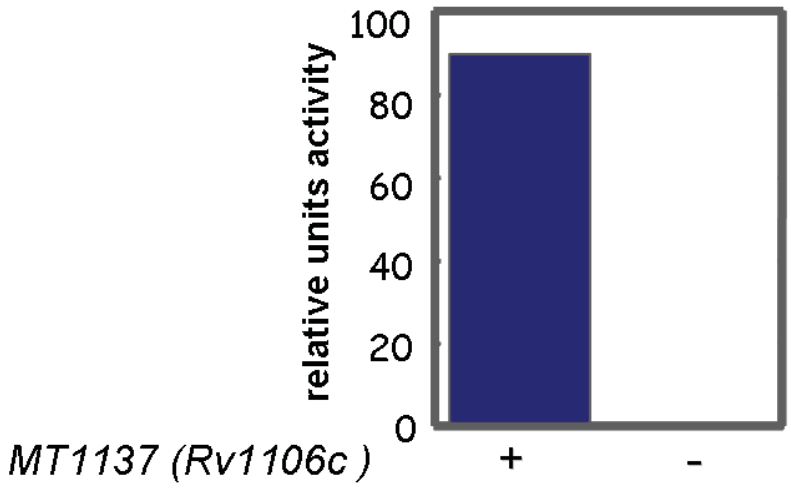


Figure 2-4. Sterol oxidation activity in *M. tuberculosis* wild type (+) and *Rv1106c* transposon mutant (-) cell lysate supernatants using the cholesterol dehydrogenase assays with cholesterol as a substrate, NAD^+ as cofactor, and cholesterol and putative products extracted by acetyl acetate were analyzed by a C_{18} reverse phase column on HPLC.

Myc11:*pMtb100* still accumulated L1334 in the cell envelope by mass spectroscopic analysis (Figure 2-8). This could be caused by using *M. tuberculosis Rv3409c* instead of *M. smegmatis MSMEG1604*. The identity between *MSMEG1604* and *Rv3409c* ortholog is 83%. In addition, *M. tuberculosis Rv3409c*'s putative

promoter within its 999 bp upstream nucleotide sequence may not be to be recognized in *M. smegmatis* system. The further Myc11 complementary studies utilized the two integrative plasmids pMs100 and pMs101.

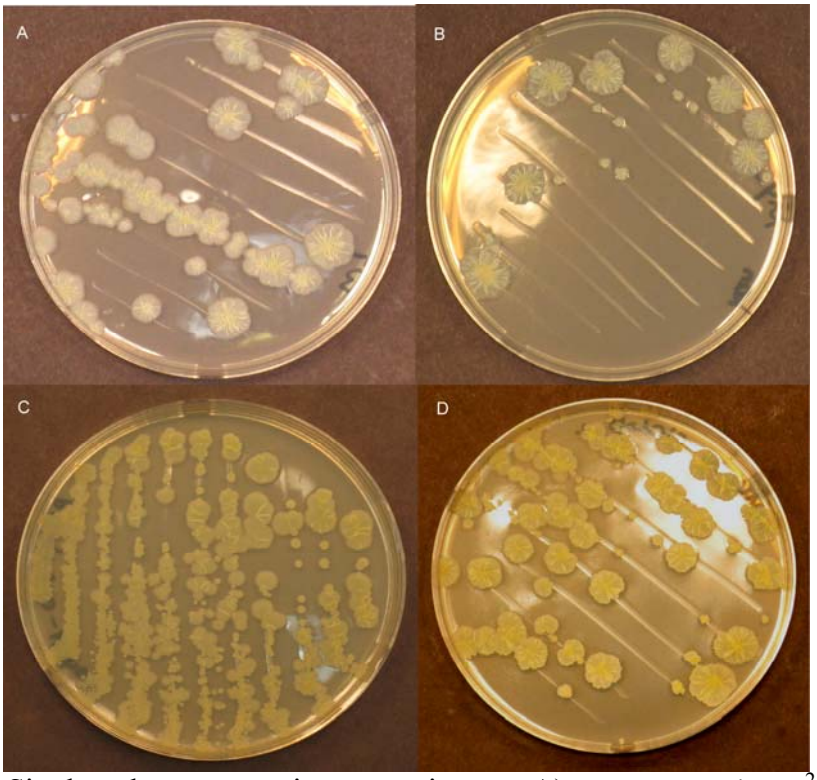


Figure 2-5. Single colony comparison experiments. A) *M. smegmatis* mc^{2}155; B) *M. smegmatis* mc^{2}155 *MSMEG1604* transposon mutant (Myc11); C) *Myc11*:*pMs100* strain; D) *Myc11*:*pMs101* strain.

Colony morphology characterization. Wild-type *M. smegmatis* mc^{2}155 *and MSMEG1604* transposon mutant (Myc11) colony morphologies were different. As reported by Sonden and co-authors, Myc11 is a RR phenotype while wild-type is a smooth red (SR) phenotype (*59*). For more extensive characterization, both strains were inoculated in 7H10 agar medium plates enriched with glycerol and proper antibiotics. After two weeks' incubation at 37 °C, the single colonies of wild-type *M. smegmatis* mc^{2}155 and Myc11 grew to about 1.00 cm diameter on the agar plate surface. The wild-type and Myc11 colony skin surfaces showed differences in the

area of wrinkled and smooth partition (Figure 2-5). The results are consistent in the paper by Sonden and coworkers (*59*).

On average, single colonies of complemented Myc11 had smaller diameters than wild type, about 0.89 cm for *Myc11*:*pMs101* and about 0.76 cm for *Myc11*:*pMs100*. The *Myc11*:*pMs100* colony surface was very smooth and *Myc11*:*pMs101* colony surface was very similar to wild type. Under 100 times magnification, the colony surfaces of both complemented strains still had the "branch"-like structure.

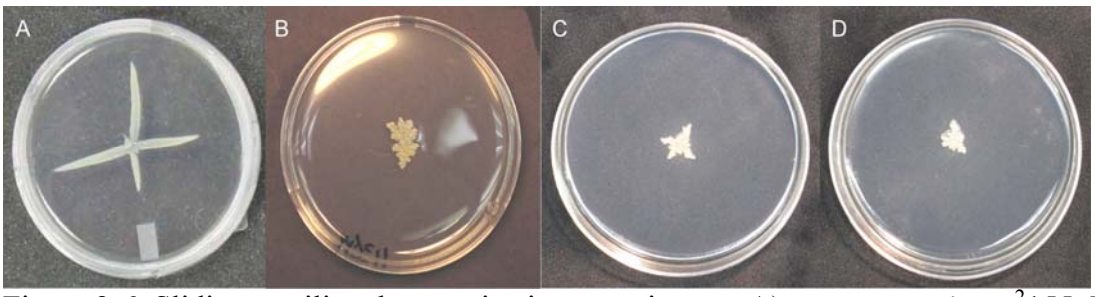


Figure 2-6. Sliding motility characterization experiments. A) *M. smegmatis* mc^2155; B) *M. smegmatis* mc^2155 *MSMEG1604* transposon mutant (Myc11); C) *Myc11*:*pMs100* strain; D) *Myc11*:*pMs101* strain.

Totally, the colony surface of Myc11 got rougher and forms a more wrinkled structure. Wild-type *M. smegmatis* colony surface had less wrinkled structures. Myc11 had more "branch" like the small branched stick to the surface. This cell colony surface difference between wild-type *M. smegmatis* and Myc11 cell morphology could indicate the lipid component difference in the cell wall. Rv3409c could be directly or indirectly related to cell wall lipids formation or cell wall assembling. One of the complemented strains, *Myc11*:*pMs100* has a smooth colony phenotype which was smoother than wild-type does.

Sliding motility characterization. Myc11, wild-type *M. smegmatis* and complemented Myc11 were inoculated in the middle of sliding motility agar plates. After incubation at 37 °C for about 2 weeks, wild-type *M. smegmatis* colonies are able to translocate on agar medium by forming a star-like colony and reaching the edge of plates on the agar plates. The Myc11 colonies could not spread on the sliding agar plates and the sliding motility of Myc11 was totally destroyed (Figure 2-6). Under 100 × magnification, Myc11 colony was thicker, the edges were rough, and halos were not obvious compared with that WT smoother edges and obvious halos. The two complemented Myc11 strains did not recover the sliding motility. The colony edges were very different from both wild type and Myc11.

The translocation ability of *M. smegmatis* is defined as sliding motility (*81*). The sliding motility not only shows the colonization ability of mycobacterium in nature environments but also in the host. For *M. smegmatis*, sliding motility is related to the surface-exposed lipid–glycopeptidolipids in the cell wall (*81*). The *M. smegmatis* strains isolated with inhibited sliding motility always have rough colony phenotypes and produce abnormal GPLs in the cell wall. Glycopeptidolipids could affect the sliding motility through a friction mechanism. Whether the GPLs in Myc11 cell envelope change the colony morphology and sliding motility will be further studied.

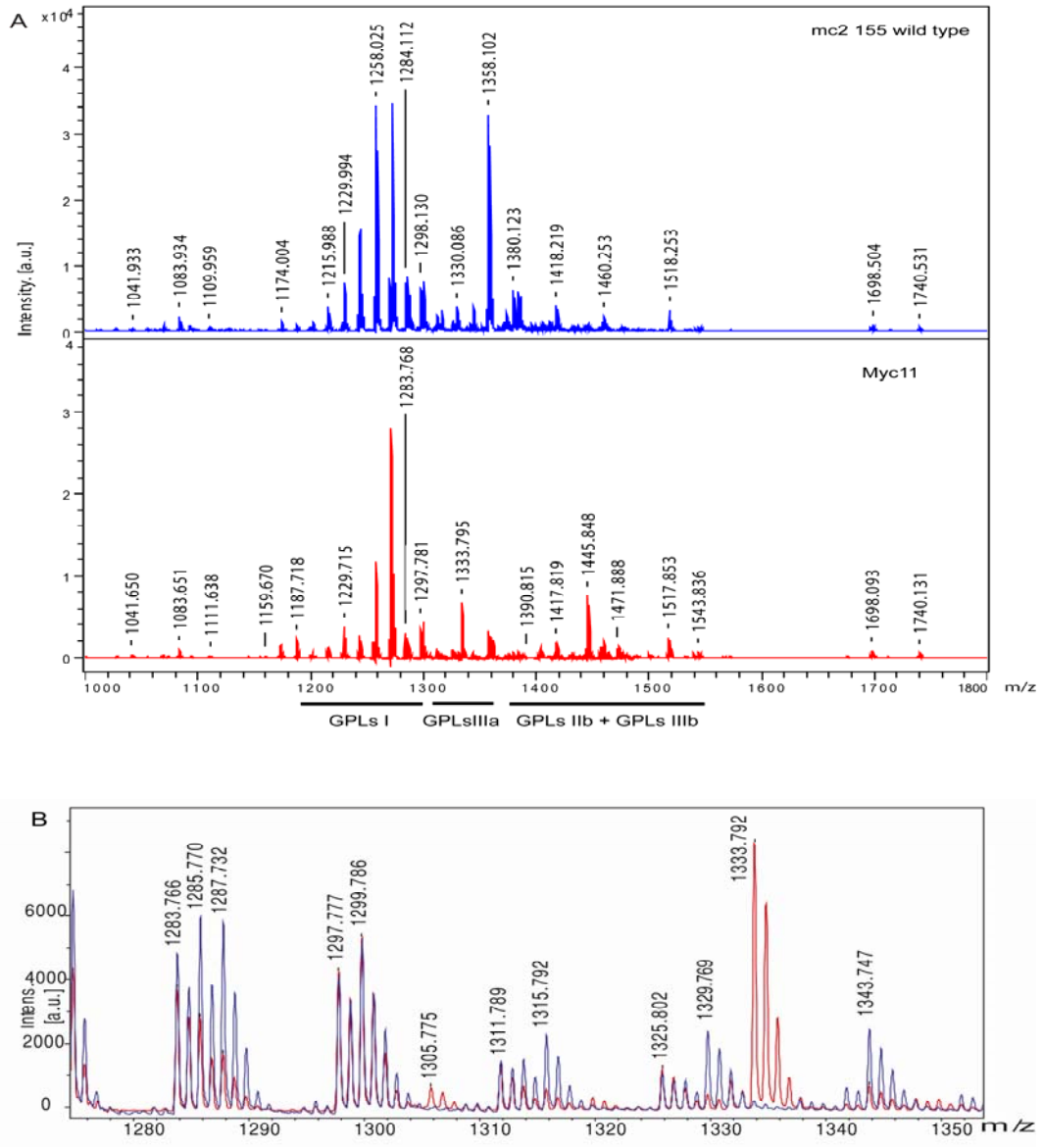


Figure 2-7. A) MALDI-TOF/MS spectrum of total lipids from wild-type *M. smegmatis* and Myc11. B) Overlaid mass spectra of wild-type *M. smegmatis* and Myc11.

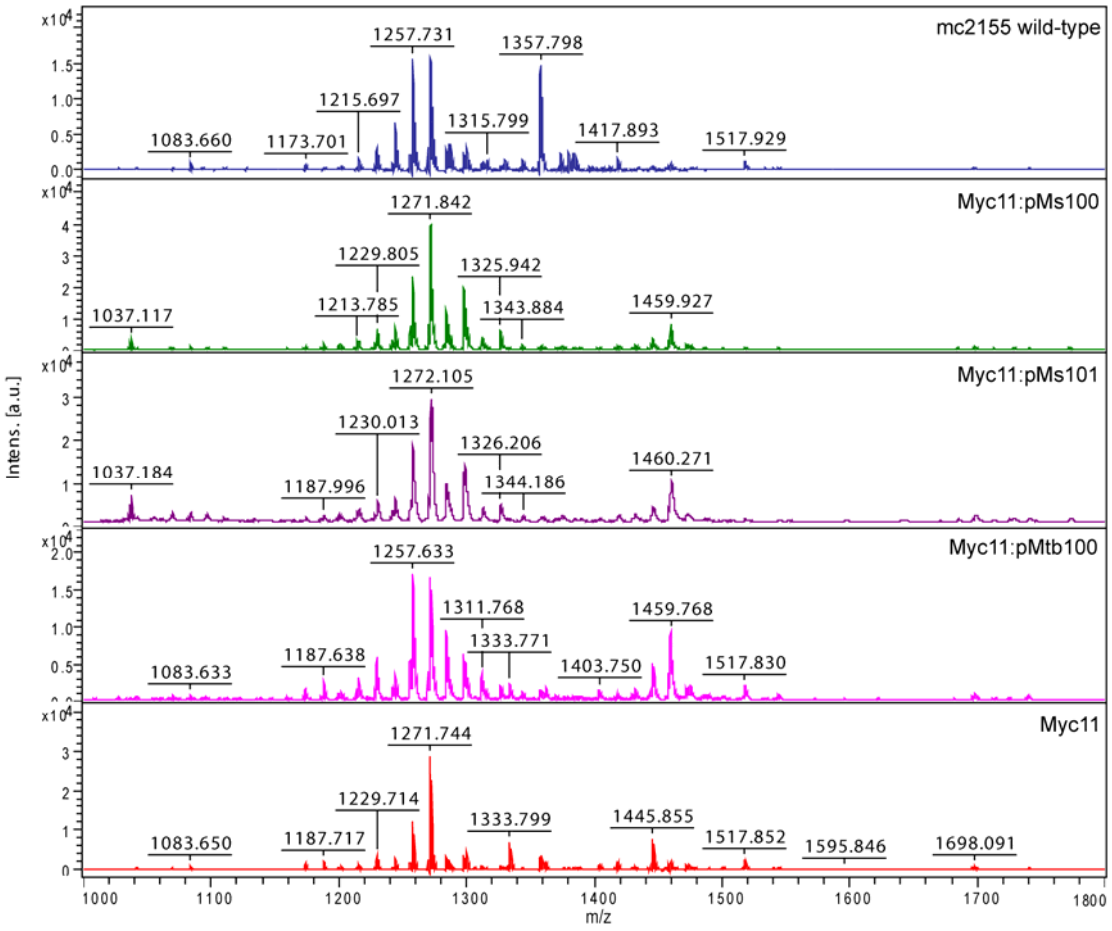


Figure 2-8. MALDI-TOF/MS of wild-type *M. smegmatis*, Myc11, and complemented strains m/z from 1000 to 1800.

MALDI-TOF/MS analysis of crude total lipids. The total lipids of late stationary phase wild-type *M. smegmatis*, Myc11 and complemented Myc11 were extracted by a modified Bligh-Dyer method (*90*) and the total lipid mass spectra were acquired on Bruker MALDI-TOF/MS equipment. In the Myc11 total lipids, one unique lipid compound (L1334) accumulated with m/z at 1334 (Figure 2-7). This molecular weight is similar to that of GPLs. No L1334 appeared in the Myc11 complemented strains' total lipid profiles (Figure 2-8 and Figure 2-9). The lipidomic profiles of Myc11 complemented strains are more like wild-type *M. smegmatis*, but the colony morphology and sliding motility are more like Myc11. The partly complemented

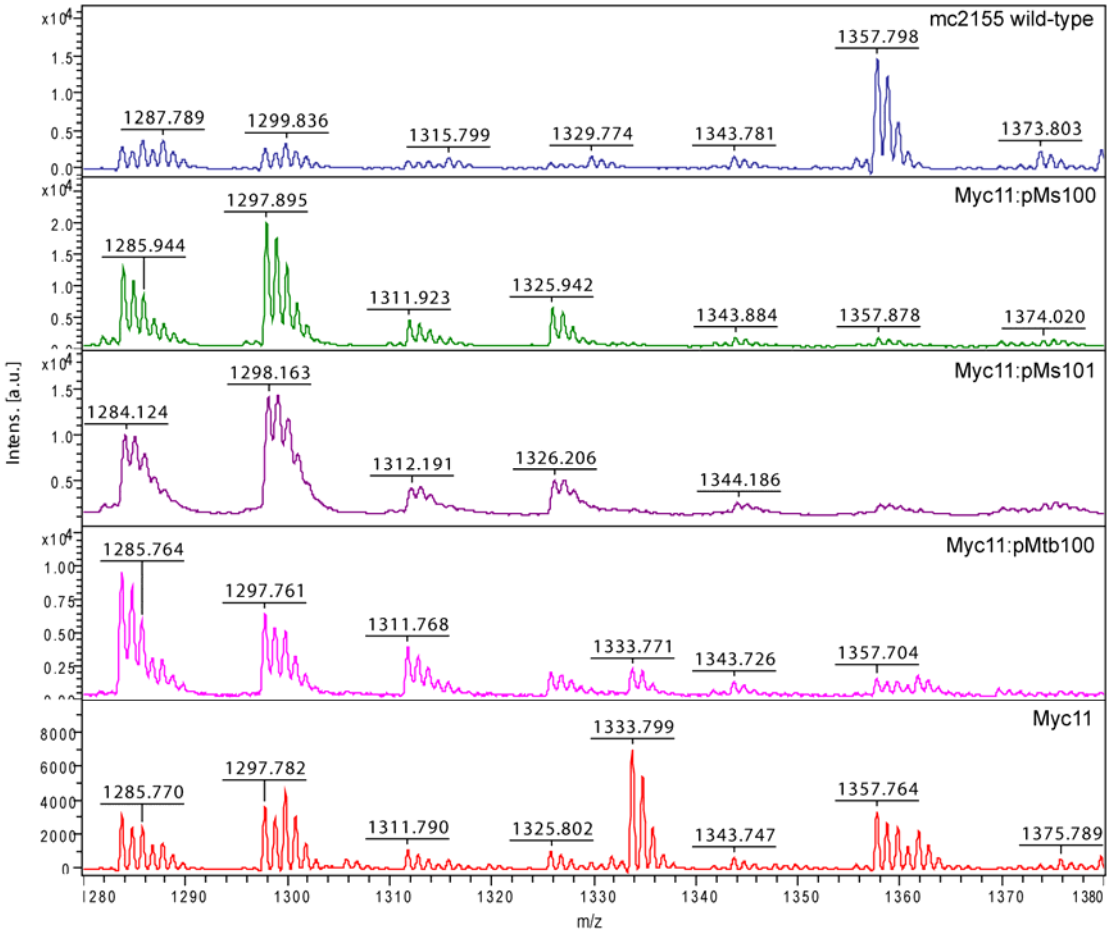


Figure 2-9. MALDI-TOF/MS of wild-type *M. smegmatis*, Myc11, and complemented strains m/z from 1280 to 1380.

strains could be caused by several reasons. The *M. smegmatis MSMEG1604* transposon mutant might lead to termination of downstream genes' transcription. Another putative reason is the *MSMEG1604* expression level by integrative plasmids pMV306.hygro (Figure S9) or pNIP/40b (Figure S10) is different from its native expression level. The MS spectra also showed that one lipid compound with m/z=1357.8 and the GPL-5 (m/z=1518), were reduced in intensity compared to the wild-type and Myc11 spectra (Figure 2-9). In contrast, one new lipid compound with m/z=1037.2 accumulated in the total lipids of Myc11 complemented strains. The L1358 and L1518 are most likely GPLs whereas L1037 is not known. How these

lipids are related to L1334 remains unclear. The TLC analysis of wild type, Myc11 and *Myc11*:*pMs100* total lipids is shown in Figure 2-10. One lipid compound labeled by an arrow is missing from the complemented strain. Some lipids of complemented Myc11 migrate differently from wild type and Myc11 on the TLC plates. These lipid changes in total lipids of the complemented Myc11 strain correlate with the MS results and may explain why the sliding motility and cell morphology are not recovered. The integrative plasmid carrying *MSMEG1604* with its upstream 300 bp nucleotide sequence could result in the enzyme expression in a different level from the natural situation, which indicates *MSMEG1604* enzyme function does regulate the cell wall lipids in an unclear mechanism. The main differences between wild type, Myc11 and Myc11 complemented strains are summarized in Table S1.

Overall, Myc11's colony morphology changes to a rough phenotype. The single colony experiments showed that Myc11 has a more rough and wrinkled colony skin that wild type. The sliding motility of Myc11 is also retarded. The colony phenotypes indicate Myc11 may have changes in the cell wall structure or a lipid component. The interesting results include: 1) Myc11 could still grow on cholesterol as the only carbon source; 2) Myc11 has a single lipid L1334 accumulated, which contributes to the colony morphology changes. The further studies about the differences of wild-type *M. smegmatis* and Myc11 will be conducted in the next Chapter and *in vitro* enzyme activities of this enzyme will be performed in Chapter 4.

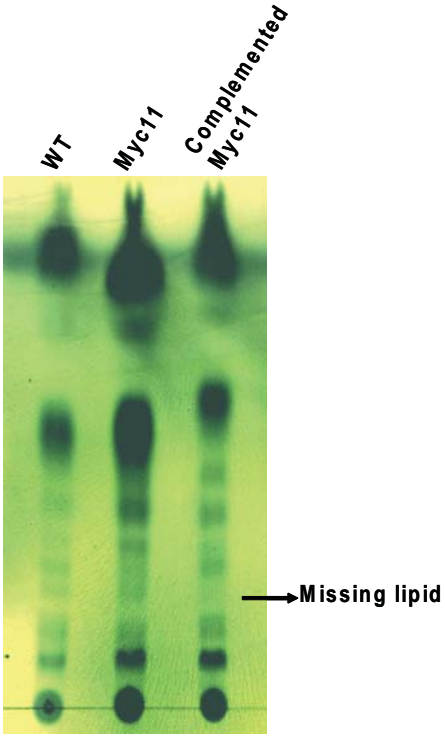


Figure 2-10. TLC experiment with crude total lipids of *M. smegmatis* mc^2155 and Myc11 and *Myc11:pMs100* using developing system (chloroform:methanol:water, 90:10:1).

Chapter 3

Structural Characterization of an Unusual Glycopeptidolipid Isolated from *M. smegmatis MSMEG1604*

I. Introduction

The mycobacteria are rod-shaped, obligate aerobes. About 50 mycobacteria species has been identified with most of them saprophytes (*91*). *Mycobacterium tuberculosis (M. tuberculosis)* is one of the famous mycobacterium pathogens and causes tuberculosis in humans. *Mycobacterium leprae (M. leprae)* can cause leprosy and *Mycobacterium avium intracellulare* mostly infects the immunocompromised human like HIV-infected patients. *M. tuberculosis* infects about one third of the world's population and kill 2 million people every year. The drug-resistant *M. tuberculosis* and multi-drug-resistant *M. tuberculosis* cause the studies on tuberculosis emergent and important.

Mycobacterial cell walls are extremely waxy, including long fatty acids (C_{78}-C_{90}), mycolic acids, polysaccharides and proteins and are specifically rich in lipid components (about 30-60% dry weight of total) (*68*). The extremely thick and waxy cell walls not only protect the mycobacteria against harsh environment and bactericides, but could also functionally be involved in the mycobacterial pathogens' virulence process. *Mycobacterium tuberculosis* cell wall has been extensively studied for almost 50 years to understand the lipids' structures and their putative roles in virulence (*69, 92*). *M. tuberculosis* can survive the early phagosome and phagosome-lysosome fusion and replicate in the macrophages. Although the tuberculosis bacillus pathogenicity mechanism is not clear, cell wall lipids especially the polyketides like the phenolphthiocerol and phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs) could be tuberculosis virulence factors.

Mycobacterium smegmatis (M. smegmatis) is a fast-growing mycobacterium found in soil and water. Isolated as a efficient-plasmid-transformation (ept) mutant, it is a nice protein over-expression system (*46*). The cell envelope lipid components are

very similar to *M. tuberculosis* but simpler. *M. smegmatis* mc^2155 is a good model for mycobacterium cell wall lipids studies. Like *M. tuberculosis*, *M. smegmatis* are composed of three layers: a plasma membrane layer, a cell wall core layer with peptidoglycan, arabinogalactan, and long chained fatty acids covalently associated to each other and glycolipids non-covalently linked outside and the outmost cell capsule layer containing polysaccharide, proteins and lipids (*68*). The outmost layer of *M. smegmatis* is mainly composed of glycopeptidolipids (GPLs). Glycopeptidolipids are produced by non-tuberculosis mycobacteria. They have been found in human infection pathogens like *Mycobacterium avium* complex (MAC) (*73*), *M. abscessus* and *M. chelonae* (*65, 93*). Some animal pathogens like *M. porcinum* and *M. senegalense* also produce GPLs in their cell envelope (*73*). Glycopeptidolipids affect GPL-producing mycobacteria cell properties such as cell morphology, sliding motility and biofilm formation. All studies indicate GPLs alter cell wall properties through GPLs structures (glycosylation, acetylation and methylation) and their concentrations.

