

**A GLOBAL ANALYSIS OF THE ADAPTATIONS REQUIRED FOR STEROL  
CATABOLISM IN MYCOBACTERIUM TUBERCULOSIS**

A Dissertation Presented

By

JENNIFER ELIZABETH GRIFFIN

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 20, 2011

Program in Molecular Genetics and Microbiology

A Dissertation Presented

By

Jennifer Elizabeth Griffin

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

Christopher Sasseti, Ph.D., Thesis Advisor

Jon Goguen, Ph.D., Member of Committee

Victor Boyartchuk, Ph.D., Member of Committee

Bert Van den Berg, Ph.D., Member of Committee

Eric Rubin, M.D., Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

John Leong, M.D., Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

Anthony Carruthers, Ph.D.,  
Dean of the Graduate School of Biomedical Sciences

Program in Molecular Genetics and Microbiology

May 20, 2011

## Acknowledgements

First and foremost I need to thank my advisor, Chris. I could have never known how fortunate I would be to call him my mentor. Not only exceptionally intelligent, but also deeply grounded, Chris naturally brings out the best in people with his seemingly limitless amount of patience and optimism. And even as I leave the bench, I am immensely grateful that he remains one of my biggest supporters.

I'm also grateful for the fabulous group of people he has assembled. Thank you to the many members, past and present, of the Sassetti Lab. You have made it easy to come to lab everyday. Thank you for the laughs, too numerous to count, and for broadening my horizons past the Massachusetts state line. 3pm will always be teatime in my book.

This work is truly a testament to the large support network I am so lucky to have in my life. To my husband, Joe, thank you for giving me a sense of perspective, when frustration and impatience would have otherwise consumed me. You are my rock. To my parents, thank you for raising me to believe there was nothing in this world I couldn't do. To my sister, Amanda, here's to earning your own Ph.D. in less than the 6 years it took me. I'm glad I'll be there for you along the way.

Lastly, thank you to the dedicated members of my thesis and dissertation examination committee. In addition to running your own successful laboratories, you never failed to find time for advice, support and to show genuine interest in my studies. Your selflessness has not gone unnoticed and is deeply appreciated.

## Abstract

Systems biology approaches have allowed for comprehensive understanding of complicated biological processes. Here, we've developed a global phenotypic profiling method by improving upon transposon mutagenesis methods for identifying genes required for bacterial growth in various conditions. By using the massively parallel power of Illumina sequencing, we precisely redefined the genes required for the growth of *Mycobacterium Tuberculosis* (Mtb) in vitro. This adapted technique provided more informative data with both increased dynamic range and resolution. As a result, we quantitatively assessed the fitness of individual mutants, as well as identified sub-genic essentiality. Mtb is well adapted to its nutrient-limiting intracellular niche. One important and novel adaptation is its ability to consume cholesterol for both energy and carbon. A combination of this genome-wide phenotypic analysis and global metabolite profiling was used to define the dedicated cholesterol catabolic pathway, as well as important transcriptional and metabolic adaptations required for the consumption of this carbon source. We identified the methylcitrate cycle (MCC) and an unexpected gluconeogenic route as essential pathways. Furthermore, we found that the cholesterol-dependent transcriptional induction of these metabolic enzymes was also essential for growth on this substrate, a function mediated by the Rv1129c regulatory protein. Using a combination of genetic and chemical methods to inhibit these pathways, we show that cholesterol represents a significant source of carbon during intracellular growth in macrophages.

Finally, we have begun to define the mechanism by which lipids, such as cholesterol, are imported into the cell by investigating the assembly of the ABC-like lipid transporter, Mce1. The subunits of this system are localized to the cell wall and data is provided to support a novel mechanism for Mce-dependent import of lipids, such as cholesterol. In sum, this global analysis of host cholesterol utilization during infection provides insight into each step of this complicated process; import into the bacterial cell, the degradation of the molecule into primary metabolites, and the transformation of these metabolites into carbon and energy.

## Table of Contents

Acknowledgements .....	iii
Abstract.....	v
Table of Contents .....	vii
List of Tables .....	x
List of Figures .....	xi
Preface .....	xii
<b>Chapter I. Introduction .....</b>	<b>1</b>
Overview .....	1
TB Disease.....	2
Mtb Pathogenesis .....	4
The Intracellular Environment .....	5
The Lipid Diet of Mtb .....	6
Mtb and Cholesterol .....	7
<b>Chapter II. Whole genome phenotypic profiling identifies essential genes in Mtb.....</b>	<b>13</b>
Summary.....	13
Introduction .....	14
Materials and Methods.....	16
Results .....	20
Discussion.....	30
<b>Chapter III: Whole genome phenotypic profiling identifies the entire set of genes required for utilizing cholesterol .....</b>	<b>33</b>

Summary.....	33
Introduction .....	34
Materials and Methods.....	36
Results .....	39
Discussion .....	50
<b>Chapter IV. Transcriptional and metabolic adaptations to intracellular growth on cholesterol .....</b>	<b>54</b>
Summary.....	55
Introduction .....	56
Materials and Methods.....	58
Results .....	61
Discussion.....	76
<b>Chapter V. The Mce lipid importers require a complex formation within the cell wall of Mtb.....</b>	<b>81</b>
Summary.....	81
Introduction .....	82
Materials and Methods.....	87
Results .....	92
Discussion.....	96
<b>Chapter VI. Discussion .....</b>	<b>101</b>
The bioavailability of cholesterol .....	101
Cholesterol and redox balance .....	103
Cholesterol and secondary metabolites .....	104



Cholesterol and pyruvate .....	106
Cholesterol and TB susceptibility .....	107
Appendix.....	109
Bibliography .....	115

## List of Tables

Table 3.1 Genes predicted to be specifically required for growth on cholesterol .....	46
Table 4.1 Metabolomic profile of cholesterol catabolism .....	65

## List of Figures

Figure 1.1 Cholesterol .....	8
Figure 2.1 Transposon insertions are randomly distributed, and gaps in transposon coverage correspond to predicted essential ORFs.....	21
Figure 2.2 Defining essential genes with deep sequencing is more sensitive and precise than previous approaches.....	24,25
Figure 2.3 Illumina sequencing predicts subgenic essentiality .....	28
Figure 2.4 Illumina sequencing identifies an essential MviN domain in FlpA .....	29
Figure 2.5 Homologous recombination validates subgenic essentiality .....	30
Figure 3.1 Defining genes that are specifically required for growth on cholesterol .....	40,42
Figure 3.2 Phenotypic predicts the cholesterol catabolic pathway .....	44
Figure 4.1 Metabolic profiling during growth on cholesterol .....	63,64
Figure 4.2 Cholesterol-derived propionate is incorporated into SL-1 and increases both its mass and abundance .....	67
Figure 4.3 The MCC genes are required for growth on cholesterol .....	69
Figure 4.4 The transcriptional activation of MCC genes by Rv1129c is required for growth on cholesterol .....	72
Figure 4.5 Cholesterol is a significant intracellular source of propionate ...	74
Figure 5. Mce loci of Mtb .....	86
Figure 5.1 Mce1 proteins localize to the cell wall fraction of Mtb .....	93
Figure 5.2 Mce1 proteins are differentially accessible to membrane impermeable biotin .....	94
Figure 5.3 MceG-dependent ATP hydrolysis is required for Mce1 complex stabilization.....	96

## Preface

Portions of this thesis have appeared in the following publications:

J.E. Griffin, J.D. Gawronski, M.A. DeJesus, T.R. Ioerger, B.J. Akerley, C.M. Sassetti *Quantitative high-resolution phenotypic profiling defines genes essential for Mycobacterium tuberculosis survival and cholesterol catabolism*, submitted

J.E. Griffin, A.K. Pandey, S.A. Gilmore, V. Mizrahi, J.D. McKinney, C.A. Bertozzi, C. Sassetti *Transcriptional and metabolic adaptation is required for cholesterol catabolism by Mycobacterium tuberculosis during intracellular growth*, manuscript in preparation

C.L. Gee, K.G. Papavinasasundaram, S.R. Blair, C. Baer, A.M. Falick, D.S. King, J. Griffin, H. Venghatakrishnan, J. Wei, E.J. Rubin, C.M. Sassetti and T. Alber, *A Ser/Thr phosphosignaling system controls cell-wall synthesis in mycobacteria*, submitted

## Chapter I. Introduction

### Overview

Tuberculosis (TB) has plagued mankind since the early days of human civilization. Sequence diversity analysis suggests that TB existed during the early human migrations out of Africa 50,000 years ago (Hershberg et al. 2008). Evidence for various manifestations of TB, including skeletal abnormalities like Pott's disease, can be found in both ancient Egyptian mummies and artwork from 5,000 years ago (Cave 1939, Daniel 2006, Morse et al. 1964, Zimmerman 1979). Archaeological evidence places TB not only in ancient Egypt, but also ancient Peru and all across Europe throughout the Middle Ages (Daniel 2006, Salo et al. 1994). By the early part of the 19<sup>th</sup> century, TB was devastating European and North American cities in epidemic proportions claiming between 800-1,000/100,000 lives per year (Grigg 1958a, Grigg 1958b, Krause 1928). In 1804, René Théophile Hyacinthe Laennec argued that the various presentations and pathologies of TB were in fact, one disease. It was not until 1882 that Robert Koch identified the etiologic agent (Daniel 2004, Sakula 1982). Koch's seminal work not only identified *Mycobacterium tuberculosis* (Mtb) as the causative agent of TB, but also defined what are still known today as Koch's postulates, which determined the criterion of infectious disease and gave rise to the science of bacteriology. Almost 130 years later, one third of the world is latently infected with Mtb and close to 2 million people die annually from TB (WHO 2010). Though modern issues, such as HIV co-infection and drug-resistant strains

complicate the current TB landscape, the root of the issue remains the same: a lack of understanding of the basic physiology of such a complicated killer.

Mtb is an acid-fast, rod shaped bacillus with close to 4000 predicted open reading frames (ORF) (Cole et al. 1998). Phylogenetically classified as gram-positive, the organization of its complex and unique, lipid-laden cell wall mimics that of a gram-negative organism (Hoffmann et al. 2008, Zuber et al. 2008). Mtb contains a single circular chromosome that is 65.6% G+C rich (Cole et al. 1998). With a generation time of 19-24 hours in vitro, an extremely hydrophobic cell wall and a biosafety level 3 classification, Mtb plagues both patients and scientists alike.

### **TB Disease**

Mtb is a successful pathogen in part, because it elicits a chronic disease. Only approximately 10% of infected individuals will develop active pulmonary TB characterized by weight loss and persistent productive cough (Flynn and Chan 2001, Kaufmann 2001). The majority of the remaining individuals develop a clinically defined latent disease, characterized by infection without symptoms. Latently infected individuals are non-infectious, but skin test positive and make up a huge reservoir of potentially deadly future infections. Latent infections carry a 10% lifetime risk of reactivation (Horsburgh 2004). This risk increases dramatically to a 10% annual risk among latently infected human

immunodeficiency virus (HIV) positive individuals (Whalen et al. 1997, Selwyn et al. 1992, Selwyn et al. 1989). The requirements and mechanisms for reactivation of TB disease are not well defined, but it is clear that an immunocompromised state favors reactivation.

Current methods for treatment of TB include both prophylactic and therapeutic measures. In the 1930s the development of the bacillus Calmette-Guérin (BCG) vaccine for TB was generated using the zoonotic pathogen *M. bovis*, the causative agent of TB in cattle and many other mammals. *M. bovis* was passed hundreds of times on bile potatoes to render an immunogenic, but avirulent organism that was and is still used in some countries today to vaccinate infants against Mtb (Calmette 1931). Unfortunately, BCG is ineffective against adult pulmonary TB and elicits a varied protective response to infant pulmonary TB (Fine 1995, Kaufmann 2010). BCG vaccination now occurs only in countries where TB is endemic and the population is at high risk of serious complications. With the varied protectiveness of BCG, the main burden of TB treatment lies on therapeutic measures. Current antitubercular therapies to treat active, drug-sensitive TB include a cocktail of first-line drugs administered for six to nine months (LoBue et al. 2010). These include isoniazid (INH), rifampin (RIF), pyrazinamide (PZA) and ethambutol (EMB). In patients where these drugs fail, second-line drugs, which include fluoroquinolones and aminoglycosides are administered. These drugs are toxic, broad-spectrum antibiotics saved for use

against drug-resistant strains. Recent surveys estimate 440,000 new cases of multiple drug-resistant tuberculosis (MDR-TB) globally in 2008 and have identified extensively drug-resistant (XDR-TB) strains in 58 countries and territories (WHO 2010). Like MDR-TB, XDR-TB is resistant to RIF and INH, but also to many less-effective second line drugs. It is clear that new antitubercular measures are needed to combat an increasing epidemic of untreatable TB.

### **Mtb Pathogenesis**

Despite our inability to tackle this human pathogen on a global scale, the past century has generated a wealth of insight into the molecular pathogenesis of Mtb. We understand that following inhalation, the bacterium is engulfed by alveolar macrophages, where it is uniquely adapted for survival (Leemans et al. 2001). Following phagocytosis, macrophages employ an arsenal of molecular mechanisms that result in destruction of foreign invaders, including acidification of the phagosome, an increase in reactive oxygen and nitrogen species, antimicrobial peptides and hydrolytic enzymes (Liu and Modlin 2008). The main strategy employed by Mtb against the macrophage is one of evasion. By halting the fusion of mycobacterial phagosomes with lysosomes, Mtb generates a protective niche in which it evades further macrophage and immune insults (Armstrong and Hart 1971). Mtb is believed to accomplish this by inhibiting the recruitment of the vacuolar proton ATPase pump and maintaining a pH of 6.4 within the infected phagosome (Armstrong and Hart 1971, Sturgill-Koszycki et al.



1994). During this time, the mycobacterial phagosome is maintained within the early endosomal recycling network and is thought to have access to cellular membranes and the extracellular milieu (Clemens and Horwitz 1996, Russell et al. 1996). During this acute phase of infection, Mtb replicates unabated and it is not until the onset of the adaptive immune response that bacterial growth is halted.

### **The Intracellular Environment**

The ability of Mtb to initially thrive in the face of host insults is remarkable considering its hostile intracellular environment. In addition to the toxic molecules resulting from the oxidative burst and low pH, the phagosome also presents a nutritional challenge for Mtb. Depriving pathogens of nutrients is an effective antimicrobial mechanism and while challenging to analyze directly, it has been indirectly demonstrated by auxotrophic mutants and bacterial transcriptional responses. The inability of several mycobacterial auxotrophs to survive *in vivo* suggest that purines, certain amino acids such as leucine, lysine and proline, and even pantothenate are limiting in the phagosome (Hondalus et al. 2000, Jackson et al. 1999, McAdam et al. 1995, Pavelka et al. 2003, Sambandamurthy et al. 2002). In addition, several groups have investigated the *in vivo* transcriptional response of Mtb and their results all confirmed that Mtb was mounting a response to a nutrient-limiting, specifically a carbohydrate-limiting, environment (Schnappinger et al. 2003, Homolka et al. 2010, Dubnau et

al. 2002, Timm et al. 2003). Schnappinger and colleagues in particular, found that Mtb upregulates the expression of 18 genes all predicted to encode  $\beta$ -oxidation functions, suggesting Mtb shifts its intracellular metabolism to a lipid-based diet.

### **The Lipid Diet of Mtb**

The idea that Mtb consumes lipids was first suggested in the 1950s when Segal and Bloch demonstrated that Mtb isolated from lungs of infected mice could preferentially use fatty acids over carbohydrates to stimulate cellular respiration (Bloch and Segal 1955). In addition, a single nucleotide polymorphism in *pykA* (pyruvate kinase) in the closely related species *M. bovis*, renders this bacterium unable to fully catabolize sugars, requiring the metabolism of non-carbohydrate carbon sources, such as lipids to support growth (Keating et al. 2005). Further support was generated in 1998 when the Mtb genome was sequenced and over 250 genes predicted to have roles in lipid metabolism were identified (Cole et al. 1998). Not only were many of these lipid metabolism genes upregulated in vivo (Schnappinger et al. 2003), but some are also required for virulence (Brzostek et al. 2007, Chang et al. 2007, McKinney et al. 2000, Muños-Elías et al. 2005, Sassetti et al. 2003). By studying one such gene, *icl*, the metabolic switch to lipids became further delineated.

*icl* encodes the enzyme isocitrate lyase, which cleaves isocitrate to glyoxylate and succinate and represents the first step of the glyoxylate cycle (McKinney et al. 2000). The glyoxylate cycle is an important anapleurotic modification to the TCA cycle and is required for growth on carbon sources that are converted to acetyl CoA (C2 compounds).  $\beta$ -oxidation of fatty acids and other lipids generate C2 compounds and therefore, require the glyoxylate cycle. Interestingly, *Mtb* specifically requires *icl* during the chronic phase of disease in mice (McKinney et al. 2000). Since the hallmark of chronic disease is the activation of the macrophage via the cytokine interferon gamma ( $\text{IFN}\gamma$ ), this suggested that *Mtb* metabolism was modified by the activation state of the macrophage. Lipid metabolism therefore, may be a necessary adaptation to the nutrient-limiting activated macrophage.

### **Mtb and Cholesterol**

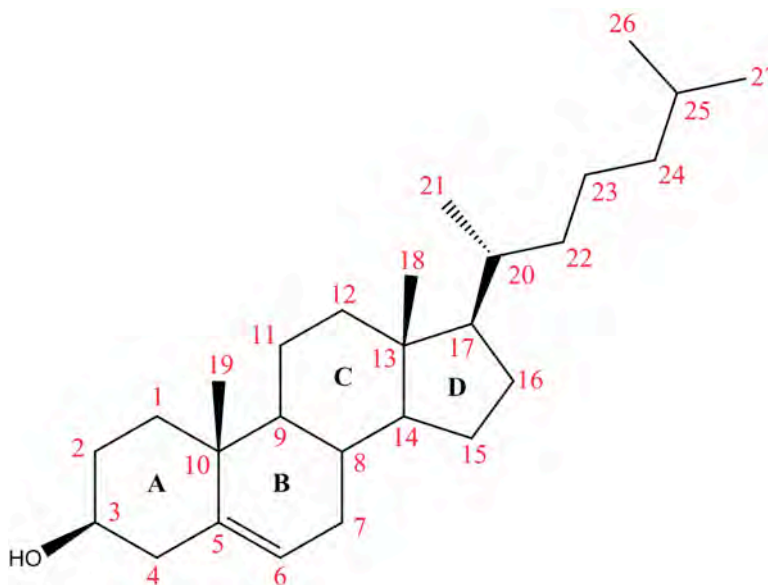
The evidence that *Mtb* shifted its metabolism intracellularly from carbohydrates to lipids was mounting. Clues to the identity of the lipid(s) were generated in 2007, when Van der Geize and colleagues identified a region in the chromosome of a related actinomycete, *Rhodococcus* sp. strain RHA1 (RHA1), that was specifically upregulated following growth on cholesterol (Van der Geize et al. 2007). RHA1 upregulates a cluster of 51 genes in response to cholesterol that all localize to a 235kb stretch in the chromosome. These genes encode proteins predicted to be involved in  $\beta$ -oxidation, as well as genes with high sequence

identity to enzymes involved in the degradation of steroid rings A and B.

Interestingly, this cluster is highly homologous to a similar stretch of 82 genes in *Mtb*. This region, called the Cho region, included genes known to be required for *Mtb* virulence, like *hsaC* (Yam et al. 2009), *igr* (Chang et al. 2007) and *Mce4* (Sasseti et al. 2003, Joshi et al. 2006) and strongly suggested that like RHA1, *Mtb* had the ability to metabolize the lipid cholesterol.

Less than a year later, our lab showed that the addition of cholesterol to a defined medium that otherwise did not support growth, allowed for *Mtb* replication (Pandey and Sasseti 2008). This was the first evidence that cholesterol served as a carbon source for *Mtb* growth and also highlighted *Mtb*'s ability to utilize cholesterol for multiple purposes. By supplementing the defined medium with differentially radiolabeled cholesterol, it was demonstrated that the carbon at

**Figure 1.1**  
**Cholesterol.** Steroid rings A-D and carbons 1-27 are labeled.



position 4 (C-4) in steroid ring A was specifically transformed into CO<sub>2</sub>, while the carbon at position 26 in the side chain was specifically incorporated into mycobacterial cell wall lipids, including phthiocerol dimycocerosate (PDIM). The transformation of C-4 into CO<sub>2</sub> suggested that Mtb was able to metabolize the steroid nucleus via the TCA cycle. Together, this demonstrated that Mtb had the ability to utilize the steroid nucleus of cholesterol to generate energy and the side chain of cholesterol as a biosynthetic precursor for lipid anabolism. Pandey and Sassetti further demonstrated that cholesterol uptake was mediated by the ABC-like transport system, Mce4.

The four Mce operons of the Mtb chromosome all share homology to ATPase Binding Cassette (ABC) transporters and due to demonstrated genetic interactions with lipid metabolism genes were hypothesized to play roles in bacterial lipid transport (Casali and Riley 2007, Joshi et al. 2006). Mce4 encodes a ten-gene operon at the beginning of the Cho region. In the absence of Mce4 or the related ATPase, MceG (Joshi et al. 2006), Mtb has a reduced growth rate in cholesterol, a decreased rate of cholesterol association, and a slower rate of <sup>14</sup>C, C-4 conversion to CO<sub>2</sub>; all of which supports the role of Mce4 in cholesterol transport (Pandey et al. 2008). In addition, the orthologous Mce4 operon in RHA1 was shown to be required for growth in, and uptake of, cholesterol (Mohn et al. 2008). The demonstrated involvement of Mce4 in cholesterol utilization had additional implications for the role of cholesterol during the course of TB disease.

Mce4 has been shown to be specifically required for virulence during the chronic phase of disease in mice; a phenotype that has been recapitulated in bone marrow derived macrophages (BMDM) following activation with IFN $\gamma$  (Sassetti et al. 2003, Pandey et al. 2008). This phenotype echoes the phenotype of *icl* mutants and strongly supports the idea that IFN $\gamma$ -induced changes to the mycobacterial phagosome during the chronic phase of infection, result in a metabolic shift to lipids, specifically cholesterol. Thus, it appears that Mtb is highly adapted to the nutrient limitations of an ever-changing phagosome environment.

By studying another operon located within the Cho region, Chang and colleagues then demonstrated that cholesterol was being metabolized during the acute phase of infection as well (Chang et al. 2009). The *igr* locus encodes six genes that are upregulated during growth in macrophages, required in vivo during the acute phase of a mouse infection and predicted to be involved in cholesterol catabolism (Schnappinger et al. 2003, Rengarajan et al. 2005, Sassetti et al. 2003, Van der Geize et al. 2007, Chang et al. 2007). Interestingly, some of the *igr* products have been shown to genetically interact with Mce4 and like Mce4, these mutants cannot grow on cholesterol as a primary carbon source (Joshi et al. 2006, Chang et al. 2009). In addition, this in vitro attenuation cannot be restored by the addition of a supplemental carbon source such as glycerol,

suggesting that a toxic intermediate of cholesterol catabolism accumulated in *igr* mutants resulting in their subsequent attenuation (Chang et al. 2009). Moreover, *Mce4* and *igr* double knockouts display an almost complete restoration of growth in acutely infected mice, indicating that blocking cholesterol uptake inhibited the accumulation of the toxic intermediate (Chang et al. 2009). This important finding demonstrates that *Mtb* metabolizes cholesterol throughout the course of infection, but following IFN $\gamma$  activation carbon sources including cholesterol become limiting and/or required for survival.

Although cholesterol is clearly an important nutrient in vivo, we knew relatively very little about how *Mtb* utilized this carbon source and the physiological adaptations required for metabolizing it. Specifically, I set out to accomplish the following goals:

- I. Identify the entire set of bacterial genes required for cholesterol utilization.
- II. Further elucidate cholesterol catabolism in *Mtb*.
- III. Understand the metabolic adaptations of using cholesterol as a carbon source.
- IV. Using the *Mce* systems as a model, understand lipid import in *Mtb*.

It was obvious that the development of a high-throughput assay was necessary to accomplish these goals in a timely and efficient fashion. By combining recent advances in next-generation sequencing technology with existing transposon mutagenesis libraries, I was able to make significant contributions to our understanding of cholesterol utilization in Mtb.



## **Chapter II. Whole genome phenotypic profiling identifies essential genes in Mtb**

### **Acknowledgements**

This chapter is excerpted from *High-resolution phenotypic profiling defines genes essential for mycobacterial survival and cholesterol catabolism*, which is currently under peer review at PLoS Pathogens. In addition to writing the manuscript, I conducted all molecular biology and genetic experiments. Data analysis was conducted in collaboration with Jeffrey D. Gawronski, Michael A. DeJesus and Tom R. Ioerger. In addition, *flpA* sub-genic data contributed to, *A Ser/Thr phosphosignaling system controls cell-wall synthesis in mycobacteria* which is currently under peer review at PLoS Biology. Kadamba Papavinasasundaram performed all homologous recombination efforts with *flpA*.

### **Summary**

The pathways that comprise cellular metabolism are highly interconnected, and alterations in individual enzymes can have far-reaching effects. As a result, global profiling methods that measure gene expression are of limited value in predicting how the loss of an individual function will affect the cell. In this chapter, we employed a new method of global phenotypic profiling to directly define the genes required for the growth of Mtb. A combination of high-density mutagenesis and next-generation sequencing was used to characterize the composition of complex mutant libraries exposed to different conditions. This

allowed for the unambiguous identification of the genes that are essential for Mtb to grow in vitro and proved to be a significant improvement over previous microarray approaches. Specifically, these improvements included increased resolution and dynamic range. Increased resolution allowed for the identification of essential genes containing transposon insertions, as well as essential regions within genes that correspond to essential functional domains. In addition, the increased dynamic range allowed for the differentiation between genes required for growth and optimal growth in Mtb. Overall the predictions made by both the prior microarray and current next-generation sequencing methods are functionally similar, but the differences highlight the technical advantages of the method described herein.

## **Introduction**

Mtb is an extremely successful pathogen. This bacterium must adapt to a number of different microenvironments over the course of a chronic infection (Russell et al. 2010), and understanding the physiological state of the pathogen in these sites is central to the design of more effective antitubercular drugs. Defining these metabolic adaptations requires a holistic understanding of the cellular pathways that are critical for survival in each specific environment. While whole genome methods to profile mRNA, protein, or transcription factor binding are valuable for understanding cellular responses, gene regulation has proven to be a poor predictor of essentiality (Rengarajan et al. 2005, Badarinarayana et al. 2001).

Therefore, genome-wide approaches to accurately and directly assess the relative contribution of each gene to the growth and survival of *M. tuberculosis* are needed.

The genome-scale methods initially used to predict gene essentiality in bacteria employed either microarray hybridization (Sasseti et al. 2001, Sasseti et al. 2003,) or conventional sequencing (Lamichhane et al. 2003, Jacobs et al. 2003, Hutchison III et al. 1999) to map the sites of transposon insertions in large libraries of random mutants and identify regions of the chromosome that were unable to sustain mutation. Both of these methods suffer from significant limitations. Microarrays are not able to precisely map insertions, resulting in ambiguity regarding the precise location of an essential region. While conventional sequencing can surmount this problem, it is both cumbersome and expensive to apply this method to the very complex libraries necessary for the unambiguous identification of essential genes. Finally, neither of these approaches can be used to quantitatively compare the compositions of mutant pools over more than a very small dynamic range.