Normal glycopeptidolipids are composed of three parts: a tetrapeptide core, saccharidyl units and one long chain fatty acid unit. The glycopeptidolipids varies depending on mycobacterium species, servor types and growth conditions. The tetrapeptide core is conserved in most GPLs structure with the sugar residues extended except the GPL-I isolated from *Mycobacterium. xenopi* (*M. xenopi*) strain NCTC 10042 (*52-53*). GPLs were found to be implicated in the virulence process. The GPLs-deficient *M. avium* was greatly attenuated in the virulence. Two *M. avium* mutants with modified genes involved in GPLs synthesis has been studied. *mtfD*-defiecient *M. avium* mutant produce an unnormal nonmethylated GPLs (*94*). The TNF-α and RANTES of *mtfD* mutatnt infected murine macrophases level were much higher than wild-type infected ones. *rtfA*-deficient *M. avium* mutant lost the

ability of producing the serotype 8 GPL and the infected J774 cells had a lower TNF-α expression level (*63*).

The GPLs structures are important in the proinflammatory mediators releasing process. Some serotype specific GPLs could stimulate the proinflammation while some could not. The changes in GPLs structures in the mycobacterium could affect the virulence abilities. It is believed to be related to interaction between GPLs and different host cell receptors. Kano and co-worker has lately found that the GPLs were implicated in the phagosome-lysosome fusion delay mechanism (*95*). The protection of GPLs made MAC survive in macrophasges. Although the clear mechanism of GPL modulating the immunology through those receptors remains to be elucidated in the future, the importance of GPLs has been significantly proved.

M. tuberculosis does not contain GPLs in the cell wall. GPLs only exist in many non-tuberculosis mycobacterium organism cell walls. However, phenoglycolipids (PGLs) and related molecules (DIMs and p-HBADs) are present in the *M. tuberculosis* cell wall. Both of GPLs and PGLs are surface-exposed and interact with environments and host immune system. The mycobacterium species which produce GPLs does not produce PGLs (*64*). Phenoglycolipids (PGL) also named glycosylated phenolphtiocerol dimycocerosates is structurally related to phthiocerol dimycocerostaes (DIMs). PGLs and DIMs have a similar lipid core. PGLs are terminated at the phtiocerol chain with a glycoylated phenolic moiety. Gene locus for biosynthesis of DIMs and PGLs are about 73 kbp large size utilizing 1.7% of *M. tuberculosis* whole genome and more than 20 genes (*71*). DIMs and PGLs are limitedly produced by slow-growing infective human pathogenic mycobacterium species (*96*). PGLs is mainly found existing in the M. leprae (PGL-1), some *M. tuberculosis* strains (PGL-tb), some *Mycobacterium bovis* and other slow-growing

mycobacteria like *Mycobacterium ulcerans*, *Mycobacterium marinum* et al. (*66*). But *M. tuberculosis* H37Rv is not included due to *pks15/1* gene mutation disrupting the PGL biosynthesis (*70*).

Some mycobacterial lipids are important for mycobacteria virulence. Besides PGLs, the extractable *M. tuberculosis* cell wall lipids include mycolic acids such as cord factor trehalose dimycolate (TDM), trehalose monomycolate (TM) and three types of phosphatidylinositol based glycolipids phosphatidylinositol mannosides (PIMs), lipomannans (LMs) and surface exposed lipids including phthiocerol dimycocerosates (PDIMs) and lipooligosaccharides (LOSs). The "cord factor" trehalose dimycolate can stimulate NADase activity and affect a broad range of NAD-dependent enzymes activity by lowering the host NAD level (*78*). DIMs-defiecient *M. tuberculosis* H37Rv attenuated in the infection and LAM is able to bind to certain receptors and stimulate signal pathway in the TB-infected host (*92*).

In this chapter, a new lipid L1334, accumulated in Myc11 strains, was identified with a structure similar to normal GPL-5. However there was no acetylation on the D-allo-6-deoxytalose and the L-allo-alcohol-rhamnose was over rhamnosylated. The GPLs components were altered by this accumulated lipid L1334, which causes the colony phenotype change and sliding motility retardance. With the help of NMR, high resolution tandem MS, and TLC experiments, the identity of L1334 was determined. *MSMEG1604* putatively encodes a FAD dependent oxidase and is an ortholog to *Rv3409c* of *M. tuberculosis*. This cholesterol oxidizing ability of Rv3409c has not been confirmed so far (*45*). The findings in our lab could indicate the Rv3409c plays other important function in lipid biosynthesis/modification. The supportive results are also found by cholesterol metabolism pathway studies in our

research group (*33, 36, 97*). This compound accumulation and structure difference from normal GPLs could help us studying the enzyme biofunctions in the future.

II. Experimental procedures

Bacterial strains, medium and growth. M. smegmatis strain mc^2155 wild type and *MSMEG1604* transposon mutant (Myc11) were grown in 7H9 Middlebrook medium supplemented with 0.2% glycerol, 0.2% glucose and 0.05% Tween 80. Antibiotics were added as required at the following concentrations: kanamycin 30 μg/mL; ampicillin 200 μg/mL; hydromycin 50 μg/mL; and cycloheximide 10 μg/mL.

Extraction and purification of total lipids from mycobacteria. Myc11 or *M. smegmatis* mc^2155 wild-type were cultured in the 7H9 Middle-brook broth supplemented with 10% ADS (8% NaCl, 10% BSA fraction V, and 4% glucose) for 3 days. Wet cells were collected by centrifugation at 3,500 rpm and then washed with PBS 3 times. Chloroform:methanol (v:v, 2:1) was added to pellets (3 mL organic solvent:1 gram wet cells) and the mixture was incubated at 55 °C for 15 minutes. The extraction was repeated and the organic phase was combined. Two washes with DDI water (1 mL DDI water:1 gram wet cells) were applied to remove the hydrophilic compounds from the extracts. The organic phase was then evaporated to provide dry lipids that were stored at -20 °C for long-term use.

L1334 purification. The Myc11 total lipid sample was dissolved in 1-2 mL chloroform and loaded onto prepared silica gel column (3×40 cm). After absorption of the sample onto the silica, chloroform:meathnol:water (v:v:v, 90:10:1) was applied to elute lipids from the column. If the compound collected was not sufficiently pure, a second silica gel column was used with the same elution system. The fractions were monitored by both TLC on silica gel coated plates (Merck) and MALDI-TOF/MS.

The collected fractions containing compound L1334 were combined and dried by a rotary evaporator followed by high vacuum. Less than 0.5 mg L1334 could also be purified by Combiflash (4 g or 12 g Silica RediRediSep Column) with the starting material chloroform and the gradient slowing growing up to 10% methanol in chloroform at the flow rate of 10 mL/min.

Matrix-assisted laser desorption ionization time-of flight mass spectrometry. 10 mg/mL 2,5-dihydroxybenzoic acid (DHBA) was prepared in $CHCl_3$:CH_3OH (v/v, 1:1) as a matrix solution. Lipid samples were dissolved in choloroform to make a 1 mg/mL lipid solution. 1 μL of sample was mixed with 1 μL of matrix solution and then loaded onto the target plate for measurement at positive ion mode by MALDI-TOF/MS on a Bruker Ultraflex TOF/TOF.

L1334 NMR analysis and orbitrap tandem mass experiments. About 3 mg L1334 were dissolved in 99.8% $CDCl_3$. The ^{1}H-^{13}C heteronuclear multiple-quantum NMR spectra was acquired on a 800 MHz NMR instrument with cryoprobe at the New York Structural Biology Center (NYSBC), The ^{1}H NMR spectra was acquired on a Bruker Avance 700 NMR Spectrometer. The mass spectra were obtained with an LTQ-Orbitrap spectrometer (Thermo Scientific) in Stony Brook University Proteomics Center.

LC/MS analysis. The purified L1334 compound was injected into a Waters UPLC system (UPLC, Waters, Milford, MA, USA) and eluted through a ACQUITY UPLC BEH C18 column (1.7μm, 2.1 × 100mm). The column was equilibrated with 98% solvent A (methanol with 0.1% acetic acid) and 2% B (methanol: ethanol 1:1). The lipid compound was eluted with a linear gradient of 100% solvent A to 70% solvent A in 20 min at a flow rate of 0.5 ml/min.

III. Results and Discussion

Lipid profile of wild-type M. smegmatis and MSMEG1604 transposn mutant. The total lipid samples of wild-type *M. smegmatis* and *M. smegmatis* MSMEG1604 transposon mutatnt (Myc11) (*59*) were characterized by MALDI-MS/TOF Bruker equipment and TLC experiments. Cell cultures of *M. smegmatis* mc^2155 wild-type and Myc11 were harvested at stationary phase OD_{600}=2.1. Cell wall total lipids were extracted using the modified Bligh-Dyer method (*90, 98*). The TLC plate indicates that one unique lipid compound accumulated in Myc11 total lipids with chloroform methanol (v:v, 9:1) (Figure 3-1). Its polarity is close but slightly higher than normal GPLs isolated from wild-type *M. smegmatis* strains prior to de-acetylation treatment. The MALDI- TOF/MS (m/z ranging from 500 to 3000) showed one lipid compound with M.W. 1334 accumulated in the total lipids extracted from Myc11 cell wall. The molecular weight of L1334 is similar to GPLs. There is no other lipid missing from Myc11. L1334 production was turned off in the Myc11 complemented strains.

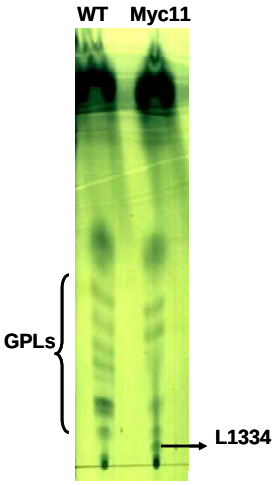


Figure 3-1. TLC experiment with crude total lipids of *M. smegmatis* mc^2155 and Myc11 with developing system (chloroform:methanol:water, 90:10:1).

L1334 has a similar polarity to de-acetylated GPL-5 according to the TLC experiment. There are 6 non-specific types of GPLs isolated from *M. smegmatis*

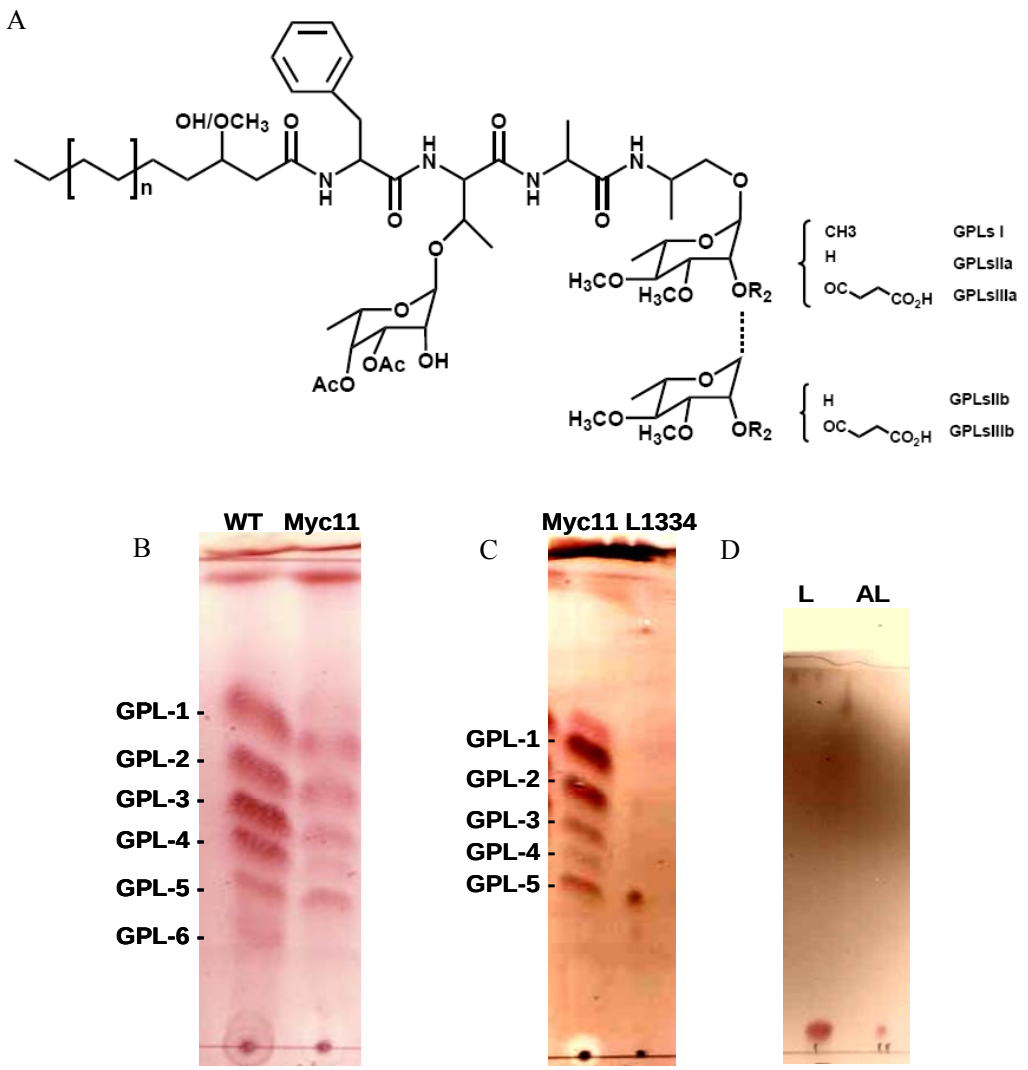


Figure 3-2. A) *M. smegmatis* glycopeptidolipids modified from reference (*54, 61*) and GPL-1 to GPL-6 are labeled according to reference (*99*); B) TLC experiment of alkali-treated total lipids from *M. smegmatis* mc^2155 and Myc11; C) TLC experiment of alkali-treated total lipids from Myc11 and purified L1334; D) purified L1334 (L) and de-acetylated L1334 (AL); the TLC developing system is chloroform:methanol:water (v:v:v, 90:10:1).

mc^{2}155 and their structures are shown in Figure 3-2. Those GPLs are isolated and characterized after mild-alkali treatment of crude total lipids. Crude total lipids of wild-type and Myc11 were also characterized and characterized by TLC experiment.

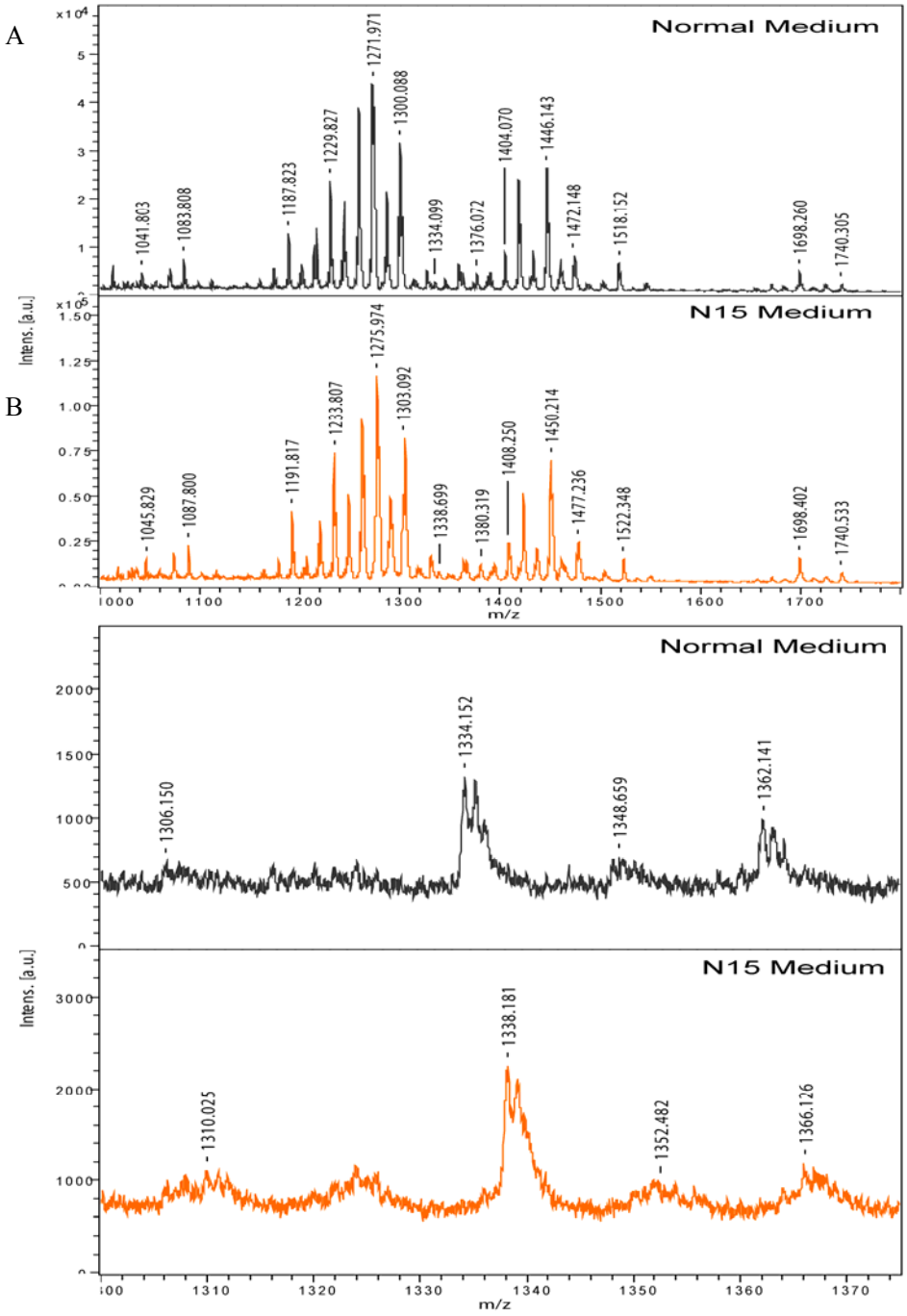


Figure 3-3. Total lipids of ^{15}N medium cultured *M. smegmatis* mc^2 155 and Myc11. A) the mass region from 1000 m/z to 1800m/z; B) the details of L1334 m/z region from 1300 to 1380.

All six types of GPLs are present in the WT lipid profile. Compared with the WT, GPL-6 is barely produced in Myc11. Deacetylated GPL-5 from wild type has the same m/z=1334 peak in the mass spectrum. The deacetylated GPLs were compared with native L1334. L1334 has the same polarity as de-acetylated GPL-5. L1334 is stable after alkali-treatment consistent with a lack of acetyl group (Figure 3-2D). The ^{15}N labeling experiments showed there are four nitrogen atoms in each L1334 molecule. Myc11 were cultured in specially prepared ^{15}N minimal medium for ^{15}N labeled L1334. The total lipids were extracted and purified by Combiflash instrument. The ^{15}N-L1334 was eluted with the same program and same solvent with the control of normal L1334 isolated from Myc11 in normal medium (Figure 3-3).

The total lipids of both ^{15}N labeled Myc11 and wild type show that there is no peak with m/z=1334 (Figure 3-3). The chromatography isolation of this ^{15}N labeled L1334 showed that there is a 4 mass shift higher compared with normal L1334. The 4 higher in mass indicate there are 4 nitrogen atoms in the molecule of L1334. Lipid ^{15}N labeling experiments established that L1334 contains 4 nitrogens. The surface-exposed glycopeptidolipids are important to *M. smegmatis* cell morphology and sliding motility. Their structures all have a common tetrapeptide core d-phenylalanine-d-*allo*-threonine-d-alanine-l-alaninol. There are four nitrogen atoms per molecule. L1334 properties are very close to GPLs in cell morphology changes, molecular weight and polarity. It is highly possible that L1334 could be a mutant

Table 3-1. Tandem MS^n fragment peaks of L1334 by Orbitrap.

Tandem spectrum	**m/z of peaks**			
	$[M\text{-sugar+Na}]^+$	**$[M+Na-H_2O]^+$**	**$[M+Na-HOCH_3]^+$**	**$[M+Na-HOCH_3-H_2O]^+$**
MS^2 (1333.86)	1169.79	1315.86	1301.86	1283.82
MS^3 (1169.77)	1009.71	1151.77	1137.79	1119.75
MS^4 (1009.71)	835.62	991.70	977.70	959.68

nsGPL compound.

Tandem mass spectrometry of L1334. Purified L1334 was dissolved in chloroform:methanol (v:v, 2:1) and injected in Thermo Fisher Scientific LTQ Orbitrap XL ETD for structure analysis. The exact mass of sodium salt of L1334 is 1333.8617 (3 ppm tolerance). The tandem MS spectra are shown in Figure 3-4. The MS^2 fragment of ion at m/z = 1333.86 yielded ions with m/z=1169.79, losing mass of 164.

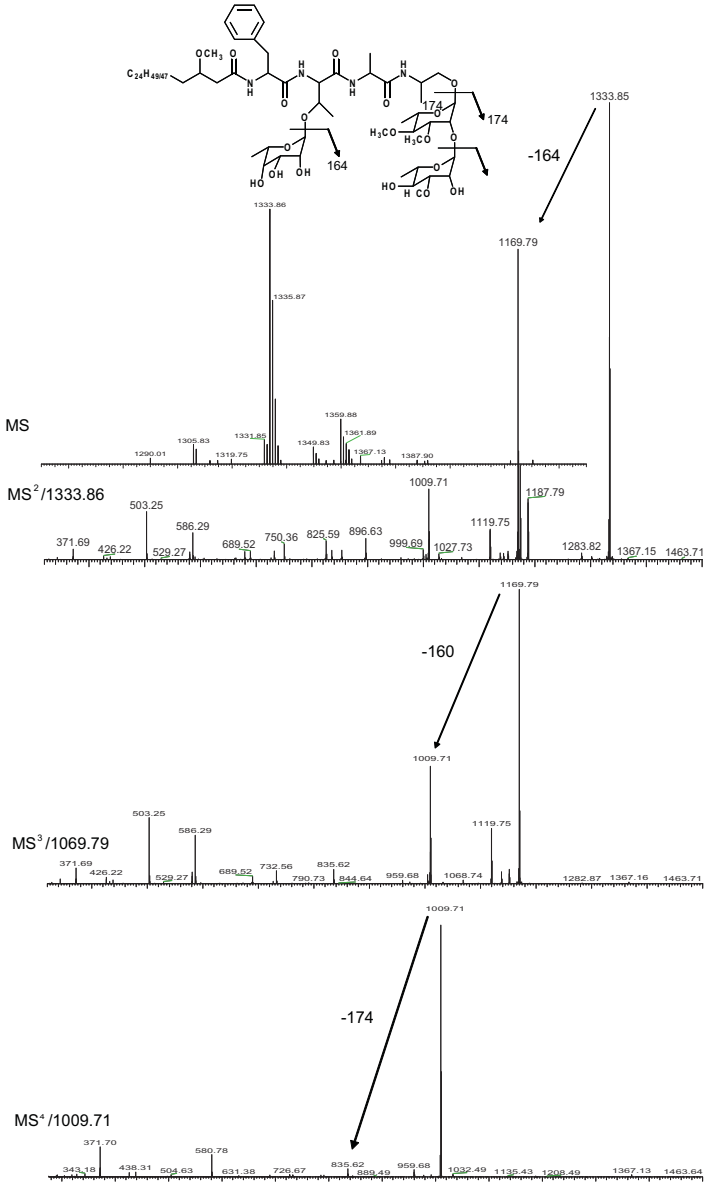


Figure 3-4. Orbitrap tandem mass spectra of purified L1334. The arrows indicate the three sugar residues fragments from the parent peak as indicated in the spectra.

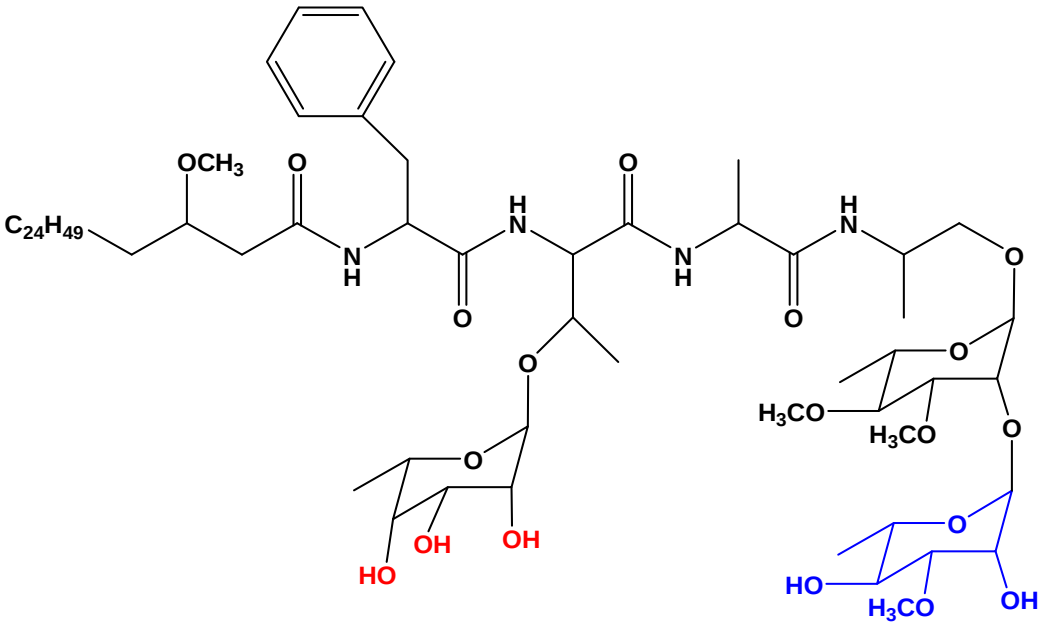


Figure 3-5. Proposed structure of L1334.

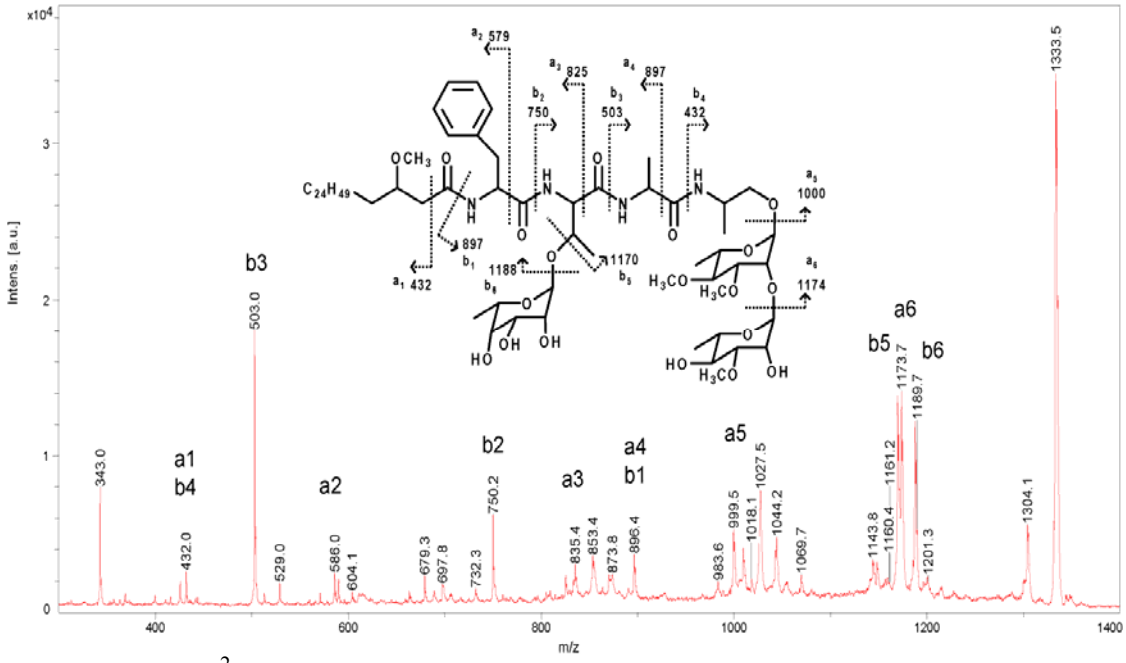


Figure 3-6. MS^2 spectrum of L1334 by MALDI-TOF/MS. The proposed L1334 fragmentations are shown.

The MS^3 fragment of the ion at m/z=1169.79 yielded ion with m/z=1009.71, losing mass of 160. The MS^4 fragment of the ion at m/z=1009.71 yielded ion with m/z=835.62, losing mass of 174. The three sequential fragments of 164, 160 and 174 are three sugar residues from the parent molecules. Also, in each fragment MS there

are series ions by mass loss of 18, 32 and 50 from the parent molecules. They are peaks of $[M+Na-H_2O]^+$, $[M+Na-HOCH_3]^+$, $[M+Na-HOCH_3-H_2O]^+$ (Table 3-1), and they indicate the L1334 structure contains the hydroxyl and methoxyl groups, which are possibly on rhamnose residues.

The best molecular formula for L1334 is $C_{69}H_{122}N_4O_{19}Na$, on the basis of the exact mass (1333.8617 with 3 ppm tolerance) and nitrogen number information. The calculated exact mass of $C_{69}H_{122}N_4O_{19}Na$ is 1333.8601, close to the exact mass analyzed by the high-resolution Orbitrap. This molecular formula has a similar carbon/oxygen number and same nitrogen number with GPLs compound. The proposed structure of L1334 is shown in Figure 3-6. This GPL-like structure of L1334 was confirmed by MALDI-TOF/MS^2 (Figure 3-5).

NMR analysis of L1334. The structure of L1334 was analyzed by NMR spectroscopy. Both ^1H NMR and ^1H-^{13}C HSQC NMR spectra were acquired. The ^1H-NMR spectrum is very similar to NMR signals of the typical GPLs isolated from

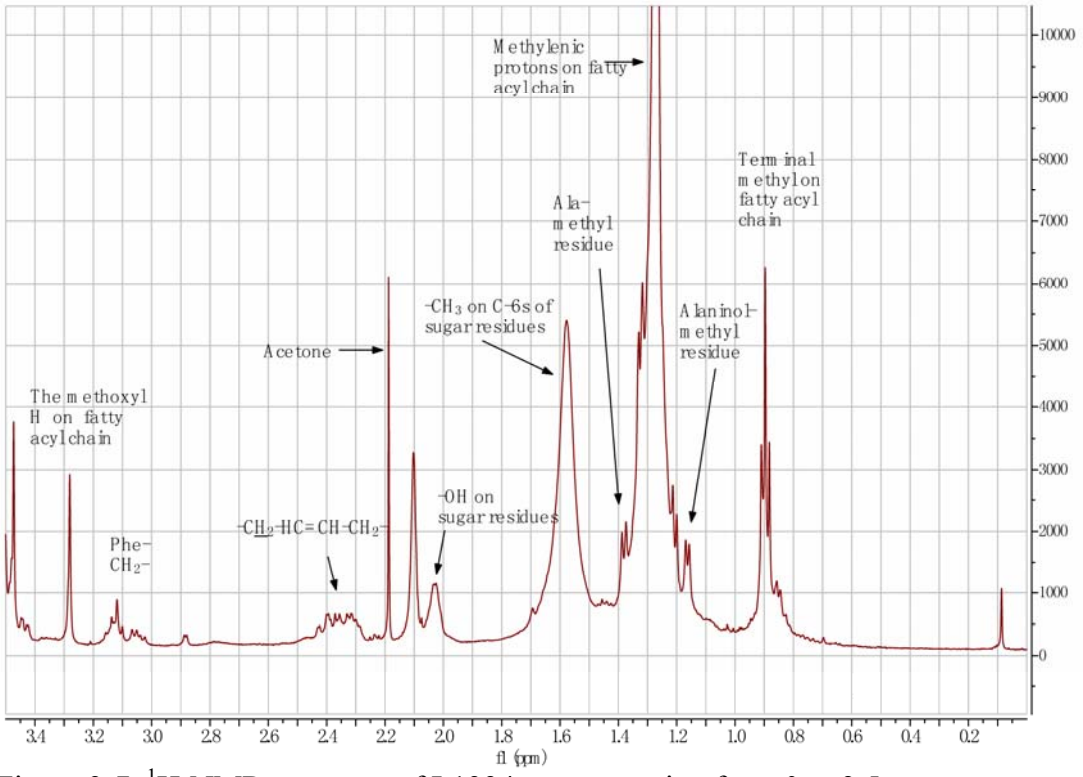


Figure 3-7. ^1H-NMR spectrum of L1334; proton region from 0 to 3.5 ppm.

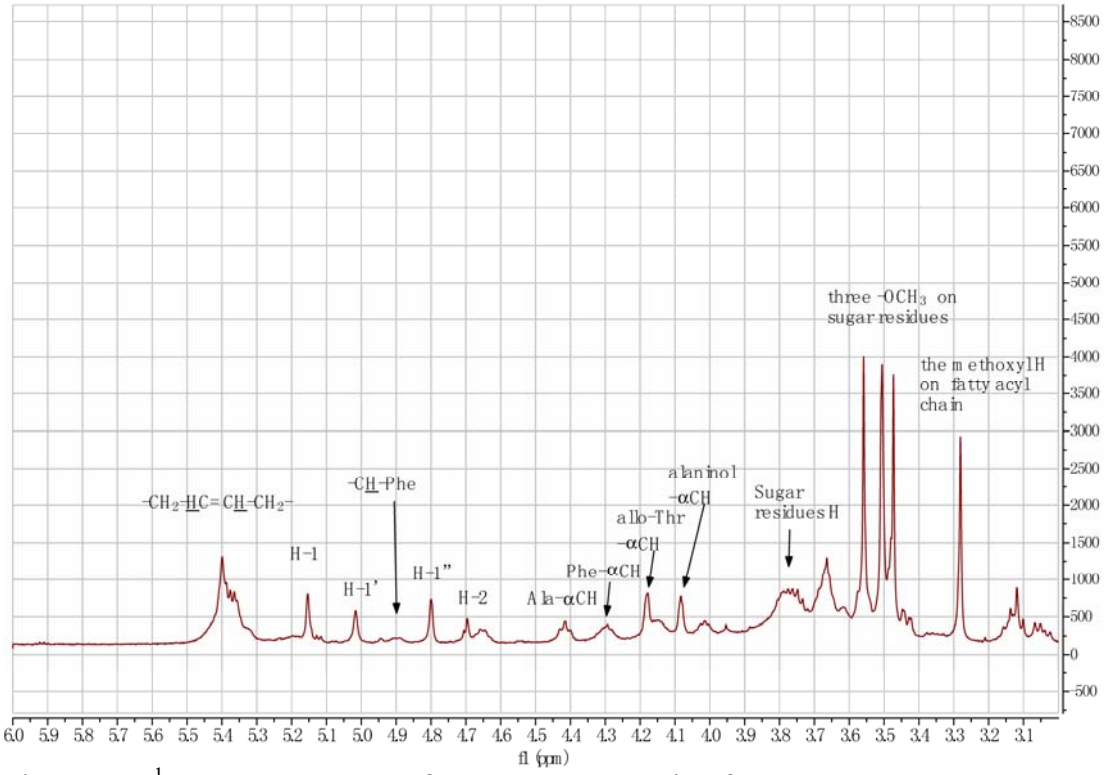


Figure 3-8. ^{1}H-NMR spectrum of L1334; proton region from 3.5 to 6 ppm.

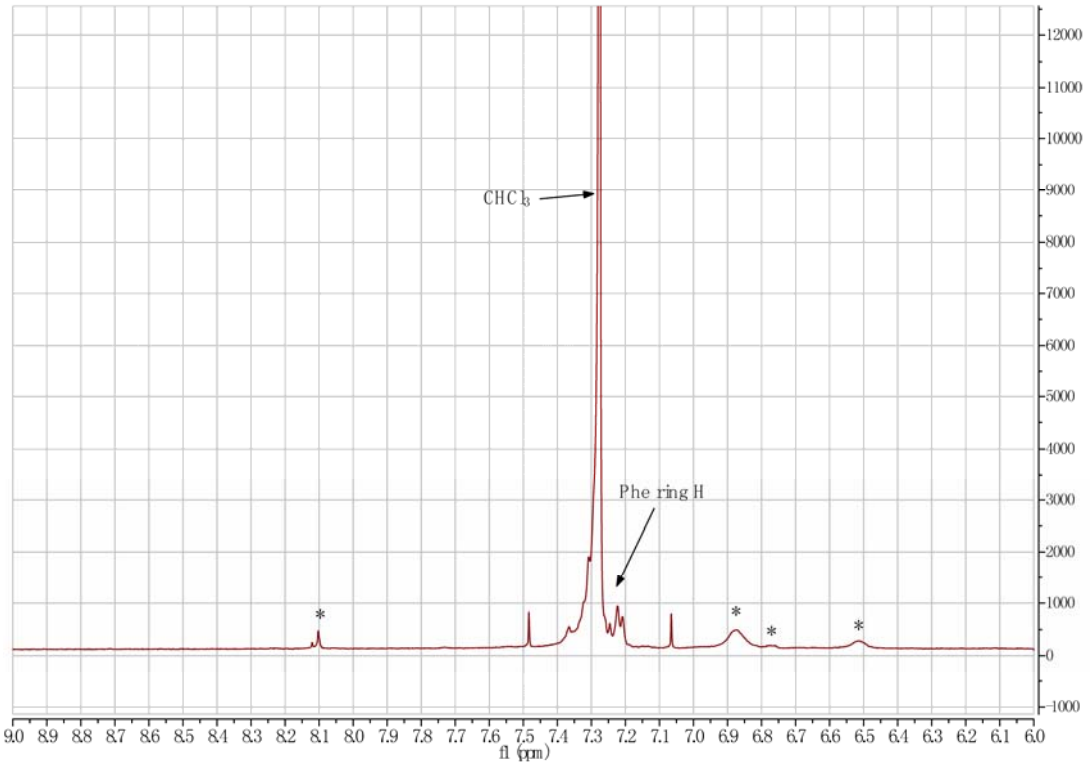


Figure 3-9. ^{1}H-NMR spectrum of L1334; proton region from 6 to 9 ppm (* are amide protons from tetrapeptide core of L1334).

M. smegmatis. Both ^{1}H-NMR and ^{1}H-^{13}C HSQC confirmed three most important parts in the L1334 structure: a) four amino residues including aromatic protons; b) a long fatty acyl chain; c) three sugar residues including the three anomeric carbons (Figure 3-6, Figure 3-7, Figure 3-8 and Figure 3-9). The full spectra are found in the Appendix Figure S1. The NMR spectrum confirmed the deacetylated glycopeptidolipid structure of L1334, consistent with the mass spectra.

Conclusively, L1334 is accumulated as a deacetylated glycopeptidolipid in *M. smegmatis MSMEG1604* transpson mutant cell wall. The accumulation of L1334 alters Myc11 into a rough phenotype from a smooth phenotype. Myc11's sliding motility was retarded. NMR and mass spectrometry confirmed the purified L1334 has a structure of deacetylated GPL-5. The structure of L1334 is shown in Scheme 1-1. Because PGLs are the *M. tuberculosis* functional homologs of GPLs, whether a unique PGL variant is accumulated upon *Rv3409c* disruption should be studied in the future. The differences of L1334 from normal GPL especially GPL-5 is the acetylation on the 6-deoxytalose residue on the GPL. According to TLC experiments and MS spectra, Myc11 does not produce GPL-6, but does produce both deacetylated GPL-5 and normal GPL-5 (m/z=1518) (Figure 3-2). The acetylation of other GPLs (1 to 5) is all normal. Rv3409c is not within the GPL biosynthesis gene cluster (*65*). The Atf (acetyltransferase) from *M. smegmatis* has not been characterized in vitro. The atf mutant produce all unacetylated GPLs in the cell wall (*50*). There is one hyperglycosylated GPLs found in the carbon starvation culture *M. smegmatis*. The hyperglycosylated GPL accumulats due to carbon source starvation (*100*). Therefore, there is a relationship between sugar metabolism and cell wall lipid structures. The reason that one GPL compound is not acetylated and is triglycosylated upon *Rv3409c* disruption is unclear for us. It is highly possible that the carbon metabolism might be

affected due to Rv3409c disruption. Rv3409c is not a cholesterol oxidase and our results will help in understanding the real function of Rv3409c in future.

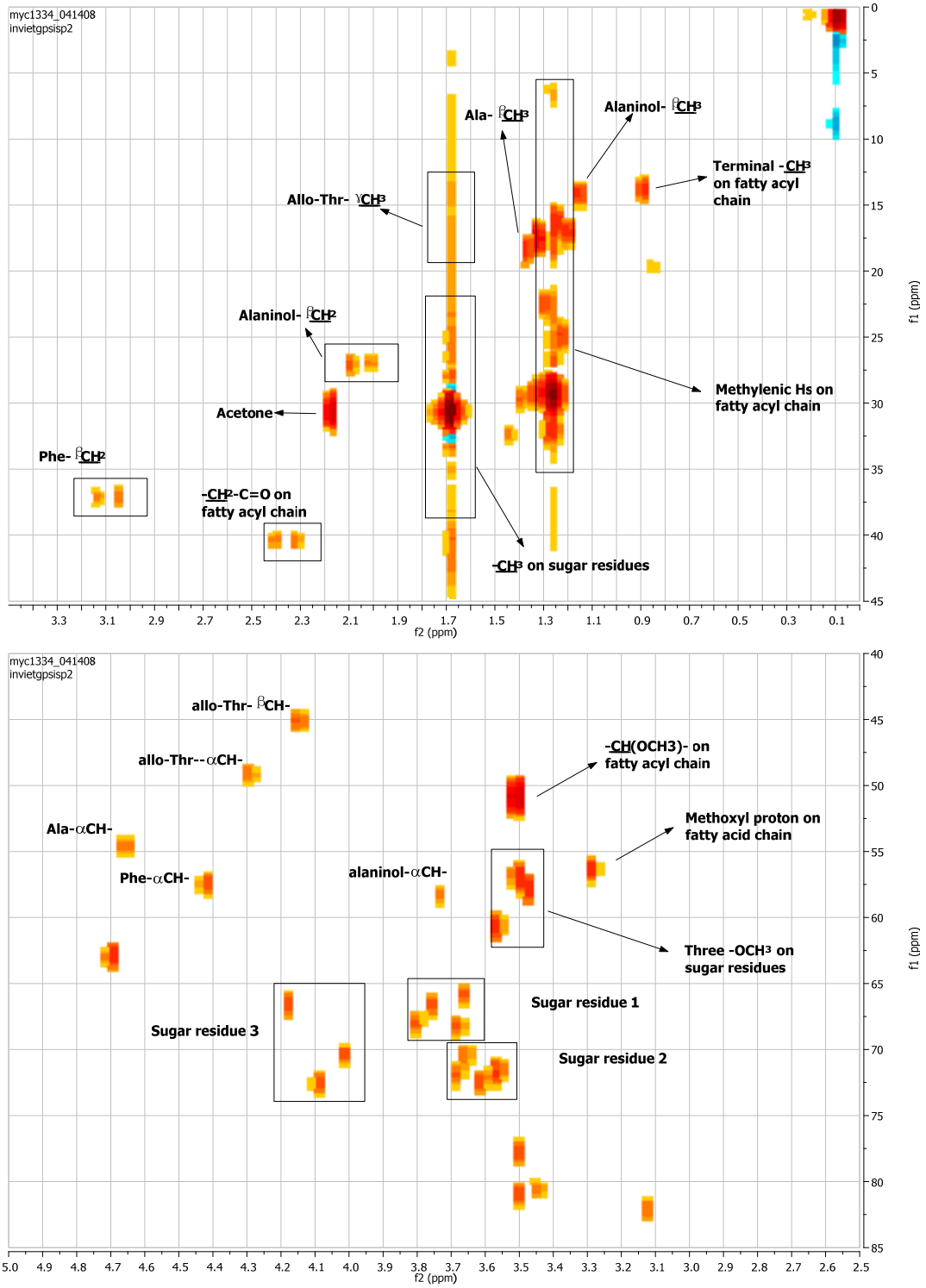


Figure 3-10. ^{1}H-^{13}C HSQC NMR of L1334.

Chapter 4

Rv3409c Purification, Reconstitution and Activity Characterization

I. Introduction

Mycobacterium tuberculosis (*M. tuberculosis*) has successfully infected one third of the whole world's population and every year about 2 million people die because of tuberculosis according to the data from World Health Organization (WHO) (*1-2*). Only 10% immuno-competent *M. tuberculosis* infected individuals could develop active TB diseases (*4*). Latent tuberculosis can be reactivated when the host immune system is compromised, for example due to HIV or malnutrition. About 15% of HIV infected people are co-infected with tuberculosis and one third of them die every year (*3*).

M. tuberculosis can survive and replicate in phagosome. There are several pathways by which *M. tuberculosis* can survive. It can inhibit phagosome-lysosome (P-L) fusion, delay phagosome maturation or interrupt IFN-γmediated signaling or adapt the host nutrition supply or alter cell wall lipids (*6-15*). But all the mechanisms remain to be studied for more details.

Cholesterol is converted to cholest-4-en-3-one as the first step of sterol ring degradation (*34*). On the basis of bioinformatic analysis, *Rv3409c* was annotated as a putative cholesterol oxidase. Cholesterol oxidase is a flavoenzyme with one molecule FAD cofactor covalently or non-covalently bound. The FAD non-covalently bound cholesterol oxidases is the type I and FAD covalently bound cholesterol oxidases is the type II (Figure 4-1). Although these two types of enzyme catalyze the same reaction, the protein structure and the sequence are completely different from each other (*34*). Rv3409c is grouped into the Glucose-Methanol-Choline (GMC) oxidoreductase superfamily on the basis of its sequence including FAD binding domain and putative conserved active site amino acid residues. The GMC

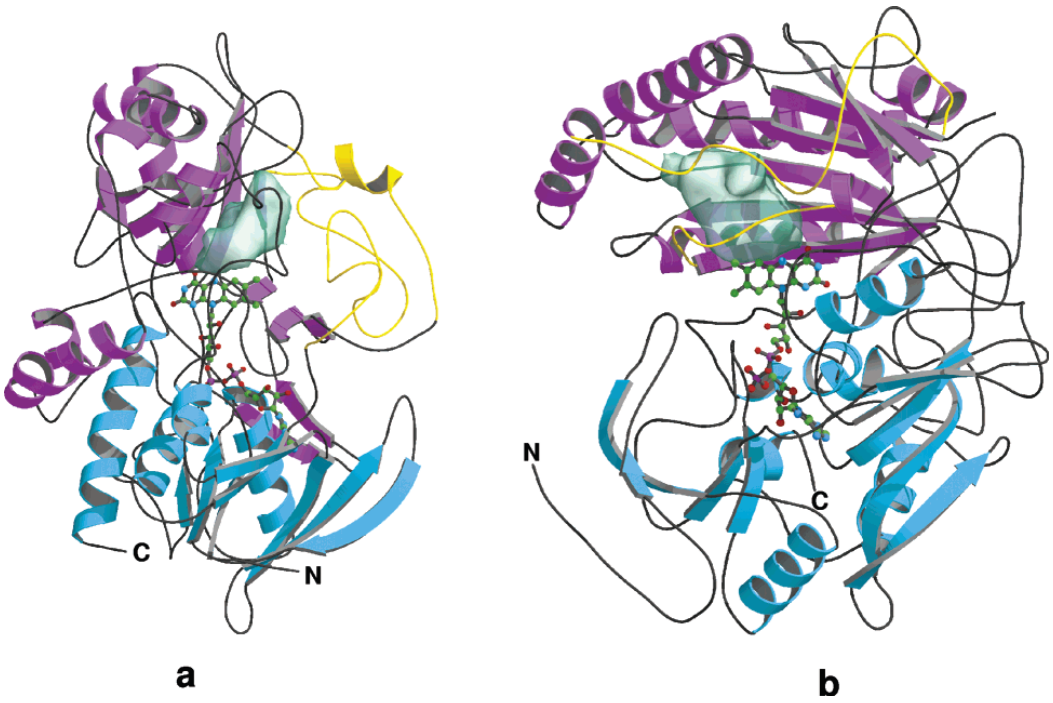


Figure 4-1. Secondary structure representations of (a) type I cholesterol oxidase (PDB entry 1B4V) and (b) type II cholesterol oxidase (PDB entry 1I19) showing the buried substrate cavity, copied from reference (*34*).

oxidoreductase family includes a large number of flavoenzymes like alcohol oxidase, glucose oxidase, choline oxidase, and cholesterol oxidase (*101-104*). The common function of enzymes of GMC family is oxidizing an alcohol into a ketone or an aldehyde. The conserved residues are a histidine (His447) for oxidation and an asparagine (Asn485) or histidine for substrate oxidation and oxygen gating (*101-102, 104-107*). Cholesterol oxidase has 1,3-allylic isomerization activity in addition to alcohol oxidation. In cholesterol oxidase choA from *Streptomyces*, the glutamic acid (Glu361) residue in the active site is required for isomerization of cholest-5-en-3-one to cholest-4-en-3-one. Glu361 is not conserved in the GMC superfamily (*108*).

Rv3409c has 24% identity and 53% similarity with the cholesterol oxidase (choE) from *Rhodococcus equi* strain. Rv3409c contains the highly conserved FAD

binding motif of the GMC superfamily, which indicates Rv3409c is a flavoenzyme and utilizes one FAD. The active site residues His447 and Asn485 are conserved in Rv3409c, but a conserved Glu361 could not be identified. *M. tuberculosis Rv3409c* also has orthologs other mycobacteria like *Mycobacterium bovis*, *Mycobacterium smegmatis*, and *Mycobacterium leprae*.

It has been reported that cell lysate supernatants of *M. smegmatis* over-expressing *Rv3409c* have cholesterol oxidation activity (*45*). However, the *Rv3409c* protein product has not been purified and assayed *in vitro*. The cholesterol oxidizing ability of *M. tuberculosis* cell lysate supernatant was not abrogated after disruption of *Rv3409c* (Eugenie Dubnau, personal communication, Figure 2-4).

M. tuberculosis Rv3409c is an important gene playing a key function in the *M. tuberculosis*. The growth of the *Rv3409c* mutant is attenuated in mouse lungs and spleens (*45*). The *Rv3409c* mRNA level is increased 20-fold 2 weeks after infection and 200-fold 3 months post-infection in mouse lungs compared to cell culture (*109*). In this chapter, *Rv3409c* from *M. tuberculosis* H37Rv has been cloned and its gene product was over-expressed and purified. Enzyme Rv3409c cholesterol oxidizing ability and flavoenzyme characteristics were analyzed. The enzyme activity with possible GMC oxidoreductase enzyme substrates was screened.