By employing next-generation sequencing, specifically Illumina sequencing technology, these deficiencies were overcome. Illumina sequencing technology allows for massively parallel sequencing of complex samples by using novel reversible terminator sequencing chemistry (Ansorge 2009, Turcatti et al. 2008). Briefly, libraries of adapter flanking DNA fragments are immobilized on a glass flow cell. Bridge amplification then creates clusters or colonies, of approximately 1000

identical sequences of a single library-derived fragment. Sequencing by synthesis then occurs by combining four reversible terminator nucleotides each labeled with a unique fluorescent dye with a polymerase and a charge-coupled device (CCD) camera. By sequencing clusters of identical molecules the fluorescence intensities are increased over other methods. In addition, the ability to simultaneously sequence millions of sequences on the same flow cell drastically reduces the time required to sequence complex libraries.

In order to precisely define the genes that are important for the growth of Mtb, massively parallel Illumina sequencing technology was used to characterize transposon libraries. This approach achieves single base pair resolution of insertion sites in very complex libraries, resulting in the precise identification of essential open reading frames.

## **Materials and Methods**

### **Mtb growth and selection**

The transposon library was generated in the H37Rv background as previously described (cite Sassetti 2001), and consisted of approximately  $10^5$  independent insertion events.  $10^6$  cfu of library were inoculated into 200mls of minimal media (asparagine 0.5g/L,  $\text{KH}_2\text{PO}_4$  1.0g/L,  $\text{Na}_2\text{HPO}_4$  2.5g/L ferric ammonium citrate 50mg/L,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0.5g/L,  $\text{CaCl}_2$  0.5g/L,  $\text{ZnSO}_4$  0.1mg/L), 25ug/ml Kanamycin, 0.2% tyloxapol, 0.2% ethanol and either 0.1% glycerol or 0.01% cholesterol.

Selections were carried out in duplicate for glycerol and triplicate for cholesterol.

Libraries were grown in roller bottles at 37°C. Cultures were diluted as necessary to maintain the optical density below 0.2. The number of cell generations was monitored by cfu enumeration.

### **Genomic library preparation**

Genomic DNA was isolated from each library, partially digested with a combination of restriction enzymes,, and ligated to asymmetric adapters. Chromosome transposon junctions were amplified as previously described (Sasseti et al. 2001). An additional nested PCR was used to incorporate Illumina attachment and sequencing sites ( AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCTCTTC CGATCTCGGGGACTTATCAGCCAACC and CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCT CTTCCGATCTGTCCAGTCTCGCAGATGATAAGG). Standard PCR conditions (denaturation at 94°C for 30 seconds, annealing at 57.5°C for 30 seconds and amplification at 72°C for 30 seconds) were used and cycled for 9 cycles.

Amplified fragments between 250-400bp were purified and sequenced using either the primer: CCGGGGACTTATCAGCCAACC (complementary to the transposon inverted terminal repeat, ITR), or

AACTCTTTCCCTACACGACGCTCTTCCGATCT (complementary to the Illumina adapter) using an Illumina GA2 instrument.

### **Sequence analysis.**

The Illumina sequencing reads that contained the Himar1 ITR sequence and the adjacent TA insertion site were identified in the raw fasta files and trimmed of the ITR sequence. The sequences were aligned to the *M. tuberculosis* H37Rv reference genome (Cole et al.1998) using SOAPv1.11 alignment software (Li et al. 2008) at default settings (2 mismatches allowed per read). A custom PERL script was used to extract the TA dinucleotide insertion site coordinates from the SOAP output file (SI Computer Script). For reads aligning to the plus strand of the genome, the genome coordinate at position 1 of the trimmed read was determined. For reads aligning to the minus strand, the genome coordinate at position 2 of the read was calculated to represent the TA coordinate position with respect to the plus strand. For each TA insertion site detected by alignment, the total number of reads and the strand orientation was determined. Sequence reads that aligned to more than one chromosomal position were randomly assigned to one of the positions. If less than 10% of the reads corresponding to an insertion site could be assigned to this single position, the TA site was removed from all further analyses. Insertion site coordinates were mapped to positions within protein coding genes annotated in protein table RefSeq file NC\_000962.ptt (from the National Center for Biotechnology Information: <ftp://ftp.ncbi.nih.gov/>). For each gene, the number of insertion sites identified and the total number of sequencing reads in the gene were determined.

### **Essential Gene Analysis**

Genes were scored for essentiality based on the length of the longest contiguous run of TA dinucleotide sites lacking observed insertions within the coding region. Read counts at each TA site were pooled from both strands and summed over all the sequencing runs for each library/growth condition (3 runs for growth on glycerol, 2 runs for growth on cholesterol), and sites with 5 or fewer observed reads were treated as non-insertions. The longest consecutive sequence of non-insertion sites  $k$  in each gene was identified. The expectation for the length of the longest run of TA sites lacking insertion in a gene with  $n$  TA sites is given by the following statistics for a sequence of Bernoulli trials with success probability  $q$ :  $m = \log_{(1/q)}(n(1-q)) + g/\ln(1/q)$ ,  $s = [p^2/(6 \ln^2(1/q) + 1/12)]^{-1/2}$ , where  $g = 0.577...$  is the Euler-Mascheroni constant (Schilling, 1990). In this case,  $q$  represents the probability of insertions being observed at TA sites in non-essential genes, which was estimated based on the total number of TA sites across the genome containing insertions divided by the total number of TA sites (excluding those in genes completely devoid of insertions between the 5' 5% and 3' 20% ends of the ORF) in each dataset ( $q = 37,849/63,307 = 0.598$  for glycerol;  $q = 35,902/62,888 = 0.571$  for cholesterol). To assess the statistical significance of the observed maximum run of sites without insertions, this was compared to the expected length of the longest run using the cumulative function of the Extreme Value Distribution to calculate a p-value:  $p(k;n,m,s) = 1 - \exp(-\exp(-(k-m)/s))$ , where  $m$  and  $s$  are the statistics of the expected distribution given above.

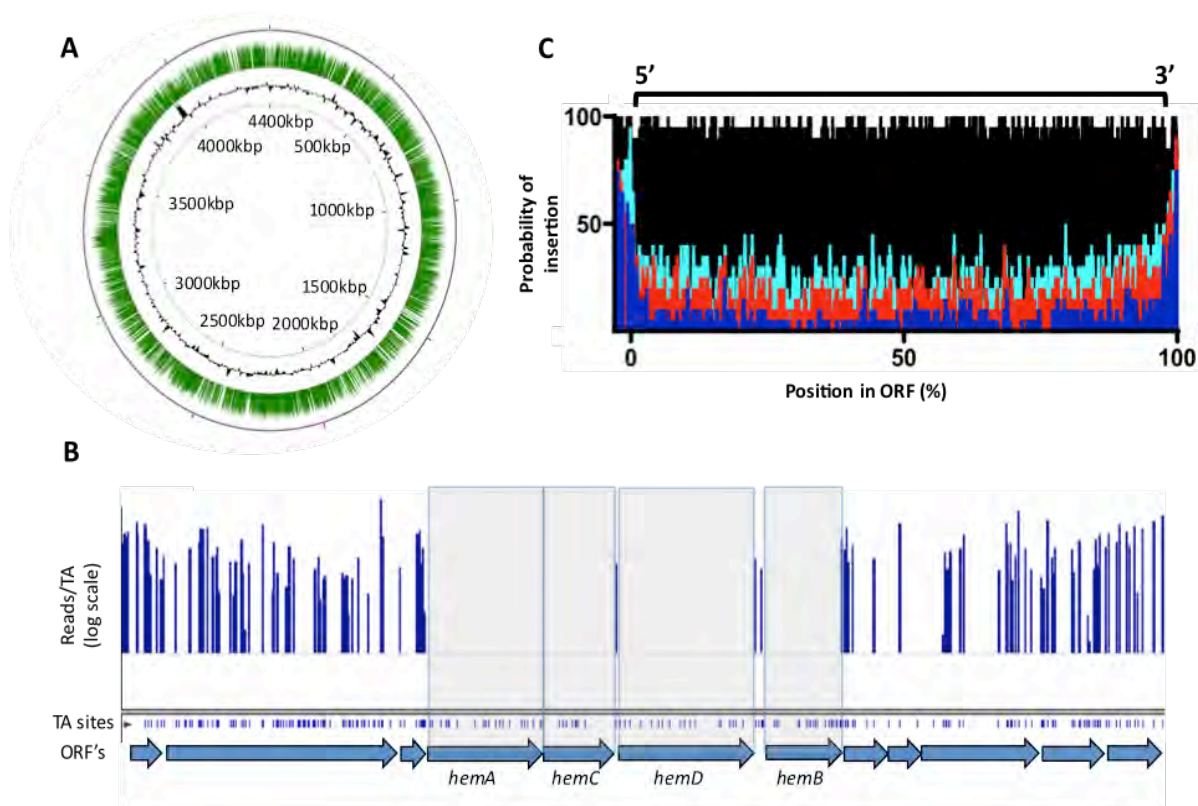
## Results

### Transposon mapping and identification of genes required for Mtb growth in vitro

A library of transposon mutants consisting of approximately 105 independent insertion events was generated in the H37Rv strain of Mtb using a modified himar1 based transposon (Sasseti et al. 2001). Replicate portions of this library were expanded for 12 generations in defined media containing different carbon sources, and chromosomal DNA was extracted from each pool. This DNA was randomly fragmented, ligated to asymmetric adapters, and transposon chromosome junctions were amplified using PCR. Illumina sequencing was then used to determine the sequences of 6 to 8 million transposon:chromosome junctions per library. 95-99% of these sequences represented the specific amplification of transposon insertions (see methods) and were used for subsequent analyses.

The transposon used in these studies requires a TA dinucleotide insertion site. We identified transposon insertions at 44,350 of the 74,605 possible TA sites in the genome (Supplemental Table 2.1). This corresponded to approximately one insertion every 100 base pairs on average and was consistent with the previously described random insertion of this element (Lamichhane et al. 2003, Rubin et al. 1999). The identified insertions were distributed throughout the chromosome (Fig. 2.1A), with the exception of small gaps, many of which corresponded to known essential genes in which we expected insertions to be lethal. For example, a small





**Figure 2.1 Transposon insertions are randomly distributed, and gaps in transposon coverage correspond to predicted essential ORFs.** (A) Transposon:chromosome junctions from high density mutant libraries were identified by deep sequencing. The number of sequence reads corresponding to each insertion site is represented as green bars mapped onto the circular chromosome of *M. tuberculosis* H37Rv. Black contour represents the GC content of the Mtb chromosome (G+C content greater or less than 50% are represented as contours outside or inside the ring, respectively). (B) Essential heme biosynthetic genes are devoid of insertions. The number of sequence reads (“reads/TA”) is shown for the indicated region of the H37Rv chromosome. Potential TA dinucleotide insertion sites are indicated. (C) Each potential TA insertion site in the genome was assigned a position relative to the nearest ORF. The probability of detecting an insertion at each ORF position (per 20 TA window) is plotted on the y-axis. Genes predicted to be nonessential ( $p=0.9-1.0$ , shown in black) show no position-dependent insertional bias. In contrast, genes predicted to be required for growth with different degrees of confidence (cyan:  $p<0.2$ , red:  $p<0.05$ , blue:  $p<0.0005$ ) show a strong bias against insertions in the open reading frame. The insertion density at the extreme ends of these ORFs approximates that of the nonessential genes.

region of the chromosome encodes several of the enzymes required for the synthesis of the essential cofactor, heme, (Figure 2.1B) and we found that gaps in transposon coverage precisely mapped to these open reading frames (ORFs).

In contrast, we found that other likely essential genes could harbor a small number of insertions (not shown), consistent with previous observations (Akerley et al. 2002). Therefore, we globally defined essential ORFs by searching for genes with statistically significant gaps in transposon insertion coverage, instead of simply compiling those that were absolutely devoid of insertions. To do this, we identified the longest sequence of TA sites lacking insertions within a given gene, and determined the likelihood of a non-essential gene harboring a gap of that length (modeled as a Bernoulli process) as a function of the total number of TA sites in the ORF. We then calculated the probability of the longest observed gap arising by chance, relative to the Extreme Value Distribution (Supplemental Table 2.2). This analysis method exploited the value of the high-resolution data provided by deep sequencing, while allowing for permissive insertion sites in essential genes.

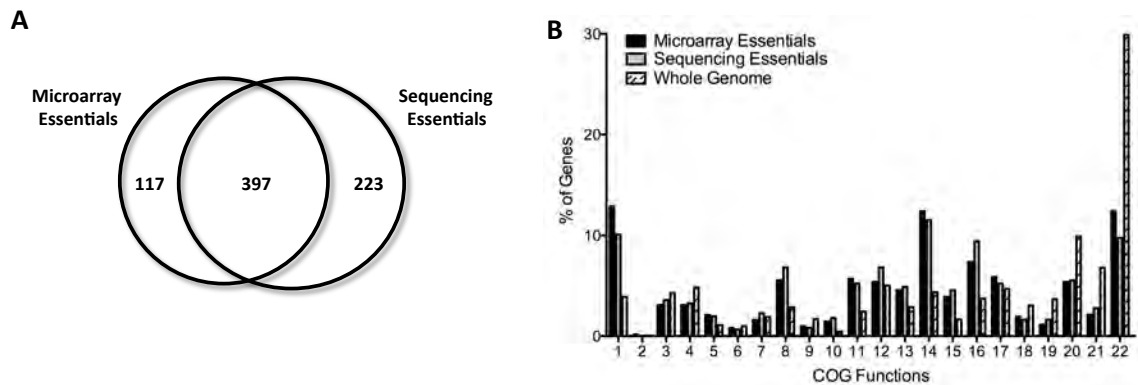
While the *himar1* transposon used in these studies has been shown to integrate relatively randomly (Lamichhane et al. 2003, Rubin et al. 1999), it remained possible that some regions of the chromosome were resistant to transposition, and genes in these regions would erroneously appear to be required for viability. The use of a high-resolution sequencing method allowed us to address this possibility

for the first time by determining if the identified gaps in transposon coverage corresponded to protein-coding regions. As shown in Figure 2.1C, the probability of identifying a transposon insertion in or near a putative essential region is highly correlated with the position of the corresponding predicted ORFs. In addition, the probability of identifying permissive sites within essential genes is higher at the extreme 3' end, likely because these insertions do not fully disrupt gene function. Together, these data indicate that regions devoid of sequenced transposon insertions represent protein-coding genes that are important for the growth or survival of the organism and not regions that are less accessible to transposition.

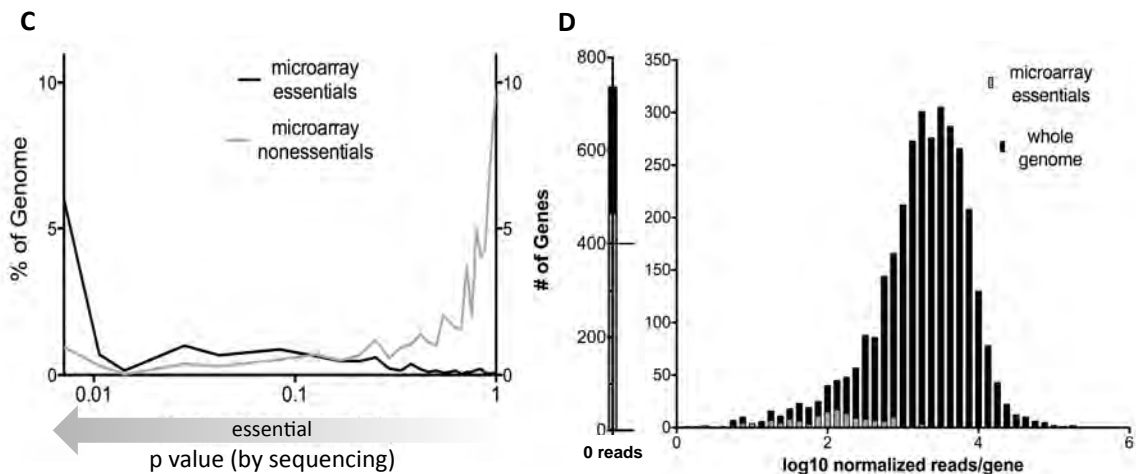
### **Comparison with previous studies**

This deep sequencing-based approach for identifying growth-attenuated mutants both validates and improves upon our previous microarray-based studies. About 15% of the Mtb genome or 614 genes, were previously predicted to be required for optimal growth in vitro using microarray hybridization to map insertions sites (Sassetti et al. 2003). Using a p value cutoff of 0.05, our current deep sequencing approach predicts a similar number of 620 essential genes. These gene sets overlap to a large extent and encode a similar distribution of cellular functions, which are quite different from the genome as a whole (Fig. 2.2A, 2.2B).

Despite the gross similarities between the two gene sets, we also found differences (Fig. 2.2A, 2.2C). While some of these differences may be attributable to minor



**Figure 2.2 Defining essential genes with deep sequencing is more sensitive and precise than previous approaches.** (A) Genes predicted to be essential via microarray (left) are compared with genes predicted to be essential via deep sequencing (right). Our false discovery rate allowed us to make confident predictions regarding essentiality for genes with more than 7 TAs. Thus, genes with fewer TAs were removed from this analysis. (B) Microarray and deep sequencing identify essential genes sets consisting of similar predicted functions, which differ from the genome as a whole. COG functions associated with essential genes predicted by microarray (Sasseti et al. 2003) or deep sequencing are plotted along with the predictions for the entire genome. COG functions are: 1) Translation, ribosomal structure, biogenesis, 2) RNA processing, modification, 3) Transcription, 4) Replication, recombination, repair, 5) Cell cycle control, cell division, chromosome partitioning, 6) Defense mechanisms, 7) Signal transduction mechanisms, 8) Cell wall/membrane/envelope biogenesis, 9) Cell motility, 10) Intracellular trafficking, secretion, vesicular transport, 11) Posttranslational modification, protein turnover, chaperones, 12) Energy production, conversion, 13) Carbohydrate transport, metabolism, 14) Amino acid transport, metabolism, 15) Nucleotide transport, metabolism, 16) Coenzyme transport, metabolism, 17) Lipid transport, metabolism, 18) Inorganic ion transport, metabolism, 19) Secondary metabolites biosynthesis, transport, catabolism, 20) General function prediction only, 21) Function unknown, 22) No assignment.



**Figure 2.2 Defining essential genes with deep sequencing is more sensitive and precise than previous approaches.** (C) Microarray predictions are generally, but not absolutely, consistent with deep sequencing. The probability of essentiality defined via deep sequencing is plotted on the x-axis (as log p value) for genes previously predicted by microarray to be either essential (black) or nonessential (gray). (D) Mutations in nonessential genes previously predicted to be required for growth cause quantitative fitness defects. Normalized sequence reads per gene (log scale) are plotted for all genes (black) and genes predicted to be essential by microarray (gray). Previously defined essential genes with detectable insertions are associated with low sequence read counts, suggestive of quantitative underrepresentation in the library.

alterations in the growth media used in these studies, it is likely that most discrepancies are due to both the increased resolution and dynamic range of the current method. For example, approximately one half of the genes that deep sequencing identified as essential ( $p < .05$ ) could sustain a small number of insertions (Fig. 2.1C and Supplemental Table 2.2). The presence of these permissive insertion sites likely caused many of these genes to be deemed nonessential by lower resolution methods.

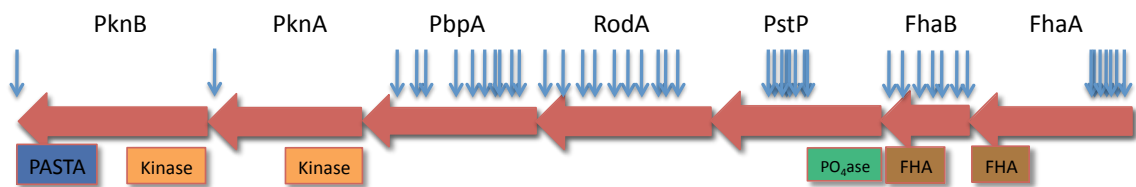
The second major difference between these datasets is a result of the sequencing depth that was employed. Because of the limited dynamic range of the microarray method (typically 3-10 fold), it was difficult to differentiate between mutants that were truly nonviable and those that were merely significantly underrepresented. The increased dynamic range of deep sequencing-based mapping allowed for this differentiation. Indeed, the insertions we identified in genes previously predicted to be essential were associated with a significantly lower number of sequence reads than the genome as a whole (Fig. 2.2D). This suggested that these mutants suffered from a fitness defect but remained viable. Based on the average number of sequence reads we detected in nonessential regions (173), we estimate that genes containing statistically significant gaps in coverage correspond to mutants that are more than 100 fold underrepresented in the pool, and are therefore essentially nonviable.

### **Investigating essentiality at the subgenic level**

The increased resolution of Illumina sequencing also allowed the identification of essential regions at the subgenic level. Microarray hybridization only probed one segment of each ORF. In contrast, by direct sequencing we individually probed the effects of insertions every 100 bp on average, and found several examples of differential phenotypes caused by mutations in the same gene. For example, two of the eleven Ser/Thr kinases (Cole et al. 1998, Av-Gay and Everett 2000) of *Mtb*, PknA and PknB, are encoded in the same operon with the only known phospho-Ser/Thr phosphatase, PstP (Figure 2.3). All three of these genes have been predicted or demonstrated to be essential for viability. This high-resolution data verifies that both of the kinases are required for in vitro growth, and that neither the intracellular kinase domain nor the extracellular sensor domains of PknB can sustain insertions.

In contrast, only the region of the *pstP* gene encoding the predicted cytosolic phosphatase domain was devoid of transposon insertions, despite the presence of 11 TA sites. Unlike the phosphatase domain, mutations were tolerated throughout the regions of the *pstP* ORF encoding the C-terminal transmembrane and extracellular domains (Figure 2.3). The absence of transposon insertions in the phosphatase domain could not be explained by chance alone, as the probability of not detecting an insertion in 11 contiguous nonessential TAs is less than  $10^{-6}$ . Similarly but somewhat paradoxically, the *Rv0020c* gene was disrupted throughout

its 5' end but could not sustain insertions in the 3' domain that encodes a phosphothreonine-binding forkhead-associated domain. Thus, it appears that insertions disrupting distinct protein domains may have different effects on growth, and these phenotypes can be distinguished in the whole-genome dataset.



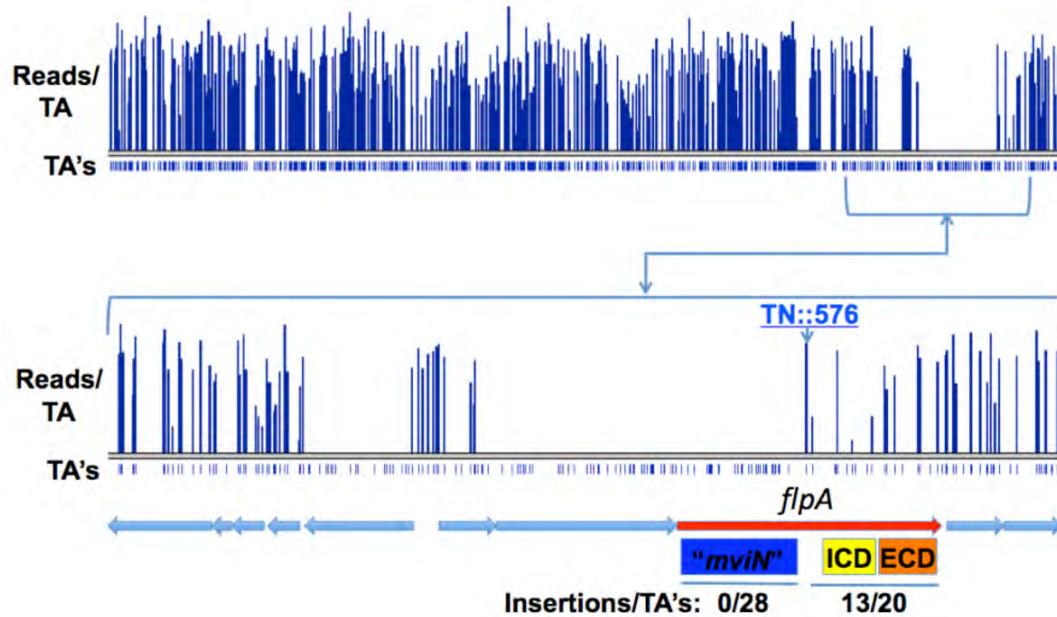
**Figure 2.3 Illumina sequencing predicts subgenic essentiality.**

Arrows indicate sequenced transposon insertions. While both *pknB* and *pknA* are completely devoid of sequenced insertions, *pstP* is devoid of insertions only throughout a predicted phosphatase domain indicating its functionally required.

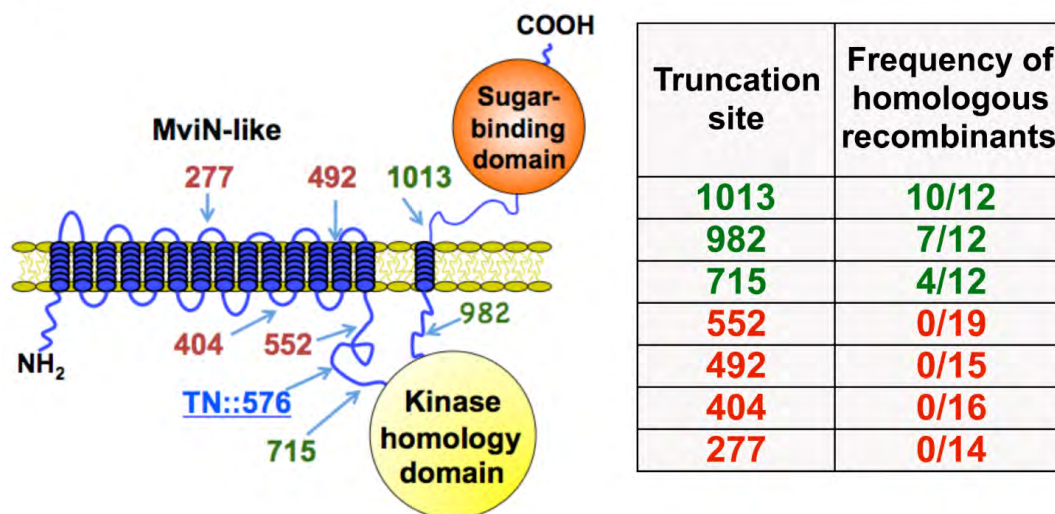
Similar phenotypes can be observed in a unique *Mtb* locus encoding a lipid II flippase protein, *flpA* (Gee et al. 2011). When the *flpA* locus was examined in detail, we found that transposon insertions are not tolerated in the 1728 basepair region homologous to flippase-like *MviN* genes or in the upstream gene, *Rv3909* (Figure 2.4). However, 3' to the *MviN*-like region, insertions were detected in 65% of the TA insertion, which is approximately the same density as the genome as a whole. This distribution suggested that the *MviN*-like domain is essential for viability, and the C-terminal domains are less important for *in vitro* growth. Gee and colleagues confirm this observation by probing this region by allelic exchange



mutagenesis, proving that differential phenotypes observed using whole-genome phenotypic profiling accurately predicts essential subgenic domains (Figure 2.5).



**Figure 2.4 Illumina sequencing identifies an essential MviN domain in FlpA.** Transposon insertion mapping in *Mtb* indicates that the MviN domain of FlpA is essential for viability. The 28 TA sites in the MviN domain were devoid of insertions, indicating that disruption of this domain produced nonviable mutants. The probability of this distribution occurring by chance in a gene of this size is  $<10^{-6}$ . Insertions in the regulatory domains of *flpA* were found at a similar density as the rest of the genome (65% of TA sites). "TN::576" indicates the amino acid position of the first detected transposon insertion.