II. Experiments and Methods

Materials and general methods. M. tuberculosis H37Rv total genomic DNA was obtained from the TB research Materials Facility at Colorado State University (Fort Collins, CO) (NIAD NO1-AI40091). All oligonucleotides were purchased from Stony Brook University DNA Sequencing Facility (Stony Brook, NY). *M. smegmatis* mc^2155 strains carrying expression plasmids were grown in 7H9 Middlebrook broth

medium supplemented with 0.2% (w/v) glucose and 0.05% (v/v) Tween 80. Antibiotics were added as needed at the following concentrations: kanamycin 30 μg/mL; ampicillin 200 μg/mL; hydromycin 50 μg/mL; and cycloheximide 10 μg/mL. Restriction endonucleases and trypsin protease enzyme were purchased from New England Biolabs (Beverly, MA). The Rapid DNA Ligation Kit was purchased from Roche (Mannheim, Germany). Cloned Pfu DNA polymerase was purchased from Stratagene (La Jolla, CA). 7H9 Middlebrook broth medium was purchased from Sigma-Aldrich (St. Louis, MO). Superdex 200 16/60 column was purchased from GE Healthcare Biosciences Corp. (Piscataway, NJ). Mono Q 5/5 column was purchased from Pharmacia Biotech (Piscataway, NJ). HisBind Resin Ni-charged was purchased from Invitrogen (Carlsbad, CA). Mass spectra were acquired on a Bruker Autoflex II MALDI-TOF/TOF instrument. DNA sequencing data was acquired by Stony Brook University DNA Sequencing Facility (Stony Brook, NY). All other chemicals and solvents are from Fisher Scientific (Pittsburgh, PA). The buffers used include: Binding buffer: 20 mM TrisHCl, 10 mM imidzole, 0.5 M NaCl, 10% (v/v) glycerol, pH 7.4; Washing buffer: 20 mM TrisHCl, 5 mM imidzole, 0.5 M NaCl, 10% (v/v) glycerol, pH 7.4; Elution buffer: 20 mM TrisHCl, 150 mM imidzole, 0.5 M NaCl, 10% (v/v) glycerol, pH 7.4; Assay buffer: 50 mM sodium phosphate, pH 7.0 and size-exclusion chromatography elution buffer: 50 mM sodium phosphate, pH 7.4; Mono Q starting buffer: 20 mM sodium phosphate, 10% glycerol, pH 7.4; Mono Q elution buffer: 20 mM sodium phosphate, 10% glycerol, 1 M sodium chloride, pH 7.4. Cholesterol (3 mM) in 2-propanol stock. Cholest-4-en-3-one (3 mM) in 2-propanol stock. Cholest-5-en-3-one (3 mM) in 2-propanol stock. FAD (5 mM) in DDI water stock. Horseradish peroxidase (1000 unit/mL) in sodium phosphate buffer, pH 7.0, stock. 2,2′-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) (10 nM) in

sodium phosphate buffer pH 7.0.

Rv3409c subclone and plasmids construction. All plasmids were propagated in *E. coli* strain XL1Blue. Plasmids names and primers are listed in Table S1 in the Appendix.

Construction of Rv3409c in pET-28b(+) plasmid with N terminal Histidine tag. *Rv3409c* clone insert was cloned with pMCO-100M as template and forward primer JGNN-1 (5'- cgcCATATGaagccggattacgacgtcctg-3') and reverse primer JGNN-2 (5'- ccg AAGCTTctagcccgcgttgctgaccgg-3'). The 50 μL PCR mixture includes: 5 μL 10× Pfu turbo cloned buffer, 4μL 10 mM dNTP, 1 μL 2mM forward primer JGNN-1, 1 μL 2mM reverse primer JGNN-2, 0.5 μL Pfu turbo polymerase and 36 μL DDI water. The PCR program is

Segment	Cycles	Temperature	Time
1	1	95 °C	2 min
2	35	95 °C	45 sec
		65 °C	45 sec
		72 °C	2.5 min
3	1	72 °C	10 min

Both the *Rv3409c* clone purified from finished PCR reaction and pET-28b(+) vector was double digested with proteases NdeI and HindIII. The ligation was completed with the Rapid DNA Ligation kit by Roche (Mannheim, Germany) with the standard procedures. The colony with pET-28b-NChoD was isolated and amplified for pET-28b-NChoD storage.

Construction of Rv3409c in pFPCA1 vector. pMCO-100 vector containing *Rv3409c* with acetamide promoter was constructed by Paigee Chou and kanamycin resistant gene cassette were stored at -80 ^{o}C. For pMCO-100M plasmid construction, the pMCO-100 was used as gene clone template. The QuikChange® Site-Directed

Mutagenesis protocols was followed. Partially overlapping primers JG005 (5'-cgcgggc**T**agcttgcggccgcactcgagc-3') and JG006 (5'-cgcaagct**A**gcccgcgttgctgaccggatc -3') were designed. The 50 µL PCR reaction mixture include 4 µL dNTP's (2.5 mM each), 5 µL 5× iProof GC buffer, 2 µL 10 µM primer JG005 and 2 µL 10 µM primer JG006, 84 ng pMCO-100 as parent plasmid, and 0.5 µL 2 unit/µL iProof High-Fidelity DNA Polymerase. PCR reaction was conducted on the PTC-100 Programmable Thermal Controller (MJ Research, Inc) with the program below:

Segment	Cycles	Temperature	Time
1	1	94 °C	3 min
2	16	94 °C	1 min
		56 °C	1 min
		68 °C	16 min
3	1	68 °C	1 h

1 µL 20 U/ µL DpnI restriction enzyme was added to 50 µL PCR reaction mix. The total mixture was incubated at 37 °C 60 min. 4 µL DNA products were picked for transformation and amplification in XL1Blue cells. The mutated pMCO-100M vector was isolated using the Wizard Plus SV Minipreps DNA Purification System. The pMCO-100M vector was transformed into *M. smegmatis* mc^2155 by electroporation.

Construction of Rv3409c in pVV16.hygro plasmid. The pET-28bNChoD was used as a template to prepare the N-terminal Histidine tagged Rv3409c (N-Rv3409c) in pVV16.hygro plasmid. The forward primer JGNN-5 (5'-cgcggatcccatcatcatcatcacagcagc-3') and backward primer (5'-ccgaagcttctagcccgcgttgctgaccgg-3') were designed to clone the *Rv3409c* insert for pVV16.hygro vector. The PCR reaction mixture includes 5 µL 10× Pfu turbo cloned buffer, 4 µL 10 mM dNTP, 1.25 µL 2mM forward primer JGNN-1, 1.25 µL 2mM

reverse primer JGNN-2, 2 μL DMSO, 1 μL Pfu turbo polymerase and 33 μL DDI water. The PCR program is

Segment	Cycles	Temperature	Time
1	1	95 °C	45 sec
2	35	95 °C	45 sec
		65 °C	45 sec
		72 °C	3 min
3	1	72 °C	10 min

The *Rv3409c* insert was isolated from the PCR reaction mixture on a 1% agarose DNA gel and then both pVV16.hygro vector and the *Rv3409c* clone were sequentially digested by BamHI and HindIII. The ligation was completed with the Rapid DNA Ligation kit by Roche (Mannheim, Germany) with the standard procedures The PCR mixture was transferred into *E. coli* competent cells for isolation. The colony with pET-28bNRv3409c was cultured and amplified for pET-28bNRv3409c storage.

Transformation and electroporation. For *E. coli* over-expression system, purified plasmids were incubated with *E. coli* competent cells for 20 min and then heated to 42 °C for 45 seconds and then cooled on ice for another 2 min. 2 mL fresh LB medium were added into the mixture and the culture shaked at 250 rpm for another 20 min. Then 100 μL cell culture was inoculated to agar plates with proper antibiotics. After 17 h incubation at 37°C, a single colony was chosen for plasmid isolation. For *M. smegmatis* mc^2155 over-expression system, the shuffle plasmids were transformed into *M. smegmatis* mc^2155 competent strain using an electro cell Manipulator 600 (BTX Electroporation System) and 1 mm gap cuvette. 25 μL *M. smegmatis* mc^2155 electrocompetent cells with 10 ng of plasmid were gently mixed

and aliquoted in the prechilled gap cuvettes. The parameters are mode: HV; capacitance: 50 μF; resistance: R5; charging voltage: 1.7 kV; desired filed strength: 17 kV/cm and desired pulse length: 4-5 msec.

N-Rv3409c protein expression and purification. Culture was started in a 50 mL tube with shaking at 250 rpm, 37 °C until OD_{600}=0.6-0.8. The cell pellets were collected and inoculated into 1 L protein expression medium in a 4 L flask. 0.088% acetamide was added to the culture to induce enzyme over-expression when OD_{600}=0.6-0.8. After 10 hours, the cells were harvested and re-suspended in Binding buffer for cell lysis.

Step-1: The frozen cells from 3 L *M. smegmatis* mc^2155 over-expressing N-Rv3409c were thawed and resuspended in 50 mL Binding buffer. After three times passage through a French press at 10,000 psi, the cell lysate supernatant was collected by ultracentrifugation at 135, 000 g for 90 min, 4 °C.

Step-2: The cell lysate supernatant was loaded onto a pre-equilibrated Ni^{2+} resin column (1×5 mL) and the column was further washed with binding buffer and washing buffer to remove impurity proteins. Finally, the protein was eluted with Elution buffer and the fractions were collected and analyzed by 10% SDS-PAGE.

Step-3: Fractions contain N-Rv3409c were combined for size-exclusion chromatography. They were concentrated by ultrafiltration (Centricon, NMWCO=10,000). A Superdex 200 16/60 column (Pharmacia Biotech) was pre-equilibrated with 20 mM sodium phosphate, pH 7.4, 4 °C. After loading N-Rv3409c, elution buffer (20 mM sodium phosphate buffer, pH 7.4) was applied at a flow rate of 0.5 mL/min, with UV detecting at 280 nm. All fractions absorbing at 280 nm were further analyzed by 10% SDS-PAGE. N-Rv3409c containing fractions were collected and combined.

Step-4: A Mono Q 5/5 column was pre-equilibrated with starting buffer (20 mM sodium phosphate, 10% glycerol, pH 7.4). The N-Rv3409c sample was loaded onto the Mono Q 5/5 column and 5 column volumes (CV) starting buffer were applied to wash the column. A slow linear gradient from starting buffer to elution buffer (20 mM sodium phosphate, 10% glycerol, 1 M sodium chloride, pH 7.4) over 10 CV at a flow rate of 0.5 mL/min was used for elution with UV detecting at 280 nm. All fractions absorbing at 280 nm were analyzed further by 10% SDS-PAGE. The fractions containing N-Rv3409c were combined for further analysis.

In-gel trypsin digestion experiment. In-gel trypsin digestion combined with MALDI-TOF/MS was applied to confirm N-Rv3409c's identity. About 5 μg N-Rv3409c from the SDS-PAGE gel was treated with trypsin and the digested fragments were extracted three times with 60% CH_3CN/ 0.1% trifluoroacetic acid (TFA) in water. The total extracts were lyophilized and the residue was dissolved in 5 μL 0.1% TFA. A saturated α-cyano-4-hydroxycinnamic acid solution was freshly prepared as the matrix. MALDI-TOF/mass spectra were acquired in positive ion mode.

Cholesterol oxidase activity assays.

Kinetic Assays:

1. A240: The assay mixtures were incubated at 37 °C. The production of cholest-4-en-3-one was monitored by UV absorbance using ε_{240} = 12,100 $M^{-1}cm^{-1}$. The standard assay mixture contained 50 mM sodium phosphate buffer, pH 7.0, 0.05% (w/v) Triton X-100, and 150 μM cholesterol, which was added as a cholesterol stock dissolved in iso-propanol, and 1 μM enzyme. 30 μM to 100 μM FAD was added as necessary.

2. UV-based ABTS-HRP coupled Assay: The assay was performed at 37 ^{o}C. The standard assay mixture contained 50 mM sodium phosphate buffer, pH 7.0, 0.05% (w/v) Triton X-100, 0.32 nM ABTS and 1 unit/ mL Horseradish peroxidase (HRP), 150 μM cholesterol, which was added as a cholesterol stock dissolved in iso-propanol (3 mM), and 1 μM enzyme. The formation of ABTS oxidized form was monitored at 600 nm. 30 μM to 100 μM FAD was added as necessary. Choline, glucose, myo-inositol, methanol, ethanol, benzyl alcohol, 2-phenyl ethanol, rhamnose, and deoxytalose were prepared 100 mM in water. 1-pentanol was prepared 100 mM in acetone.

3. DCIP/PMS electron transfer assay (*110*): The assay was performed at 37 ^{o}C. 50 mM sodium phosphate buffer (pH 7.0) is the standard assay buffer. The final concentration of 2,6-dichloroindophenol (DCIP) is 0.048 mM, phenazine methosulfate (PMS) is 0.9 mM, and the substrate final concentration is 150 μM and all other substrates final concentrations are 10 mM. 30 μM to 100 μM FAD was added as necessary. The enzyme sample was added with a final concentration 1 μM. Formation of the PMS oxidized form was monitored at 600 nm. Cholesterol was 3 mM in isopropanol. Choline, glucose, myo-inositol, methanol, ethanol, benzyl alcohol, 2-phenyl ethanol, rhamnose, and deoxytalose were prepared 100 mM in water. 1-pentanol was prepared 100 mM in acetone.

Isolation and characterization of FAD from the N-Rv3409c. Purified N-Rv3409c was boiled at 100 °C for 20 min until protein precipitated from the solution. The supernatant was collected after centrifugation at 13,000 rpm for 5 min. The supernatant was mixed with 10 mg/mL 2, 3-dihydroxybenzoic acid (DHBA) as matrix for MALDI-TOF/MS analysis.

Reconstitution of N-Rv3409c. Buffers and Reagents: Denaturing buffer: 0.05 M

potassium phosphate, 6 M urea, 3.5 M KBr, pH 7.4; Elution buffer: 0.03 M potassium phosphate, 6 M urea, pH 7.4; 4 mM FAD in Elution buffer; Dialysis Buffer: 0.03 M potassium phosphate, 0.01 M KCl, pH 7.4; and a G25 column (1×9cm). Purified enzyme was concentrated by Ultrafiltration (Centricon, NMWCO=10,000) to about 1.5 mg/mL into Denaturing buffer and incubated at 25 °C for 4 h in the dark. 4 mM FAD was loaded onto a G25 column pre-equilibrated with Elution buffer and allowed to migrate 3 cm before loading the denatured N-Rv3409c sample. Fractions from the G25 column were collected and the N-Rv3409c containing fractions were combined and dialyzed (NMWCO=10,000). Dialysis was performed against 2 L Dialysis buffer.

Fluorescence emissions scan of reconstituted N-Rv3409c. A free FAD standard solution with the same UV absorbance at 350 nm was prepared. The fluorescence emission scan of reconstituted N-Rv3409c samples was acquired with a blank buffer as a control. The emission scans were acquired from 375 nm to 425 nm with an excitation wavelength of 350 nm.

Glucose oxidase assays. The glucose standards (glucose in DDI H_2O) were: 0 mg/mL, 0.1 mg/mL, 0.25 mg/mL, 0.5 mg/mL and 0.75 mg/mL, and 1 mg/mL. ABTS-HRP coupled glucose oxidase assay mixtures include 888 μL of 50 mM sodium phosphate buffer, pH 7.0; 82 μL of ABTS (10 nM); 15 μL of horseradish peroxidase (1000 unit/mL) and 50 μL of glucose standards or sample for testing. 15 μL of glucose oxidase (1000 unit/mL) was added as the last reagent and the UV absorbance at 595 nm was monitored at 25 °C. 1 mL of *M. smegmatis* wild type or *M. smegmatis MSMEG1604* transposon mutant culture was centrifuged to remove the cell pellets. The clear supernatants were analyzed in the glucose oxidase assay.

III. Results and Discussion.

3.1 Rv3409c over-expression and purification

Rv3409c was expressed in 5 different *E. coli*/plasmids combinations (Table 4-1). In all systems, inclusion bodies were obtained. *M. smegmatis* mc^2155 was used as the host to express *Rv3409c* to solve the solubility problem. All the *M. smegmatis* mc^2155 expression systems provided soluble proteins (Table 4-1).

3.1.1 C-terminus His-tagged Rv3409c (C-Rv3409c) over-expression

For simple purification, a His_6 tag was constructed at C-terminal of Rv3409c (C-Rv3409c). The optimum induction conditions were explored as listed in Table 4-1. The highest yield was about 3 mg/L. The soluble protein was colorless which indicated the FAD cofactor was missing. The identity of C-Rv3409c was confirmed by in-gel trypsin digestion for tryptic peptides mass analysis. About 28% peptide sequences were confirmed by MADLI-TOF/MS analysis. C-Rv3409c was assayed with the A240 assay method with extra FAD added in the assay and no cholesterol oxidizing ability was identified.

Table 4-1. Plasmids and expression systems tested for Rv3409c protein expression. (* construction and purification work done by Paigee Chou.)

Vector	Strain	Affinity Tag	Induction	Protein status	yield	Activity
pET-20b(+)*	*E.coli* BL21(DE3)	C-terminus His-tag	25 °C, 0.4 mM IPTG, 14.5 h	Inclusion bodies	N/A	ND
pET-20b(+)*	*E.coli* BL21(DE3)pLYsS	C-terminus His-tag	10 °C, 0.1 mM IPTG, 61 h	Inclusion bodies	N/A	ND
pET-20b(+)*	*E.coli* C41(DE3)	C-terminus His-tag	10 °C, 0.1 mM IPTG, 30 h	Triton X-100 recovered limited solubility	N/A	ND
pET-20b(+)*	*E.coli* BL21(DE3) Codon plus-RIL	Non-tag	No Colony	No Colony	N/A	N/A
pET 28b (+)	*E.coli* BL21DE3, Rosetta, RosettaDE3	N-terminus His-tag	4mM IPTG, 18 °C, 20 h	Inclusion bodies	N/A	N/A
pFPCA1	*M. smegmatis* mc^2155	N-terminus His-tag	35.5 °C 0.088%acetmide	Soluble/colorless	1 mg/L	ND
pFPCA1	*M. smegmatis* mc^2155	Non-Tag	35.5 °C 0.088%acetmide	Soluble	N/A	ND
pFPCA1	*M. smegmatis* mc^2155	N-terminus His-tag	35.5 °C 0.088%acetmide	Soluble/yellow	1 mg/L	ND
pVV16.hygro	*M. smegmatis* mc^2155	N-terminus His-tag	35.5 °C to 45 °C	Soluble/yellow	3 mg/L	ND

3.1.2 Non-tagged Rv3409c for activities

Non-tagged Rv3409c was over-expressed in *M. smegmatis* mc^2155 transformed with the plasmid pMCO-100M. The same over-expression and induction protocol used for C-Rv3409c was applied to this over-expression system. It was assumed that the non-tagged Rv3409c would be over-expressed in the *M. smegmatis* mc^2155 system. The cell lysate supernatant was analyzed using the A240 assay. There was no cholesterol oxidizing activity detected.

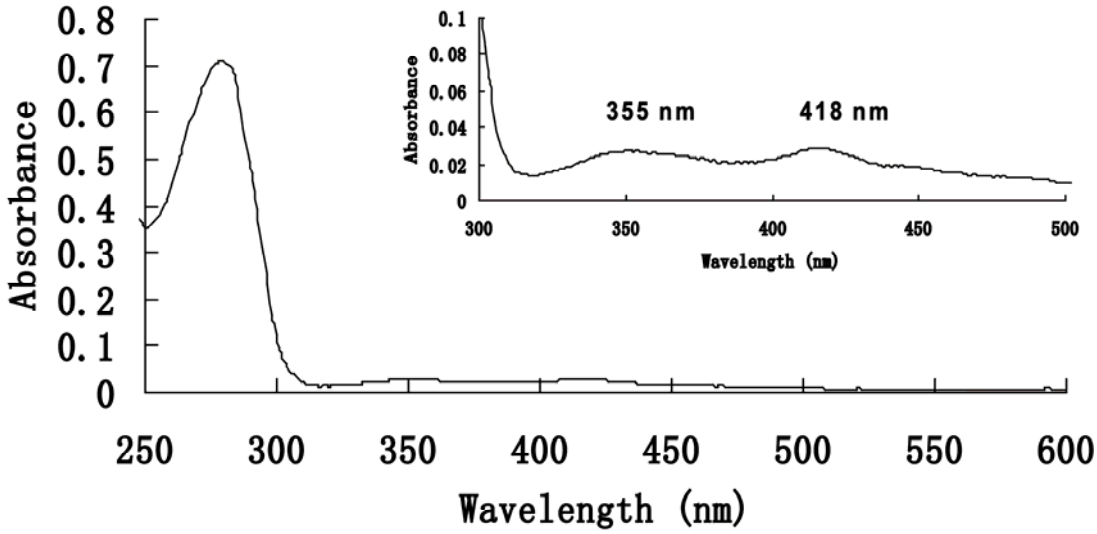


Figure 4-2. UV-vis spectrum of N-Rv3409c enzyme purified by Ni Hisbind column.

3.1.3 N-terminal Histagged Rv3409c (N-Rv3409c) over-expression and purification

Rv3409c was also purified with an N-terminal His_6 tag from *M. smegmatis* mc^2155 (N-Rv3409c). The isolated N-Rv3409c was soluble and a yellow color was present. The UV-vis spectrum is shown in Figure 4-2. With the pVV16.hygro-NRv3409c plasmid construction with heat shock promoter *hsp60*, N-Rv3409c yield increased from 1 mg/L to 3 mg/L. The optimized induction conditions required increasing the temperature from 35.5 °C to 45 °C when

OD_{600}=1.0.

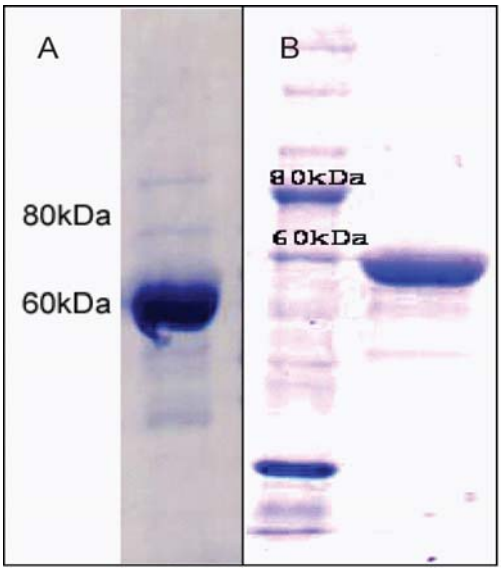


Figure 4-3. 10% SDS-PAGE analysis with N-Rv3409c by A) the size-exclusion chromatography and B) the cation exchange chromatography.

N-Rv3409c eluted from the Ni Hisbind column with impurities as judged by 10% SDS-PAGE and further purification was required. Further purification by size exclusion chromatography (Superdex 200 column) and cation exchange (MonoQ column) chromatography were employed. N-Rv3409c was concentrated and injected into pre-equilibrated Superdex 200 column on AKTA system. By SDS-PAGE analysis, N-Rv3409c was mainly eluted at the second peak in the elution spectrum. All fractions containing N-Rv3409c were injected into Mono Q 5/5 column on AKTA system. After cation exchange purification step, N-Rv3409c's purity get greatly improved (Figure 4-3). The N-Rv3409c band on SDS-PAGE gel were cut from SDS-PAGE gels and analyzed with the In-gel trypsin digestion. The tryptic peptides mixtures were analyzed by MALDI-TOF/MS. The covered peptides were confirmed to up 40% (Table 4-2).

Table 4-2. Tryptic peptides (in bold) of N-Rv3409c confirmed by MADLI-TOF/MS analysis.

Residues	Sequences
1-50	MGSSHHHHHHSSGLVPRGSHMKPDYDVLIIGSGFGGSVTALRLTEK**GYRV**
51-100	**GVLEAGRRFSDEEFAKTSWDLRKFLWAPRLGCYGIQR**IHPLRNVMILAGA
101-150	GVGGGSLNYANTLYVPPEPFFADQQWSHITDWR**GELMPHYQQAQRMLGVV**
151-200	**QNPTFTDADR**IVKEVADEMGFGDTWVPTPVGVFFGPDGTKTPGKTVPDPY
201-250	FGGAGPARTGCLECGCCMTGCRHGAKNTLVK**NYLGLAESAGAQVIPMTTV**
251-300	**KGFERRSDGLWEVRTVRTGSWLRRDRRTFTATQLVLAAGTWGTQHLLFK**M
301-350	RDRGRLPGLSKRLGVLTRTNSESIVGAATLKVNPDLDLTHGVAITSSIHP
351-400	TADTHIEPVR**YGKGSNAMGLLQTLMTDGSGPQGTDVPRWRQLLQTASQDP**
401-450	**RGTIRMLNPRQWSERTVIALVMQHLDNSITTFTKR**GKLGIRWYSSK**QGHG**
451-500	**EPNPTWIPIGNQVTRRIAAK**IDGVAGGTWGELFNIPLTAHFLGGAVIGDD
501-550	PEHGVIDPYHRVYGYPTLYVVDGAAISANLGVNPSLSIAAQAERAASLWP
551-598	NKGETDRRPPQGEPYRRLAPIQPAHPVVPADAPGALRWLPIDPVSNAGXX

In summary, the soluble and stable Rv3409c was purified as a recombinant protein with an N-terminus His_6 tag. With *hsp60* promoter (*111*), the N-Histagged recombinant Rv3409c (N-Rv3409c) production yield increased three times (from 1 mg/L to 3 mg/L). Rv3409c with the His6 tag on the N-terminus was purified with a yellow color with a cofactor. Application of three types of chromatography (affinity chromatography, size-exclusion chromatography and ion-exchange chromatography) in sequence provides N-Rv3409c in about 99% purity for further studies. The protein sequence was confirmed by trypsin enzyme digestion coupled with MALDI-TOF/MS method. The following work is all about the recombinant N-Rv3409c purified from the pVV16NRv3409c vector and *M. smegmatis* mc^2155 over-expression system unless stated otherwise.