**Figure 2.5 Homologous recombination validates subgenic essentiality.**

Homologous recombination confirms the differential essentiality of FlpA domains. Progressive deletions from the 3' end of the Mtb *flpA* gene were generated using the mycobacteriophage che9c recombination system. Deletions removing any transmembrane segments in the MviN-like module (red residue numbers) were not obtained. Deletions were isolated within the sensor kinase-like module (green residue numbers). The number of deletion mutants identified over the total number of *hyg*<sup>r</sup> colonies obtained following transformation with linear recombineering substrates is indicated.

## Discussion

In this work, we have significantly refined our understanding of the cellular functions necessary for the viability of Mtb. While others have used similar deep sequencing methods to map insertion sites in other organisms (Gawronski et al. 2009, Goodman et al. 2009, Langridge et al. 2009, van Opijnen et al. 2009, Gallagher et al. 2010) our new analytical tools allowed the statistically rigorous prediction of the genes essential for the viability of Mtb. The majority of these essential genes are consistent with those found by previous microarray-based methods. However, very significant differences in these predictions were also noted, and the majority of

these are attributable to technical and analytical refinements. As a result, this work provides a much more precise and statistically rigorous assessment of essentiality than previously possible.

The advantages of this method are apparent in the data generated. The ability to identify essential genes containing insertions allowed for the refinement of the definition of an essential gene. In addition, the increased resolution also allowed for the analysis of subgenic essentiality. The power of this method was demonstrated here in two *Mtb* loci. Larger scale analysis of this type would require more functional domains to be described and mapped, as well as non-coding RNAs. It also requires several assumptions regarding the transposon and expression of the downstream chromosome. For example, it is likely that the strong promoter controlling expression of the Kanamycin resistance cassette within the transposon is also driving transcription through the transposon:chromosome junction. Effective expression would also require a functional translational start site and possibly a ribosome binding site after the end of the transposon, which contains 2 translational stop sites. In addition, it certainly requires an absence of polarity resulting from transposition events. Our data and others' supports this claim (Sasseti et al. 2001). In particular, insertions upstream of essential genes, *pknA* and *pknB* are tolerated indicating the absence of polar effects associated with this transposon.

As with any whole-genome profiling method, this method is bound by its limitations. While Illumina sequencing is becoming more routine and less costly than traditional Sanger sequencing, it remains out of reach for some. This cost can be reduced by either engineering unique transposons or modifying amplified libraries with barcodes, or signature sequences, that identify them from other libraries. Either of these modifications would allow multiple libraries to be run in the same lane of a flow cell, significantly decreasing the cost associated with high-throughput sequencing. In addition to cost, bias can be introduced during PCR amplification of output libraries. Current efforts are underway to implement emulsion based PCR to reduce this bias. Similar to other next-generation sequencing techniques, emulsion PCR allows the amplification of a single template in an oil droplet within a liquid phase. This decreases bias, as each template is not in competition with other templates for nucleotides, enzyme or cofactors.

It is likely that as high-throughput sequencing technologies continue to emerge and improve, further applications of this method will arise. The increased dynamic range of the current method not only allowed for the identification of mutants with reduced fitness *in vitro*, but also the identification of mutants required in various conditions. By using the abundance of each mutant as an indicator for relative fitness, the next chapter focuses on using this method to identify genes required for growth in cholesterol, a critical carbon source *in vivo*.

### **Chapter III: Whole genome phenotypic profiling identifies the entire set of genes required for utilizing cholesterol**

#### **Acknowledgements**

This chapter is excerpted from *Quantitative high-resolution transposon mapping defines genes essential for mycobacterial survival and cholesterol catabolism*, which is currently under peer review at PLoS Pathogens. In addition to writing the manuscript, I conducted all molecular biology and genetic experiments. Data analysis was conducted in collaboration with Jeffrey D. Gawronski.

#### **Summary**

To explore functions that are required for cholesterol utilization, a combination of high-density mutagenesis and Illumina sequencing was used to characterize the composition of complex mutant libraries grown on different carbon sources, including cholesterol. The cellular pathways specifically required for cholesterol utilization were then defined. Few of the genes we identified had previously been implicated in this adaptation by transcriptional profiling, and only a fraction was encoded in the chromosomal region known to encode sterol catabolic functions. These genes could be assigned to distinct steps in cholesterol utilization, and comprise an unexpectedly large percentage of genes previously shown to be required for bacterial growth in mouse tissue. The nutritional adaptation to cholesterol by Mtb accounts for a significant fraction of the adaption to the host.

This work provides the first comprehensive characterization of a sterol catabolic pathway and suggests putative roles for uncharacterized virulence genes.

## **Introduction**

Despite work from our lab and others identifying cholesterol as an important carbon source during chronic Mtb infection (Pandey and Sassetti 2008, Van der Geize et al. 2007), the cholesterol catabolic pathway remained to be clearly defined. Much of the Cho region was predicted to encode enzymes with functions in cholesterol catabolism and yet, only a handful of those had been demonstrated in vitro to catalyze their predicted reactions (Yang et al. 2007, Capyk et al. 2009, Dresen et al. 2010, Yam et al. 2009, Lack et al. 2010, Ouellet et al. 2010, Nesbitt et al. 2010). Moreover, of the genes encoding these enzymes, few had demonstrated phenotypes either in vitro or in vivo (Nesbitt et al. 2010, Yam et al. 2009, Chang et al. 2007, Hu et al. 2010).

Cholesterol catabolism is better defined in the related actinomycete, *Rhodococcus* strain RHA1, where it occurs in two simultaneous phases;  $\beta$ -oxidation of the side chain and sterol ring cleavage and opening of the steroid nucleus (Van der Geize et al. 2007). It is currently unclear which pathway begins cholesterol catabolism in Mtb, though it is likely started with the oxidation of cholesterol to cholest-4-en-3-one, a reaction mediated in other organisms by both cholesterol oxidases and dehydrogenases. Mtb encodes both a cholesterol oxidase, ChoD, and a  $3\beta$ -

hydroxysteroid dehydrogenase, Rv1106c, and both have been predicted to be responsible for this obligatory first step (Brzostek et al. 2007, Yang et al. 2007).

The cytochrome P450, Cyp125, initiates side-chain degradation in RHA1 by mediating the oxidation of carbon at position 26 to the subsequent carboxylic acid (Rosloniec et al. 2009). Paradoxically, the orthologous Cyp125 is required for growth in cholesterol for *M. bovis* and Mtb strain CDC1551, but not for the Mtb lab strain H37Rv (Capyk et al. 2009, Ouellet et al. 2010, Johnston et al. 2010). However, Johnston and colleagues have recently shown that Cyp125 and an additional cytochrome P450, Cyp142 are functionally redundant. It remained unclear which steroid monooxygenase was required for cholesterol catabolism in Mtb H37Rv.

The depth of sequencing provided by Illumina technology allowed for the relative abundance of individual mutants to be accurately assessed, enabling the quantitative comparison of libraries grown under different conditions. Using this approach, genes required for growth in cholesterol were defined as genes that were significantly underrepresented in the cholesterol pool relative to the glycerol pool. This analysis identified genes involved in both cholesterol catabolism and the metabolism of the resulting metabolites.

In addition to identifying the cholesterol catabolic pathway in Mtb, this chapter serves as an introduction into the metabolic adaptations to growth on cholesterol, which are further investigated in chapter IV.

## **Materials and Methods**

### **Mtb growth and selection**

The transposon library was generated in the H37Rv background as previously described (cite Sassetti 2001), and consisted of approximately  $10^5$  independent insertion events.  $10^6$  cfu of library were inoculated into 200mls of minimal media (asparagine 0.5g/L,  $\text{KH}_2\text{PO}_4$  1.0g/L,  $\text{Na}_2\text{HPO}_4$  2.5g/L ferric ammonium citrate 50mg/L,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0.5g/L,  $\text{CaCl}_2$  0.5g/L,  $\text{ZnSO}_4$  0.1mg/L), 25ug/ml Kanamycin, 0.2% tyloxapol, 0.2% ethanol and either 0.1% glycerol or 0.01% cholesterol. Selections were carried out in duplicate for glycerol and triplicate for cholesterol. Libraries were grown in roller bottles at 37C. Cultures were diluted as necessary to maintain the optical density below 0.2. The number of cell generations was monitored by cfu enumeration.

### **Genomic library preparation**

Genomic DNA was isolated from each library, partially digested with a combination of restriction enzymes,, and ligated to asymmetric adapters. Chromosome transposon junctions were amplified as previously described (Sassetti et al. 2001). An additional nested PCR was used to incorporate Illumina attachment and sequencing sites (



AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC  
CGATCTCGGGGACTTATCAGCCAACC and  
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCT  
CTTCCGATCTGTCCAGTCTCGCAGATGATAAGG). Standard PCR conditions  
(denaturation at 94°C for 30 seconds, annealing at 57.5°C for 30 seconds and  
amplification at 72°C for 30 seconds) were used and cycled for 9 cycles.  
Amplified fragments between 250-400bp were purified and sequenced using  
either the primer: CCGGGGACTTATCAGCCAACC (complementary to the  
transposon inverted terminal repeat, ITR), or  
ACACTCTTTCCCTACACGACGCTCTTCCGATCT (complementary to the  
Illumina adapter) using an Illumina GA2 instrument.

### **Sequence analysis.**

The Illumina sequencing reads that contained the Himar1 ITR sequence and the adjacent TA insertion site were identified in the raw fasta files and trimmed of the ITR sequence. The sequences were aligned to the *M. tuberculosis* H37Rv reference genome (Cole et al.1998) using SOAPv1.11 alignment software (Li et al. 2008) at default settings (2 mismatches allowed per read). A custom PERL script was used to extract the TA dinucleotide insertion site coordinates from the SOAP output file (SI Computer Script). For reads aligning to the plus strand of the genome, the genome coordinate at position 1 of the trimmed read was determined. For reads aligning to the minus strand, the genome coordinate at position 2 of the read was calculated to represent the TA coordinate position with

respect to the plus strand. For each TA insertion site detected by alignment, the total number of reads and the strand orientation was determined. Sequence reads that aligned to more than one chromosomal position were randomly assigned to one of the positions. If less than 10% of the reads corresponding to an insertion site could be assigned to this single position, the TA site was removed from all further analyses. Insertion site coordinates were mapped to positions within protein coding genes annotated in protein table RefSeq file NC\_000962.ptt (from the National Center for Biotechnology Information: <ftp://ftp.ncbi.nih.gov/>). For each gene, the number of insertion sites identified and the total number of sequencing reads in the gene were determined.

### **Differential Growth Analysis**

Samples were normalized such that the mean number of sequence reads per insertion site was equal. The relative representation of each mutant was determined by calculating the fold change (sequence reads/insertion in cholesterol divided by sequence reads/insertion in glycerol) for each gene. Statistical significance was determined by T test. Data sets used for comparison consisted of each normalized sequence read value within a given gene (each insertion site in each sample was treated as a separate data point). The hyperbola used for defining genes specifically required for growth in cholesterol was defined by the formula,  $y=3.8/x+0.7$ . Genes above this line are annotated as required for growth on cholesterol in Table 3.1.

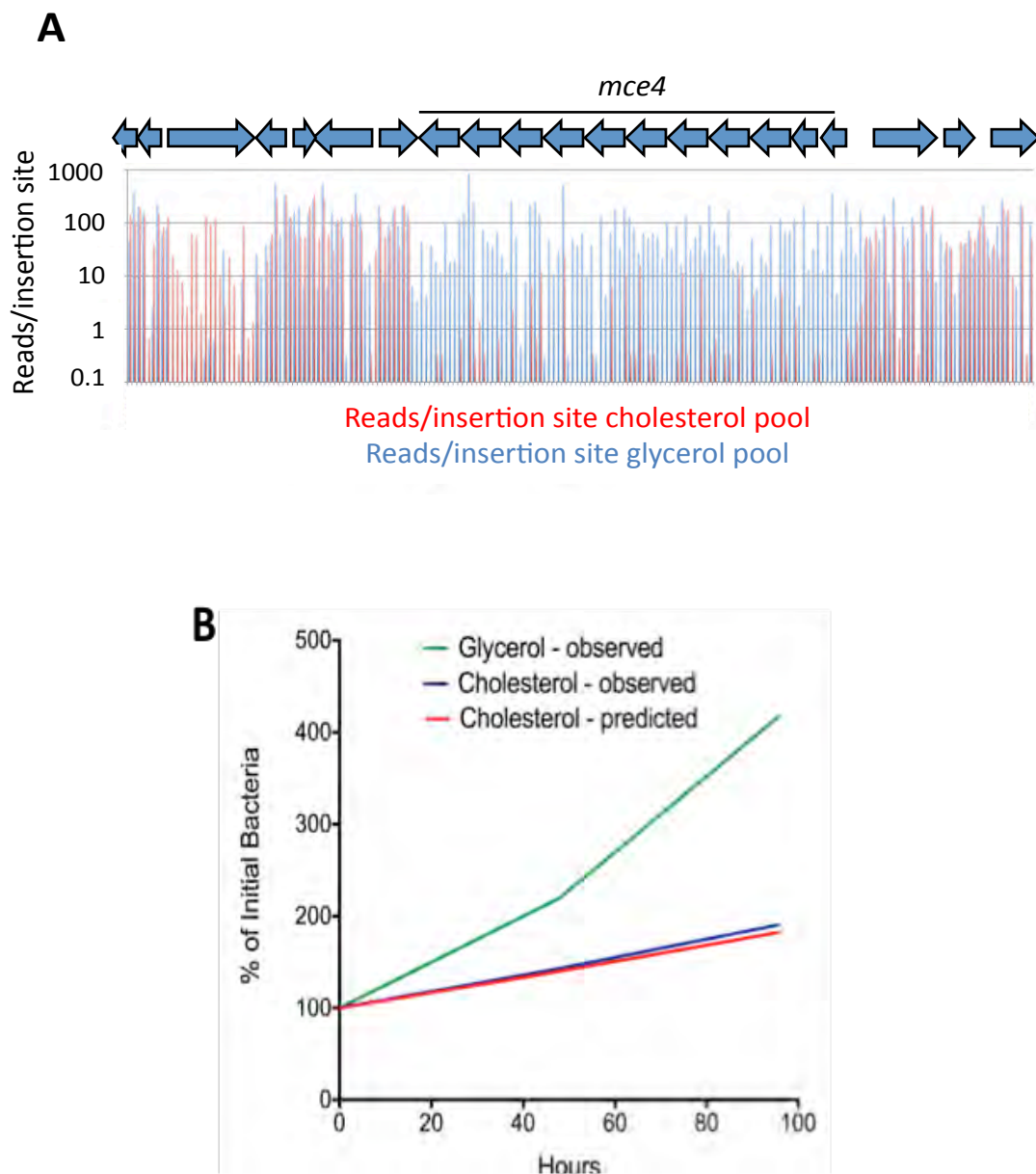
## Results

### Genes specifically required for growth on cholesterol

Illumina sequencing provided a rapid, accurate, and quantifiable assessment of the composition of mutant libraries, indicating that the quantitative comparison of mutant pools selected under different conditions should be possible. We compared pools grown in glycerol, a standard in vitro carbon source, or cholesterol, a critical carbon source during infection (Pandey et al. 2008), in order to define the discrete genes and pathways required for the catabolism of cholesterol.

In principle, the number of sequence reads associated with a specific insertion should be proportional to the relative abundance of the corresponding mutant in the library. We predicted that the sequencing depth used in our study would allow the relative abundance of each mutant in different libraries to be compared over a 100-1000 fold dynamic range. To verify this prediction, we specifically analyzed the genes encoding the previously characterized cholesterol uptake system encoded by the ten-gene *mce4* operon. Disruption of Mce4 function by deleting an essential transmembrane component has been shown to cause specific defects in cholesterol uptake and growth in this carbon source (Pandey et al. 2008).

Consistent with these observations, we found that mutations in every gene of this operon, as well as the associated ATPase (Joshi et al. 2006), resulted in severely impaired growth in cholesterol relative to glycerol (Fig 3.1A). The degree of this selective underrepresentation predicted that these mutants suffered a 31% growth

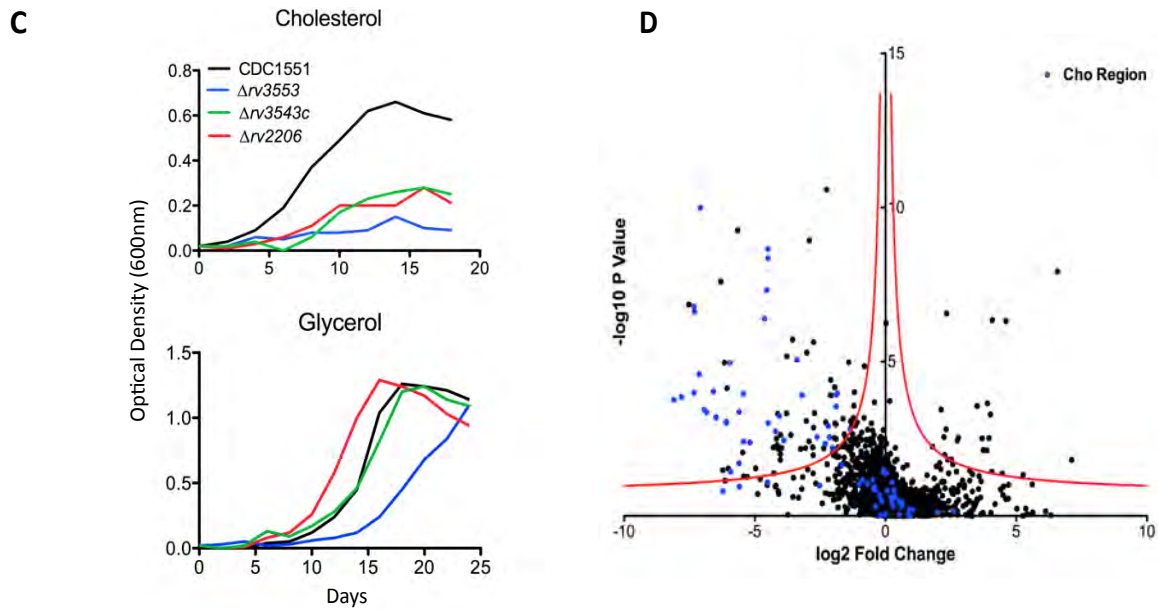


**Figure 3.1 Defining genes that are specifically required for growth on cholesterol.** (A) *Mce4* mutants are specifically underrepresented in the cholesterol-grown pool. Transposon libraries grown for 12 generations in media with either glycerol or cholesterol as a primary carbon source were compared by deep sequencing. Normalized sequence reads for individual insertions sites throughout the *Mce4* operon are shown following growth in glycerol (blue) and cholesterol (red). The average underrepresentation of *Mce4* mutants in the cholesterol-grown pool corresponds to a predicted 31% growth defect. (B) The number of sequence reads per TA provides an accurate estimate of relative growth rates. The experimentally determined growth curves of a *Mce4* deletion mutant in the indicated media are compared to the growth rate of *Mce4* mutants predicted in panel A. Log phase growth is plotted as percentage of initial bacterial number.

disadvantage per generation, which is very similar to the experimentally observed doubling time of an isolated Mce4 mutant in these media (Fig. 3.1B). In addition, growth curves for 3 additional transposon mutants were experimentally verified and found to correlate with their predicted phenotypes in each culture condition (Fig. 3.1C). Thus, the number of sequence reads associated with a given mutant could be used to accurately estimate relative growth rate.

Using the number of sequence reads per insertion site as an estimate of relative abundance, we compared the genome-wide data sets. As expected, we found that the representation of most mutants was similar in both pools (Fig. 3.1D). Due to the pre-selection of the library on complex glycerol-containing media before passage in single carbon sources, relatively few mutants were found to be specifically required for growth in glycerol. In contrast, a much larger number of mutants were underrepresented in the cholesterol-grown pool. To define differentially represented mutants, we employed a cutoff using a continuous function that equally weighted statistical significance and the magnitude of the change in representation (Fig. 3.1D). 96 genes met these criteria and were therefore predicted to be important for growth on cholesterol (Table 3.1 and Supplemental Table 3.1).

While the cholesterol catabolic pathway of Mtb has only been partially defined, all of the known and predicted catabolic enzymes are encoded in a distinct ~50kb region. This “Cho region” comprises over 80 genes and includes the mce4 operon (Van der



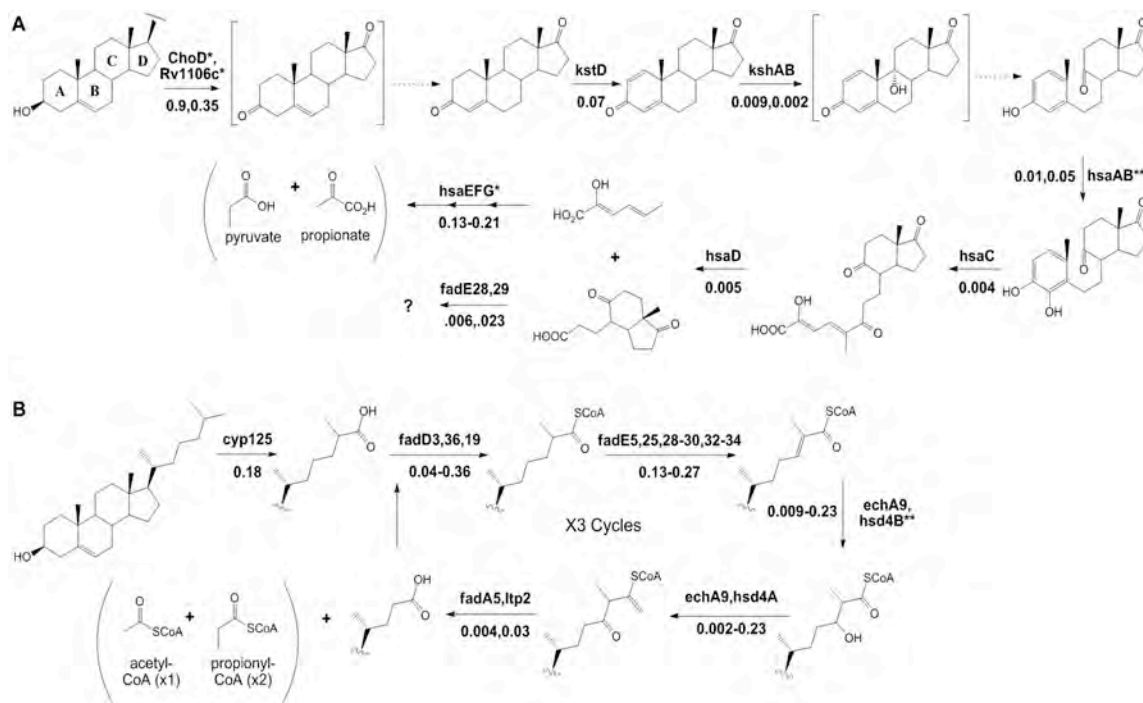
**Figure 3.1 Defining genes that are specifically required for growth on cholesterol.** (C) Mutants predicted to be required for cholesterol utilization display a specific growth defect in this media. Transposon insertion mutants were grown in minimal media with the indicated primary carbon sources and the rate of growth during log phase in each condition was plotted. (D) Identification of genes that are differentially required for growth. For each gene, the ratio of normalized sequence reads per insertion site (cholesterol pool/glycerol pool) is plotted on the x-axis ("fold change"). Y-axis represents the significance of each of these changes in representation (p value). A hyperbolic function was used to define genes that were differentially represented. The asymptotes of these curves are 0 for the log2 fold change and -0.07 for the log10 p value. Genes containing fewer than 2 TA sites were excluded from the analysis. Genes in blue represent those within the predicted cholesterol region.

Geize et al. 2007; Nesbitt et al. 2010). While the genes we found to be required for growth in cholesterol were enriched in this region, the majority (>60%) were distributed throughout the chromosome (Fig. 3.1D). These genes encoded a wide variety of predicted enzymatic activities that could be categorized into the following three functional groups associated with cholesterol utilization.

### **Sterol ring degradation**

Following import, cholesterol is degraded via  $\beta$ -oxidation of the side-chain, and ring cleavage to open the steroid nucleus (Fig. 3.2). Our phenotypic data supports and augments the current predicted pathway for sterol ring degradation. The first step in cholesterol catabolism requires the oxidation of the 3 $\beta$ -hydroxyl group and isomerization of the resulting cholest-5-en-3-one to cholest-4-en-3-one. Two distinct enzymes have been suggested to be involved in this transformation; a hydroxysteroid dehydrogenase, Rv1106c, and a cholesterol oxidase, ChoD (Yang et al. 2007, Brzostek et al. 2007). We found that only mutations in the ketosteroid dehydrogenase gene caused a significant growth defect, whereas choD mutants appeared to grow at a similar rate in both carbon sources (Fig. 3.2A, and Supplemental Table 3.1).

While all of the subsequent predicted steps of sterol ring degradation were critical for growth in cholesterol, the degree of importance varied depending on the position of the enzyme in the pathway. We found that hsaEFG mutants, which are predicted



**Figure 3.2 Phenotypic profiling predicts the cholesterol catabolic pathway.**

Proposed chemical transformations necessary for cholesterol degradation are divided into sterol ring cleavage and opening (A) and side-chain degradation (B). Enzymes known to or predicted to function at each step are noted with their respective fold change (sequence reads in cholesterol/sequence reads in glycerol) below. Dotted lines represent spontaneous reactions. We did not find enzymes with marked \* to be significantly required for growth on cholesterol. Enzymes marked with \*\* indicate those where we are unable to report significance due to insufficient data. The specific  $\beta$ -oxidation steps mediated by fadE28 and fadE29 are difficult to predict based on their homology to testosterone degrading enzymes and therefore, we have assigned them to both equally likely steps.



to be unable to further degrade 2-hydroxyhexa-2,4-dieneoic acid (HHD) to propionate and pyruvate, were only modestly defective in cholesterol media, in contrast to the drastic growth attenuation of mutants lacking functions at earlier steps. This is likely because HsaEFG act distal to a branchpoint in the pathway and the mutant bacteria could grow, albeit slowly, by fully catabolizing the portion of the molecule containing rings C and D.

### **Side-chain degradation**

In the related actinomycete, *Rhodococcus jostii* RHA1, the initial hydroxylation of C26 during side-chain degradation is mediated by the cytochrome P450 encoded by the *cyp125* gene (Rosloniec et al. 2009). In *Mtb*, the Cyp125 ortholog serves a similar function (Capyk et al. 2009, Ouellet et al. 2010), although Cyp142 has been shown to serve a functionally redundant role (Johnston et al. 2010). Here, we identify Cyp125 as the sole monooxygenase significantly required for growth on cholesterol (Table 3.1 and Supplemental Table 3.1). However, the modest 5.5 fold underrepresentation of *cyp125* mutants in cholesterol supports the previously observed redundancy between cytochrome P450 enzymes.

Many enzymes resembling those predicted to be required for the  $\beta$ -oxidation of the cholesterol side-chain are encoded in the Cho region. Surprisingly, we found significant functional clustering even within this region. None of the predicted  $\beta$ -oxidation genes encoded within the first half of the Cho region were required for

**Table 3.1 Genes predicted to be specifically required for growth on cholesterol.**

<b>Locus</b>	<b>synonym</b>	<b>Ratio glycerol/cholesterol</b>	<b>p value</b>
Rv0009	ppiA	0.269	0.003
Rv0153c	ptbB	0.023	0.053
Rv0202c	mmpL11	0.122	0.001
Rv0244c	fadE5	0.133	0.000
Rv0362	mgtE	0.005	0.000
Rv0391	metZ	0.073	0.000
Rv0450c	mmpL4	0.210	0.000
Rv0485	-	0.020	0.000
Rv0495c	-	0.081	0.002
Rv0655	mkl	0.149	0.001
Rv0693	pqqE	0.062	0.002
Rv0694	lldD1	0.058	0.007
Rv0695	-	0.147	0.011
Rv0696	-	0.062	0.015
Rv0761c	adhB	0.051	0.006
Rv0805	-	0.105	0.007
Rv0876c	-	0.421	0.001
Rv1071c	echA9	0.347	0.003
Rv1084	-	0.428	0.001
Rv1096	-	0.120	0.006
Rv1129c	-	0.015	0.000
Rv1130	-	0.013	0.000
Rv1131	gltA1	0.056	0.005
Rv1183	mmpL10	0.149	0.000
Rv1193	fadD36	0.064	0.006
Rv1428c	-	0.124	0.000
Rv1432	-	0.023	0.009
Rv1608c	bcpB	0.145	0.006
Rv1626	-	0.014	0.000
Rv1627c	-	0.085	0.000
Rv1798	-	0.056	0.003
Rv1906c	-	0.062	0.010
Rv1919c	-	0.194	0.006
Rv1963c	mce3R	0.057	0.000
Rv2048c	pks12	0.152	0.000
Rv2118c	-	0.323	0.001
Rv2206	-	0.013	0.033
Rv2239c	-	0.015	0.029
Rv2416c	eis	0.396	0.001

Rv2462c	tig	0.260	0.000
Rv2506	-	0.233	0.000
Rv2668	-	0.205	0.001
Rv2681	-	0.571	0.000
Rv2684	arsA	0.073	0.000
Rv2710	sigB	0.132	0.004
Rv2799	-	0.262	0.001
Rv2914c	pknI	0.437	0.000
Rv2985	mutT1	0.334	0.002
Rv3050c	-	0.026	0.055
Rv3274c	fadE25	0.268	0.001
Rv3419c	gcp	0.032	0.029
Rv3421c	-	0.052	0.042
Rv3492c	-	0.027	0.004
Rv3493c	-	0.015	0.001
Rv3494c	mce4F	0.007	0.000
Rv3495c	lprN	0.045	0.001
Rv3496c	mce4D	0.043	0.000
Rv3497c	mce4C	0.045	0.000
Rv3498c	mce4B	0.060	0.001
Rv3499c	mce4A	0.044	0.000
Rv3500c	yrbE4B	0.016	0.000
Rv3501c	yrbE4A	0.009	0.000
Rv3502c	fabG	0.004	0.000
Rv3515c	fadD19	0.268	0.000
Rv3526	kshA	0.006	0.000
Rv3531c	-	0.279	0.000
Rv3534c	hsaF	0.225	0.003
Rv3536c	hsaE	0.202	0.005
Rv3537	kstD	0.109	0.000
Rv3540c	ltp2	0.015	0.001
Rv3542c	-	0.054	0.002
Rv3543c	fadE29	0.023	0.004
Rv3544c	fadE28	0.006	0.000
Rv3545c	cyp125	0.269	0.000
Rv3546	fadA5	0.040	0.000
Rv3548c	-	0.010	0.000
Rv3549c	-	0.004	0.000
Rv3551	-	0.007	0.000
Rv3553	-	0.021	0.000
Rv3559c	-	0.016	0.037
Rv3560c	fadE30	0.006	0.000
Rv3561	fadD3	0.218	0.001
Rv3563	fadE32	0.096	0.000

Rv3564	fadE33	0.021	0.020
Rv3568c	hsaC	0.011	0.001
Rv3569c	hsaD	0.023	0.000
Rv3570c	hsaA	0.067	0.004
Rv3571	kshB	0.008	0.000
Rv3572	-	0.392	0.002
Rv3573c	fadE34	0.245	0.006
Rv3575c	-	0.234	0.000
Rv3779	-	0.403	0.001
Rv3820c	papa2	0.376	0.000
Rv3824c	papa1	0.431	0.000
Rv3825c	pks2	0.400	0.000
Rv3911	sigM	0.258	0.003

---

"Ratio glycerol/cholesterol" = mean of sequence reads per detected insertion in cholesterol pools / mean of sequence reads per detected insertion in glycerol pools

"p value" = significant of the ratio measurement. Determined as described in materials and methods.

growth on cholesterol, except *hsd4A* (Supplementary Table 3.1). Instead, we found that almost all of those encoded in the second half were required (including *ltp2*, *fadE29*, *fadE28*, *fadA5*, *fadE30*, *FadE32*, *fadE33*, *fadE34*, and *hsd4B*). Notably, we also identified predicted  $\beta$ -oxidation genes encoded outside of the *Cho* region that had not yet been implicated in cholesterol degradation (including *fadE5*, *echA9*, *fadD36*, and *fadE25*) (Figure 3.2B). Together, these genes could perform the three rounds of  $\beta$ -oxidation predicted to be required for full catabolism of the C17 side-chain, although we cannot exclude the participation of other functionally-redundant enzymes, or the possibility that some of these genes participate in degradation of the steroid nucleus.

### **Intermediary metabolism**

Cholesterol catabolism is thought to result in the production of an unusual mixture of metabolites including the C2 unit acetyl CoA, the C3 unit propionyl CoA, and pyruvate (Pandey et al. 2008, Van der Geize et al. 2007, Yang et al. 2009). Under the conditions used in our study, incorporation of propionyl CoA into the TCA cycle of *Mtb* requires the activity of the methylcitrate cycle (MCC) (Gould et al. 2006). Consistent with these predictions, *prpD*, *prpC* mutants lacking the dedicated steps in the MCC were defective for growth in cholesterol (Supplemental Table 3.1). The isocitrate lyase gene product (*lcl*) functions both in the MCC and in a more traditional role in the anapleurotic glyoxylate bypass, by converting isocitrate to succinate and glyoxylate during growth on C2-generating substrates (Gould et al.

2006). We found *icl* mutants to be 125-fold underrepresented in the cholesterol-grown pool (Supplemental Table 3.1), verifying an important functional role. In addition, our data suggests that the expression of this multifunctional enzyme may be rate limiting during growth on cholesterol, as the mutation of *RamB*, a negative regulator of *icl* expression (Micklinghoff et al. 2009), resulted in the opposite effect (a nine-fold overrepresentation in the cholesterol-grown pool).

Growth on lipids as a primary carbon source requires gluconeogenesis, and different metabolites can be used to fuel this pathway. Following growth on cholesterol, we found that *ppdK* mutants were severely underrepresented (Supplemental Table 3.1). This gene encodes pyruvate phosphate dikinase, which can mediate the conversion of pyruvate to phosphoenolpyruvate (PEP), the first committed step of gluconeogenesis. Together, these observations define discrete pathways required for the assimilation of all three of the metabolites that most likely result from cholesterol catabolism. These include the MCC, the glyoxylate bypass and PpdK-mediated gluconeogenesis.

## **Discussion**

In addition to the qualitative analysis in chapter II, here quantitative analysis is used to compare mutant pools grown in different carbon sources in order to understand how *Mtb* metabolizes cholesterol. A number of recent studies have demonstrated that *Mtb* mutants lacking the capacity to acquire or degrade cholesterol are

defective for growth in animal models of TB (Chang et al. 2009, Hu et al. 2010, Nesbitt et al. 2007, Pandey et al. 2008, Yam et al. 2009). However, our ignorance of the pathways required for cholesterol metabolism limited our ability to assess the importance of this nutritional adaptation during infection. In order to define a discrete set of the predicted 3987 ORFs in *Mtb* that was required for growth on cholesterol, we applied a continuous function cutoff to our comparative data, which placed equal importance on fold change in representation and statistical significance. More traditional one-dimensional cutoffs would have excluded genes that were only moderately underrepresented but exceptionally significant, which our analysis includes. Nonetheless, in order to avoid false positive predictions, we set a relatively stringent cutoff, and it is likely that even more than the 96 identified genes contribute to growth in cholesterol.

In addition to the dedicated cholesterol catabolic functions, we found that the use of this compound as a source of both cellular energy and biosynthetic carbon requires a variety of central metabolic pathways. While some of these requirements might have been predicted based on the shift from a glycolytic substrate to one that relies heavily on  $\beta$ -oxidation, the precise pathways utilized appear specific to the unusual mixture of metabolites derived from cholesterol. For example, gluconeogenesis under these conditions appears to be initiated by the conversion of pyruvate to PEP via the action of pyruvate phosphate dikinase (PpdK), whereas this pathway relies exclusively on phosphoenolpyruvate carboxykinase (PckA) during growth on acetate

(Marrero et al. 2010). In other bacteria, PpdK-mediated gluconeogenesis is favored during growth in the presence of pyruvate (Benziman et al.1971; Østerås et al.1997). As pyruvate is produced both as a direct product of sterol catabolism and through the activity of the methylcitrate cycle, we speculate that the relative abundance of cholesterol-derived pyruvate favors the PpdK-mediated pathway. These observations indicate that different gluconeogenic pathways may be preferentially used by Mtb depending on the relative abundance of precursor metabolites.

Cholesterol acquisition is predominantly required for bacterial persistence during the chronic stage of Mtb infection in mice (Pandey et al. 2008), indicating that the requirements for many of these central metabolic pathways are likely to change as infection proceeds. Identifying the full complement of cellular functions necessary to utilize cholesterol has also revealed the scale of this nutritional adaptation during infection. When we compared these data to previous genome-wide screens for mutants attenuated in mouse models of infection (Sassetti et al. 2003) we found that a full ten percent of genes specifically required for bacterial growth in vivo are also required for the utilization of cholesterol in vitro (Supplemental Table 3.1). These genes encode both dedicated sterol catabolic functions, as well as important steps in central metabolism. Thus, while it is clear that Mtb must adapt to a variety of host-specific conditions to sustain a productive infection, our data suggest that a single nutritional change is responsible for a significant portion of this adaptation.



These cholesterol catabolic functions, in conjunction with the hundreds of other genes that we found to be essential for bacterial viability, both expand and refine the repertoire of targets for new TB therapies.

Whole genome profiling techniques have proven to be useful tools for understanding complex pathways, such as those required for cholesterol utilization. Most of these approaches rely on determining the relative transcript or protein abundance. However, these strategies make a major assumption – that critical genes are tightly regulated in response to metabolic changes. While this may be true in many cases, it is often not. In order to avoid this assumption, we have directly identified the genes required for growth, rather than those that are regulated. While every approach has its own inherent strengths and weaknesses, the phenotypic profiling strategy described here provides a powerful complement to other systems biology approaches used to understand bacterial physiology.

## **Chapter IV. Transcriptional and metabolic adaptations to intracellular growth on cholesterol.**

### **Acknowledgements**

This chapter comprises a paper currently in final stages of preparation. In addition to writing the manuscript, I performed all metabolomic profiling, the generation and all characterization of the Rv1129c mutant including, growth curves, QPCR and macrophage infection experiments. In addition, I characterized the in vitro growth phenotype of the prpDC mutant in cholesterol and glycerol. Other characterization of prpDC, icl and mce4 mutants, as well as all other macrophage experiments were carried out by Amit K. Pandey. In addition, TLC and mass spectrometry of sulpholipid-1 were performed by AKP in collaboration with Sarah Gillmore and Carolyn Bertozzi.

**Transcriptional and metabolic adaptation is required for cholesterol catabolism by *Mycobacterium tuberculosis* during intracellular growth.**

**Jennifer E. Griffin<sup>1</sup>, Amit K. Pandey<sup>1</sup>, Sarah Gillmore<sup>4</sup>, Valerie Mizrahi<sup>2</sup>, John D. McKinney<sup>3</sup>, Carolyn Bertozzi<sup>4,5</sup>, and Christopher M. Sassetti<sup>1,5</sup>**

<sup>1</sup>Department of Microbiology and Physiological Systems, University of Massachusetts, Worcester, MA

<sup>2</sup>Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa

<sup>3</sup>Rockefeller University, New York, NY

<sup>4</sup>University of California Berkeley, Berkeley, CA

<sup>5</sup>Howard Hughes Medical Institute

**Summary**

To understand the adaptation of *Mycobacterium tuberculosis* to the intracellular environment, we used comprehensive metabolite profiling to identify the metabolic pathways utilized during growth on cholesterol, a critical carbon source during infection. Metabolic alterations observed during cholesterol catabolism centered on propionyl-CoA and pyruvate pools, metabolites that are produced directly by the catabolism of cholesterol. Growth on cholesterol required the transcriptional induction of the propionyl CoA assimilating methylcitrate cycle

(MCC) enzymes, a function mediated by the Rv1129c protein. We show that both Rv1129c and the MCC enzymes were required for intracellular growth in macrophages, and the growth defect of MCC mutants was directly attributable to the degradation of host-derived cholesterol. Together, these observations define a coordinated transcriptional and metabolic adaptation that is required for scavenging carbon during intracellular growth.

## **Introduction**

The physiological state of intracellular pathogens is inextricably linked to the host cell in which they reside, and each pathogen is metabolically adapted to its specific niche. Nutrition is a primary factor driving this adaption, as each pathogen is forced to subsist on locally-available host-derived compounds. Intraphagosomal pathogens, like *Mycobacterium tuberculosis*, must acquire these nutrients while surrounded by a host-derived lipid membrane (Appelberg 2006). Perhaps because of this sequestration, *M. tuberculosis* is thought to primarily catabolize these surrounding host lipids as carbon sources when growing in this compartment (Russell et al. 2010).

The preferential use of lipid carbon sources by *M. tuberculosis in vivo* has been inferred from the transcriptional induction of genes encoding fatty acid degradation enzymes (Schnappinger et al. 2003), and from the requirement for specific pathways in central carbon metabolism, such as the glyoxylate cycle

(McKinney et al. 2000) and gluconeogenesis (Marrero et al. 2010), that are often essential for the utilization of nonglycolytic carbon sources by bacteria. The identity of specific host lipids used by *M. tuberculosis* remained elusive until a dedicated cholesterol uptake and catabolic pathway was identified in the bacterium (Van der Geize et al. 2007). The demonstration that these functions were necessary for the growth and survival of *M. tuberculosis* in chronically-infected mice (Pandey and Sasseti 2008) verified that this single component of host membranes represents an essential nutrient during infection.

However, it remained unclear whether cholesterol catabolism contributes significantly to the observed metabolic requirements for growth in the complex host environment. Indeed, while cholesterol is catabolized to some extent throughout infection (Chan et al. 1995), cholesterol uptake mutants grow normally during the early preimmune phase of murine tuberculosis and in the relatively unstimulated macrophages that characterize this stage of infection (Pandey and Sasseti 2008). These observations indicate that additional carbon sources must be available to the bacterium *in vivo*. Fatty acids, sugars, and amino acids may all be available. In addition, a variety of other environmental conditions encountered during infection, including altered O<sub>2</sub>, CO, NO, and H<sup>+</sup> concentrations (Chan et al. 1995, Kumar et al. 2008, Rohde et al. 2007, Via et al. 2008), could contribute to metabolic alterations. To begin to understand how nutrition contributes to the metabolic adaptation to the host, we sought to

understand how the use of cholesterol, the only specifically defined carbon source *in vivo*, influenced the physiology of the bacterium during infection.

To do this, we used comprehensive metabolite profiling to characterize the metabolic state of *M. tuberculosis* during growth on defined carbon sources including cholesterol. Utilization of this compound profoundly altered the abundance of a variety of primary metabolites, most notably the intermediates of a propionyl CoA catabolic pathway known as the methylcitrate cycle (MCC). The activity of this pathway appeared to be rate-limiting for growth under these conditions, as transcriptional induction of the dedicated enzymatic steps by a divergently-encoded regulatory protein was essential. While the three-carbon unit, propionyl CoA, can be derived from a variety of host components other than cholesterol, we demonstrate the requirement for this metabolic pathway *in vivo* is directly attributable to the utilization of this single host carbon source.

## **Materials and Methods**

### **Growth and maintenance of bacterial strains**

All bacteria were maintained in complete Middlebrook 7H9 medium containing 0.05% Tween 80 and albumin-dextrose-catalase supplement. Kanamycin, Hygromycin, and Zeocin were used at 25 ug/ml, 50 ug/ml and 25 ug/ml, respectively. For cholesterol experiments bacteria were grown in minimal media containing asparagine 0.5g/L,  $\text{KH}_2\text{PO}_4$  1.0g/L,  $\text{Na}_2\text{HPO}_4$  2.5g/L ferric ammonium

citrate 50mg/L,  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  0.5g/L,  $\text{CaCl}_2$  0.5g/L,  $\text{ZnSO}_4$  0.1mg/L), 0.2% tyloxapol, 0.2% ethanol and either 0.1% glycerol or 0.01% cholesterol. Where noted, Vitamin B<sub>12</sub> (cyanocobalamin; Sigma) was added to a final concentration of 10 ug/ml. Growth was monitored by optical density at 600nm.

### **Genetic manipulation of strains**

Marked Rv1129c was generated by “recombineering” essentially as described (van Kessel and Hatfull 2007), except the recombination functions were expressed from the pNIT::ET plasmid (GenBank no. GU459073). Briefly, phage che9c recombination genes were induced in *M. tuberculosis* strain H37Rv by overnight incubation with 10 $\mu$ M isovaleronitrile. Bacteria were then electroporated with a linear substrate containing 500bp of homology to both the upstream and downstream regions of *rv1129c* flanking a hygromycin-resistance cassette. The presence of both recombination junctions were verified by PCR and sequencing. The absence of the *rv1129c* open reading frame was also confirmed by PCR. Nucleotides 1253102-1254505 were deleted from the genome. A complemented strain was generated by transforming the mutant with pJEB402 (Guinn et al. 2004) containing full length Rv1129c sequence (nucleotides 1253074-1254564 in the H37Rv sequence) cloned between HindIII and KpnI sites. This plasmid integrates in single copy into the phage L5 *attB* site (Lee et al. 1991). The construction of *lcl1/2* mutants in *M. tuberculosis* strain Erdman was described in (Munoz-Elias and McKinney 2005). *prpDC* mutants

were generated in H37Rv or an  $\Delta mce4$  mutant of H37Rv (Joshi et al. 2006) essentially as described in (Munoz-Elias et al. 2006).

### **Metabolomic profiling**

Strains were grown in quadruplicate in 50mls of complete 7H9 medium until OD<sub>600</sub> of 0.5 was reached. Cultures were then washed in PBS twice and passed into minimal media with either 0.01% cholesterol or 0.1% glycerol for 48 hours at 37°C. Cells were harvested at 4000g for 5 minutes at 4°C and lysed by vortexing in 10 ml chloroform:methanol (2:1). Samples were transferred to glass vials and dried under a stream of nitrogen, and subjected to LC/MS and GC/MS analysis as described (Evans et al. 2009).

### **Macrophage infections**

Bone marrow-derived macrophages (BMMs) were isolated by culturing bone marrow cells from C57BL/6 mice in DMEM containing 10% FBS, 2 mM glutamine, 10% L929-conditioned medium, and 10 µg/ml ciprofloxacin for 5 days; ≈24 h before infection, differentiated BMMs were detached and seeded on a 24-well tissue culture plate at  $5 \times 10^5$  cells per well in the same medium lacking antibiotic. Macrophages were infected with different strains of *M. tuberculosis* at an moi of 1 for 4 h at 37°C. Extracellular bacteria were killed by incubating the cells with DMEM containing 100 µg/ml amikacin for 1 h followed by washing three times with warm PBS. Intracellular bacteria were quantified by lysing the cells with PBST at the indicated time points and plating dilutions on 7H10 agar.

### **Q-PCR**



Bacterial strains were grown in complete 7H9 medium until log phase, washed twice in PBS, and cultured in minimal medium with either 0.01% cholesterol or 0.1% glycerol for 48 hours at 37°C. Bacteria were centrifuged and resuspended in Trizol and lysed by bead beating. RNA was isolated according to the manufacturers directions and DNase treated (TURBO DNA-free Kit, Applied Biosystems). cDNA was generated using equal volumes of RNA from all samples (SMARTScribe Reverse Transcriptase, Clontech). QPCR was conducted using iQ SYBR Green Supermix from BioRad and the following PCR parameters; 95°C for 10 minutes, 40 cycles of 95°C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds, read plate, 76°C for 2 seconds, read plate, 78°C for 2 seconds, read plate, 80°C for 2 seconds, read plate, 81°C for 2 seconds, read plate then 72°C for 1 minute, followed by a standard melting curve from 60°C-95°C. PCR primers for prpD forward were TGACTTTCACGACACGTTTCTGGC, prpD reverse were TGTGGATCTCATAGGCGGTTACCA, prpC forward were ATGGCTGGACATCTACCAGGTGTT and prpC reverse were TGTCGAATCCCATCAGGTAGTACG. Relative abundance of message was calculated by comparing all amounts for each gene relative to a standard curve from the H37Rv strain in cholesterol for the corresponding gene. In addition, all samples were normalized to sigA under the same conditions.

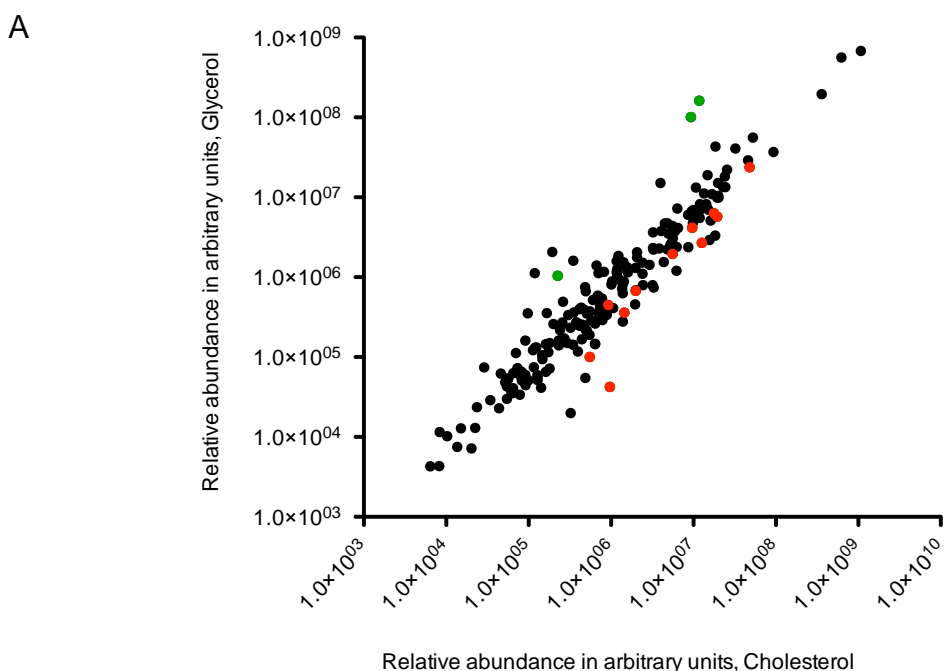
## **Results**

### **Metabolomic profile of cholesterol catabolism**

The dedicated pathway responsible for the degradation of cholesterol into primary metabolites remains to be completely elucidated, and the catabolism of this compound is predicted to produce an ill-defined mixture of propionyl CoA, acetyl CoA, and pyruvate (Van der Geize et al. 2007). In order to characterize the central metabolic consequences of growth on this compound, *M. tuberculosis* was grown in minimal media with either glycerol or cholesterol as a primary carbon source. Liquid- and gas-chromatography coupled mass spectrometry was then used to globally profile metabolite pools in each population. Individual metabolites were identified by comparison to a library of retention times and masses of authentic standards, and the relative abundance of each metabolite was then determined with respect to carbon source.

This analysis unambiguously identified 153 distinct metabolites, which were distributed throughout amino acid, carbohydrate, lipid, nucleotide, cofactor, and central carbon metabolism (Supplementary Table 4.1). Overall, metabolite levels were comparable between these two cultures (Figure 4.1A) consistent with the similar doubling times observed in these two media (Pandey and Sassetti 2008). Further statistical analysis of these data identified 35 metabolites that accumulated to a significant degree in the cholesterol-grown cells. (Table 4.1 and Supplementary Table 4.1). Among these compounds were all of the measurable intermediates of the MCC, a variant of the tricarboxylic acid cycle

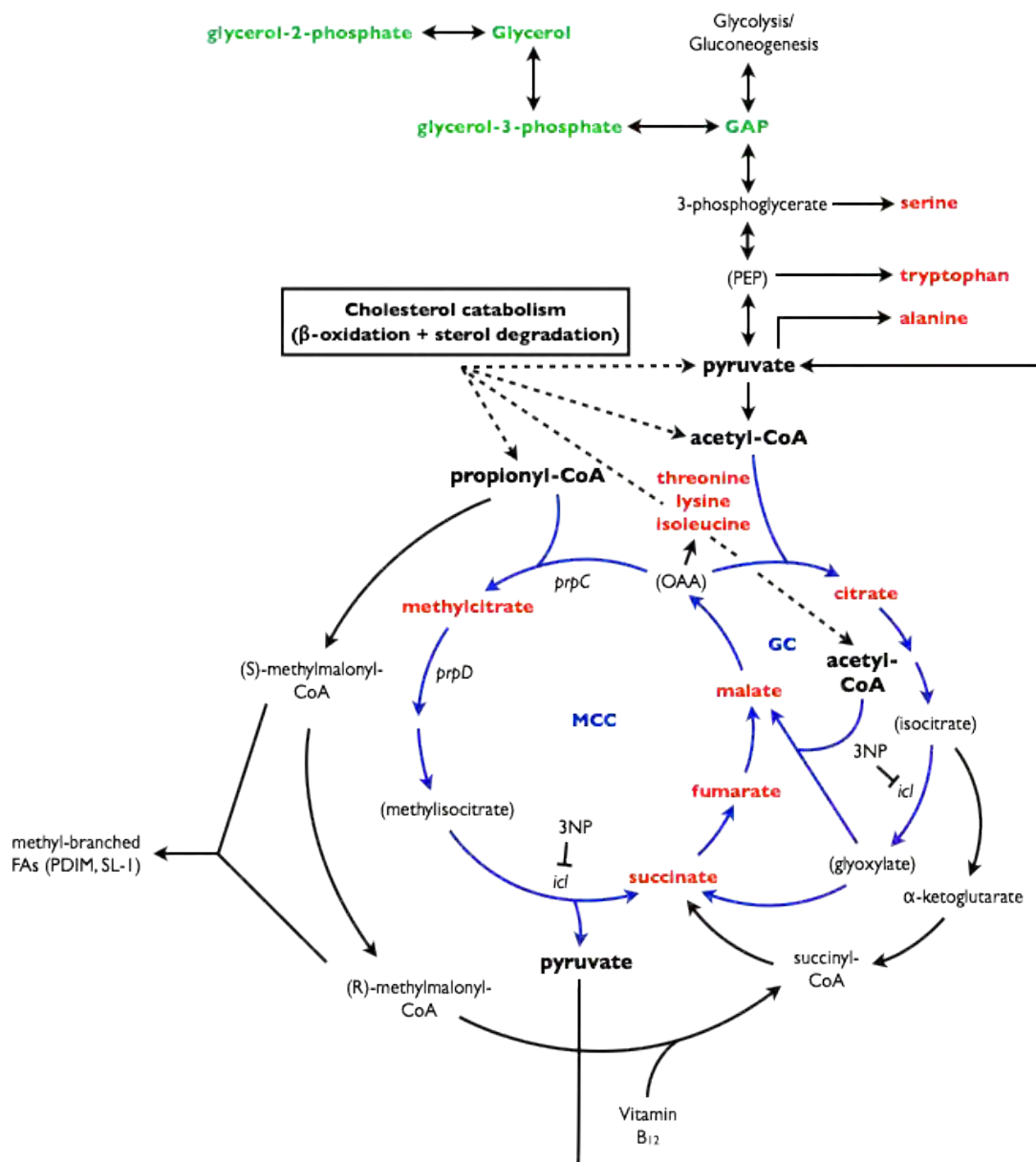
used for the conversion of the three carbon unit propionyl-CoA into pyruvate (Pronk et al. 1994, Textor et al. 1997).