3.2 N-Rv3409c reconstitution

3.2.1 The FAD isolated from N-Rv3409c Analysis

N-Rv3409c's activity could be inhibited by the FAD adduct. Recombinant N-Rv3409c purified by Ni^{2+} resin column appeared as a brownish yellow in solution. The UV-vis spectrum of N-Rv3409c displayed a significant absorbance at 355 nm and 418 nm (Figure 4-2). Compared to ChoA, a cholesterol oxidase from *Streptomyces*, N-Rv3409c's two peaks by FAD cofactor were blue-shifted to lower wavelengths. To analyze the cofactor, N-Rv3409c was denatured by addition of concentrated HCl or boiling at 100 °C for 20 min.

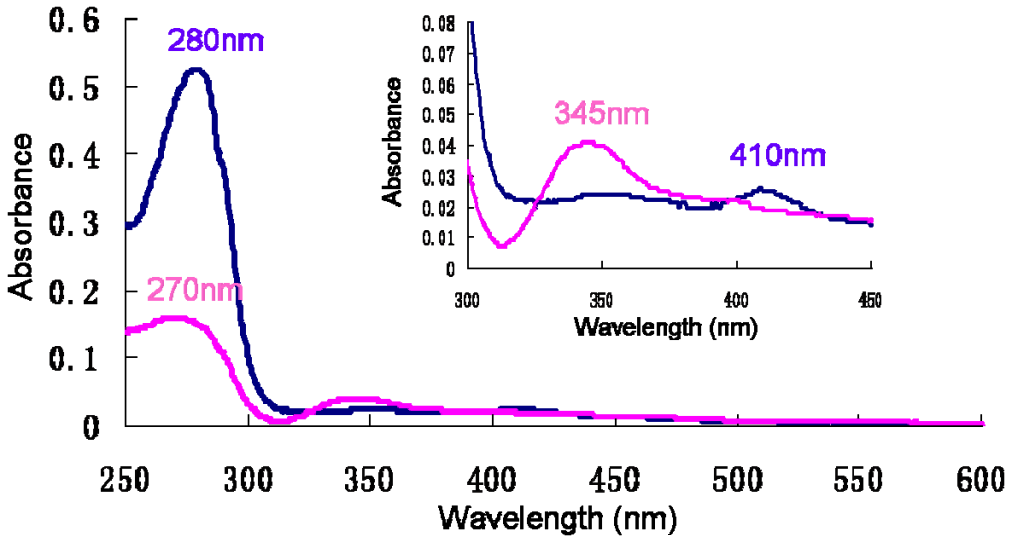


Figure 4-4. UV-vis spectrum of the supernatant of N-Rv3409c sample before (blue) and after (purple) HCl treatment.

The HCl addition was performed dropwise into protein sample until protein precipitation ceased. After centrifugation, the supernatant was yellow with a white precipitant on the bottom. The supernatant was further analyzed by UV-vis spectroscopy. The absorbance at 280 nm (A280) disappeared due to protein denaturation by acid treatment. In contrast, the intensity of the peak associated with the FAD adduct at 345 nm increased (Figure 4-4). The peak at 270 nm was a result of the adenine group of FAD cofactor. The A345 indicated the isolated FAD was a

reduced form (*112*). These UV peaks indicated the isolated FAD from N-Rv3409c could be an adduct.

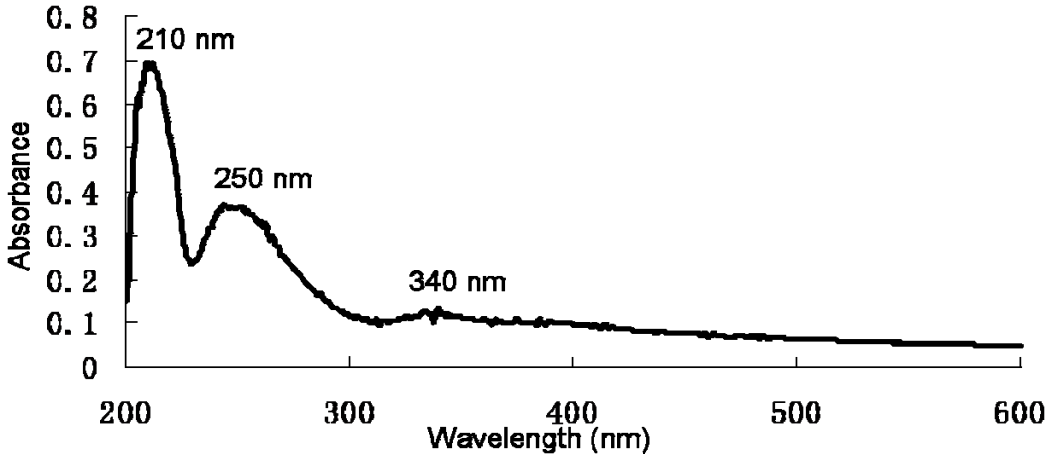


Figure 4-5. UV-vis spectrum of supernatant of denatured N-Rv3409c by heating at 100 °C for 20 min.

The concentrated sample was dialyzed into 50 mM sodium phosphate buffer (pH 7.0) boiled at 100 °C for 20 min. The N-Rv3409c precipitates were removed by centrifugation. The supernatant was analyzed by UV-vis spectroscopy (Figure 4-5) and MALDI-TOF/MS methods (Figure 4-6). The supernatant had three absorbance peaks in the UV spectrum: A210, A250 and A340, instead of an authentic FAD absorbance (A360 and A450). The A250 was a result of the adenine group of FAD cofactor. The A340 indicated the isolated FAD was a reduced form (*112*). The A210 could be the imidazole residues in the sample supernatant. The mass spectra of FAD isolated from Rv3409c were obtained in negative mode. Authentic FAD was used as control. An ionic peak at m/z=783 is 1 mass lower than standard at m/z=784 indicated the isolated FAD fragmented differently from native FAD. There was a significant mass peak at m/z=857, which could be from the FAD adduct. The FAD adduct was not identified in our studies. We found that after size exclusion chromatography

purification, the N-Rv3409c sample does not contain FAD adduct peaks by UV-vis analysis. The purified N-Rv3409c contained no yellow color. The "FAD adduct" could be caused by complicated reasons. Why the C-Rv3409c did not have the "FAD adduct" and yellowish color still remains unclear for us.

3.2.2 Reconstituiton of N-Rv3409c

The protocol for reconstituting methanol oxidase has been followed for N-Rv3409c reconstitution (*113*). In the last ultrafiltration concentration step, the buffer went through the Centricon filter unit (NMWCO=10, 000) was colorless indicating the adduct remained with the protein. After centrifugation, the protein sample, which was retained above the membrane, was concentrated and the intensity of the yellow color increased. Fluorescence spectroscopy was used to characterize the reconstituted

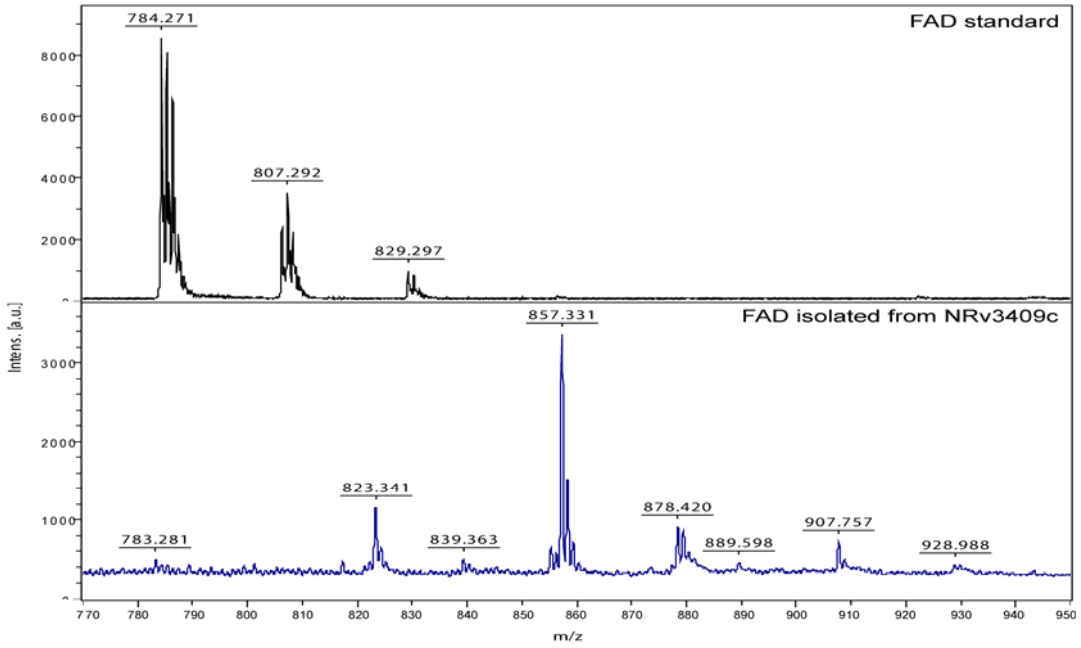


Figure 4-6. MALDI-TOF/MS analysis of FAD adduct isolated from N-Rv3409c.

N-Rv3409c with an authentic FAD solution as the control. As shown in Figure 4-7, the flavin emission peak was greatly quenched due to binding with the protein. There was a slight blue shift of the peak at 525 nm to 523 nm. The UV-vis spectra of

reconstituted N-Rv3409c and apo-N-Rv3409c were compared in the same figure (Figure 4-8) and there were two typical flavoenzyme peaks at 375 nm and 455 nm of the reconstituted N-Rv3409c.

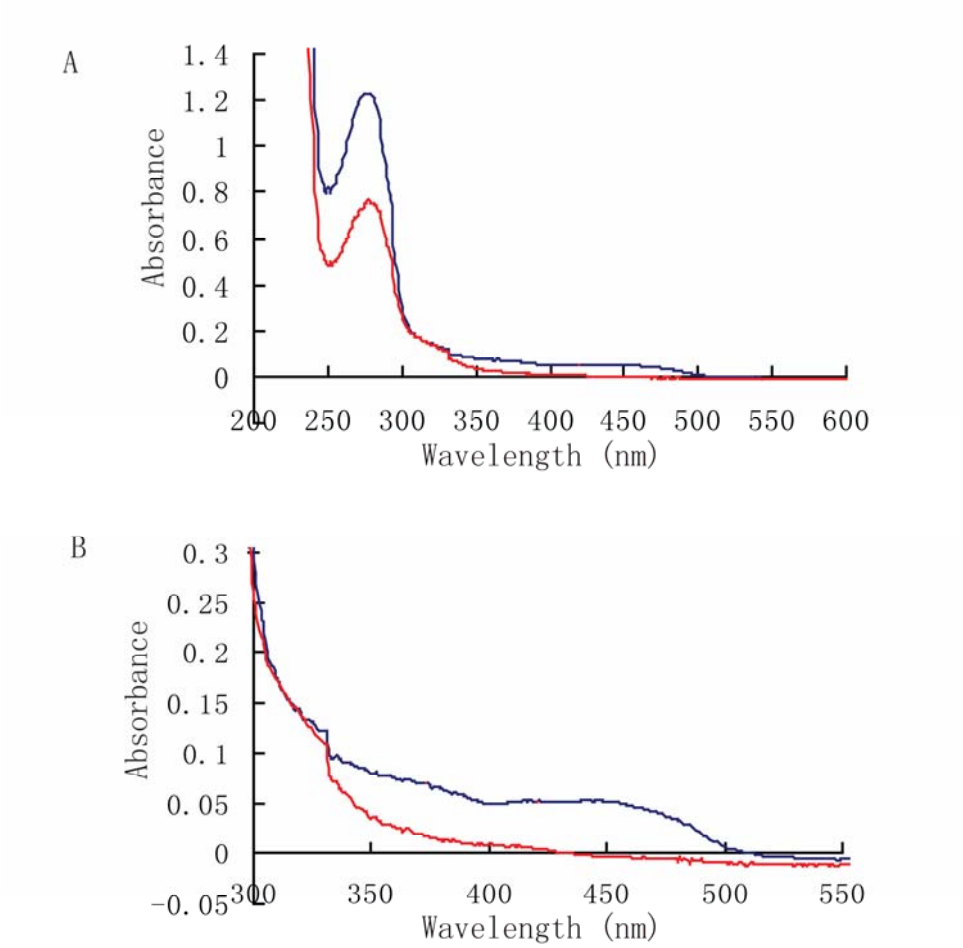


Figure 4-7. UV-vis analysis of the reconstituted N-Rv3409c. Blue line is the reconstituted N-Rv3409c and red line is the holo-N-Rv3409c. Spectrum A) The UV scan from 200 nm to 600 nm; spectrum B) The magnified region of UV absorbance from 300 nm to 600 nm.

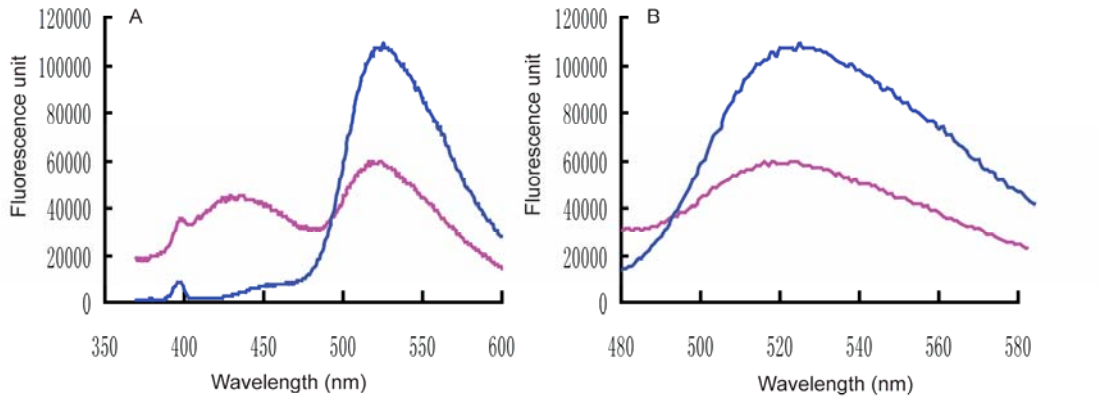


Figure 4-8. Fluorescence emission spectrum of reconstituted N-Rv3409c. The emission scan of free FAD is purple and the spectrum of reconstituted N-Rv3409c is blue; A) the emission spectra from 375 nm to 600 nm, with excitation at 350 nm; B) Magnified emission spectrum from A.

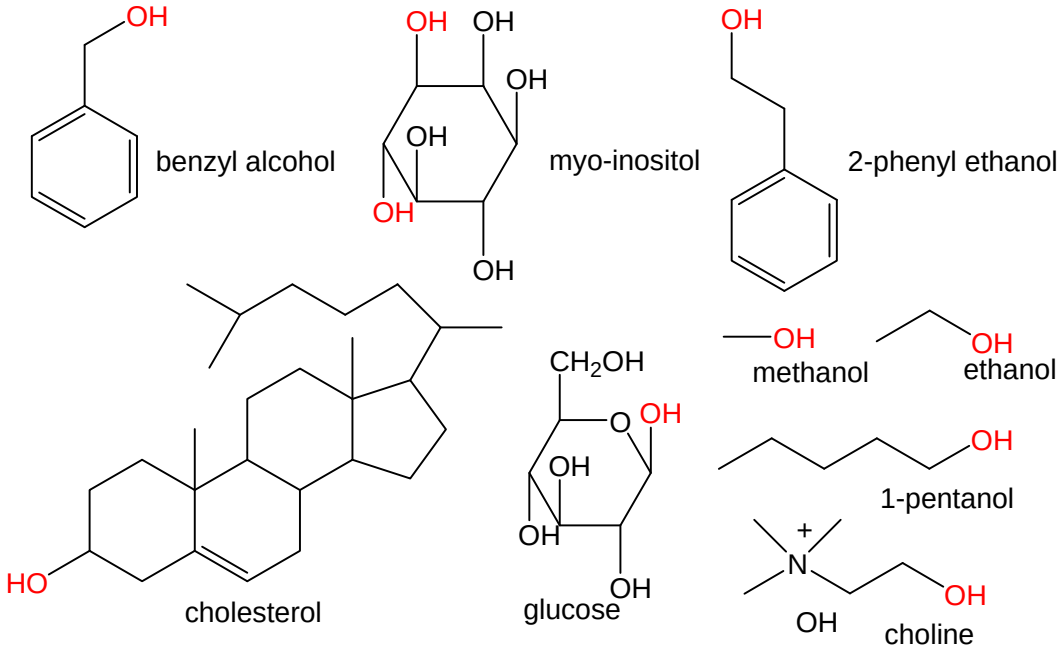
The reconstituted N-Rv3409 finally bound FAD with a FAD/enzyme monomer ratio of about 0.25. Theoretically, each monomer binds a FAD cofactor with an ideal FAD/monomer ratio of 1. The reconstitution works would be improved in the future by optimizing the denaturing conditions and incubation conditions of the apo-N-Rv3409c with FAD. The activity of reconstituted N-Rv3409c was assayed for activities as a cholesterol oxidase or a cholesterol dehydrogenase.

3.3 N-Rv3409c activity assays

3.3.1 Oxidizing activity on cholesterol

Both apo-C-Rv3409c enzyme and reconstituted N-Rv3409c enzyme activity were tested with extra free FAD added at different pH conditions. C-Rv3409c is stable at pH 7 at 37 °C, which was used as the assay conditions. Cholesterol and the potential intermediate cholest-5-en-3-one and product cholestel-4-en-3-one were extracted from the reaction mixtures for TLC analysis. There was no

cholest-5-en-3-one or cholesten-4-en-3-one produced under the assays tested. No cholesterol oxidizing activity was observed after incubation for 17 h overnight time at 37 °C. The ABTS-HRP coupled assay, which monitors the intermediate production of H_2O_2, was also applied to explore the oxidizing ability. To date, no cholesterol oxidizing enzyme activity has been identified with reconstituted N-Rv3409c even with extra FAD present.



Scheme 4-1. Substrates from GMC superfamily.

3.3.2 Oxidizing ability on substrates from GMC superfamily

The bioinformatic analysis indicated that Rv3409c is a putative cholesterol oxidase with a FAD binding domain and that Rv3409c belongs to the GMC superfamily. Rv3409c might have the oxidation activity on substrates from the GMC superfamily. In this study over ten substrates including choline, methanol, and ethanol were incubated with reconstituted N-Rv3409c (Scheme 4-1). For oxidase activity, ABTS-HRP coupled assay was used, whereas, DCIP/PMS coupled assay was applied for analyzing dehydrogenase activity. The reaction mechanisms are shown in Figure 4-9. Our results showed that no oxidizing activity for the listed substrates was

observed. There was no oxidizing activity detected with rhamnose and 6-deoxytalose (sugar residues on L1334 lipid compound) either.

Recombinant N-Rv3409c was demonstrated to have no cholesterol oxidizing ability *in vitro*. We know that the mycobacteria *Rv3409c* homologs have low similarity and identity to other isolated cholesterol oxidase, as discussed in Chapter 1. Combining our studies on lipids and cell morphology of wild-type *M. smegmatis* and *M. smegmatis MSMEG1604* transposon mutant and the assay studies of reconstituted N-Rv3409c, Rv3409c enzyme is not a cholesterol oxidase. Although the *Rv3409c* gene level in the *Rhodococcus* strain was slightly induced by addition of cholesterol versus pyruvate, *Rv3409c* disruption greatly affect on *M. tuberculosis* survival in IFN-γ-activated macrophages (*32*). The up-regulation by cholesterol could be that Rv3409c putatively be involved in the sugar metabolism or lipid synthesis. In *Rhodococcus equi* genome, there is one *Rv3409c* ortholog, and an identified

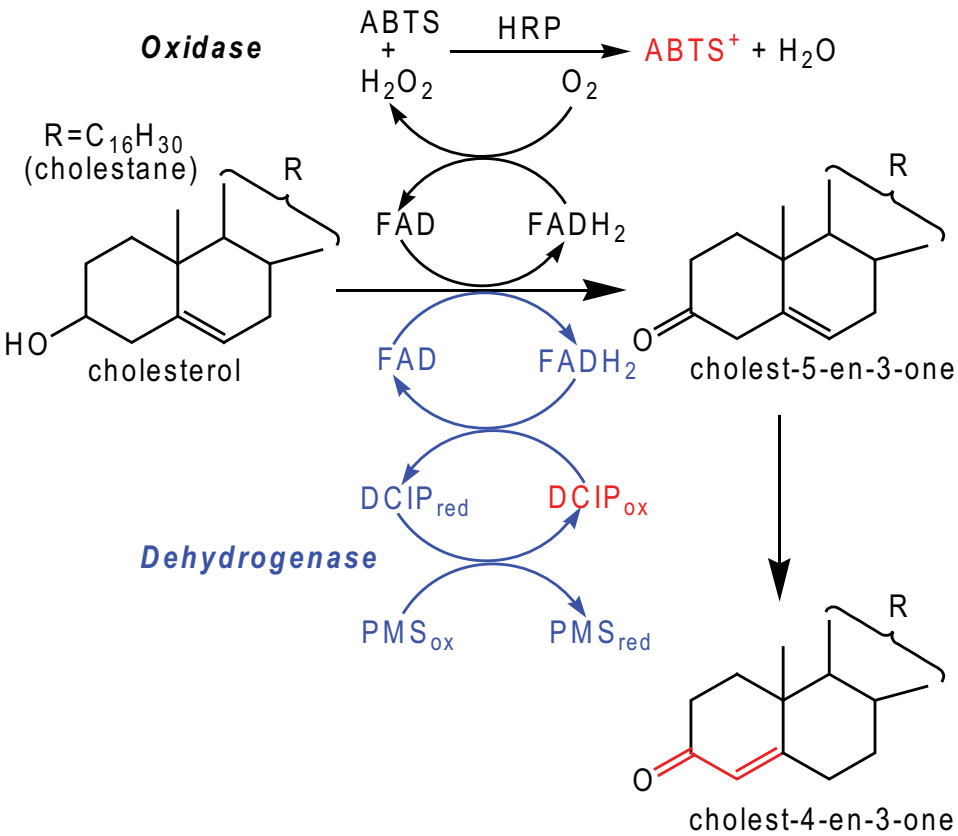


Figure 4-9. ABTS-HRP and DCIP-PMS assay mechanisms.

cholesterol oxidase (ChoE). In *M. tuberculosis* genome, there is Rv3409c and an activity identified 3β-hydroxysteroid dehydrogenase. Whether those bacteria need to contain two enzymes with identical functions needs to be clarified by future studies.

3.3.3 N-Rv3409c assay with lipids

N-Rv3409c could be an enzyme functionally involved in lipid biosynthesis/regulation. Our studies indicate that an unusual GPL compound L1334 was accumulated in the cell envelope of *M. smegmatis MSMEG1604* transposon mutant (Myc11). Myc11 changed to a rough phenotype. The cell walls of Myc11 complemented strains contain no L1334. Oxidation activity of Rv3409c as a cholesterol oxidase/dehydrogenase or with GMC substrates was not observed. It is possible that L1334 was the substrate of N-Rv3409c and a lipid-enzyme assay was carefully designed for reconstituted N-Rv3409c. The MALDI-TOF/MS mass were used to monitor the lipid molecular weight changes in the assay mixtures. Purified L1334 was dissolved in methanol and slowly injected into aqueous solution with fast stirring, which allow the formation of L1334 vesicles in the aqueous buffer. N-Rv3409c was carefully added into the assay mixture. Aliquots of assay mixture were extracted by chloroform at certain time points (0 h, 2 h, 4 h and overnight) and extracted lipids were analyzed by MALDI-TOF/MS. The sample extracted at 0 h was the lipid mass profile control. All MALDI-TOF/MS spectra are shown in Figure 4-10. Comparison with controls indicates that there is no new peak/compound produced as a function of time.

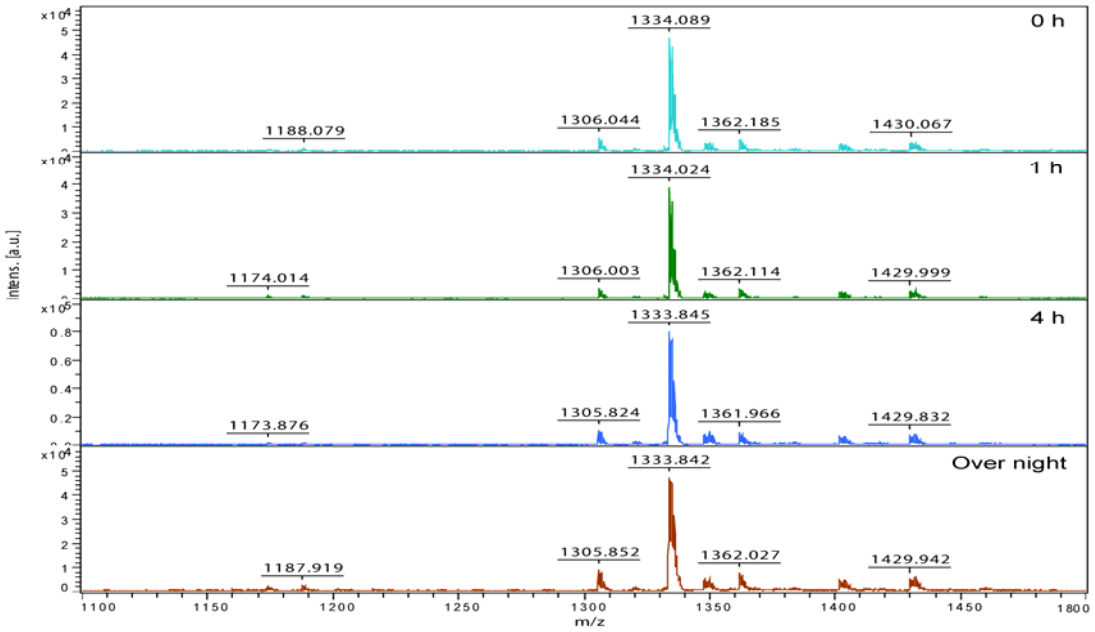


Figure 4-10. MALDI-TOF/MS spectra of chloroform extracts from N-Rv3409c lipid assay; purified L1334 and N-Rv3409c were incubated at 37 °C in 50 mM sodium phosphate buffer (pH 7.0).

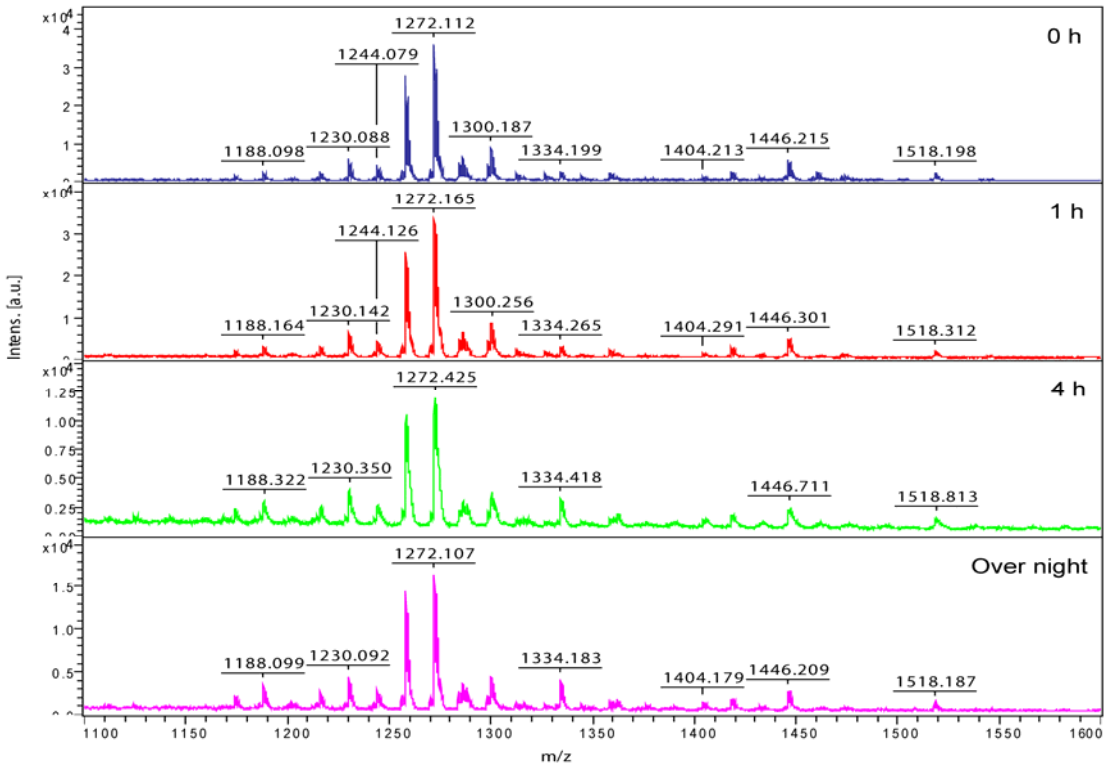


Figure 4-11. MALDI-TOF/MS spectra of chloroform extracts from N-Rv3409c lipid assay; Myc11 total lipids including L1334 and N-Rv3409c were incubated at 37 °C in 50 mM sodium phosphate buffer (pH 7.0).

Total lipids-enzyme assay were prepared in the same way of L1334. All the mass spectra are shown in Figure 4-11. There are more new peaks after two hours, but they are not present in the mass spectra of the four hour and overnight samples and this is not reproducible. Some of the peaks at m/z about 2300 could be peptides resulting from degradation of the N-Rv3409c sample. The homogeneous Myc11 crude cell lysate including cell wall total lipids and proteins will be prepared by Bead-beating equipment. Reconstituted N-Rv3409c will be added into the crude cell lysate mixture and the assay mixture will be incubated for overnight. Then the total lipids will be extracted by chloroform and methanol for MALDI-TOF/MS analysis with enzyme untreated total lipids as control.

In summary, N-Rv3409c was purified and reconstituted with a FAD cofactor bound with the ratio of FAD/monomer N-Rv3409c about 0.25. The ABTS-HRP coupled assay and DCIP/PMS assays showed that N-Rv3409c lacks enzyme activity for oxidizing a hydroxyl group to a ketone or an aldehyde group with about ten different substrates for enzymes from GMC superfamily. The incubation of purified L1334 or total lipids containing L1334 with reconstituted N-Rv3409c does not make any change to L1334 or produce any lipid by MALDI-TOF/MS. Rv3409c is not located within a functional identified gene cluster neither. The real function of Rv3409c needs to be carefully studied.

According to our latest results, Myc11 growth rate is significantly dependant on glucose concentration in the medium. Myc11 was cultured in medium with different glucose concentrations with wild-type culture as control. Three types of glucose medium are: 7H9 + 0.2 % glycerol; 7H9 + 0.2 % glycerol + 0.2% glucose; and 7H9 + 0.2 % glycerol+1% glucose. The growth rates were monitored by measuring OD_{600}. Myc11 culture showed significant difference in growth rate on different glucose

concentrations. The growth curve is shown in the Figure 5S. Considering the glucose-dependent growth property and an abnormal deacetylation and over-rhamnosylation GPL lipid accumulated due to *MSMEG1604* disruption, we propose that Rv3409c might be involved in sugar/glycerol metabolism for lipid biosynthesis in a certain mechanism. In *M. smegmatis*, Rv3409c function could be specifically involved in GPL biosynthesis.

References

1. Kochi, A. (1991) The global tuberculosis situation and the new control strategy of the world-health-organization, *Tubercle 72*, 1-6.
2. Raviglione, M. C., Snider, D. E., and Kochi, A. (1995) Global epidemiology of tuberculosis - morbidity and mortality of a worldwide epidemic, *JAMA-J. Am. Med. Assoc. 273*, 220-226.
3. WHO. (2009) Global tuberculosis control - epidemiology, strategy, financing, in who report 2009, *The World Health Organization*.
4. WHO Fact Sheet No. 104. (August 2004).
5. Vivek T. Natarajan, D. M., and Rajesh S. Gokhale. (2008) Biosynthesis of mycobacterial lipids by multifunctional polyketide synthases, *The Mycobacterial Cell Envelope Chapter 15*, 235-248.
6. Armstron.Ja, and Hart, P. D. (1971) Response of cultured macrophages to mycobacterium-tuberculosis, with observations on fusion of lysosomes with phagosomes, *J. Exp. Med. 134*, 713-740.
7. Malik, Z. A., Denning, G. M., and Kusner, D. J. (2000) Inhibition of Ca^{2+} signaling by *Mycobacterium tuberculosis* is associated with reduced phagosome-lysosome fusion and increased survival within human macrophages, *J. Exp. Med. 191*, 287-302.
8. Malik, Z. A., Iyer, S. S., and Kusner, D. J. (2001) *Mycobacterium tuberculosis* phagosomes exhibit altered calmodulin-dependent signal transduction: contribution to inhibition of phagosome-lysosome fusion and intracellular survival in human macrophages, *J. Immunol. 166*, 3392-3401.
9. Dechastellier, C., Lang, T., and Thilo, L. (1995) Phagocytic processing of the macrophage endoparasite, *Mycobacterium avium*, in comparison to

phagosomes which contain bacillus-subtilis or latex beads, *Eur. J. Cell. Biol. 68*, 167-182.

10. Clemens, D. L. (1996) Characterization of the *Mycobacterium tuberculosis* phagosome, *Trends in Microbiology 4*, 113-118.
11. Via, L. E., Deretic, D., Ulmer, R. J., Hibler, N. S., Huber, L. A., and Deretic, V. (1997) Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7, *J. Biol. Chem. 272*, 13326-13331.
12. Vergne, I., Chua, J., and Deretic, V. (2003) *Mycobacterium tuberculosis* phagosome maturation arrest: Selective targeting of PI3P-dependent membrane trafficking, *Traffic 4*, 600-606.
13. Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., Davis, A., de Haro, S., Naylor, J., Lee, H. H., and Vergne, I. (2006) *Mycobacterium tuberculosis* inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism, *Cell. Microbiol. 8*, 719-727.
14. Ting, L. M., Kim, A. C., Cattamanchi, A., and Ernst, J. D. (1999) *Mycobacterium tuberculosis* inhibits IFN-gamma transcriptional responses without inhibiting activation of STAT1, *J. Immunol. 163*, 3898-3906.
15. Lafuse, W. P., Alvarez, G. R., Curry, H. M., and Zwilling, B. S. (2006) *Mycobacterium tuberculosis* and *Mycobacterium avium* inhibit IFN-gamma-induced gene expression by TLR2-dependent and independent pathways, *J. Interf. Cytok. Res. 26*, 548-561.
16. Granich, R. M. (2005) Multidrug resistance among persons with tuberculosis in California, 1994-2003, *JAMA-J. Am. Med. Assoc. 294*, 44.
17. Jackson, M., Stadthagen, G., and Gicquel, B. (2007) Long-chain multiple

methyl-branched fatty acid-containing lipids of *Mycobacterium tuberculosis*: Biosynthesis, transport, regulation and biological activities, *Tuberculosis 87*, 78-86.

18. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature 393*, 537-544.

19. Smith, I. (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence, *Clinical Microbiology Reviews 16*, 463-496.

20. Slayden, R. A., and Barry, C. E. (2000) The genetics and biochemistry of isoniazid resistance in *Mycobacterium tuberculosis*, *Microbes. Infect. 2*, 659-669.

21. Abdallah, A. M., Van Pittius, N. C. G., Champion, P. A. D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E., Appelmelk, B. J., and Bitter, W. (2007) Type VII secretion - mycobacteria show the way, *Nat. Rev. Microbiol. 5*, 883-891.

22. Daffe, M. (2008) The global architecture of the mycobacterial cell envelope, *The Mycobacterial Cell Envelope Chapter 1*, 3-11.

23. Jackson, C. J., Lamb, D. C., Marczylo, T. H., Parker, J. E., Manning, N. L., Kelly, D. E., and Kelly, S. L. (2003) Conservation and cloning of CYP51: A

sterol 14 alpha-demethylase from *Mycobacterium smegmatis*, *Biochem. Biophys. Res. Commun. 301*, 558-563.

24. Gatfield, J., and Pieters, J. (2000) Essential role for cholesterol in entry of mycobacteria into macrophages, *Science 288*, 1647-1650.

25. Cardona, P. J., Llatjos, R., Gordillo, S., Diaz, J., Ojanguren, I., Ariza, A., and Ausina, V. (2000) Evolution of granulomas in lungs of mice infected aerogenically with *Mycobacterium tuberculosis*, *Scand. J. Immunol. 52*, 156-163.

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

27. Munoz-Elias, E. J., and McKinney, J. D. (2006) Carbon metabolism of intracellular bacteria, *Cell. Microbiol. 8*, 10-22.

28. Boshoff, H. I. M., and Barry, C. E. (2005) Tuberculosis - metabolism and respiration in the absence of growth, *Nat. Rev. Microbiol. 3*, 70-80.

29. Pandey, A. K., and Sassetti, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol, *Proc. Nat. Acad. Sci. U.S.A. 105*, 9130-9130.

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

31. Li, L. L., Bannantine, J. P., Zhang, Q., Amonsin, A., May, B. J., Alt, D., Banerji, N., Kanjilal, S., and Kapur, V. (2005) The complete genome sequence of *Mycobacterium avium* subspecies paratuberculosis, *Proc. Nat. Acad. Sci. U.S.A. 102*, 12344-12349.

32. Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M.

C., Sim, E., Dijkhuizen, L., Davies, J. E., Mohn, W. W., and Eltis, L. D. (2007) A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages, *Proc. Nat. Acad. Sci. U.S.A. 104*, 1947-1952.

33. Kreit, J., and Sampson, N. S. (2009) Cholesterol oxidase: Physiological functions, *FEBS Journal 276*, 6844-6856.

34. Sampson, N. S., and Vrielink, A. (2003) Cholesterol oxidases: A study of mature's approach to protein design, *Acc. Chem. Res. 36*, 713-722.

35. Yang, X. X., Dubnau, E., Smith, I., and Sampson, N. S. (2007) Rv1106c from *Mycobacterium tuberculosis* is a 3 beta-hydroxysteroid dehydrogenase, *Biochemistry 46*, 9058-9067.

36. Yang, X. X., Nesbitt, N. M., Dubnau, E., Smith, I., and Sampson, N. S. (2009) Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*, *Biochemistry 48*, 3819-3821.

37. van der Geize, R., Hessels, G. I., van Gerwen, R., van der Meijden, R., and Dijkhuizen, L. (2002) Molecular and functional characterization of *kshA* and *kshB*, encoding two components of 3-ketosteroid 9 alpha-hydroxylase, a class ia monooxygenase, in rhodococcus erythropolis strain sq1, *Mol. Microbiol. 45*, 1007-1018.

38. Horinouchi, M., Hayashi, T., Koshino, H., Kurita, T., and Kudo, T. (2005) Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in comamonas testosteroni TA441, *Appl. Environ. Microb. 71*, 5275-5281.

39. Horinouchi, M., Kurita, T., Yamamoto, T., Hatori, E., Hayashi, T., and Kudo, T.

(2004) Steroid degradation gene cluster of comamonas testosteroni consisting of 18 putative genes from meta-cleavage enzyme gene *tesB* to regulator gene *tesR*, *Biochem. Biophys. Res. Commun. 324*, 597-604.

40. Andor, A., Jekkel, A., Hopwood, D. A., Jeanplong, F., Ilkoy, E., Konya, A., Kurucz, I., and Ambrus, G. (2006) Generation of useful insertionally blocked sterol degradation pathway mutants of fast-growing mycobacteria and cloning, characterization, and expression of the terminal oxygenase of the 3-ketosteroid 9 alpha-hydroxylase in *Mycobacterium smegmatis* mc^2155, *Appl. Environ. Microb. 72*, 6554-6559.

41. Pollegioni, L., Piubelli, L., and Molla, G. (2009) Cholesterol oxidase: biotechnological applications, *FEBS Journal 276*, 6857-6870.

42. Weinstock, D. M., and Brown, A. E. (2002) *Rhodococcus equi*: an emerging pathogen, *Clin. Infect. Dis. 34*, 1379-1385.

43. Reading, P. C., Moore, J. B., and Smith, G. L. (2003) Steroid hormone synthesis by vaccinia virus suppresses the inflammatory response to infection, *J. Exp. Med. 197*, 1269-1278.

44. Cowley, S. C., and Av-Gay, Y. (2001) Monitoring promoter activity and protein localization in mycobacterium spp. using green fluorescent protein, *Gene 264*, 225-231.

45. Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J., and Dziadek, J. (2007) Cholesterol oxidase is required for virulence of mycobacterium tuberculosis, *FEMS Microbiology Letters 275*, 106-112.

46. Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., and Jacobs, W. R. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*, *Mol. Microbiol. 4*, 1911-1919.

47. Jacobs, W. R. (2000) *Mycobacterium tuberculosis*: a once genetically intractable organism, *In Molecular Genetics of Mycobacteria Edited by G.F. Hatfull & W.R. Jacobs, Jr.*, 1-16.

48. Etienne, G., Laval, F., Villeneuve, C., Dinadayala, P., Abouwarda, A., Zerbib, D., Galamba, A., and Daffe, M. (2005) The cell envelope structure and properties of *Mycobacterium smegmatis* mc^2155: Is there a clue for the unique transformability of the strain?, *Microbiol-SGM 151*, 2075-2086.

49. Deshayes, C., Laval, F., Montrozier, H., Daffe, M., Etienne, G., and Reyrat, J. M. (2005) A glycosyltransferase involved in biosynthesis of triglycosylated glycopeptidolipids in *Mycobacterium smegmatis*: Impact on surface properties, *J. Bacteriol. 187*, 7283-7291.

50. Recht, J., and Kolter, R. (2001) Glycopeptidolipid acetylation affects sliding motility and biofilm formation in *Mycobacterium smegmatis*, *J. Bacteriol. 183*, 5718-5724.

51. Smith, D. W. H. M. r., A. P. Maclennan, and E. Lederer. (1960a) Mycosides: A new class of type-specific glycolipids of mycobacteria, *Nature 186*, 887-888.

52. Besra, G. S., Mcneil, M. R., Rivoire, B., Khoo, K. H., Morris, H. R., Dell, A., and Brennan, P. J. (1993) Further structural definition of a new family of glycopeptidolipids from *Mycobacterium xenopi*, *Biochemistry 32*, 347-355.

53. Riviere, M., Auge, S., Vercauteren, J., Wisingerova, E., and Puzo, G. (1993) Structure of a novel glycopeptidolipid antigen containing a O-methylated serine isolated from *Mycobacterium xenopi* complete H-1-NMR and C-13-NMR assignment, *Eur. J. Biochem. 214*, 395-403.

54. Patterson, J. H., McConville, M. J., Haites, R. E., Coppel, R. L., and Billman-Jacobe, H. (2000) Identification of a methyltransferase from

Mycobacterium smegmatis involved in glycopeptidolipid synthesis, *J. Biol. Chem. 275*, 24900-24906.

55. Aspinall, G. O., Chatterjee, D., and Brennan, P. J. (1995) The variable surface glycolipids of mycobacteria: structures, synthesis of epitopes, and biological properties, *Adv. Carbohyd. Chem. Bi. 51*, 169-242.

56. Belisle, J. T., Klaczkiewicz, K., Brennan, P. J., Jacobs, W. R., and Inamine, J. M. (1993) Rough morphological variants of *Mycobacterium avium* - characterization of genomic deletions resulting in the loss of glycopeptidolipid expression, *J. Biol. Chem. 268*, 10517-10523.

57. Daffe, M., Laneelle, M. A., and Puzo, G. (1983) Structural elucidation by field desorption and electron-impact mass-spectrometry of the c-mycosides isolated from *Mycobacterium smegmatis*, *Biochim. Biophys. Acta. 751*, 439-443.

58. Schorey, J. S., and Sweet, L. (2008) The mycobacterial glycopeptidolipids: structure, function, and their role in pathogenesis, *Glycobiology 18*, 832-841.

59. Sonden, B., Kocincova, D., Deshayes, C., Euphrasie, D., Rhayat, L., Laval, F., Frehel, C., Daffe, M., Etienne, G., and Reyrat, J. M. (2005) Gap, a mycobacterial specific integral membrane protein, is required for glycolipid transport to the cell surface, *Mol. Microbiol. 58*, 426-440.

60. Pourshafie, M., Ayub, Q., and Barrow, W. W. (1993) Comparative effects of *Mycobacterium avium* glycopeptidolipid and lipopeptide fragment on the function and ultrastructure of mononuclear cells, *Clinical and Experimental Immunology 93*, 72-79.

61. Villeneuve, C., Etienne, G., Abadie, V., Montrozier, H., Bordier, C., Laval, F., Daffe, M., Maridonneau-Parini, I., and Astarie-Dequeker, C. (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*, 51291-51300.

62. Tereletsky, M. J., and Barrow, W. W. (1983) Postphagocytic detection of glycopeptidolipids associated with the superficial-L1 layer of *Mycobacterium intracellulare*, *Infect. Immun. 41*, 1312-1321.

63. Irani, V. R., and Maslow, J. N. (2005) Induction of murine macrophage TNF-alpha synthesis by *Mycobacterium avium* is modulated through complement-dependent interaction via complement receptors 3 and 4 in relation to *Mycobacterium avium* glycopeptidolipid, *FEMS Microbiology Letters 246*, 221-228.

64. Caroline Deshayes, D. K., Gilles Etienne, and Jean-marc Reyrat. (2008) Glycopeptidolipids: a complex pathway for small pleiotropic molecules, *The Mycobacterial Cell Envelope Chapter 21*, 345-366.

65. Ripoll, F., Deshayes, C., Pasek, S., Laval, F., Beretti, J. L., Biet, F., Risler, J. L., Daffe, M., Etienne, G., Gaillard, J. L., and Reyrat, J. M. (2007) Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and *Mycobacterium chelonae*, *BMC Genomics 8*, 114.

66. Christophie Guilhot, C. C., and Mamadou Daffe. (2008) Biosynthesis and the roles of phenolic glycolipids and related molecules in *Mycobacterium tuberculosis*, *The Mycobacterial Cell Envelope Chapter 17*, 273-289.

67. Daffe, M., and Laneelle, M. A. (1988) Distribution of phthiocerol diester, phenolic mycosides and related-compounds in mycobacteria, *J. Gen. Microbiol. 134*, 2049-2055.

68. Daffe, M., and Draper, P. (1998) The envelope layers of mycobacteria with

reference to their pathogenicity, *Adv. Microb. Physiol. 39*, 131-203.

69. Camacho, L. R., Constant, P., Raynaud, C., Laneelle, M. A., Triccas, J. A., Gicquel, B., Daffe, M., and Guilhot, C. (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis* - evidence that this lipid is involved in the cell wall permeability barrier, *J. Biol. Chem. 276*, 19845-19854.

70. Constant, P., Perez, E., Malaga, W., Laneelle, M. A., Saurel, O., Daffe, M., and Guilhot, C. (2002) Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex - evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks15/1* gene, *J. Biol. Chem. 277*, 38148-38158.

71. Onwueme, K. C., Vos, C. J., Zurita, J., Ferreras, J. A., and Quadri, L. E. N. (2005) The dimycocerosate ester polyketide virulence factors of mycobacteria, *Prog. Lipid. Res. 44*, 259-302.

72. Ripoll, F., Deshayes, C., Pasek, S., Laval, F., Beretti, J. L., Biet, F., Risler, J. L., Daffe, M., Etienne, G., Gaillard, J. L., and Reyrat, J. M. (2007) Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and *Mycobacterium chelonae*, *BMC Genomics 8*, 114-123.

73. Sweet, J. S. S. a. L. (2008) The mycobacterial glycopeptidolipids: structure, function, and their role in pathogenesis, *Glycobiology 18*, 832-841.

74. Neill, M. A., and Klebanoff, S. J. (1988) The effect of phenolic glycolipid-1 from *Mycobacterium leprae* on the antimicrobial activity of human macrophages, *J. Exp. Med. 167*, 30-42.

75. Tsenova, L., Ellison, E., Harbacheuski, R., Moreira, A. L., Kurepina, N., Reed,

M. B., Mathema, B., Barry, C. E., and Kaplan, G. (2005) Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli, *J. Infect. Dis. 192*, 98-106.

76. Manca, C., Reed, M. B., Freeman, S., Mathema, B., Kreiswirth, B., Barry, C. E., and Kaplan, G. (2004) Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis, *Infect. Immun. 72*, 5511-5514.

77. Perez, E., Constant, P., Laval, F., Lemassu, A., Laneelle, M. A., Daffe, M., and Guilhot, C. (2004) Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycoserosate in the *Mycobacterium tuberculosis* complex, *J. Biol. Chem. 279*, 42584-42592.

78. Goren M.B. and Brennan, P.J. (1979) *Tuberculosis (Youmans, G.P., ed)*, 62-193.

79. Reyrat, J. M., and Kahn, D. (2001) *Mycobacterium smegmatis*: an absurd model for tuberculosis?, *Trends in Microbiology 9*, 472-473.

80. Recht, J., Martinez, A., Torello, S., and Kolter, R. (2000) Genetic analysis of sliding motility in *Mycobacterium smegmatis*, *J. Bacteriol. 182*, 4348-4351.

81. Martinez, A., Torello, S., and Kolter, R. (1999) Sliding motility in mycobacteria, *J. Bacteriol. 181*, 7331-7338.

82. Etienne, G., Villeneuve, C., Billman-Jacobe, H., Astarie-Dequeker, C., Dupont, M. A., and Daffe, M. (2002) The impact of the absence of glycopeptidolipids on the ultrastructure, cell surface and cell wall properties, and phagocytosis of *Mycobacterium smegmatis*, *Microbiol-SGM 148*, 3089-3100.

83. Rosenberg, M., and Kjelleberg, S. (1986) Hydrophobic interactions - role in bacterial adhesion, *Adv. Microb. Ecol. 9*, 353-393.

84. Gopalaswamy, R., Narayanan, S., Jacobs, W. R., and Av-Gay, Y. (2008)

Mycobacterium smegmatis biofilm formation and sliding motility are affected by the serine/threonine protein kinase pknF, *FEMS Microbiology Letters 278*, 121-127.

85. Eckstein, T. M., Belisle, J. T., and Inamine, J. M. (2003) Proposed pathway for the biosynthesis of serovar-specific glycopeptidolipids in *Mycobacterium avium* serovar 2, *Microbiol-SGM 149*, 2797-2807.

86. Cangelosi, G. A., Palermo, C. O., Laurent, J. P., Hamlin, A. M., and Brabant, W. H. (1999) Colony morphotypes on congo red agar segregate along species and drug susceptibility lines in the *Mycobacterium avium intracellulare* complex, *Microbiology-UK 145*, 1317-1324.

87. Beggs, M. L., Crawford, J. T., and Eisenach, K. D. (1995) Isolation and sequencing of the replication region of *Mycobacterium avium* plasmid pLR7, *J. Bacteriol. 177*, 4836-4840.

88. Mederle, I., Bourguin, I., Ensergueix, D., Badell, E., Moniz-Peireira, J., Gicquel, B., and Winter, N. (2002) Plasmidic versus insertional cloning of heterologous genes in *Mycobacterium bovis* BCG: impact on *in vivo* antigen persistence and immune responses, *Infect. Immun. 70*, 303-314.

89. Murry, J., Sassetti, C. M., Moreira, J., Lane, J., and Rubin, E. J. (2005) A new site-specific integration system for mycobacteria, *Tuberculosis 85*, 317-323.

90. Rosenkrands, I., Agger, E. M., Olsen, A. W., Korsholm, K. S., Andersen, C. S., Jensen, K. T., and Andersen, P. (2005) Cationic liposomes containing mycobacterial lipids: a new powerful th1 adjuvant system, *Infect. Immun. 73*, 5817-5826.