**Figure 4.1 Metabolic profiling during growth on cholesterol.** Mtb was grown with either cholesterol or glycerol as the primary carbon sources and LC/MS and GC/MS were used to quantify the relative abundance of primary metabolites. (A) All 153 metabolites measured are plotted with abundance in glycerol and cholesterol on the y and x axis, respectively. Green and red metabolites accumulated during growth on glycerol and cholesterol, respectively and correspond to green and red labeled metabolites on figure 4.1B.

Consistent with increased activity of the MCC, we also found an accumulation of amino acids derived from both the MCC intermediate oxaloacetate (OAA) and its product, pyruvate (Figure 4.1B). Pyruvate may also be directly liberated from the cholesterol steroid nucleus during catabolism (Van der Geize et al. 2007) further increasing the abundance of this metabolite. This analysis suggested that

B



**Figure 4.1 Metabolic profiling during growth on cholesterol.** *Mtb* was grown with either cholesterol or glycerol as the primary carbon sources and LC/MS and GC/MS were used to quantify the relative abundance of primary metabolites. (B) Metabolites in red indicate those that significantly accumulated following growth on cholesterol relative to growth in glycerol. Both the methylcitrate cycle, MCC and the glyoxylate cycle, GC are highlighted in blue.

**Table 4.1 Metabolomic profile of cholesterol catabolism**

<b>BIOCHEMICAL NAME</b>	<b>Ratio Cholesterol / Glycerol</b>	<b>P-VALUE</b>
tryptophan	2.07	0.0226
2-aminoadipate	5.54	0.0022
alanine	2.02	0.027
beta-alanine	4.44	0.0098
lysine	3.39	0.0424
pipecolate	2.15	0.023
S-adenosylhomocysteine (SAH)	1.85	0.0432
threonine	2.86	0.0218
isoleucine	2.33	0.0376
methysuccinate	16.28	< 0.001
histidine	2.9	0.0195
N-acetylornithine	2.19	0.0079
serine	2.81	0.0259
gamma-glutamylthreonine*	1.9	0.0205
adenosine-5'-diphosphoglucose	3.45	0.0103
mannose	4.02	0.0494
mannose-1-phosphate	5.34	0.0125
trehalose 6-phosphate	5.17	0.0224
2-methylcitrate	22.81	< 0.001
citrate	4.02	0.0054
fumarate	2.95	0.0371
malate	5.49	0.012
succinate	4.67	0.0022
7-alpha-hydroxycholesterol	8.9	0.006
adenosine 5'-monophosphate (AMP)	1.73	0.0305
thymidine 5'-monophosphate	3.38	0.0294
pantothenate	5	0.0024
nicotinate	3.66	0.0264
nicotinate ribonucleoside*	2.91	0.0338
nicotinic acid mononucleotide (NaMN)	2.2	0.0011

“Ratio Cholesterol/Glycerol” = relative abundance of metabolite in cholesterol /  
relative abundance of metabolite in glycerol

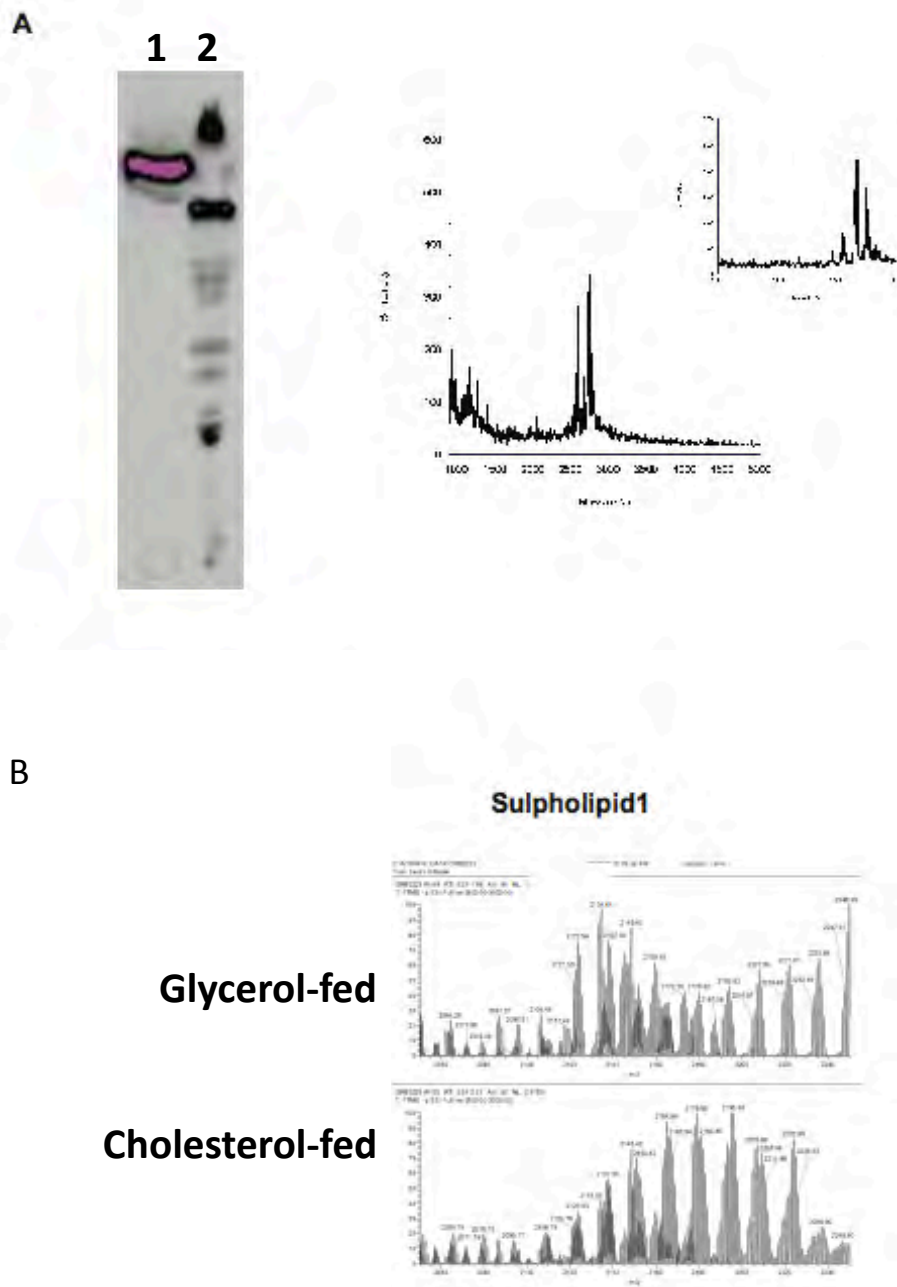
“P-value” = calculated using student t-test from quadruplicate samples.

propionyl-CoA and pyruvate metabolism were central for the adaptation of *M. tuberculosis* to growth on cholesterol, and these metabolites were linked via the MCC.

### **Propionyl CoA is liberated from the sidechain of cholesterol.**

In addition to fueling the MCC, propionyl CoA can also be used as a precursor for methyl branched chain lipid synthesis after conversion to methylmalonyl CoA (Jain et al. 2007). We have previously shown that carbon derived from the branched sidechain of cholesterol was efficiently incorporated into mycobacterial lipids (Pandey and Sassetti 2008), making the sidechain terminus a likely source of propionyl CoA. To more specifically characterize the metabolic fate of this moiety, we thoroughly analyzed the composition of the cellular lipids labeled after growth in [26-<sup>14</sup>C]-cholesterol. Upon profiling the total polar lipid extract, we found that a single species incorporated the vast majority of the <sup>14</sup>C derived from [26-<sup>14</sup>C]-cholesterol, in contrast to [1,2-<sup>14</sup>C]-acetate which labeled a variety of other lipids. Mass spectrometry was used to identify this species as sulfolipid-1 (SL-1), one of the prominent methyl branched chain lipids of *M. tuberculosis* (Figure 4.2A).

Since both the abundance and chain length of SL-1 is directly related to availability of the propionyl CoA precursor (Jain et al. 2007), we used these metrics to estimate the abundance of this precursor metabolite in cholesterol-grown cells. We found that both the molecular weight of SL-1 and its relative



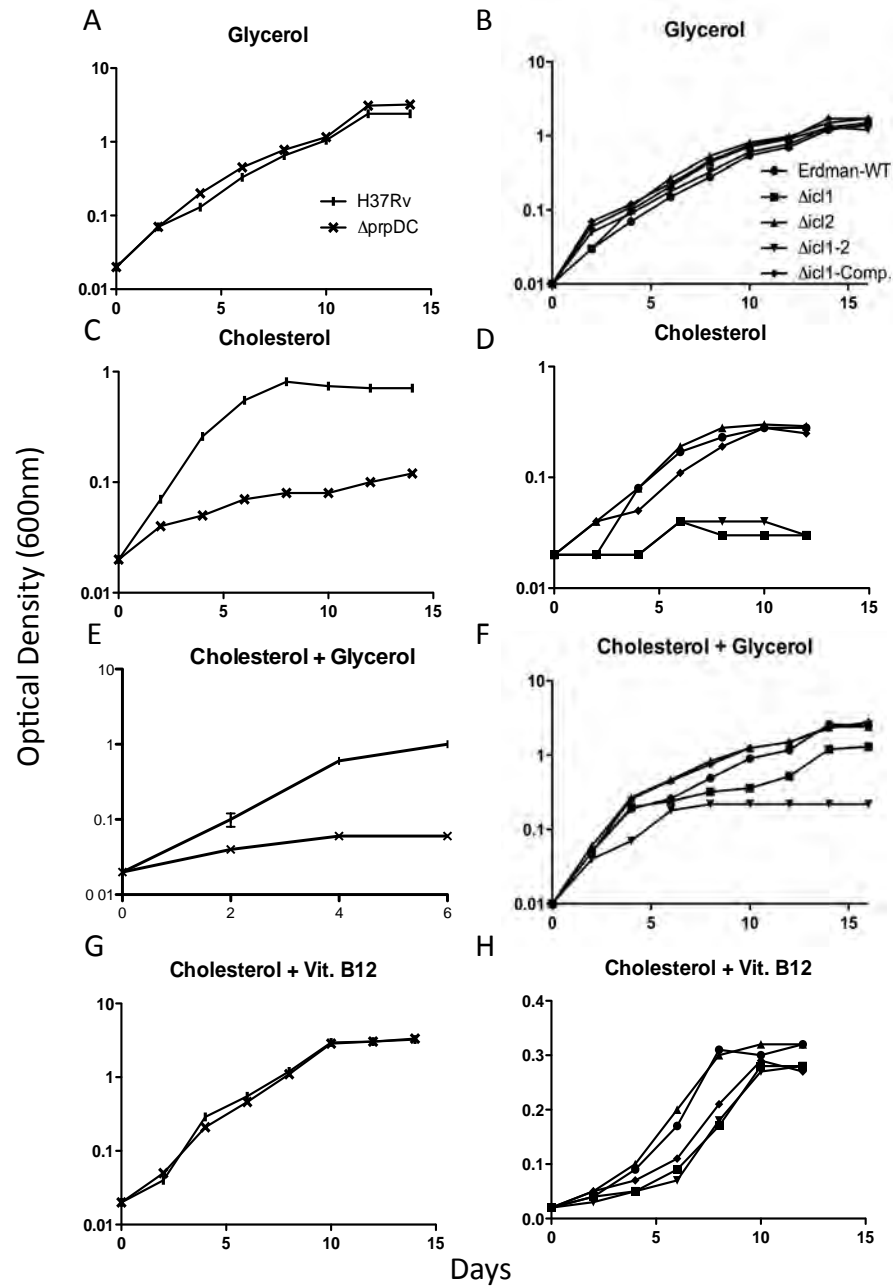
**Figure 4.2 Cholesterol-derived propionate is incorporated into SL-1 and increases both its mass and abundance.** (A) Total polar lipids were extracted from cholesterol-fed (lane 1) and acetate-fed (lane 2) bacteria and a cholesterol specific species was identified using MS as sulpholipid-1 (SL-1). (B) MS was used to characterize the SL-1 from glycerol and cholesterol-fed bacteria.

abundance increased in cholesterol-grown cultures, compared to those grown in glycerol (Figure 4.2B). This is consistent with the observed cholesterol-induced increase in length and abundance of a distinct methyl branched chain lipid, phthiocerol dimycocerosate (Yang et al. 2009 and data not shown), which is synthesized from the same precursor pool. The preferential flux of cholesterol sidechain-derived carbon into SL-1 and the inferred increase in branched chain lipid precursors in cholesterol-grown cells both indicate that cholesterol is a rich source of propionyl CoA and that at least a portion of this metabolite is derived from the sidechain terminus.

### **Functional MCC enzymes are required for growth in the presence of cholesterol.**

To determine the importance of propionyl CoA metabolism during growth on cholesterol, we tested the ability of MCC-deficient mutants to grow using this carbon source. We found that the deletion of the *prpC* and *prpD* genes, encoding the two dedicated steps of this pathway, specifically impaired the bacterium's ability to grow in media containing cholesterol as a primary carbon source (Figure 4.3A-D). Similarly, deletion of the *icl1* gene, which encodes a bifunctional methylisocitrate- and isocitrate-lyase that is necessary for both the MCC and glyoxylate cycles (Gould et al. 2006), caused a similar growth defect. Deletion of the partially-redundant *icl2* gene, either alone or in combination with the *icl1* mutation, had little effect.





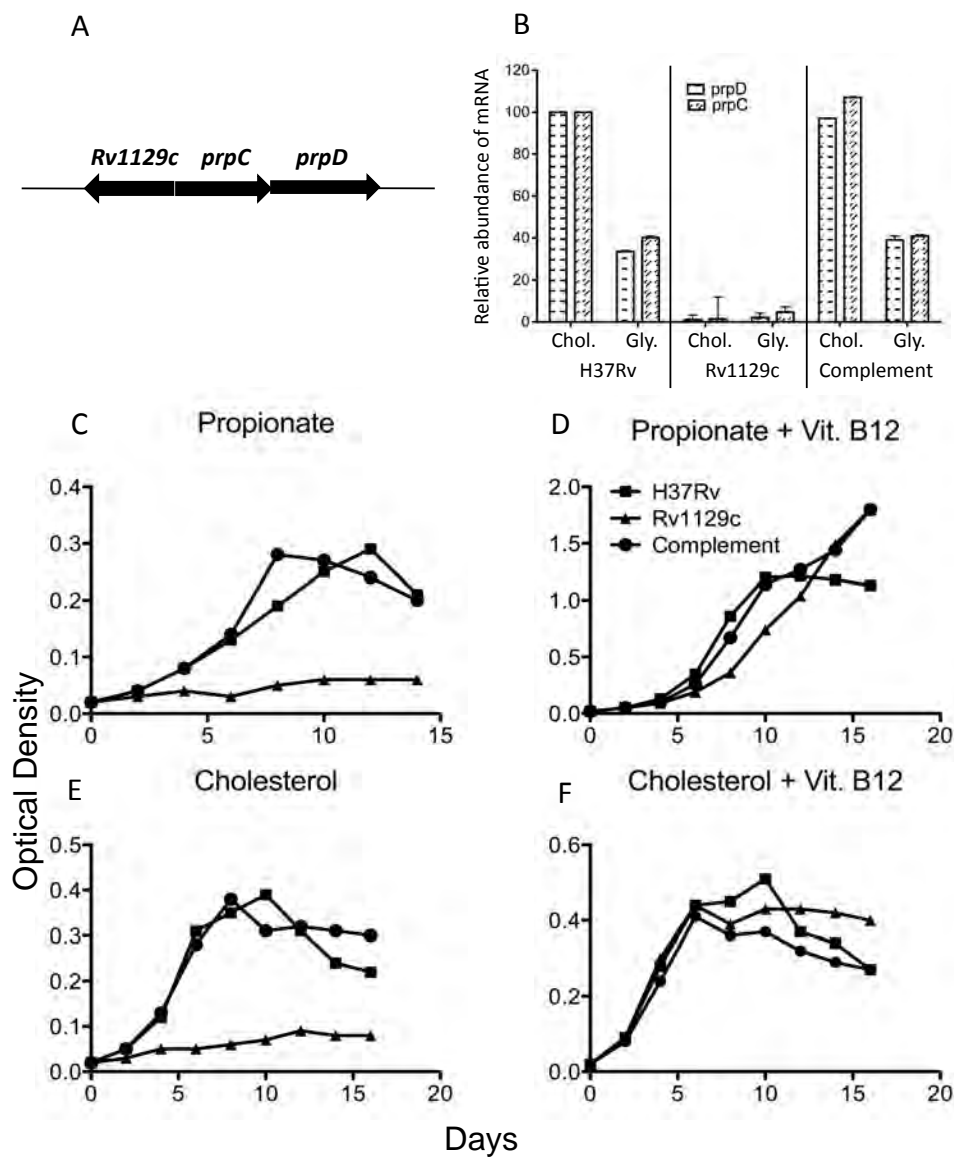
**Figure 4.3 The MCC genes are required for growth on cholesterol.** prpDC mutants (A,C,E,G) and icl mutants (B,D,F,H) were grown in minimal media with the indicated carbon sources in the presence or absence of Vitamin B12. Optical density was measured at the indicated times.

In many bacteria, including mycobacteria, the inability to assimilate propionyl CoA inhibits growth even in the presence of other carbon sources (Munoz-Elias et al. 2006, Savvi et al. 2008). This is likely due to the accumulation of intermediates that inhibit other essential pathways (Rocco and Escalante-Semerena 2010). Indeed, when we cultured  $\Delta prpDC$  or  $\Delta icl1$  mutants in mixed carbon sources containing cholesterol, we observed growth defects that were comparable to the single carbon source media (Figure 4.3E-F). To verify that this growth defect was due to propionyl CoA-related toxicity, we enabled an alternative assimilation system, the methylmalonyl pathway (MMP). This pathway converts propionyl CoA into succinyl CoA via the B12-dependent methylmalonyl CoA mutase enzyme (Figure 4.1B), allowing MCC mutants to grow on propionyl CoA-generating substrates (Savvi et al. 2008). Since *M. tuberculosis* is unable to synthesize B12 under these conditions, the MMP is only active upon B12 supplementation. As anticipated, the addition of B12 to these mixed carbon source cultures reversed the growth defect imparted by MCC mutations (Figure 4.3G-H). The modest residual defect of the  $\Delta icl1$  mutant in media containing only cholesterol may be attributable to the additional role played by this enzyme in another variant TCA cycle, the glyoxylate cycle. We conclude that propionyl CoA metabolism is essential for growth in the presence of cholesterol.

**Transcriptional induction of MCC genes by Rv1129c is required for propionyl CoA metabolism *in vitro*.**

The accumulation of MCC intermediates during the growth of *M. tuberculosis* on cholesterol suggested that the activity of the dedicated MCC enzymes were likely increased under these conditions. Indeed, we found that the abundance of the mRNA encoding both *prpC* and *prpD* was increased by 5 fold in cholesterol-grown cultures. Encoded divergently from this transcript is the gene *rv1129c* (Figure 4.4A), which encodes a protein homologous to RamB, the regulator of *icl* gene transcription in actinobacteria (Micklinghoff et al. 2009, Gerstmeir et al. 2004). To determine if Rv1129c performed a similar function, we compared the level of *prpDC* mRNA in wild type bacteria with a strain in which the *rv1129c* gene had been deleted. We found that transcript levels of both *prpD* and *prpC* were dramatically reduced in the absence of *rv1129c* regardless of carbon source, and that this phenotype could be complemented by expressing a wild type copy of the gene (Figure 4.4B).

To determine the physiological relevance of this regulation we determined if the deletion of *rv1129c* altered growth on propionyl CoA generating substrates. We found that the *rv1129c* mutant was unable to grow using either propionate or cholesterol as a primary carbon source (Figure 4.4C,E). This phenotype was rescued either by genetic complementation or by adding B12 to circumvent the MCC requirement (Figure 4D,F), demonstrating that transcriptional induction of



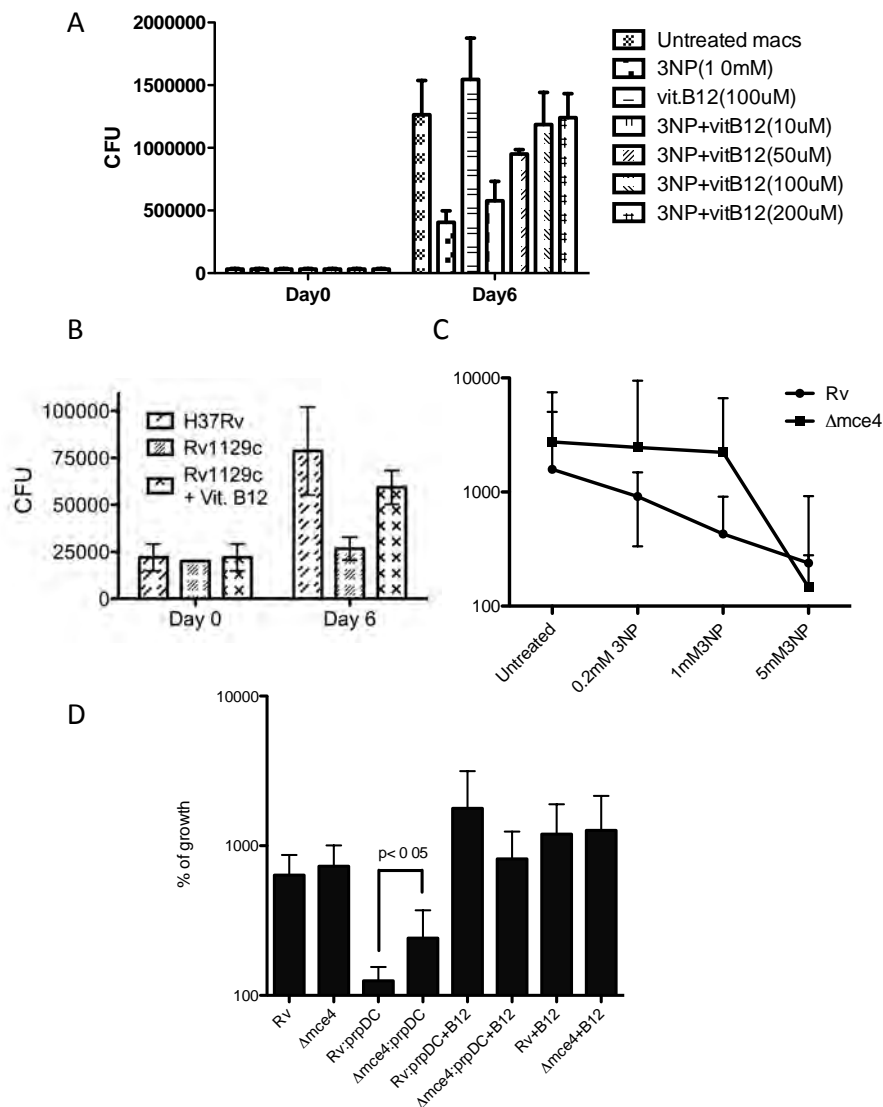
**Figure 4.4 The transcriptional activation of MCC genes by Rv1129c is required for growth in cholesterol.** (A) Genomic organization of *rv1129c* and the MCC genes. (B) Relative abundance of *prpD* and *prpC* mRNA was measured 48hrs following growth in vitro with either cholesterol or glycerol as the primary carbon sources. (C-F) Bacteria were grown in the defined carbon sources with or without vitamin B12 and monitored by optical density.

*prpDC* via Rv1129c is required for the assimilation of cholesterol-derived propionate.

**Cholesterol is a significant source of the propionyl CoA generated during intracellular growth.**

The dedicated MCC components PrpD, PrpC, as well as the multifunctional ICL/MCL enzymes are required for intracellular growth in macrophages (Munoz-Elias et al. 2006, Munoz-Elias and McKinney 2005), indicating that propionyl CoA metabolism is critical in this niche. However, it has remained unclear whether this requirement could be attributed directly to cholesterol catabolism. To address this, we investigated the impact of cholesterol uptake on this phenotype.

Initially, we sought to determine whether the observed intracellular growth defect of MCC mutants was due to propionyl CoA-related toxicity. To do this we infected bone marrow-derived macrophages with *M. tuberculosis* and inhibited the MCC either by genetic deletion of *prpDC* or chemical inhibition of Icl1/2 with 3-nitropropionate (3-NP) (Schloss and Cleland 1982). Both MCC inhibition strategies resulted in a significant inhibition of intracellular growth. Addition of B12 to enable the alternative methylmalonyl pathway caused a dose dependent reversal of this phenotype (Figure 4.5A). Similarly, we found that the *rv1129c* regulatory mutant was unable to grow in BMDM unless supplemented with B12,



**Figure 4.5 Cholesterol is a significant intracellular source of propionate.** (A) BMDM were infected with *M. tuberculosis* and icl was inhibited by the addition of 3NP with or without vitamin B12 at the indicated concentrations. Surviving bacteria were cultured at day 6 post-infection and resulting colony forming units (cfu) are plotted. (B) BMDM were infected with *M. tuberculosis* H37Rv or Rv1129c mutant and cultured in the presence or absence of vitamin B12 for 6 days. Surviving bacteria were plated and cfu were plotted. (C) BMDM were infected with *M. tuberculosis* H37Rv or Mce4 mutant and icl was inhibited with the indicated concentrations of 3NP. Surviving bacteria are plated as % of growth. (D) BMDM were infected with the indicated mutants in the presence or absence of vitamin B12 and surviving bacteria were enumerated via cfu. Data is plotted as percent of growth.

suggesting that propionyl CoA toxicity was the basis for the growth defect observed for MCC-deficient strains (Figure 4.5B).

We then investigated whether cholesterol contributed significantly to the bacterial propionyl CoA pool during intracellular growth. To do this, we employed *M. tuberculosis* mutants lacking the Mce4 cholesterol uptake system, which are unable to import cholesterol but grow normally in resting BMDM (Pandey and Sassetti 2008, Mohn et al. 2008). To estimate the propionyl CoA pool in wild type and *mce4* mutants during intracellular growth, we compared the relative sensitivity of these strains to Icl1/2 inhibition with 3-NP. The cholesterol uptake mutants were significantly more resistant to 3-NP, indicating that less ICL/MCL activity was necessary, and therefore that less propionyl CoA was generated in the absence of cholesterol catabolism (Figure 4.5C). To verify this conclusion, we determined if the *mce4* deletion was also able to suppress the growth defect of the  $\Delta prpDC$  mutant. Indeed, we observed that  $\Delta mce4/\Delta prpDC$  double mutants grew significantly more rapidly than  $\Delta prpDC$  single mutants in BMDM, and the residual growth defect of the double mutants could be fully reversed by the addition of B12 (Figure 4.5D). We conclude that MCC mutants are unable to grow intracellularly due to propionyl CoA toxicity and a significant fraction of this metabolite is produced via the catabolism of host cholesterol.

## Discussion

Many approaches have been used to determine how the intracellular environment influences the physiological state of a pathogen during infection. Most rely on identifying biochemical pathways that are expressed or essential during infection, and using this information to infer both the metabolic state of the cell and the environment that imposes this state. However, the complexity of metabolic networks makes these inferences difficult, as a single enzyme can participate in multiple biochemical pathways, and the reliance on any pathway can be influenced by a number of different environmental factors. This work employed a unique combination of direct metabolomic profiling and genetic epistasis analysis to functionally connect a distinct feature of the intracellular environment with its metabolic consequence in the pathogen.

In order to understand the central metabolic adaptations required for the utilization of the host-derived carbon source, cholesterol, we globally surveyed the steady state levels of primary metabolites during growth on this compound. This approach is fundamentally different from more commonly used profiling of mRNA or proteins levels, which may not correlate with the metabolite levels (Jozefczuk et al. 2010). We found that the differential utilization of a pathway could be predicted by the abundance of both its specific intermediates and its more stable biosynthetic products. Labeling of proteinogenic amino acids has been used extensively to predict the flux through central carbon metabolism, and our data suggest that pool sizes of individual amino acids are also likely stable



enough to serve this purpose. Together, these data provided a holistic view of cellular metabolism and allowed us to identify pathways that were critical for cholesterol catabolism.

The relatively dramatic accumulation of the PrpD substrate, methylcitrate, suggested that the constitutive levels this enzyme might be insufficient for the flux of propionyl-CoA derived from cholesterol. Indeed, it appears that several propionate units are liberated from each cholesterol molecule. In addition to the terminal sidechain carbons at least two other propionyl-CoA molecules may be produced (Van der Geize et al. 2007), suggesting that as much as one third of the carbon derived from cholesterol could enter the MCC. While cholesterol appears to be a rich source of propionyl CoA, the incomplete suppression of the  $\Delta prpDC$  mutant's intracellular growth defect indicates that alternative carbon sources, such as odd chain fatty acids or branched chain amino acids, may also contribute to the propionyl-CoA pool during infection.

The increased reliance on the MCC pathway during cholesterol catabolism was confirmed by demonstrating that transcriptional induction of the *prpDC* genes via the Rv1129c protein was required for growth. This protein is highly homologous to the RamB regulator of both *Mycobacteria* and *Corynebacteria*, which controls the expression of the glyoxylate cycle enzymes necessary for growth on acetate. It has been proposed that the DNA binding function of the RamB protein of

*Corynebacterium glutamicum* is posttranslationally regulated (Micklinghoff et al. 2009, Cramer et al. 2007), perhaps via binding of an intermediary metabolite. Thus, mycobacteria appear to use very similar regulatory proteins to control the pathways necessary for propionyl- and acetyl-CoA catabolism, and differential expression of these systems could be mediated via distinct allosteric regulation.

Consistent with previous observations, we found that the growth of mutants lacking sufficient MCC activity was inhibited by propionate, even in the presence of an additional carbon source. A similar phenomenon in *Salmonella* is due to the accumulation of a specific 2-methylcitrate isomer, which inhibits the gluconeogenic enzyme fructose 1,6 biphosphatase (Rocco and Escalante-Semerena 2010). While a similar mechanism could underly propionate toxicity in mycobacteria, the type II FBP expressed by Mtb bears little if any resemblance to the *Salmonella* enzyme, suggesting that alternative mechanisms may be responsible.

These metabolic insights provided the information necessary to test whether the MCC requirement during intracellular growth was due to host cholesterol utilization. The ability of Mce4 mutation to alleviate the intracellular growth defect produced by either prpDC mutation or ICL/MCL inhibition implicate cholesterol as a major source of propionate *in vivo* even under conditions, such as growth in resting BMDM, in which cholesterol uptake is not absolutely required for growth.

This MCC mutant phenotype was reminiscent of strains lacking either the *hsaC* gene (Yam et al. 2009) or the “*igr*” locus (Chang et al. 2009, Chang et al. 2007), which encode functions dedicated to earlier steps in cholesterol degradation. Like MCC mutants, these strains are intoxicated in the presence of cholesterol, and are unable to grow in resting macrophages or acutely infected animals. Together, these observations strongly argue that *M. tuberculosis* obtains a significant amount of its carbon requirements from cholesterol throughout infection.

In contrast to their absolute essentiality during intracellular growth in macrophages, the *prpDC* genes have been found to be dispensable for bacterial growth in the mouse model of TB (Munoz-Elias et al. 2006). This was particularly surprising as cholesterol catabolism occurs throughout infection, and cell wall lipid alterations consistent with propionyl-CoA assimilation are evident in bacteria isolated from mouse tissue (Jain et al. 2007). This apparent contradiction could be explained if the B12-dependent methylmalonyl pathway (MPP) were active under these conditions, either because the bacterium acquires B12 from the host, or some aspect of the host environment stimulates B12 production by the pathogen (Savvi et al. 2008). Our observation that B12 supplementation reverses growth defect of *prpDC* mutants in macrophages, supports this hypothesis.

This work highlights propionyl-CoA assimilation as a critical feature of *M. tuberculosis* metabolism *in vivo*, and demonstrates that this requirement is a direct result of the nutritional environment in the mycobacterial phagosome. However, even within this relatively isolated compartment, the bacterium is likely to co-catabolize multiple carbon sources (de Carvalho et al. 2010), and inhibiting the bacterium's ability to utilize individual nutrients is unlikely to represent a fruitful therapeutic strategy. In contrast, interference with at least three distinct steps in cholesterol catabolism, via mutation of *hsaC*, *igr*, or MCC-encoding genes, results in metabolic toxicity that might be effectively exploited to treat TB. Thus, a fundamental understanding of mycobacterial physiology during infection is critical for identifying vulnerabilities that can be exploited for the rational design of more effective TB therapies.

### **Acknowledgements**

We thank Branch Moody for metabolomic profiling advice and sample preparation protocol. This work was supported by grants to CMS from the NIH (AI073509 and AI064282) and the Howard Hughes Medical Institute.

## **Chapter V: The Mce lipid importers require a complex formation within the cell wall of Mtb**

### **Summary**

The dependence on Mce4 for cholesterol utilization demonstrates the need for effective transport of lipid substrates across the complex cell wall of Mtb. The mechanisms by which any solutes transit this impermeable mycobacterial outer membrane remain ill defined. By using the Mce systems as a model, we have begun to understand this process. Each Mce operon encodes an inner membrane complex that resembles an ABC transporter and at least six proteins that are predicted to be secreted into the periplasmic space. We hypothesized that these components may form a complex in the outer membrane that delivers substrate to the inner membrane components. Here, we investigate this hypothesis by determining the subcellular localization of the secreted components encoded by the Mce1 operon, which is highly homologous to the Mce4 cholesterol importer. By developing tools including peptide antibodies against Mce1 proteins and cell surface biotinylation conditions, we show that these Mce1 proteins are localized to the cell wall of Mtb and evidence is provided that suggests they form a complex with the inner membrane ABC-like components. This complex depends on ATP hydrolysis generated by the associated ATPase, MceG. Mechanisms for transport and the role of ATP hydrolysis are discussed.

## Introduction

It is clear that lipids play a vital role in TB infection, specifically during persistence (Pandey et al. 2008, McKinney et al. 2000). Chapters III and IV describe the lengths to which Mtb has metabolically adapted to using one particular lipid, cholesterol, as an intracellular carbon source. While Pandey and Sassetti have demonstrated that Mce4 is required for the bacterial transport of cholesterol, the mechanism for transport of lipids across the extremely complex Mtb cell wall remains elusive. In fact, transport of solutes and nutrients in general in Mtb is not well understood.

This lack of clarity stems from the complex and effective permeability barrier of the Mtb cell wall. Recent investigations using cryo-electron tomography of intact mycobacterial cell walls demonstrate that despite the phylogenetic classification of Mtb as a Gram-positive organism, its cell wall has similarities to that of a Gram-negative organism (Hoffman et al. 2008, Zuber et al. 2008). The Mtb cell envelope is comprised of a lipid bilayer plasma membrane and a complex cell wall. An insoluble cell wall core is comprised of the covalently linked glycoconjugates peptidoglycan (PG) and arabinogalactan (AG) (Nierderweis et al. 2010). Long chain (up to 90 carbons in length) mycolic acids (MA) are then esterified to AG and are a dominant component to the outer membrane, comprising between 30%-40% of the cell envelope mass (Brennan 2003, Nierderweis et al. 2010). In addition to MAs, which comprise the majority of the

outer membrane (Hoffmann 08, Zuber 08), *Mtb* contains many additional free lipids intercollated into this outer structure. These include cord factor/dimycolyltrehaolse (TDM), sulfolipids, phthiocerol dimycocerosate (PDIM), lipoarabinomannan (LAM), phoshatidylinositol mannosides (PIM), lipomannan (LM), etc. (Kaur et al. 2009). Not only do the abundant mycobacterial lipids create an extremely hydrophobic and effective permeability barrier peripheral to the cell membrane, they are also important in virulence and host immune modulation (Karakousis et al. 2004).

The unique complexity and hydrophobicity of the mycobacterial cell wall presents a challenge in understanding how nutrients and solutes are transported across it. Small hydrophilic substrates are able to diffuse through outer membrane proteins (OMPs) that form water-filled channels or porins spanning the outer membrane of gram-negative organisms. OMPs have been difficult to identify in mycobacteria (Niederweis et al. 2010). The fast-growing saprophytic mycobacterial species *M. smegmatis*, has been shown to contain the OMP MspA, the most abundant protein in the organism, which is believed to serve as the major diffusion pathway. Interestingly, expression of MspA in *M. bovis* BCG increased the growth rate of the organism, suggesting that the rate of hydrophilic nutrient uptake limits growth (Mailaender et al. 2004, Sharbati-Tehrani et al. 2004). There are no MspA homologues in *Mtb* and despite the identification of the suggested porins OmpATb and Rv1698, contradicting data precludes their

classification as true porins in Mtb (Raynaud et al. 2002, Siroy et al. 2008, Wolschendorf et al. 2011, Niederweis et al. 2010).

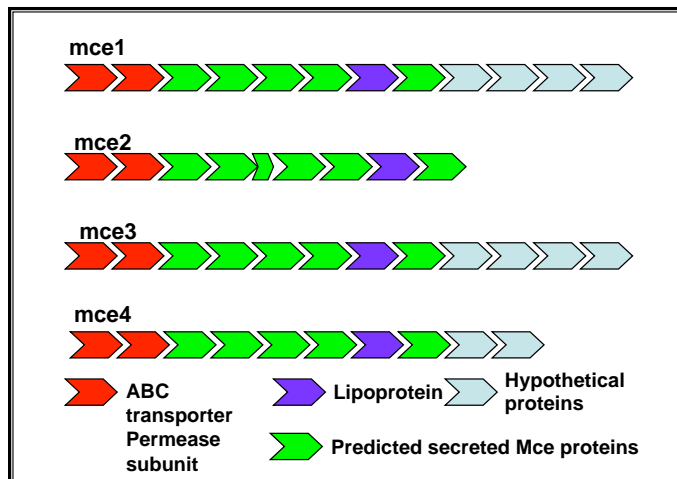
Diffusion rates through porins are extremely slow for hydrophobic solutes (Nikaido 2003). Direct diffusion of fatty acids is slow and depends on the permeability of the lipid bilayer. Differential scanning calorimetry has been used to determine the transition temperatures of mycobacterial cell wall lipids (Liu et al. 1995). This study found that mycobacterial cell walls have abnormally high melting temperatures (60°C) indicative of very low fluidity at physiological temperatures and are thus, likely to be impermeable to even hydrophobic compounds like fatty acids. Thus, spontaneous diffusion of hydrophobic nutrients like cholesterol would occur too slowly, if at all to support growth. Mycobacteria must, therefore, use additional transporters for such import. In *E. coli*, the outer membrane transporter FadL mediates uptake of fatty acids across the OM in an energy-independent manner. This process involves conformational changes resulting in import of long chain fatty acids (LCFA) to the periplasm. Following the formation of uncharged LCFA, these lipids flip into and out of the IM (van den Berg 2005). There are no known FadL homologues in Mtb and the process of lipid import is unclear.

As noted in earlier chapters, Mce4 is required for the import of the lipid cholesterol in Mtb (Pandey and Sassetti 2008, Mohn et al. 2008). There are four



Mce loci, *mce1-4*, in the genome of Mtb (Cole et al 1998). Mce operons have homology to ABC transporters, of which there are exporters and importers. ABC importers have only been identified in prokaryotes and are recognized by the presence of substrate binding proteins (SBP), nucleotide binding proteins that bind and hydrolyze ATP, and transmembrane permeases that span the inner membrane forming a pore for translocation. Mce operons contain 8-12 genes encoding proteins with homology to both transmembrane permeases and SBPs. Each operon is identically organized within the chromosome and begins with 2 *yrbE* genes each containing 6 predicted transmembrane domains with homology to permeases. 5 *mce* genes with homology to SBPs and 1 *lpr* gene encoding a predicted lipoprotein follow. Although one report has indicated that Mce proteins may contain beta-barrel structures (Pajón et al. 2006), most prediction algorithms indicate the presence of 8 alpha-helices (Casali et al. 2007). In addition, these proteins are predicted to have an N-terminal transmembrane helix followed by a cleavage site as predicted by SignalP. Thus, Mce proteins are very likely secreted past the IM. In fact, transmission electron microscopy has indicated that Mce4A is localized to the cell wall (Saini et al. 2008).

While homology and other functional evidence indicate Mce loci encode ABC importers, Mce operons themselves are lacking a dedicated adenoside triphosphatase (ATPase). Our lab first demonstrated that the unlinked gene,



**Figure 5. Mce loci of Mtb.**

The 4 mce loci of Mtb are identically organized and encode predicted subunits of a lipid importers.

*rv0655* or *mceG* functionally interacted with both Mce1 and Mce4 (Joshi et al. 2006). MceG is part of a larger Mkl family of ATPases that is tightly conserved with Mce proteins throughout the Actinomycetales order and also cluster with the SBP-dependent importer clade of ABC transporters (Casali et al. 2007). In addition, the presence of canonical Walker A and B motifs required for ATP binding and the ABC transporter family signature sequence, LSGGC, further support the role of MceG as a Mce-associated ATPase (Casali et al. 2007). ATPases physically interact with permease subunits in ABC transporters (Davidson and Chen 2004, Hollenstein et al. 2007). YrbE proteins have been shown to contain a unique and conserved EExDA motif in the second to last cytoplasmic loop believed to interact with MceG (Casali et al. 2007). We have demonstrated supporting evidence that the expression of YrbE1A-B is required for the stable expression of MceG in *M. smegmatis* (Joshi et al. 2006). It is likely that Mce proteins form transporter complexes spanning the cell membrane and cell wall of Mtb.

In this chapter, Mce1 is used as a model for lipid import. Although the substrate for Mce1 is not known, the Mce1 locus is identically organized to the Mce4 locus, which we and others have shown to be required for cholesterol import (Pandey and Sassetti 2008, Mohn et al. 2008). In addition, our lab has demonstrated that like Mce4, Mce1 functionally interacts with genes involved in lipid metabolism (Joshi et al. 2006). It is likely that like Mce4, Mce1 functions in lipid import. Like Mce4, Mce1 mutants are attenuated in mice indicating their functional relevance in vivo (Sassetti et al. 2003, Joshi et al. 2006). In addition, Mce1 is the most highly expressed mce loci during exponential growth in vitro (Kumar et al. 2003). Thus, in order to better understand lipid import in mycobacteria, the structure of this proposed complex was probed using a combination of a previously reported fractionation method for Mtb, purified peptide antibodies raised against Mce1 proteins, and a novel cell surface biotinylation assay developed for Mtb. Using these techniques, we further demonstrate the presence of an ordered Mce complex spanning the cell wall of Mtb and propose a novel role for MceG in the function of the transporter.

## **Materials and Methods**

### **Growth and maintenance of bacterial strains**

All bacteria were maintained in complete Middlebrook 7H9 medium containing 0.05% Tween 80 and albumin-dextrose-catalase supplement. Kanamycin and Hygromycin were used at 25 ug/ml and 50 ug/ml, respectively.

### **Genetic manipulation of strains**

Mutations were made in the H37Rv background as previously described (Joshi et al. 2006). Mce1, 2, 3, 4 quadruple mutant (Quad) was a generous gift from Jim Gomez and John McKinney.

### **Mtb Cell Fractionation**

*Mycobacterium tuberculosis* H37Rv  $\Delta lysA$  and  $\Delta panCD$  auxotroph (mc<sup>2</sup>6020) was used for all fractionation experiments and was created in the laboratory of Dr. Bill Jacobs (Sambandamurthy et al. 2005). Bacteria were grown to an optical density at 600nm of 1.0 in 50mls of complete 7H9 media with lysine (1mg/ml) and pantothenate (24ug/ml) then washed in PBS twice and passed into 200mls of Sauton's minimal media (KH<sub>2</sub>PO<sub>4</sub> 0.5g/L, MgSO<sub>4</sub> 7H<sub>2</sub>O 0.5g/L, citric acid 2.0g/L, ferric ammonium citrate 0.05g/L, glycerol 60ml/L, asparagine 4g.0g/L) with 0.05% Tween 80 with leucine and pantothenate supplements as above. Culture grew aerobically in a roller bottle until it reached an optical density at 600nm of 0.75-1.0 when cells were harvested at 4400g for 10 minutes at room temperature. Bacterial pellets were kept at -80°C until fractionated as previously described (Rezwan et al. 2007). Briefly, cells were washed in 0.16M NaCl, incubated with lysozyme (2.4mg/ml) for 2hrs at 37°C and sonicated in an ice water bath. Unbroken cells were removed via centrifugation and lysates were

centrifuged at 27,000g for 40 minutes at 4°C to isolate cell walls. The resulting pellet was separated by centrifugation at 100,000g for 1hr at 4°C into the bacterial membrane (pellet) and cytosol (supernatant). All fractions were resuspended in the same volume of PBS and protease inhibitors (Protease Inhibitor Cocktail tablets, Roche) and equal volumes were used in all analyses. Fractions were aliquoted and frozen at -80°C until used. Cell wall fraction was boiled in 0.1% SDS for 30 minutes and cytosol and membrane fractions were filtered using 0.2µm filter before fractions were used in biosafety level 2 conditions.

### **Cell Surface Biotinylation**

Mtb H37Rv was used for all biotinylation experiments and was grown in complete 7H9 media with 0.05% Tween 80 and albumin-dextrose-catalase supplement until an optical density at 600nm of 1.0. Culture was then washed in PBS twice and passed to an optical density at 600nm of 0.3 in complete 7H9 with or without 0.05% Tween 80. Following growth for 48hrs, cultures were washed 3x in PBS, concentrated 50x in PBS and incubated with 1mM NHS-Biotin (EZ-Link NHS-Biotin, Thermo Fisher Scientific) for 10 minutes at room temperature. Samples were washed once with 50mM Tris, pH 8.0 and then twice with PBS and 0.05% Tween 80 and resuspended in PBS and protease inhibitors. Samples were lysed via mechanical rupture using glass beads for 30 seconds each at 6.5 speed using a FastPrep FP120 machine with 1-2 minute incubations on ice in between. Samples were boiled in 1.0% SDS for 30 minutes before removal from the

biosafety level 3 facility and then spun at max speed for 5 minutes at 4°C (lysate). The supernatant was diluted to 0.06% SDS in cold PBS and incubated for 1hr at 4°C with either streptavidin (High Capacity SA Agarose Resin, Thermo Fisher Scientific) or glutathione S-transferase (GST) (GST-Bind Resin, Nalgene) coated beads and spun at max speed for 30 seconds at 4°C. The resulting supernatant was TCA precipitated by incubation with an equal volume of ice-cold 20% TCA at 4°C overnight. Samples were then washed once with ice-cold acetone and air dried for an hour at room temperature and resuspended in sample buffer and boiled 30 minutes (unbound fraction). Beads were washed once with cold PBS and protease inhibitors at 4°C and boiled in sample buffer (bound fraction). Lysate, bound and unbound fractions were then probed via western blot.

### **Western Blots**

All protein samples were boiled in sample buffer (50mM Tris-HCl, pH 6.8, 1% SDS, 0.1% bromophenol blue, 10% glycerol, 1%  $\beta$ -mercaptoethanol) for 30 minutes. Samples were then run on 10% denaturing polyacrylamides, transferred to polyvinylidene fluoride (PVDF) and membranes were blocked overnight in blocking buffer (5% milk, 0.1% Tween 20 and 1.0% goat serum in PBS). Membranes were probed via western blot by incubation with 1ug primary, purified peptide antibodies for 1hr in blocking buffer at room temperature with gentle rocking. Membranes were washed 3x for 15 minutes in PBS and incubated with anti-rabbit IgG conjugated with horseradish peroxidase (HRP) at

1:2500 dilution in blocking buffer for 1hr at room temperature with gentle rocking. Membranes were then washed 4x for 15 minutes at room temperature with gentle rocking and developed using chemiluminescent substrate for detection of HRP according to manufacturer's instructions (ECL Western Blotting Substrate, Pierce).

### **Peptide Antibody Purification**

Short (19 amino acids) peptides were generated by Open Biosystems (Thermo Fisher Scientific) and inoculated into rabbits. Crude antibody serum was then purified using immunoaffinity purification following the immobilization of peptides to solid support (HiTrap NHS-activated HP, GE Healthcare) according to manufacturers instructions. Briefly, 1mg of peptide was dissolved in coupling buffer (0.2M NaHCO<sub>3</sub>, 0.5M NaCl, pH 8) and incubated with matrix for 30 minutes at room temperature. Peptide bound matrix was then washed and deactivated with alternating washes with Buffer A (0.5M ethanolamine, 0.5M NaCl, pH 8) and Buffer B (0.1M acetate, 0.5M NaCl, pH 4). Bound matrix was stored in storage solution (0.05M Na<sub>2</sub>HPO<sub>4</sub>, 0.1% NaN<sub>3</sub>, pH 7) until serum immunoaffinity purification. Bound matrix was prepared for serum purification by washing with 10 bed volumes each of 10mM Tris (pH 7.5), 100mM acetate (pH 2.5), 10mM Tris (pH 8.8). 10 bed volumes of 100mM triethylamine (pH 11.5) made fresh was then passed through the matrix followed by one wash with 10mM Tris (pH 7.5). Polyclonal serum was passed through a 0.2µm filter and diluted 1:10 in 10mM Tris (pH 7.5) and then passed slowly over the matrix 3



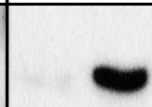
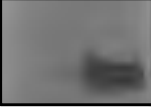
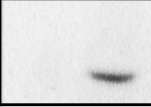

times. Matrix was then washed with 20 bed volumes each of 10mM Tris (pH 7.5) and 500mM NaCl, 10mM Tris (pH 7.5). An acid elution was conducted by passing 10 bed volumes of 100mM acetate (pH 2.5) and eluate was collected and neutralized with 1M Tris (pH 8). Matrix was then washed once with 10mM Tris (pH 8.8) and eluate was collected. A base elution was conducted by passing 10 bed volumes of 100mM triethylamine (pH 11.5) made fresh and eluate was collected and neutralized with 1M Tris (pH 8). Collected eluates were concentrated via centrifugation over a 30kD Amicon centriplus centrifugal device (Millipore) and quantified at 280nm.

## **Results**

### **Mce proteins localize to the cell wall**

The presence of predicted signal cleavage sites at the N-terminus of Mce proteins strongly suggest their localization peripheral to the IM. In order to demonstrate this, Mtb was fractionated into cytosol and cell wall fractions using a previously validated protocol (Rezwan et al. 2007). Using specific peptide antibodies developed against Mce1 proteins, cell fractions were probed for the presence of Mce1 proteins. These experiments demonstrate that all Mce1 proteins, including the lipoprotein LprK, localize preferentially to the cell wall fraction of Mtb confirming their translocation past the IM (Figure 5.1). Notably, Mce1 proteins were not detected in bacterial supernatants, suggesting that in vitro, Mce1 proteins are not secreted (data not shown).



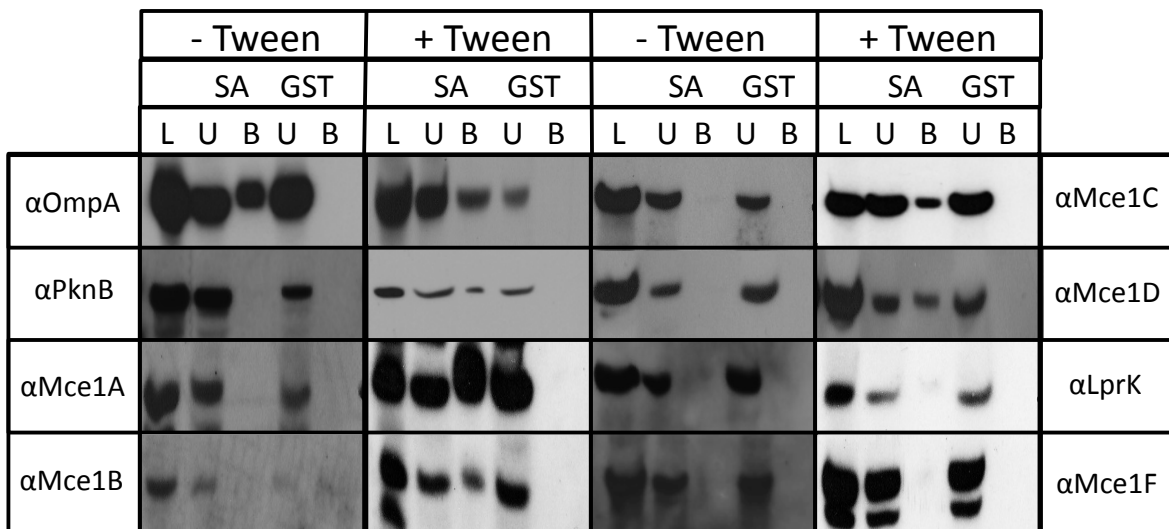
$\alpha$ Mce1A		$\alpha$ Mce1B		$\alpha$ Mce1C	
C	W	C	W	C	W
					
$\alpha$ Mce1D		$\alpha$ LprK		$\alpha$ Mce1F	
C	W	C	W	C	W
					

**Figure 5.1 Mce1 proteins localize to the cell wall fraction of Mtb.** Mtb H37Rv was grown to exponential phase, harvested and fractionated using differential centrifugation as described in materials and methods. Equal volumes of each fraction, C (cytosol) and W (cell wall) were probed via western blot using purified peptide antibodies against appropriate Mce proteins.

#### **Mce1 proteins are differentially localized within the cell wall.**

It became apparent through the above experiments that the detection of many of the Mce1 proteins was dependent on bacterial lysis via a mechanical disruption technique involving glass beads, a common method for mycobacterial cell disruption. This suggested they were buried in the cell wall and prompted investigations into the localization of Mce1 proteins within the cell wall. Despite unsuccessful attempts at localizing Mce1 proteins using transmission electron microscopy, we were able to demonstrate subcellular localization of Mce1 proteins in the cell wall by developing conditions that allowed for the specific biotinylation of cell wall proteins. By combining the labeling of surface exposed primary amines using an NHS-activated biotin molecule with streptavidin immunoprecipitation, the subcellular localization of cell wall proteins was

determined. Figure 5.2 shows that in the absence of the detergent Tween 80, the predicted IM bound serine-threonine kinase PknB (Young et al. 2003), is not accessible to the membrane-impermeable biotin molecule, but is accessible in the presence of the detergent. In contrast, the outer membrane protein OmpATb displays consistent biotin accessibility. These results demonstrate that cell surface biotinylation of Mtb in the absence and presence of Tween 80 can detect both exposed outer membrane proteins and proteins exposed to the periplasm, respectively.