91. Geo.F. Brooks, J. S. B., Karen C. Carroll and Stephen A. Morse. (2007) Chapter 24. Mycobacteria, *Medical Microbiology (24th Edition)*, 320-331.

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

93. Howard, S. T., and Byrd, T. F. (2000) The rapidly growing mycobacteria: saprophytes and parasites, *Microbes. Infect. 2*, 1845-1853.

94. Krzywinska, E., Krzywinski, J., and Schorey, J. S. (2004) Naturally occurring horizontal gene transfer and homologous recombination in mycobacterium, *Microbiol-SGM 150*, 1707-1712.

95. Kano, H., Doi, T., Fujita, Y., Takimoto, H., Yano, I., and Kumazawa, Y. (2005) Serotype-specific modulation of human monocyte functions by glycopeptidolipid (GPL) isolated from *Mycobacterium avium* complex, *Biol. Pharm. Bull. 28*, 335-339.

96. Daffe, M., and Servin, P. (1989) Scalar, dipolar-correlated and J-resolved 2D-NMR spectroscopy of the specific phenolic mycoside of *Mycobacterium tuberculosis*, *Eur. J. Biochem. 185*, 157-162.

97. Nesbitt, N. M., Yang, X. X., Fontan, P., Kolesnikova, I., Smith, I., Sampson, N. S., and Dubnau, E. (2010) A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol, *Infect. Immun. 78*, 275-282.

98. Bligh, E. G., and Dyer, W. J. (1959) A rapid method of total lipid extraction and purification, *Can. J. Biochem. Phys. 37*, 911-917.

99. Miyamoto, Y., Mukai, T., Nakata, N., Maeda, Y., Kai, M., Naka, T., Yano, I., and Makino, M. (2006) Identification and characterization of the genes involved in glycosylation pathways of mycobacterial glycopeptidolipid biosynthesis, *J. Bacteriol. 188*, 86-95.

100. Ojha, A. K., Varma, S., and Chatterji, D. (2002) Synthesis of an unusual polar

glycopeptidolipid in glucose-limited culture of *Mycobacterium smegmatis*, *Microbiol-SGM 148*, 3039-3048.

101. Zamocky, M., Hallberg, M., Ludwig, R., Divne, C., and Haltrich, D. (2004) Ancestral gene fusion in cellobiose dehydrogenases reflects a specific evolution of GMC oxidoreductases in fungi, *Gene 338*, 1-14.

102. Cavener, D. R. (1992) GMC oxidoreductases - a newly defined family of homologous proteins with diverse catalytic activities, *J. Mol. Biol. 223*, 811-814.

103. Iida, K., Cox-Foster, D. L., Yang, X. L., Ko, W. Y., and Cavener, D. R. (2007) Expansion and evolution of insect GMC oxidoreductases, *BMC Evol. Biol. 7*, 75.

104. Kiess, M., Hecht, H. J., and Kalisz, H. M. (1998) Glucose oxidase from penicillium amagasakiense - primary structure and comparison with other glucose-methanol-choline (GMC) oxidoreductases, *Eur. J. Biochem. 252*, 90-99.

105. Rotsaert, F. A. J., Renganathan, V., and Gold, M. H. (2003) Role of the flavin domain residues, His689 and Asn732, in the catalytic mechanism of cellobiose dehydrogenase from *Phanerochaete chrysosporium*, *Biochemistry 42*, 4049-4056.

106. Kass, I. J., and Sampson, N. S. (1998) Evaluation of the role of His(447) in the reaction catalyzed by cholesterol oxidase, *Biochemistry 37*, 17990-18000.

107. Yue, Q. K., Kass, I. J., Sampson, N. S., and Vrielink, A. (1999) Crystal structure determination of cholesterol oxidase from streptomyces and structural characterization of key active site mutants, *Biochemistry 38*, 4277-4286.

108. Kass, I. J., and Sampson, N. S. (1998) The importance of Glu(361) position in

the reaction catalyzed by cholesterol oxidase, *Bioorg. Med. Chem. Lett. 8*, 2663-2668.

109. Dubnau, E., Chan, J., Mohan, V. P., and Smith, I. (2005) Responses of *Mycobacterium tuberculosis* to growth in the mouse lung, *Infect. Immun. 73*, 3754-3757.

110. Ikeda, Y., Okamuraikeda, K., and Tanaka, K. (1985) Purification and characterization of short-chain, medium-chain, and long-chain acyl-coa dehydrogenases from rat-liver mitochondria - isolation of the holoenzymes and apoenzymes and conversion of the apoenzyme to the holoenzyme, *J. Biol. Chem. 260*, 1311-1325.

111. Stover, C. K., Delacruz, V. F., Fuerst, T. R., Burlein, J. E., Benson, L. A., Bennett, L. T., Bansal, G. P., Young, J. F., Lee, M. H., Hatfull, G. F., Snapper, S. B., Barletta, R. G., Jacobs, W. R., and Bloom, B. R. (1991) New use of BCG for recombinant vaccines, *Nature 351*, 456-460.

112. Gadda, G., Edmondson, R. D., Russell, D. H., and Fitzpatrick, P. F. (1997) Identification of the naturally occurring flavin of nitroalkane oxidase from fusarium oxysporum as a 5-nitrobutyl-fad and conversion of the enzyme to the active FAD-containing form, *J. Biol. Chem. 272*, 5563-5570.

113. Sherry, B., and Abeles, R. H. (1985) Mechanism of action of methanol oxidase, reconstitution of methanol oxidase with 5-deazaflavin, and inactivation of methanol oxidase by cyclopropanol, *Biochemistry 24*, 2594-2605.

Appendix

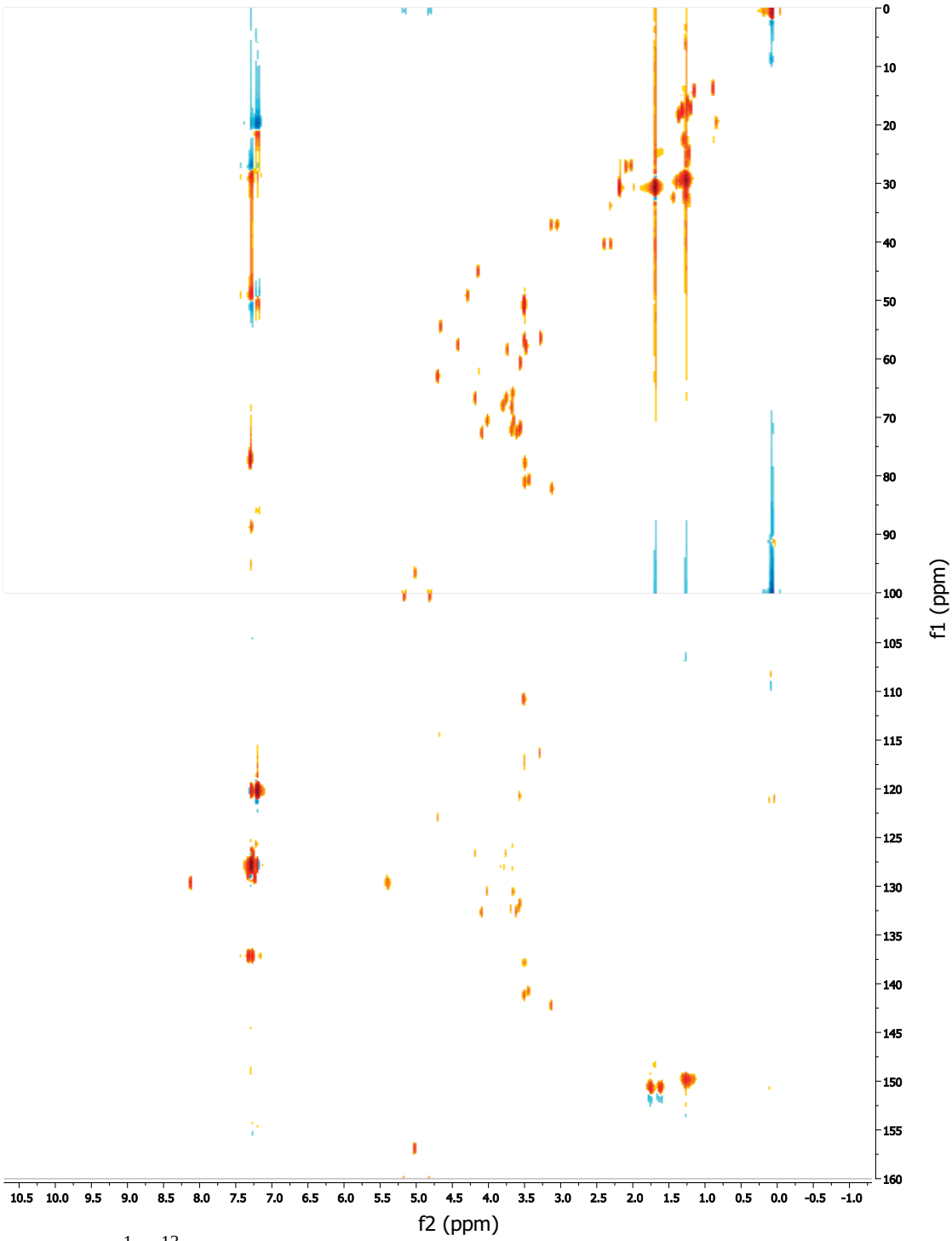


Figure S1. ^{1}H-^{13}C HSQC NMR of purified L1334.

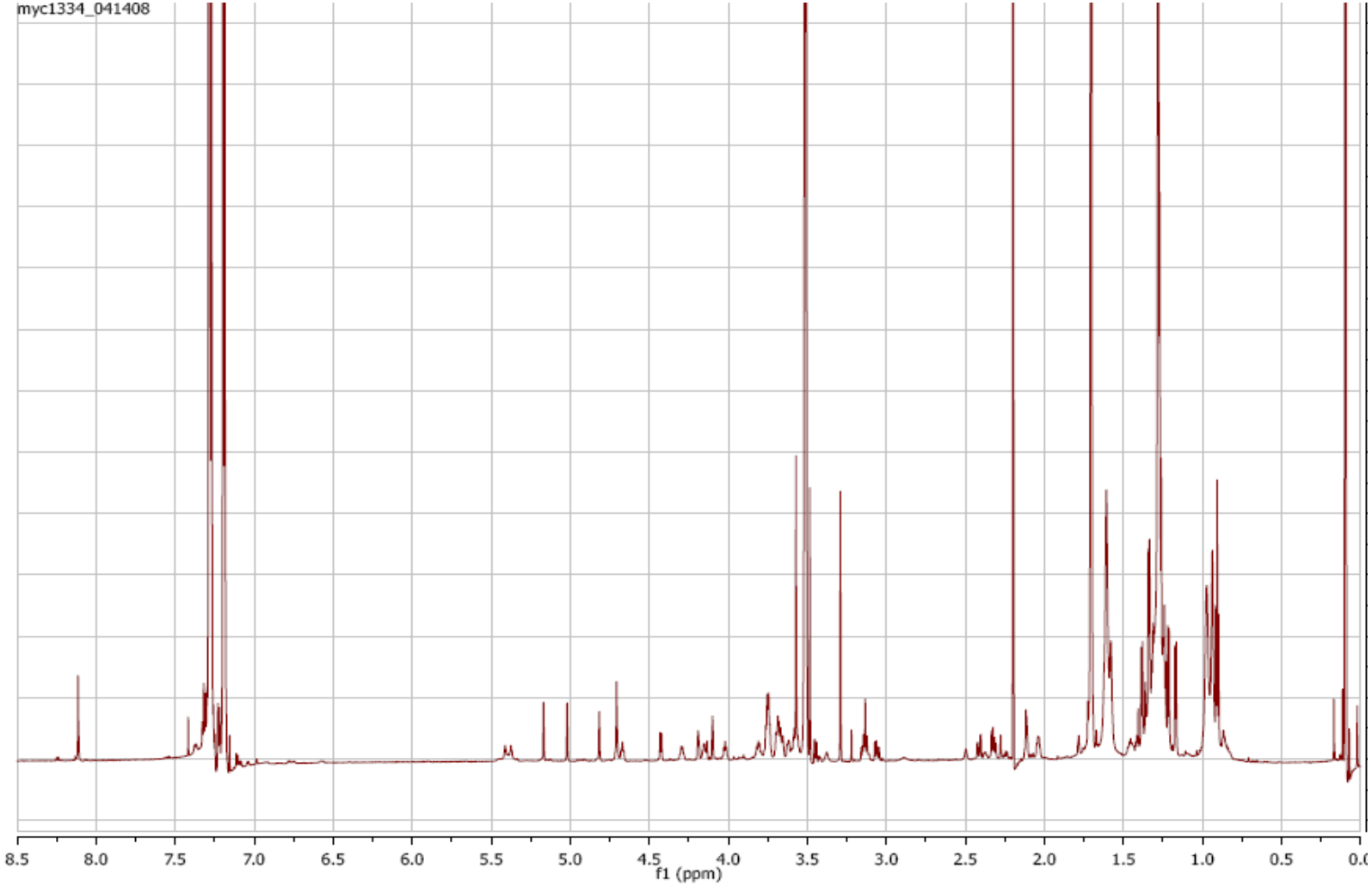


Figure S2. ^{1}H NMR spectrum of purified L1334.

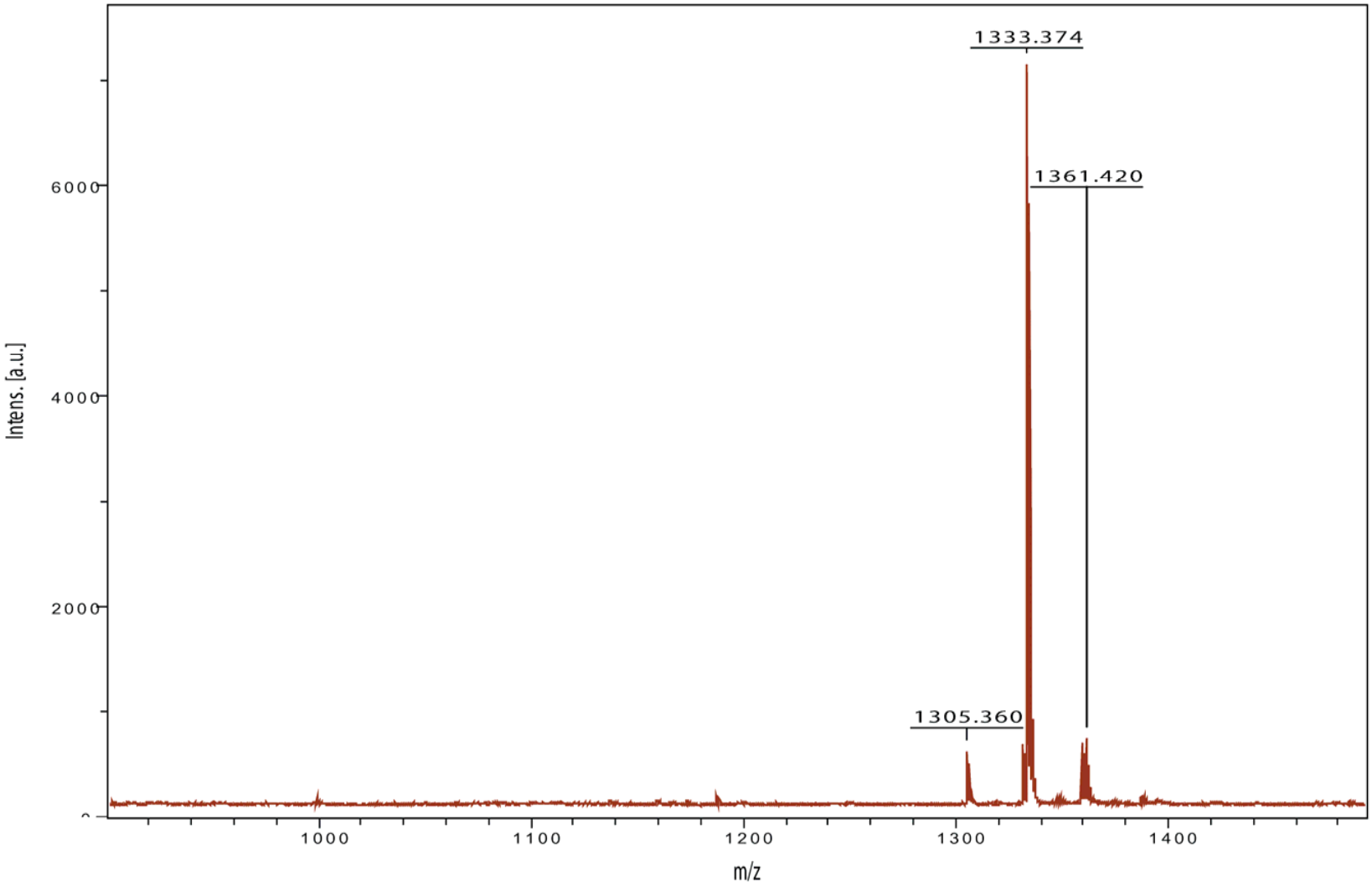


Figure S3. MALDI-TOF/MS spectrum of purified L1334.

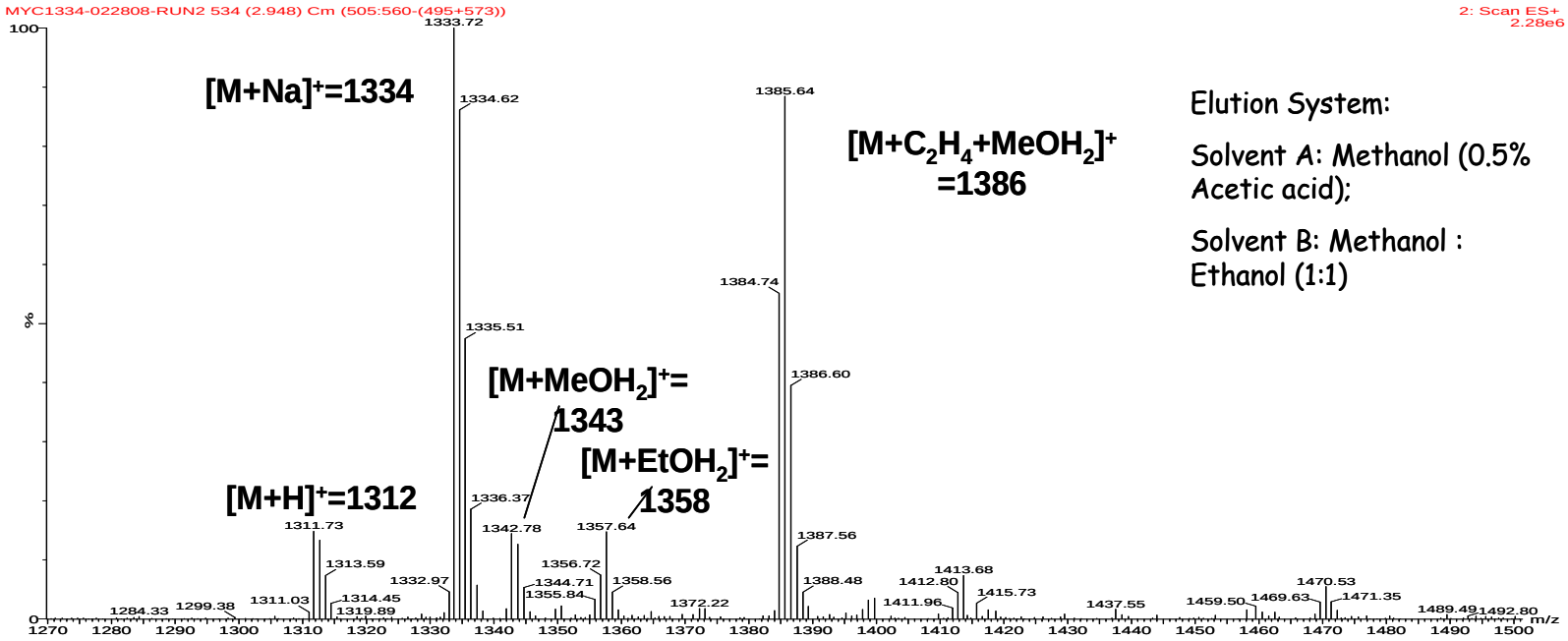


Figure S4. LC/MS ESI positive mode spectrum of purified L1334.

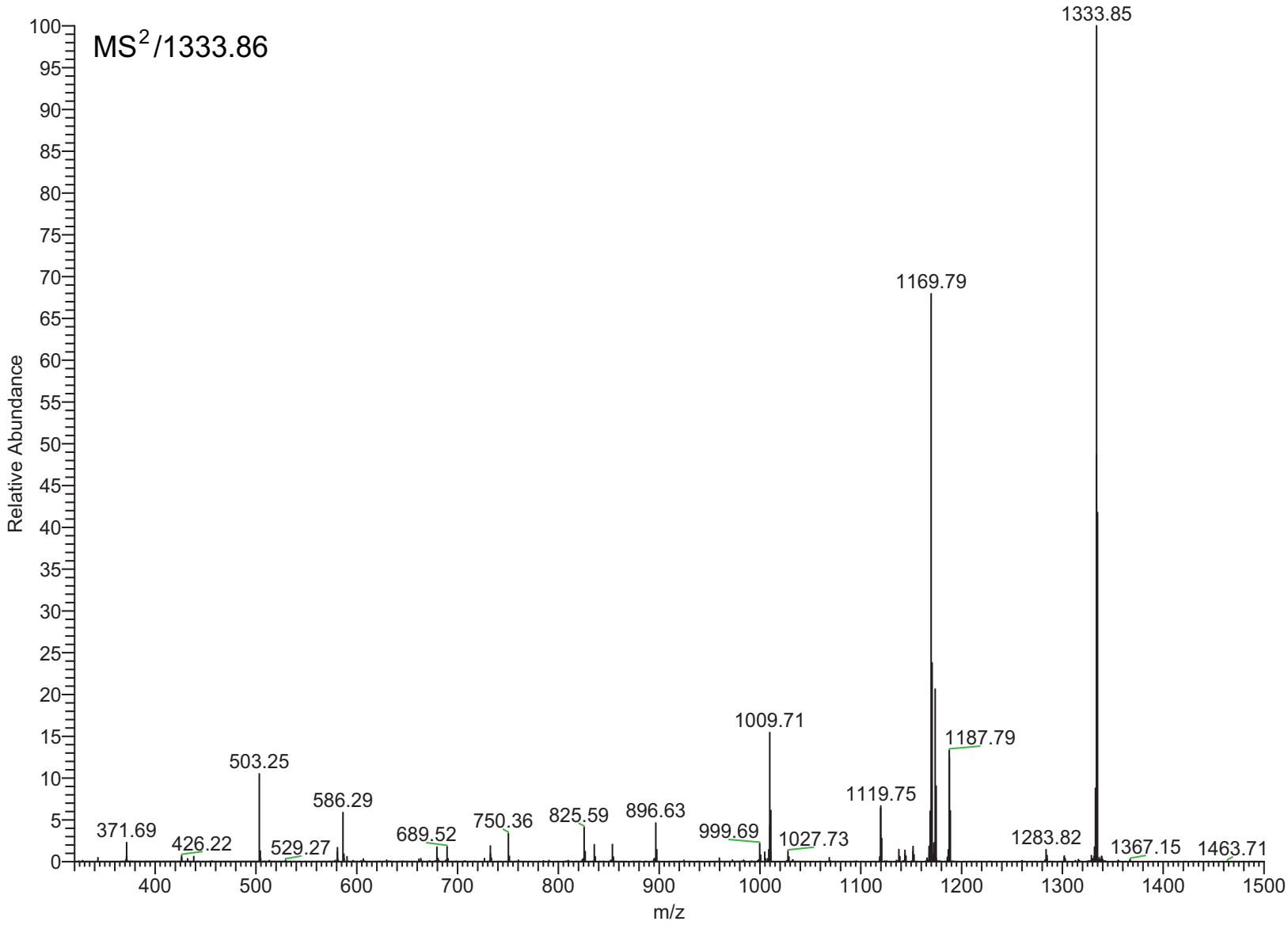


Figure S5. The MS^2 spectrum of m/z=1333.86 from MS spectrum by Orbitrap.

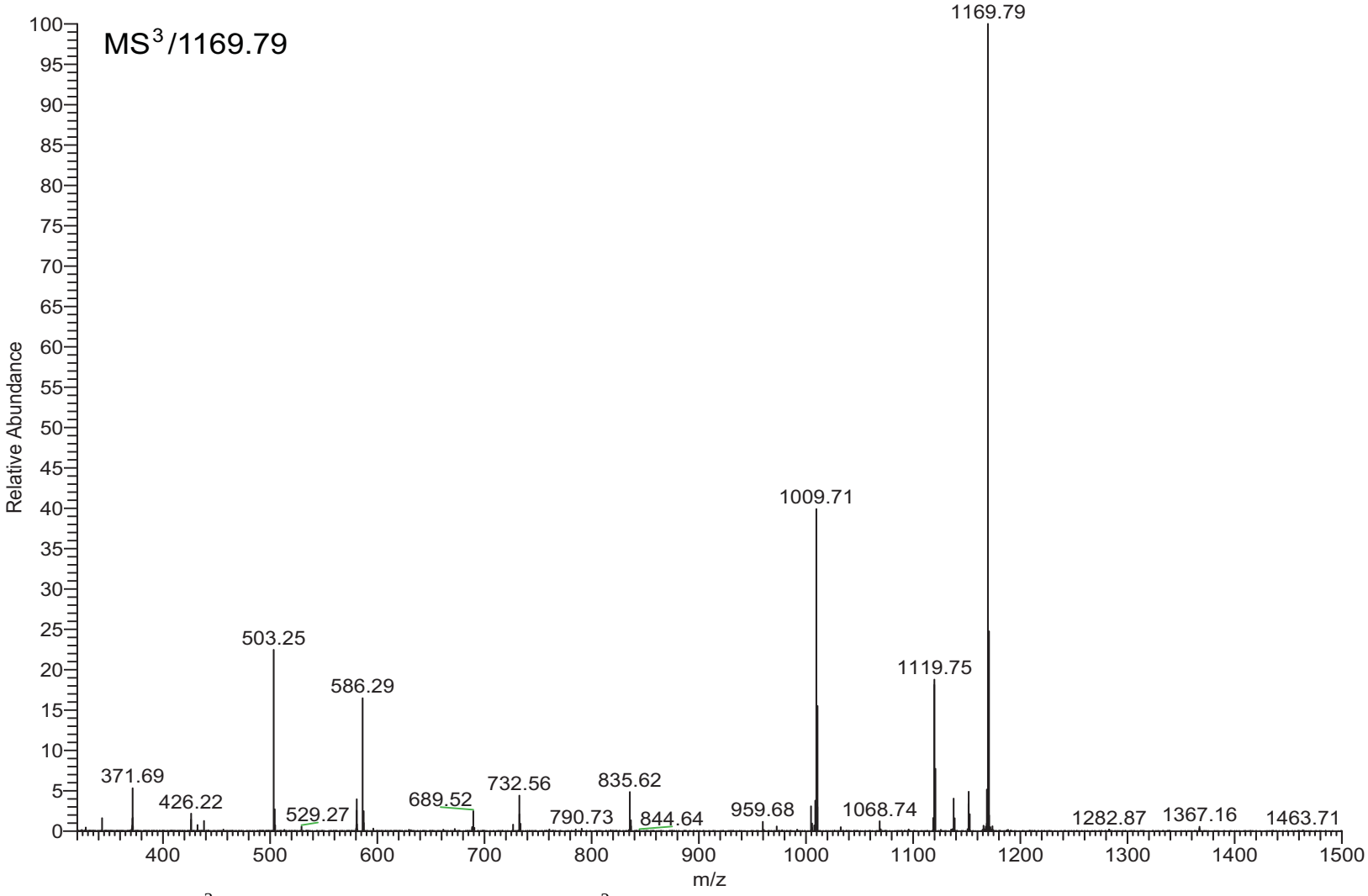


Figure S6. The MS^3 spectrum of m/z=1169.79 from MS^2 spectrum by Orbitrap.

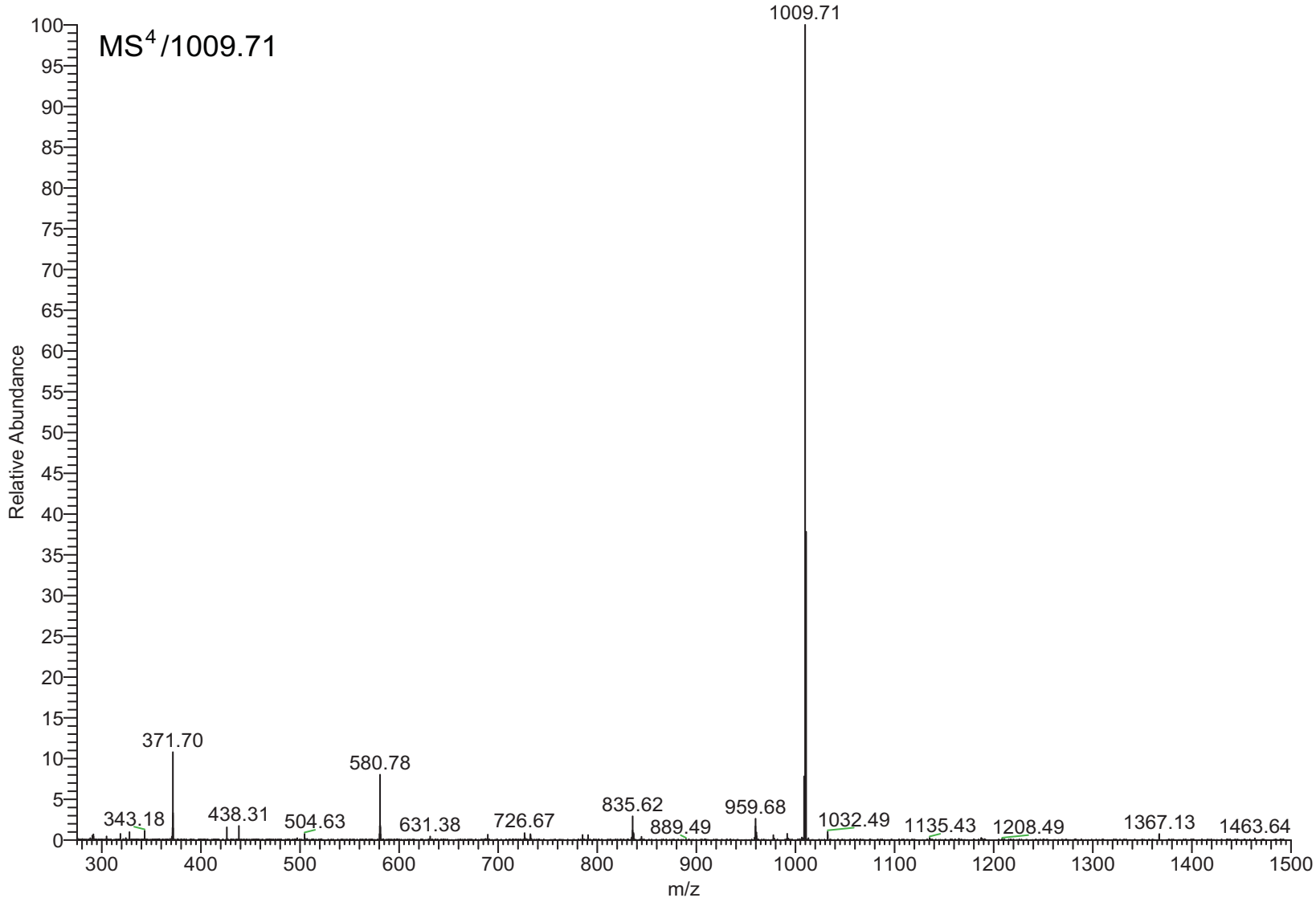


Figure S7. The MS^4 spectrum of m/z=1009.71 from MS^3 spectrum by Orbitrap.

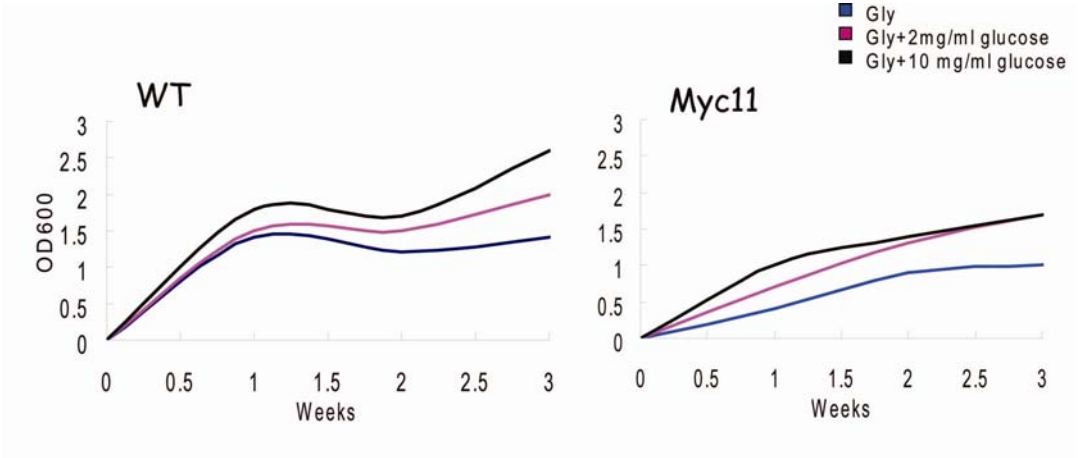


Figure S8. *M. smegmatis* mc^{2}155 wild type and Myc11 growth curve in different glucose media. The black curve is the 7H9 medium supplemented with 0.2% (v/v) glycerol; the purple curve is the 7H9 medium supplemented with 0.2% (v/v) glycerol and 0.2% (w/v) glucose; the blue curve is the 7H9 medium supplemented with 0.2% (v/v) glycerol and 1% (w/v) glucose.

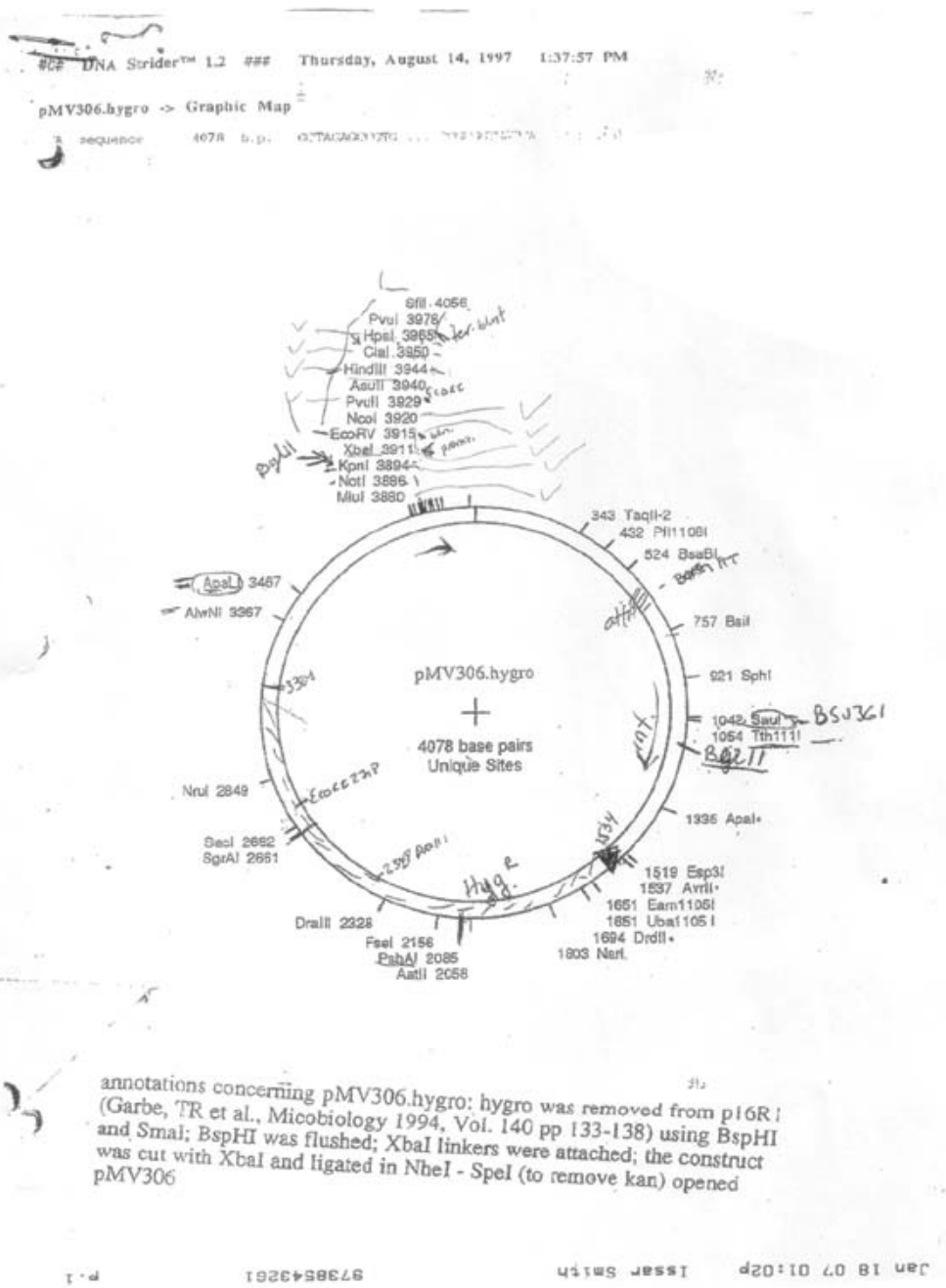


Figure S9. The pMV306.hygro vector map.

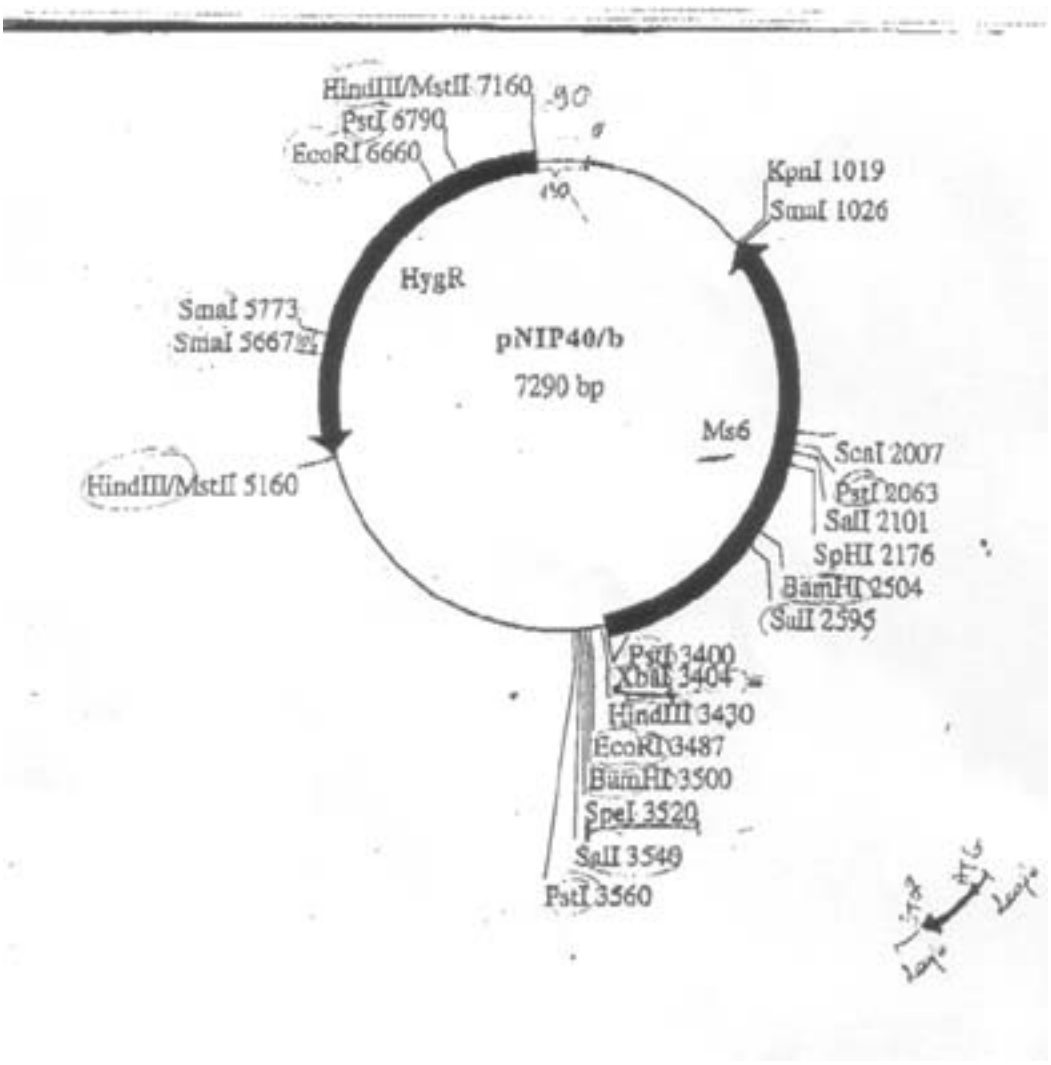


Figure S10. The pNIP40/b vector map.

pFPCA1 (*E. coli*/*M. smegmatis* shuttle vector)

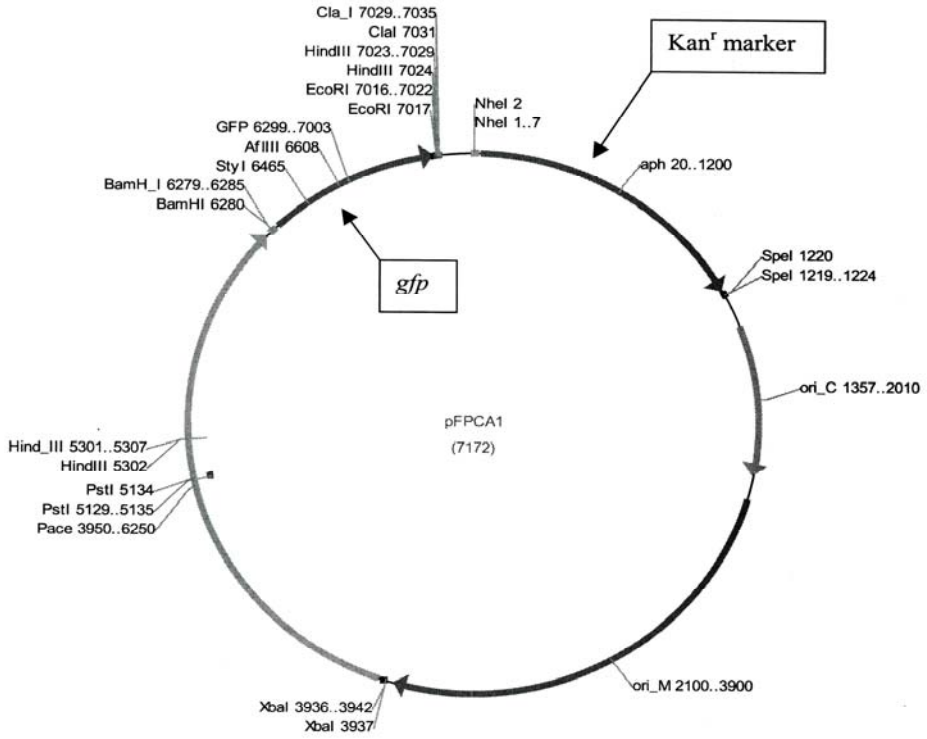
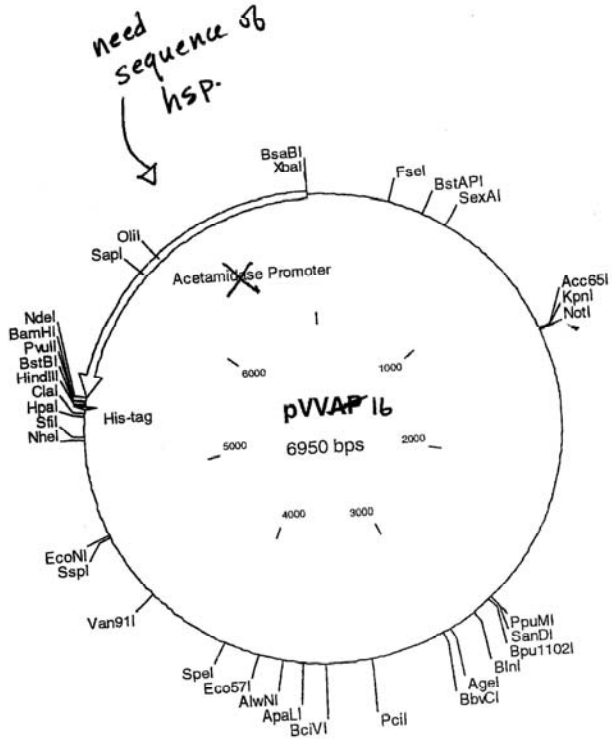


Figure S11. The pFPCA1 vector map.



hygro & kan resistant

MCS: NdeI, BamHI, PvuII, BstBI and HindIII

hsp60

I amplified the acetamidase promoter from M. smegmatis genomic DNA to get PCR product (1.5 kb) with XbaI site at 5' end and NdeI site at 3' end. After PCR product was psequenced, I cloned the acetamidase promoter (AP) to XbaI and NdeI sites of pVV16 to replace the hsp promoter of pVV16. Therefore, the multiple cloning sites (NdeI, BamHI, PvuII, BstBI and HindIII) in pVV-AP are similar with MCS of pVV16, and the C-terminal His-tag is still available in pVV-AP. The acetamdiase promoter I cloned is the same as the promoter of pJAM2 (James A. Triccas, et al. FEMS Microbiology Letter 167 (1998) 151-156). Please see the map of pVVAP and pVVAP sequence in Attachment.

Please contact me if you have any question. Good luck for your expression of proteins.

Mary Ma

pMV261

Figure S11. The pVV16 vector map.

Table S1. Summary of *M. smegmatis* wild type, Myc11 and complementary strains properties.

	WT	Myc11	*Myc11*:*pMS100*	*Myc11*:*pMS101*	*Myc11*:*pMtb100*
Phenotype	Smooth	Rough	Rough	Smooth	ND
Sliding motility	Yes	No	No	No	ND
Single colony diameter	1 cm	1 cm	0.76 cm	0.89 cm	ND
L1334 production	No	Yes	No	No	Yes
L1358 production	High	High	Low	Low	Low
L1518 production	Yes	Yes	No	No	No
L1037 production	No	No	Yes	Yes	No

Table S2. Summary of all primers applied in the Rv3409c projects.

PCR Primer	Product	Sequence and size (bp)	Restriction Digest Site	Tm (°C)
JGNN-1	pET-28(b)-NRv3409c	5’- CGC CAT ATG AAG CCG GAT TAC GAC GTC CTG -3’	NdeI	71.5°C
JGNN-2	pET-28(b)-NRv3409c	5’- CCG AAG CTT CTA GCC CGC GTT GCT GAC CGG -3’	HindIII	75.6°C
JGNN-3	pFPCA1-NRv3409c	5’- GCC AGA TCT CAT CAT CAT CAT CAT CAC AGC AGC -3’	BglII	69.6°C
JGNN-4	pFPCA1-NRv3409c	5’- GGC GAA TTC CTA GCC CGC GTT GCT GAC CGG -3’	EcoRI	75.6°C
JGNN-5	pVV16-NRv3409c	5’- CGC GGA TCC CAT CAT CAT CAT CAT CAC AGC AGC -3’	BamHI	72.1°C
JGNN-7	pMAL-C4E-Rv3409c	5’- CGC GAA TTC CAT CAT CAT CAT CAT CAC AGC AGC -3’	EcoRI	69.6°C
JG005	pMCO-100M	5’- CGC GGG CTA GCT TGC GGC CGC ACT CGA GC -3’	N/A	74.1°C
JG006	pMCO-100M	5’- CGC AAG CTA GCC CGC GTT GCT GAC CGG ATC -3’	N/A	70.3°C
JGP-1	pMV306.hygro_Mtb999Rv3409c*	5’- GGG GGT CAT ACT GCA GCG ATG AAG CCG G -3’	N/A	67.7°C
JGP-2	pMV306.hygro_Mtb999Rv3409c*	5’- GCC GAA GCT TCT AGC CCG GGT TGC -3’	HindIII	66.6°C
JGP-3	pMV306.hygro_Mtb999Rv3409c*	5’- GCC CGG ATA TCC GAT CGA GCT CG -3’	EcoRV	63.3°C
JGP-4-2	pMV306.hygro_Mtb999Rv3409c*	5’- CCG GCT TCA TCG CTG CAG TAT GAC CCC C -3’	N/A	67.7°C
JGP-20	pNIP40/b_Msmeg300MSMEG1604‡	5’- GGA CTA GTG GGC AAC GGC TGG TTC TGG CC -3’	SpeI	74.5°C
JGP-21	pNIP40/b_Msmeg300MSMEG1604‡	5’- GGA CTA GTC TAC CCC GCC GAC GAC ACG GG -3’	SpeI	75.9°C
JGP-22	pMV306.hygro_Msmeg300MSMEG1604¶	5’- CCG ATA TCG GGC AAC GGC TGG TTC TGG CC -3’	EcoRV	74.5°C
JGP-23	pMV306.hygro_Msmeg300MSMEG1604¶	5’- CGG AAG CTT CTA CCC CGC CGA CGA CAC GGG -3’	HindIII	76.9°C
ChoD1	pMCO-100, pMCO-100M	5’- GTG TTC TTC GGC CCC GAC GGC ACC -3’	Forward Sequencing	68.2°C
ChoD2	pMCO-100, pMCO-100M	5’- GTC ATG CAG CAC CTG GAC AAC TCG -3’	Forward Sequencing	62.2°C
ChoD3	pMCO-100, pMCO-100M	5’- CC CGT CGG ACC GCC GCT CAA ACC -3’	Backward Sequencing	68.7°C
ChoD4	pMV306.hygro-Mtb999Rv3409c*	5’- GCT AAA TCC CGA CCT GTT GGG TGC CGC -3’	Forward Sequencing	72.2°C
ChoD5	pMV306.hygro-Mtb999Rv3409c*	5’- CGT TAG CCG AAT CGG CCG AGG CGC -3’	Forward Sequencing	73.1 °C
ChoD6	pMV306.hygro-Mtb999Rv3409c*	5’- GGA GCC GTA GCC AAC GAT GAC GCC -3’	Backward Sequencing	71.4 °C

ChoD7	pMV306.hygro, pNIP40/b derivatives	5’- GGT GCT ACG GCA TCC AGC GCA TCC -3’	Forward Sequencing	71.4 °C
ChoD8	pMV306.hygro, pNIP40/b derivatives	5’- GGT GCC GAG GTG CAT CCG CTG C -3’	Forward Sequencing	72 °C
ChoD9	pMV306.hygro, pNIP40/b derivatives	5’- CGT GGC ACA CTG CGC CTG CTC AAC G -3’	Forward Sequencing	72.8 °C
ChoD10	pMV306.hygro, pNIP40/b derivatives	5’- GGC GCC GGC CAG GAT CAT CAC G -3’	Backward Sequencing	72°C

(*pMV306.hygro_Mtb999Rv3409c is also named pMtb100; ‡ pNIP40/b_Msmeg300MSMEG1604 is also named pMs100; ¶ pMV306.hygro_Msmeg300MSMEG1604 is also named pMs101.)