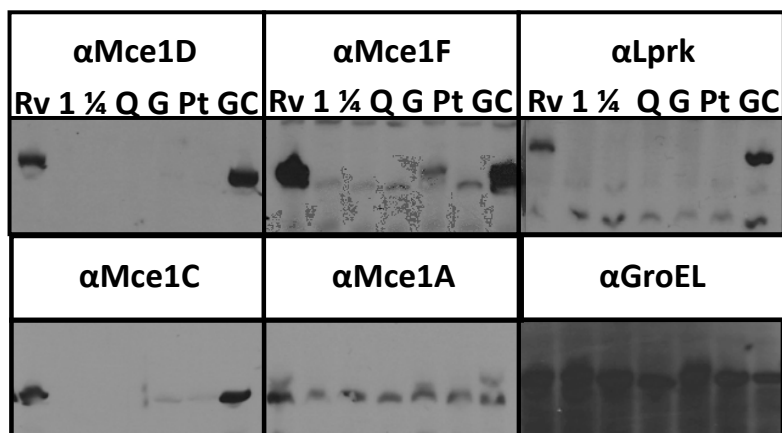
**Figure 5.2 Mce1 proteins are differentially accessible to membrane impermeable biotin.** Mtb H37Rv was grown as described in the presence or absence of 0.05% Tween 80 and biotinylated according to materials and methods. Following mechanical lysis of biotinylated cells, immunoprecipitation of biotinylated molecules using either streptavidin (SA) or GST coated agarose beads was performed. Equal volumes of each (L) lysate, (U) unbound, (B) bound fractions were probed via western blot using purified peptide antibodies.

By probing for the presence of Mce1 proteins in the above conditions, differential biotin accessibility was demonstrated for Mce1 proteins. Mce1A, 1B, 1C, and 1D

are accessible to biotinylation only in the presence of Tween 80, while LprK and Mce1F are inaccessible to biotin regardless of the presence or absence of the detergent. These results indicate that while all Mce1 proteins localize to the cell wall (Figure 5.1) a subset of these including Mce1A, 1B, 1C and 1D are peripheral to the more buried LprK and Mce1F proteins.

### **MceG function is required for Mce1 protein stabilization**

While circumstantial evidence exists suggesting Mce proteins assemble into lipid importer complexes, direct experimental evidence was lacking. The stability of bacterial transporters has been demonstrated to be dependent on the presence of proteins within that complex (Berger and Christie 1993, Fernandez et al. 1996). Consistent with that observation, Joshi and colleagues have shown that the stability of MceG is dependent on expression of Mce1 proteins in *M. smegmatis*. Here, Figure 5.3 demonstrates that in the absence of MceG or an MceG point mutant (K66N) rendering the enzyme unable to bind ATP (Henriksen et al. 2005), Mce1 proteins are not present. Transcription of Mce1 proteins is not altered in these MceG mutants (personal communication, Sassetti, CM). Therefore, the stability of Mce1 proteins is dependent on a functional copy of MceG.



**Figure 5.3 MceG-dependent ATP hydrolysis is required for Mce1 complex stabilization.** Bacteria was grown to exponential phase and mechanically lysed. Equal amounts of lysate, relative to optical density, were probed via western blot using purified peptide antibodies. (Rv) H37Rv, (1) Mce1 operon mutant, (Q) Mce1-4 quadruple mutant, (G) MceG mutant. (Pt) MceG point mutant (K66N), (GC) MceG complement.

## Discussion

In this chapter, Mce1 proteins were localized to the cell wall fraction of Mtb and data is provided to support the presence of an ordered transporter complex spanning the cell wall. The fact that a subset of Mce1 proteins are differentially biotinylated in the presence or absence of Tween 80, argues that some Mce proteins, including the lipoprotein LprK and Mce1F, are more buried in the cell wall than others. Alternatively, we cannot rule out the possibility that the detergent itself can modify the overall structure of the transporter. This caveat is difficult to control for in the highly complex cell wall of Mtb, where structural organization of any cell wall apparatus is not clear. In addition, since the

irreversible biotinylation reaction requires a free amine group, it is possible that Mce1F and Lprk are folded or interacting with other proteins or themselves in a manner that does not provide free amine groups. Although with over 100 lysine residues between the two proteins, this is highly unlikely.

As noted above, bacterial transporter complexes have been shown to display mutual stabilization (Berger and Christie 1993, Fernandez et al. 1996) and here, evidence is provided to suggest the same occurs with Mce transporters.

Although there is similarity regarding complex stabilization, data here is novel in that the energy generating subunit is functionally required for subunit stability.

ATPases are required for generating energy for translocation and in some secretions systems the assembly of the secretion machinery itself (Sagulenko et al. 2001). In this way, MceG could also be providing the energy for translocation and assembly of the Mce1 complex. This would be surprising for two reasons. First, as noted above, all Mce proteins are predicted to contain a signal peptide cleavage site indicating their likely secretion by the classical Sec secretion apparatus. Second, Mce1 is most likely, as in the case of Mce4, a lipid importer not an exporter. Thus, this would be the first example of an ATPase providing both the energy for secretion of complex subunits and also for the import of a substrate. Alternatively, ATP hydrolysis provided by MceG may only be necessary for complex assembly and lipid import across the IM may happen either in an energy-independent fashion, like the 'flip-flop' mechanism of LCFAs

following import by FadL (van den Berg 2005) or another IM transporter may be required for this process. As no other apparent transporter complex was found to be required for cholesterol utilization in chapter III, the latter is unlikely.

While it is clear that the stability of Mce1 proteins require MceG, it is unclear what might be degrading Mce1 proteins in the absence of the ATPase. Genomic analysis suggests that Mtb contains over 40 predicted proteases (Ribeiro-Guimarães and Pessolani 2007). These include ClpX-like proteases as well as periplasmic DegP-like serine proteases, HtrA1-3. Efforts using a DegP-specific irreversible protease inhibitor diisopropyl fluorophosphate (DFP) (Krojer et al. 2010) showed that in the presence of DFP Mce1 protein stability was not restored and may suggest that Mce1 proteins are degraded in the cytosol (data not shown). While the transcription of Mce1 proteins is unaltered in the MceG point mutant (personal communication, Sassetti, CM), it also remains possible that Mce1 proteins are not translated or partially translated in the absence of a functional ATPase. It is possible that similar to SecA-dependent co-translation of translocated proteins, MceG may be intimately involved in the translation/translocation of Mce1 proteins and in the absence of a functional MceG, translation is inhibited (Papanikou et al. 2007).

Mce proteins were originally identified for their ability to promote uptake of non-invasive E.coli in epithelial cells and were therefore named mycobacterial cell

entry proteins. (Arruda et al. 1993). Although many have reproduced similar results using latex beads coated with recombinant Mce proteins (Chitale et al. 2001, Lu et al. 2006, El-Shazly et al. 2007), the mechanism for this observation and its relevance in vivo is not understood. It is possible that lipid-binding outer membrane Mce proteins may interact with mammalian membranes in a manner that could promote uptake.

In vivo phenotypes of Mce mutants have generated mixed results. TraSH experiments have identified Mce1 mutants to be attenuated as early as 1 week following infection, whereas Mce4 mutants display greatest attenuation following the adaptive immune response between weeks 3 and 4 following infection (Sasseti et al. 2003). We have verified these phenotypes using transposon mutants in a competitive infection model (Joshi et al. 2006). In single strain mice infections, Mce1 mutants have been reported to demonstrate a hypervirulence phenotype by week 4 (Shimono et al. 2003). It is not clear why these results are inconsistent, although the immune status of the host might play a role in determining the outcome of Mce1 mutants.

Although we and others have demonstrated that Mce4 functions in cholesterol import, the substrate for Mce1 remains elusive (Pandey and Sasseti 2008, Mohn et al. 2008). Immediately upstream of and encoded in the same direction as the YrbE1 genes, lies FadD5. FadD5 is predicted to encode a fatty acyl-coenzyme A

synthetase, catalyzing the first step in fatty acid degradation (Cole et al. 1998). And although this mutant has been shown to display a reduced in vitro growth rate in media containing mycolic acid as the sole carbon source, this phenotype was independent of the Mce1 operon (Dunphy et al. 2010). Although Mce1 does not play a role in mycolic acid transport, it is likely that it transports another lipid utilized by Mtb. Evidence suggests that lung surfactant contains LCFAs including dipalmitoyl phosphatidylcholine and can be phagocytosed by alveolar macrophages (Gräbner and Meerbach 1991). It has been demonstrated that Mtb can utilize such a lipid in a glyoxylate cycle-dependent fashion and as mentioned above, it is likely this would require a transporter for efficient growth on such a compound. Mtb is bound by the phagosomal membrane; a lipid bilayer containing various lipid carbon sources. Utilizing these lipids would require both their transport and hydrolysis by lipases. Mtb encodes 20 putative lipases, including 4 secreted phospholipases C enzymes required for virulence (Raynaud et al. 2002, Cole et al. 1998). Lipase freed membrane lipids could serve as potential carbon sources and substrates for Mce1. In addition, Mtb can catabolize TAGs, which are freed in macrophages during phagocytosis (Alvarez and Steinbüchel 2002, Mason et al. 1972). Whatever the substrate of the Mce1 transporter, it is required during the early stages of TB infection (Sassetti et al. 2003) and will thus, shed light on the variety of lipids that are likely metabolized by Mtb during this important stage of infection.



## **Chapter VI: Discussion**

Cholesterol is a vital molecule for human biology and thus, it is no surprise that human pathogens exploit it. The role of cholesterol during bacterial and viral infections has been extensively studied. Cholesterol plays important roles in viral budding and replication, bacterial invasion and replication, and evasion of the host immune response (Ye 2007, Garner et al. 2002, Lamberti et al. 2008, Wunder et al. 2006, Golusko and Nowicki 2005, Hayward et al. 2005, Mackenzie et al. 2007). Mycobacteria use cholesterol during various aspects of its life cycle including invasion and inhibition of lysosome fusion (Gatfield and Pieters 2000, de Chastellier and Thilo 2006). Recently, the demonstration that Mtb utilizes cholesterol as a carbon source represented a novel use of this host-derived molecule during the course of TB infection (Pandey and Sassetti 2008, Mohn et al. 2008). Here, I have demonstrated how Mtb has adapted to utilizing this molecule in a rather unique capacity. It is clear that utilizing cholesterol as a carbon source requires both significant metabolic adaptations and a unique dedicated uptake system. The significant adaptations argue that cholesterol is an advantageous intracellular carbon source, but they provide little evidence as to why. And although the answer remains unclear, evidence supporting several hypotheses now exists.

### **The Bioavailability of Cholesterol**

Mtb requires cholesterol during the chronic phase of infection, but catabolizes it throughout the course of infection (Pandey and Sasseti 2008, Chang et al. 2009). Cholesterol therefore, must be available throughout infection. It is possible that Mtb has evolved to rely on cholesterol during discrete times throughout infection based on the bioavailability of various carbon sources, including cholesterol. The plasma membrane is a rich source of cholesterol early during invasion by mycobacteria (Gatfield and Pieters 2000). Because bacterial replication occurs in a phagosome-like vesicle believed to be part of the recycling network, the intracellular membranes including the phagosome membrane itself is likely a source of cholesterol during early replication (Clemens and Horwitz 1996). Filipin, a fluorophore with specific cholesterol binding activity, has been used to localize cholesterol in infected macrophages and these experiments have demonstrated that Mtb co-localizes with cholesterol in a perinuclear area independent of the activation state of the infected macrophage (Pandey and Sasseti 2008). Thus, evidence suggests that Mtb has access to cholesterol-rich intracellular membranes pre- and post-IFN $\gamma$  activation. It is likely though, that following IFN $\gamma$  activation when the phagosome is no longer part of the recycling network, carbon sources may become limiting. This carbon limitation may be driving the requirement of cholesterol during persistence.

As disease progresses, the presence of lipid bodies (LB) containing TAG, phospholipids and cholesterol esters among other lipids, become more abundant

within the infected macrophage (Russell et al. 2009). Both mouse and guinea pig models have shown that early LBs are present as soon as 2 weeks post infection and continue to develop and mature into foam cells following adaptive immunity (Cáceres et al. 2009). In addition, macrophages infected with Mtb develop into foam cells in vitro and this is dependent on stimulation of Toll-like receptor 2 (D'Avila et al. 2006). In humans, foamy macrophages can comprise the majority of infected cells in advanced caseating and cavitating granulomas (Peyron et al. 2008). Foamy macrophages consist of intracellular stores of cholesterol and cholesterol esters as a result of the dysregulation of cellular lipid metabolism. These cells, which are co-infected with Mtb represent a huge reservoir of cholesterol and thus a plentiful carbon source during later stages of infection. Electron microscopy has demonstrated that infected foamy macrophages contain Mtb-infected vesicles that are tightly aligned and in some cases within the same compartment as cholesterol-filled lipid bodies (Peyron et al. 2008, Russell et al. 2009). The cavitation of TB granulomas results in bacteria replicating extracellularly, where cholesterol exists in circulating lipoprotein particles, as well as insoluble cholesterol crystals (Hunter et al. 2007). The bioavailability of cholesterol throughout infection suggests that cholesterol and in particular cholesterol esters become increasingly available and accessible to both intracellular and extracellular bacteria.

### **Cholesterol and Redox Balance**

In addition to using cholesterol because it is available, *Mtb* may also catabolize this molecule because it serves additional purposes during infection. The catabolism of cholesterol causes an increase in propionyl-CoA. In addition to the MCC, the methylmalonyl pathway (MMP) metabolizes propionyl-CoA into methylmalonyl CoA (MMCoA), an important precursor to virulence lipids such as PDIM and SL-1 (Savvi et al. 2008). We and others have shown that these virulence lipids increase in mass and abundance during growth on cholesterol (Pandey and Sasseti 2008, Yang et al. 2009, Jain et al. 2007). It has been proposed that the consumption of NADH by the MMP may be an important redox regulator during growth on lipids, as both  $\beta$ -oxidation and the MCC generate this reducing agent. Thus, in addition to increasing the pathogenicity of the organism, lipid anabolism is thought to serve as a reductant sink for NADH being generated via cholesterol catabolism, which otherwise may cause an imbalance to redox homeostasis (Singh et al. 2009).

### **Cholesterol and Secondary Metabolites**

In addition to providing energy and biosynthetic precursors for lipid anabolism, cholesterol may also serve as a precursor for secondary metabolites such as steroid hormones. Following side-chain cleavage, cholesterol is enzymatically modified into pregnenolone, which is the precursor for glucocorticosteroids. Glucocorticosteroids are well documented as effective immunosuppressants (Nikhil et al. 2004) and other pathogens are believed to modulate cholesterol

levels in the host in order to take advantage of this phenomenon (Reading et al. 2003, Mackenzie et al. 2007). In addition to glucocorticoids, androgens can be synthesized from cholesterol and have demonstrated various effects on the immune system (Gilliver 2010). Nesbitt and colleagues have demonstrated that the precursors to androgens, androsten-4-en-3, 17-dione (AD) and androsta-1,4-dien-3,17-dione (ADD) accumulate in the supernatants of *Mtb* grown in the presence of cholesterol in vitro (Nesbitt et al. 2010).

An important enzymatic step in the conversion of cholesterol into steroid hormones is the oxidation of the 3 $\beta$ -hydroxyl group. Here, of two potential enzymes predicted to mediate this step, *rv1106c*, encoding a 3 $\beta$ -hydroxysteroid dehydrogenase is the only gene that when mutated causes a growth defect in cholesterol. Rv1106c is predicted to be intracellularly expressed (Kreit and Sampson 2009). As most intracellular 3 $\beta$ -hydroxysteroid dehydrogenases are involved in steroid hormone biosynthesis, it is possible that Rv1106c may be involved in both the catabolism of cholesterol and the synthesis of steroid hormone precursors (Kreit and Sampson 2009).

There are two reasons that make the conversion of cholesterol to steroid hormones an unlikely alternative for cholesterol utilization in *Mtb*. First, the accumulation of AD and ADD in cell culture supernatants is likely an in vitro artifact with little relevance in vivo. In fact, metabolite leakage in vitro is a well-

documented phenomenon (Shin et al. 2010). Second, the release of C<sub>14</sub> C-4 into CO<sub>2</sub>, indicates that Mtb has the ability to completely degrade the steroid nucleus. The production of both glucocorticoids (C<sub>21</sub> molecules) and androgens (C<sub>19</sub> molecules) require an intact steroid nucleus.

### **Cholesterol and Pyruvate**

Several lines of evidence from this work support the importance of metabolizing pyruvate during growth on cholesterol. First, pyruvate is a direct byproduct of sterol ring degradation (Yang et al. 2009). In addition, growth on cholesterol is dependent on a functional MCC, which converts toxic propionate into additional pyruvate. Second the accumulation of alanine, tryptophan and serine indicate increased flux through biosynthetic pathways stemming from pyruvate. Third, the requirement of *ppdK* for growth on cholesterol indicates that the metabolism of pyruvate to PEP, and more generally gluconeogenesis, is an important metabolic adaptation to growth on cholesterol. In addition, mutations to lactate dehydrogenase (*lldL1*, *rv0694*) result in a significant growth defect in cholesterol. Lactate dehydrogenase mediates the interconversion of lactate and pyruvate, consuming NADH during the conversion of pyruvate to lactate. One possible explanation for the requirement of lactate dehydrogenase is that the production of pyruvate during cholesterol catabolism requires many outlets for pyruvate metabolism, including the conversion to lactate. LldL mediated conversion of pyruvate to lactate would also consume excess NADH produced during

cholesterol metabolism helping to restore redox balance. Alternatively, the conversion of lactate to pyruvate may be necessary to feed gluconeogenesis needed to produce biosynthetic precursors essential for cell viability, including tryptophan and serine, both of which accumulated following growth on cholesterol. Further analysis is needed to differentiate between both possibilities, but the requirement of lactate dehydrogenase further supports the importance of pyruvate metabolism during growth on cholesterol.

### **Cholesterol and TB Susceptibility**

The importance of cholesterol during the course of TB infection leads one to consider if patients with altered cholesterol levels are more susceptible to disease. While studies have suggested that TB patients are hypocholesterolemic (Pérez-Guzmán and Vargas 2006, Deniz et al. 2007), it is currently unknown if pre-existing hypercholesterolemia predisposes patients to TB. In an attempt to answer this question, mice with defects in lipoprotein processing were fed high-cholesterol diets and then infected with *Mtb* via the aerosol route. Interestingly, these mice were more susceptible to *Mtb* and this was dependent on elevated serum total cholesterol levels (Martens et al. 2008). While the same has yet to be observed in human patients, these results indicate that cholesterol plays a pivotal role in TB susceptibility.

In summary, the work presented here highlights the complexity of Mtb physiology and the lengths at which this organism has gone to adapt to its unique niche. As humans are the sole reservoir for this highly infectious pathogen, it is evident that Mtb has evolved to be metabolically flexible enough to utilize its available carbon sources. By better understanding these adaptations, we may continue to elucidate novel pathways to target with unique antitubercular drugs. For example, the absence of the isocitrate lyase in the mammalian host represents a novel drug target that is vital for intracellular growth on cholesterol. In addition, the design of drugs aimed at causing the buildup of toxic metabolic intermediates such as propionate, may represent a novel avenue for drug therapy. It is clear that a better understanding of basic Mtb physiology will continue to elucidate the mechanisms used by such an effective pathogen.



## Appendix

**Supplementary Table 2.1 Sequence reads detected per TA insertion site.**

**Supplementary Table 2.2 Analysis of essentiality in vitro.**

**Supplementary Table 3.1 Differential growth in glycerol and cholesterol media.**

The above supplementary tables are too large to include here, but are available in the supplementary data of, J.E. Griffin, J.D. Gawronski, M.A. DeJesus, T.R. Ioerger, B.J. Akerley, C.M. Sassetti *Quantitative high-resolution phenotypic profiling defines genes essential for Mycobacterium tuberculosis survival and cholesterol catabolism*

**Supplementary Table 4.1 Carbon source-dependent metabolomic profiling.**

“Ratio Cholesterol/Glycerol” = relative abundance of metabolite in cholesterol / relative abundance of metabolite in glycerol

“P-Value” = calculated using student t-test on quadruplicate samples

BIOCHEMICAL NAME	Ratio Cholesterol / Glycerol	P- VALUE
5-methylthioadenosine (MTA)	2.19	0.0854
phenylalanine	1.47	0.1810
tryptophan	2.07	0.0226
tyrosine	1.39	0.2113
2-aminoadipate	5.54	0.0022
alanine	2.02	0.0270
asparagine	1.02	0.7070
aspartate	1.16	0.7792
beta-alanine	4.44	0.0098
cadaverine	0.22	0.2884
diaminopimelate	0.65	0.9722
homoserine	1.35	0.2703
lysine	3.39	0.0424

<b>N2-acetyllysine</b>	<b>0.09</b>	<b>0.1629</b>
<b>N6-acetyllysine</b>	<b>1.43</b>	<b>0.1737</b>
<b>N-6-trimethyllysine</b>	<b>1.82</b>	<b>0.0528</b>
<b>O-acetylhomoserine</b>	<b>1.17</b>	<b>0.5363</b>
<b>pipecolate</b>	<b>2.15</b>	<b>0.0230</b>
<b>S-adenosylhomocysteine (SAH)</b>	<b>1.85</b>	<b>0.0432</b>
<b>threonine</b>	<b>2.86</b>	<b>0.0218</b>
<b>isoleucine</b>	<b>2.33</b>	<b>0.0376</b>
<b>beta-hydroxyisovalerate</b>	<b>1.61</b>	<b>0.3293</b>
<b>beta-methyllevulinate</b>	<b>1.65</b>	<b>0.2664</b>
<b>leucine</b>	<b>2.15</b>	<b>0.0506</b>
<b>methylsuccinate</b>	<b>16.28</b>	<b>&lt; 0.001</b>
<b>valine</b>	<b>1.86</b>	<b>0.0681</b>
<b>2-aminobutyrate</b>	<b>1.33</b>	<b>0.2716</b>
<b>2-pyrrolidinone</b>	<b>2.07</b>	<b>0.1016</b>
<b>4-acetamidobutanoate</b>	<b>1.50</b>	<b>0.2523</b>
<b>4-hydroxybutyrate (GHB)</b>	<b>2.33</b>	<b>0.0943</b>
<b>arginine</b>	<b>1.42</b>	<b>0.3167</b>
<b>citrulline</b>	<b>0.74</b>	<b>0.1119</b>
<b>X - 11568 - retired for ergothioneine</b>	<b>1.57</b>	<b>0.1395</b>
<b>gamma-aminobutyrate (GABA)</b>	<b>1.45</b>	<b>0.2528</b>
<b>glutamate</b>	<b>1.53</b>	<b>0.2388</b>
<b>glutamate, gamma-methyl ester</b>	<b>1.45</b>	<b>0.3784</b>
<b>glutamine</b>	<b>1.83</b>	<b>0.1213</b>
<b>histidine</b>	<b>2.90</b>	<b>0.0195</b>
<b>N-acetylglutamate</b>	<b>1.45</b>	<b>0.2500</b>
<b>N-acetylorithine</b>	<b>2.19</b>	<b>0.0079</b>
<b>ornithine</b>	<b>1.81</b>	<b>0.2121</b>
<b>proline</b>	<b>0.88</b>	<b>0.8975</b>
<b>5-oxoproline</b>	<b>1.48</b>	<b>0.2714</b>
<b>gamma-glutamyl-2-aminobutyrate</b>	<b>0.74</b>	<b>0.3052</b>
<b>betaine</b>	<b>0.77</b>	<b>0.6220</b>
<b>cysteine</b>	<b>1.22</b>	<b>0.5703</b>
<b>glycine</b>	<b>0.95</b>	<b>0.7994</b>
<b>serine</b>	<b>2.81</b>	<b>0.0259</b>
<b>alanylalanine</b>	<b>1.62</b>	<b>0.0546</b>
<b>gamma-glutamylalanine</b>	<b>1.52</b>	<b>0.0674</b>
<b>gamma-glutamylglutamate</b>	<b>1.55</b>	<b>0.2489</b>
<b>gamma-glutamylglutamine</b>	<b>1.25</b>	<b>0.1966</b>
<b>gamma-glutamylisoleucine*</b>	<b>1.87</b>	<b>0.0899</b>
<b>gamma-glutamylleucine</b>	<b>1.31</b>	<b>0.3244</b>
<b>gamma-glutamylmethionine</b>	<b>1.21</b>	<b>0.3419</b>

X - 10821 - retired for gamma-glutamylmethionine	1.41	0.2843
gamma-glutamylphenylalanine	1.02	0.7569
gamma-glutamylthreonine*	1.90	0.0205
gamma-glutamyltyrosine	0.92	0.9901
gamma-glutamylvaline	1.58	0.1143
adenosine-5'-diphosphoglucose	3.45	0.0103
3-phosphoglycerate	1.00	
dihydroxyacetone phosphate (DHAP)	1.30	0.3910
2-isopropylmalate	1.57	0.3252
fructose-6-phosphate	2.04	0.1247
glucose	2.16	0.4840
glucose 1-phosphate	2.15	0.4424
glucose-6-phosphate (G6P)	3.17	0.0582
glycerate	1.20	0.4595
Isobar: fructose 1,6-diphosphate, glucose 1,6-diphosphate	1.00	
pyruvate	1.97	0.1416
inositol 1-phosphate (I1P)	0.63	0.4418
myo-inositol	0.78	0.7872
glycolate (hydroxyacetate)	1.43	0.1634
Isobar: ribulose 5-phosphate, xylulose 5-phosphate	1.10	0.8090
ribose 5-phosphate	2.70	0.1055
sedoheptulose-7-phosphate	2.50	0.1110
fructose	1.69	0.2470
GDP-mannose	1.27	0.6003
malitol	2.63	0.0826
maltose	2.82	0.0618
maltotetraose	1.64	0.3749
maltotriose	2.06	0.1699
mannose	4.02	0.0494
mannose-1-phosphate	5.34	0.0125
mannose-6-phosphate	3.06	0.0507
sorbitol	0.98	0.7253
trehalose	1.58	0.1840
trehalose 6-phosphate	5.17	0.0224
X - 11011 - retired for 2-methylcitrate	22.81	< 0.001
alpha-ketoglutarate	2.42	0.1283
cis-aconitate	1.94	0.0690
citrate	4.02	0.0054
fumarate	2.95	0.0371

malate	5.49	0.0120
succinate	4.67	0.0022
succinyl CoA	1.91	0.1835
1-octadecanol	0.63	0.7341
arachidate (20:0)	0.70	0.3207
caproate (6:0)	1.49	0.2339
caprylate (8:0)	1.18	0.3959
eicosenoate (20:1n9 or 11)	1.07	0.8025
linoleamide (18:2n6)	0.28	0.0514
linoleate (18:2n6)	1.17	0.8898
margarate (17:0)	0.94	0.8777
myristate (14:0)	1.08	0.5795
n-Butyl Oleate	2.01	0.3080
oleamide	0.11	< 0.001
oleate (18:1n9)	1.97	0.3420
palmitate (16:0)	0.94	0.8019
pentadecanoate (15:0)	1.18	0.4794
stearate (18:0)	0.79	0.3902
X - 13183 - retired for stearamide	0.47	0.0196
1-stearoylglycerol (1-monostearin)	1.04	0.9502
glycerol	0.09	0.0378
1-palmitoylglycerophosphoinositol*	0.91	0.7839
glycerol 2-phosphate	0.22	0.0089
glycerol 3-phosphate (G3P)	0.07	0.0013
glycerophosphorylcholine (GPC)	1.00	
7-alpha-hydroxycholesterol	8.90	0.0060
7-beta-hydroxycholesterol	4.41	0.0616
cholesterol	1.58	0.3172
2'-deoxyadenosine	1.39	0.2563
adenine	4.29	0.0586
adenosine	1.27	0.3263
adenosine 3',5'-cyclic monophosphate (cAMP)	1.11	0.7701
adenosine 3'-monophosphate (3'-AMP)	2.48	0.1328
adenosine 5'-monophosphate (AMP)	1.73	0.0305
guanosine	0.94	0.5504
guanosine 5'- monophosphate (GMP)	1.80	0.1436
inosine	0.72	0.8938
thymidine 5'-monophosphate	3.38	0.0294
acetyl CoA	1.73	0.2986
pantothenate	5.00	0.0024
adenosine 5'diphosphoribose	2.50	0.2568

nicotinamide adenine dinucleotide (NAD <sup>+</sup> )	1.00	0.7441
nicotinate	3.66	0.0264
nicotinate ribonucleoside*	2.91	0.0338
nicotinic acid mononucleotide (NaMN)	2.20	0.0011
methylphosphate	0.43	0.7192
phosphate	1.11	0.9654
flavin adenine dinucleotide (FAD)	2.32	0.0944
flavin mononucleotide (FMN)	1.81	0.1027
alpha-tocopherol	1.00	
diisopropanolamine	1.73	0.2027
triethyleneglycol	1.67	0.1006
X - 07887- retired for p-toluic acid	1.84	0.3311
X - 02973	2.62	0.0197
X - 03094	0.48	0.1306
X - 04033	1.66	0.1780
X - 04044	1.58	0.1119
X - 04768	1.59	0.3000
X - 05407	1.48	0.3071
X - 05437	0.39	0.0708
X - 06346	1.32	0.6263
X - 08754	2.58	0.0388
X - 08889	1.50	0.3927
X - 08988	2.65	0.2003
X - 10889	2.44	0.4393
X - 11251	1.93	0.3696
X - 11297	0.97	0.8966
X - 11391	1.74	0.3745
X - 11394	1.70	0.4539
X - 11395	1.82	0.4523
X - 11404	1.46	0.5208
X - 11406_200	2.25	0.0014
X - 11453	1.84	0.4326
X - 11455	1.43	0.6571
X - 11456	1.97	0.1755
X - 11457	1.92	0.4031
X - 11461	2.04	0.3025
X - 11509	0.74	0.3543
X - 11533	1.04	0.5873
X - 11550	1.03	0.7752
X - 11687_200	2.29	0.0524
X - 11730	1.21	0.8751
X - 11795	1.89	0.0473

<b>X - 11809</b>	<b>0.90</b>	<b>0.9005</b>
<b>X - 11840</b>	<b>1.38</b>	<b>0.5042</b>
<b>X - 11875</b>	<b>0.95</b>	<b>0.5852</b>
<b>X - 11910</b>	<b>1.18</b>	<b>0.5004</b>
<b>X - 12051</b>	<b>0.81</b>	<b>0.3731</b>
<b>X - 12660</b>	<b>1.17</b>	<b>0.6596</b>
<b>X - 12776</b>	<b>0.98</b>	<b>0.8125</b>
<b>X - 12805</b>	<b>1.68</b>	<b>0.2807</b>
<b>X - 12889_201</b>	<b>1.92</b>	<b>0.0842</b>
<b>X - 13005</b>	<b>1.26</b>	<b>0.4297</b>
<b>X - 13093</b>	<b>0.26</b>	<b>0.2625</b>
<b>X - 13149</b>	<b>0.67</b>	<b>0.2942</b>
<b>X - 13168</b>	<b>0.89</b>	<b>0.6008</b>
<b>X - 13548</b>	<b>1.73</b>	<b>0.4464</b>
<b>X - 13549</b>	<b>2.15</b>	<b>0.3927</b>
<b>X - 13645</b>	<b>2.82</b>	<b>0.1428</b>
<b>X - 13672</b>	<b>1.16</b>	<b>0.6364</b>
<b>X - 13737</b>	<b>0.57</b>	<b>0.2719</b>
<b>X - 13828</b>	<b>0.94</b>	<b>0.6603</b>
<b>X - 13872</b>	<b>1.08</b>	<b>0.6826</b>
<b>X - 14318</b>	<b>0.53</b>	<b>0.1841</b>
<b>X - 14359</b>	<b>1.19</b>	<b>0.9505</b>
<b>X - 14568</b>	<b>2.06</b>	<b>0.0233</b>
<b>X - 14575</b>	<b>1.38</b>	<b>0.8869</b>
<b>X - 14588</b>	<b>1.52</b>	<b>0.3618</b>
<b>X - 14955</b>	<b>4.40</b>	<b>0.0044</b>
<b>X - 15138</b>	<b>0.92</b>	<b>0.6365</b>
<b>X - 15177</b>	<b>1.39</b>	<b>0.3367</b>
<b>X - 15261</b>	<b>1.58</b>	<b>0.0609</b>
<b>X - 15279</b>	<b>2.53</b>	<b>0.0795</b>

## Bibliography

- Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, Mekalanos JJ. 2002. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci* **99**: 966-71.
- Alvarez HM, Steinbüchel A. 2002. Triacylglycerols in prokaryotic microorganisms. *Appl Microbiol Biotechnol* **60**: 367-76.
- Ansorge WJ. 2009. Next-generation DNA sequencing techniques. *N Biotechnol* **25**: 195-203.
- Appelberg R. 2006. Macrophage nutritive antimicrobial mechanisms. *J Leukoc Biol* **79**: 1117-1128.
- Armstrong JA, Hart PD. 1971. Response to cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. *J Exp Med* **134**: 713-40.
- Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW. 1993. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261**: 1454-7.
- Av-Gay Y, Everett M. 2000. The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**: 238-44.
- Badarinarayana V, Estep PW 3rd, Shendure J, Edwards J, Tavazoie S, Lam F, Church GM. 2001. Selection analyses of insertional mutants using subgenic-resolution arrays. *Nat Biotechnol* **19**: 1060-5.
- Benziman M, Eizen N. 1971. Pyruvate-Phosphate Dikinase and the Control of Gluconeogenesis in *Acetobacter xylinum*. *J Biol Chem* **246**: 57-61.
- Berger BR, Christie PJ. 1993. The *Agrobacterium tumefaciens* virB4 gene product is an essential virulence protein requiring an intact nucleoside triphosphate-binding domain. *J Bacteriol* **175**: 1723-34.
- Bloch H, Segal W. 1956. Biochemical differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. *J Bacteriol* **72**: 132-41.
- Brennan PJ. 2003. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* **83**: 91-7.

Brzostek A, Dziadek B, Rumijowska-Galewicz A, Pawelczyk J, Dziadek J. 2007. Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* **275**: 106-12.

Cáceres N, Tapia G, Ojanguren I, Altare F, Gil O, Pinto S, Vilaplana C, Cardona PJ. 2009. Evolution of foamy macrophages in the pulmonary granulomas of experimental tuberculosis models. *Tuberculosis* **89**: 175-82.

Calmette A. 1931. Preventive vaccination against tuberculosis with BCG. *Proc R Soc Med* **24**: 1481-90.

Capyk JK, D'Angelo I, Strynadka NC, Eltis LD. 2009a. Characterization of 3-ketosteroid 9 $\alpha$ -hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*. *J Biol Chem* **284**: 9937-46.

Capyk JK, Kalscheuer R, Stewart GR, Liu J, Kwon H, Zhao R, Okamoto S, Jacobs WR Jr, Eltis LD, Mohn WW. 2009b. Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. *J Biol Chem* **284**: 35534-42.

Casali N, Riley LW. 2007. A phylogenomic analysis of the actinomycetales mce operons. *BMC Genomics* **8**: 60

Cave AJE. 1939. The evidence for the incidence of tuberculosis in ancient Egypt. *Br J Tuberc* **33**: 142-52.

Chan J, Tanaka K, Carroll D, Flynn J, Bloom BR. 1995. Effects of nitric oxide synthase inhibitors on murine infection with *Mycobacterium tuberculosis*. *Infect Immun* **63**: 736-740.

Chang JC, Harik NS, Liao RP, Sherman DR. 2007. Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *J Infect Dis* **196**: 788-95.

Chang JC, Miner MD, Pandey AK, Gill WP, Harik NS, Sassetti CM, Sherman DR. 2009. igr Genes and *Mycobacterium tuberculosis* Cholesterol Metabolism. *J Bacteriol* **191**: 5232-9.

Chitale S, Ehrt S, Kawamura I, Fujimura T, Shimono N, Anand N, Lu S, et al. 2001. Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. *Cell Microbiol* **3**: 247-54.



Clemens DL, Horwitz MA. 1996. The Mycobacterium tuberculosis phagosome interacts with early endosomes and is accessible to exogenously administered transferrin. *J Exp Med* **184**: 1349-55.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**: 537-44.

Cramer A, Auchter M, Frunzke J, Bott M, Eikmanns BJ. 2007. RamB, the transcriptional regulator of acetate metabolism in Corynebacterium glutamicum, is subject to regulation by RamA and RamB. *J Bacteriol* **189**: 1145-1149.

Daniel TM. 2004. René Théophile Hyacinthe Laënnec and the founding of pulmonary medicine. *Int J Tuberc Lung Dis* **8**: 517-518.

Daniel TM. 2006. The History of tuberculosis. *Respir Med* **11**: 1862-70.

D'Avila H, Melo RC, Parreira GG, Werneck-Barroso E, Castro-Faria-Neto HC, Bozza PT. 2006. Mycobacterium bovis bacillus Calmette-Guérin induces TLR2-mediated formation of lipid bodies: intracellular domains for eicosanoid synthesis in vivo. *J Immunol* **176**: 3087-97.

Davidson AL, Chen J. 2004. ATP-binding cassette transporters in bacteria. *Annu Rev Biochem* **73**: 241-68.

de Carvalho LP, Fischer SM, Marrero J, Nathan C, Ehrt S, et al. 2010. Metabolomics of Mycobacterium tuberculosis reveals compartmentalized catabolism of carbon substrates. *Chem Biol* **17**: 1122-1131.

de Chastellier C, Thilo L. 2006. Cholesterol depletion in Mycobacterium avium-infected macrophages overcomes the block in phagosome maturation and leads to the reversible sequestration of viable mycobacteria in phagolysosome-derived autophagic vacuoles. *Cell Microbiol* **8**: 242-56.

Deniz O, Gumus S, Yaman H, Ciftci F, Ors F, Cakir E, Tozkoparan E, et al. 2007. Serum total cholesterol, HDL-C and LDL-C concentrations significantly correlate with the radiological extent of disease and the degree of smear positivity in patients with pulmonary tuberculosis. *Clin Biochem* **40**: 162-6.

Dresen C, Lin LY, D'Angelo I, Tocheva EI, Strynadka N, Eltis LD. 2010. A flavin-dependent monooxygenase from Mycobacterium tuberculosis involved in cholesterol catabolism. *J Biol Chem* **285**: 22264-75.

- Dubnau E, Fontán P, Manganeli R, Soares-Appel S, Smith I. 2002. Mycobacterium tuberculosis genes induced during infection of human macrophages. *Infect Immun* **70**: 2787-95.
- Dunphy KY, Senaratne RH, Masuzawa M, Kendall LV, Riley LW. 2010. Attenuation of Mycobacterium tuberculosis functionally disrupted in a fatty acyl-coenzyme A synthetase gene fadD5. *J Infect Dis* **201**: 1232-9.
- El-Shazly S, Ahmad S, Mustafa AS, Al-Attiya R, Krajci D. 2007. Internalization by HeLa cells of latex beads coated with mammalian cell entry (Mce) proteins encoded by the mce3 operon of Mycobacterium tuberculosis. *J Med Microbiol* **56**: 1145-51.
- Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. 2009. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* **81**: 6656-6667.
- Fernandez D, Spudich GM, Zhou XR, Christie PJ. 1996. The Agrobacterium tumefaciens VirB7 lipoprotein is required for stabilization of VirB proteins during assembly of the T-complex transport apparatus. *J Bacteriol* **178**: 3168-76.
- Fine PE. 1995. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* **346**: 1339-45.
- Flynn JL, Chan J. 2001. Immunology of tuberculosis. *Annu Rev Immunol* **19**: 93-129.
- Gallagher LA, Shendure J, Manoil C. 2011. Genome-Scale Identification of Resistance Functions in Pseudomonas aeruginosa Using Tn-seq. *MBio* **2**: e00315-10.
- Garner MJ, Hayward RD, Koronakis V. 2002. The Salmonella pathogenicity island 1 secretion system directs cellular cholesterol redistribution during mammalian cell entry and intracellular trafficking. *Cell Microbiol* **4**: 153-65.
- Gatfield J, Pieters J. 2000. Essential role for cholesterol in entry of mycobacteria into macrophages. *Science* **288**: 1647-50.
- Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci* **106**: 16422-7.

Gerstmeir R, Cramer A, Dangel P, Schaffer S, Eikmanns BJ. 2004. RamB, a novel transcriptional regulator of genes involved in acetate metabolism of *Corynebacterium glutamicum*. *J Bacteriol* **186**: 2798-2809.

Gilliver SC. 2010. Sex steroids as inflammatory regulators. *J Steroid Biochem Mol Biol* **120**: 105-15.

Goluszko P, Nowicki B. 2005. Membrane cholesterol: a crucial molecule affecting interactions of microbial pathogens with mammalian cells. *Infect Immun* **73**: 7791-6.

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**: 279-89.

Gould TA, van de Langemheen H, Muñoz-Elías EJ, McKinney JD, Sacchettini JC. 2006. Dual role of the isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in *Mycobacterium tuberculosis*. *Mol Microbiol* **61**: 940-7.

Gräbner R, Meerbach W. 1991. Phagocytosis of surfactant by alveolar macrophages in vitro. *Am J Physiol* **261**: L472-7.

Grigg ERN. 1958a. The arcana of tuberculosis with a brief epidemiologic history of the disease in the USA. *Am Rev Tuberc Pulm Dis* **78**: 151-72.

Grigg ERN. 1958b. The arcana of tuberculosis with a brief epidemiologic history of the disease in the U.S.A. Part III. *Am Rev Tuberc Pulm Dis* **78**: 426-53.

Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, et al. 2004. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **51**: 359-370.

Hayward RD, Cain RJ, McGhie EJ, Phillips N, Garner MJ, Koronakis V. 2005. Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol Microbiol* **56**: 590-603.

Henriksen U, Gether U, Litman T. 2005. Effect of Walker A mutation (K86M) on oligomerization and surface targeting of the multidrug resistance transporter ABCG2. *J Cell Sci* **118**: 1417-26.

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, et al. 2008. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. *PLoS Biol* **6**: e311.

- Hoffmann C, Leis A, Niederweis M, Plitzko JM, Engelhardt H. 2008. Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilary structure. *Proc Natl Acad Sci USA* **105**: 3963-7.
- Hollenstein K, Dawson RJ, Locher KP. 2007. Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* **17**: 412-8.
- Homolka S, Neimann S, Russell DG, Rohde KH. 2010. Functional genetic diversity among Mycobacterium tuberculosis complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathogens* **6**: e1000988.
- Hondalus MK, Bardarov S, Russell R, Chan J, Jacobs WR Jr. Bloom BR. 2000. Attenuation of and protection induced by a leucine auxotroph of Mycobacterium tuberculosis. *Infect Immun* **68**: 2888-2898.
- Horsburgh CR, Jr. 2004. Priorities for the treatment of latent tuberculosis infection in the United States. *N Engl J Med* **350**: 2060-2067.
- Hu Y, van der Geize R, Besra GS, Gurcha SS, Liu A, Rohde M, Singh M, Coates A. 2010. 3-Ketosteroid 9alpha-hydroxylase is an essential factor in the pathogenesis of Mycobacterium tuberculosis. *M Microbiol* **75**: 107-21.
- Hunter RL, Jagannath C, Actor JK. 2007. Pathology of postprimary tuberculosis in humans and mice: contradiction of long-held beliefs. *Tuberculosis* **87**: 267-78.
- Hutchinson CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286**: 2165-9.
- Jackson M, Phalen SW, Lagranderie M, Ensergueix D, Chavarot P, Marchal G, et al. 1999. Persistence and protective efficacy of a Mycobacterium tuberculosis auxotroph vaccine. *Infect Immun* **67**: 2867-73.
- Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, et al. 2003. Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc Natl Acad Sci* **100**: 14339-44.
- Jain M, Petzold CJ, Schelle MW, Leavell MD, Mougous JD, et al. 2007. Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. *Proc Natl Acad Sci USA* **104**: 5133-5138.

- Johnston JB, Ouellet H, Ortiz de Montellano PR. 2010. Functional Redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. *J Biol Chem* **285**: 36352-60.
- Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sasseti CM. 2006. Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc Natl Acad Sci* **103**: 11760-5.
- Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, et al. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* **6**: 364.
- Karakousis PC, Bishai WR, Dorman SE. 2004. *Mycobacterium tuberculosis* cell envelope lipids and the host immune response. *Cell Microbiol* **6**: 105-16.
- Kaufmann SH. 2001. How can immunology contribute to the control of tuberculosis? *Nature Rev Immunol* **1**: 20-30.
- Kaufmann SH. 2010. Future vaccination strategies against tuberculosis: thinking outside the box. *Immunity* **33**: 567-77.
- Kaur D, Guerin ME, Skovierová H, Brennan PJ, Jackson M. 2009. Chapter 2: Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*. *Adv Appl Microbiol* **69**: 23-78.
- Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, Gordon SV. 2005. The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: implications for in vivo growth. *Mol Microbiol* **56**: 163-74.
- Krause AK. 1928. Tuberculosis and public health. *Am Rev Tuberc* **18**: 271-322.
- Kreit J, Sampson NS. 2009. Cholesterol oxidase: physiological functions. *FEBS J* **276**: 6844-56.
- Krojer T, Sawa J, Huber R, Clausen T. 2010. HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues. *Nat Struct Mol Biol* **17**: 844-52.
- Kumar A, Bose M, Brahmachari V. 2003. Analysis of expression profile of mammalian cell entry (mce) operons on *Mycobacterium tuberculosis*. *Infect Immun* **71**: 6083-7.

Kumar A, Deshane JS, Crossman DK, Bolisetty S, Yan BS, et al. 2008. Heme oxygenase-1-derived carbon monoxide induces the Mycobacterium tuberculosis dormancy regulon. *J Biol Chem* **283**: 18032-18039.

Lack NA, Yam KC, Lowe ED, Horsman GP, Owen RL, Sim E, Eltis LD. 2010. Characterization of a carbon-carbon hydrolase from Mycobacterium tuberculosis involved in cholesterol metabolism. *J Biol Chem* **285**: 434-43.

Lamberti Y, Perez Vidakovics ML, van der Pol LW, Rodríguez ME. 2008. Cholesterol-rich domains are involved in Bordetella pertussis phagocytosis and intracellular survival in neutrophils. *Microb Pathog* **44**: 501-11.

Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR. 2003. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to Mycobacterium tuberculosis. *Proc Natl Acad Sci* **100**: 7213-8.

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G et al. 2009. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* **19**: 2308-16.

Lee MH, Pascopella L, Jacobs WR, Jr., Hatfull GF. 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc Natl Acad Sci USA* **88**: 3111-3115.

Leemans JC, Juffermans NP, Florquin S, van Rooijen N, Vervoordeldonk MJ, Verbon A, et al. 2001. Depletion of alveolar macrophages exerts protective effects in pulmonary tuberculosis in mce. *J Immunol* **166**: 4604-11.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: Short oligonucleotide alignment program. *Bioinformatics* **24**: 713-4.

Liu J, Rosenberg EY, Nikaido H. 1995. Fluidity of the lipid domain of cell wall from Mycobacterium chelonae. *Proc Natl Acad Sci USA* **92**: 11254-8.

Liu PT and Modlin RL. 2008. Human macrophage host defense against Mycobacterium tuberculosis. *Curr Opin Immunol* **20**: 371-6.

LoBue PA, Enarson DA, Thoen TC. 2010. Tuberculosis in humans and its epidemiology, diagnosis and treatment in the United States. *Int J Tuberc Lung Dis* **14**: 1226-32.

Lu S, Tager LA, Chitale S, Riley LW. 2006. A cell-penetrating peptide derived from mammalian cell uptake protein of *Mycobacterium tuberculosis*. *Anal Biochem* **353**: 7-14.

Mackenzie JM, Khromykh AA, Parton RG. 2007. Cholesterol manipulation by West Nile virus perturbs the cellular immune response. *Cell Host Microbe* **2**: 229-39.

Mailaender C, Reiling N, Engelhardt H, Bossman S, Ehlers S, Niederweis M. 2004. The MspA porin promotes growth and increases antibiotic susceptibility of both *Mycobacterium bovis* BCG and *Mycobacterium tuberculosis*. *Microbiology* **150**: 853-64.

Marrero J, Rhee KY, Schnappinger D, Pethe K, Ehrt S. 2010. Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection. *Proc Natl Acad Sci* **107**: 9819-24.

Martens GW, Arian MC, Lee J, Ren F, Vallerkog T, Kornfeld H. 2008. Hypercholesterolemia impairs immunity to tuberculosis. *Infect Immun* **76**: 3464-72.

Mason RJ, Stossel TP, Vaughan M. 1972. Lipids of alveolar macrophages, polymorphonuclear leukocytes, and their phagocytic vesicles. *J Clin Invest* **51**: 2399-407.

McAdam RA, Weisbrod TR, Martin J, Scuderi JD, Brown AM et al. 1995. In vivo growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect Immun* **63**: 1004-12.

McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, et al. 2000. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**: 735-8.

Micklinghoff JC, Breiting KJ, Schmidt M, Geffers R, Eikmanns BJ, Bange FC. 2009. Role of the Transcriptional Regulator RamB (Rv0465c) in the Control of the Glyoxylate Cycle in *Mycobacterium tuberculosis*. *J Bacteriol* **191**: 7260-9.

Mohn WW, van der Geize R, Stewart GR, Okamoto S, Liu J, et al. 2008. The actinobacterial mce4 locus encodes a steroid transporter. *J Biol Chem* **283**: 35368-74.

Morse D, Brothwell DR, Ucko PJ. 1964. Tuberculosis in ancient Egypt. *Am Rev Respir Dis* **90**: 5224-541.

Muñoz-Elías EJ, McKinney JD. 2005. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11**: 638-44.

Muñoz-Elías EJ, Upton AM, Cherian J, McKinney JD. 2006. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. *Mol Microbiol* **60**: 1109-1122.

Nesbitt NM, Yang X, Fontán P, Kolesnikova I, Smith I, Sampson NS, Dubnau E. 2010. A thiolase of Mycobacterium tuberculosis is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infect Immun* **78**: 275-82.

Niederweis M, Danilchanka O, Huff J, Hoffmann C, Engelhardt H. 2010. Mycobacterial outer membranes: in search of proteins. *Trends Microbiol* **18**: 109-16.

Nikaido H. 2003. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**: 593-656.

Nikhil S, Reider MJ, Tucker, MJ. 2004. Mechanisms of glucocorticoid-mediated anti-inflammatory and immunosuppressive action. *Paediatr Perinat Drug Ther* **7**:31-42.

Østerås M, Driscoll BT, Finan TM. 1997. Increased pyruvate orthophosphate dikinase activity results in an alternative gluconeogenic pathway in Rhizobium (Sinorhizobium) meliloti. *Microbiology* **143**: 1639-48.

Ouellet H, Guan S, Johnston JB, Chow ED, Kells PM, Burlingame AL, Cox JS, Podust LM, de Montellano PR. 2010. Mycobacterium tuberculosis CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Mol Microbiol* **77**: 730-42.

Pajón R, Yero D, Lage A, Llanes A, Borroto CJ. 2006. Computational identification of beta-barrel outer membrane proteins in Mycobacterium tuberculosis predicted proteomes as putative vaccine candidates. *Tuberculosis* **86**: 290-302.

Papanikou E, Karamanou S, Economou A. 2007. Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol* **5**: 839-51.



Pandey AK, Sassetti CM. 2008. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci USA* **105**: 4376-80.

Pavelka MS Jr, Chen B, Kelley CL, Collins FM, Jacobs WR Jr. 2003. Vaccine efficacy of a lysine auxotroph of *Mycobacterium tuberculosis*. *Infect Immun* **71**: 190-2.

Pérez-Guzmán C, Vargas MH. 2006. Hypocholesterolemia: a major risk factor for developing pulmonary tuberculosis? *Med Hypotheses* **66**: 1227-30.

Peyron P, Vaubourgeix J, Poquet Y, Levillain F, Botanch C, Bardou F, Daffé M, et al. 2008. Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for *M. tuberculosis* persistence. *PLoS Pathog* **4**: e1000204.

Pronk JT, van der Linden-Beuman A, Verduyn C, Scheffers WA, van Dijken JP. 1994. Propionate metabolism in *Saccharomyces cerevisiae*: implications for the metabolon hypothesis. *Microbiology* **140** ( Pt 4): 717-722.

Raynaud C, Guilhot C, Rauzier J, Bordat Y, Pelicic V, Manganelli R, Smith I, et al. 2002. Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **45**: 203-17.

Raynaud C, Papavinasasundaram KG, Speight RA, Springer B, Sander P, Böttger EC, Colston MJ, Draper P. 2002. The functions of OmpATb, a pore-forming protein of *Mycobacterium tuberculosis*. *Mol Microbiol* **46**: 191-201.

Reading PC, Moore JB, Smith GL. 2003. Steroid hormone synthesis by vaccinia virus suppresses the inflammatory response to infection. *J Exp Med* **197**: 1269-78.

Rengarajan J, Bloom BR, Rubin EJ. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci* **102**: 8327-32.

Rezwan M, Lanéelle MA, Sander P, Daffé M. 2007. Breaking down the wall: fractionation of mycobacteria. *J Microbiol Methods* **68**: 32-9.

Ribeiro-Guimarães ML, Pessolani MC. 2007. Comparative genomics of mycobacterial proteases. *Microb Pathog* **43**: 173-8.

Rocco CJ, Escalante-Semerena JC. 2010. In *Salmonella enterica*, 2-methylcitrate blocks gluconeogenesis. *J Bacteriol* **192**: 771-778.

Rohde KH, Abramovitch RB, Russell DG. 2007. Mycobacterium tuberculosis invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* **2**: 352-364.

Rosłoniec KZ, Wilbrink MH, Capyk JK, Mohn WW, Ostendorf M, van der Geize R, Dijkhuizen L, Eltis LD. 2009. Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1. *Mol Microbiol* **74**: 1031-43.

Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ. 1999. In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc Natl Acad Sci* **96**: 1645-1650.

Russell DG, Cardona PJ, Kim MJ, Allain S, Altare F. 2009. Foamy macrophages and the progression of the human tuberculosis granuloma. *Nat Immunol* **10**: 943-8.

Russell DG, Dant J, Sturgill-Koszycki S. 1996. Mycobacterium avium- and Mycobacterium tuberculosis- containing vacuoles are dynamic, fusion-competent vesicles that are accessible to glycosphingolipids from the host cell plasmalemma. *J Immunol* **156**: 4764-73.

Russell DG, VanderVen BC, Lee W, Abramovitch RB, Kim MJ, Homolka S, Niemann S, Rohde KH. 2010. Mycobacterium tuberculosis wears what it eats. *Cell Host Microbe* **8**: 68-76.

Sagulenko E, Sagulenko V, Chen J, Christie PJ. 2001. Role of *Agrobacterium* VirB11 ATPase in T-pilus assembly and substrate selection. *J Bacteriol* **183**: 5813-25.

Saini NK, Sharma M, Chandolia A, Pasricha R, Brahmachari V, Bose M. 2008. Characterization of Mce4A protein in *Mycobacterium tuberculosis*: role in invasion and survival. *BMC Microbiol* **8**: 200.

Sakula A. 1982. Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax* **37**: 246-251.

Salo W, Aufderheide AC, Buikstra J, Holcomb TA. 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci USA* **91**: 2091-4.

Sambandamurthy VK, Derrick SC, Jalapathy KV, Chen B, Russell RG, Morris SL, Jacobs WR Jr. 2005 Long-term protection against tuberculosis following

vaccination with a severely attenuated double lysine and pantothenate auxotroph of *Mycobacterium tuberculosis*. *Infect Immun* **73**: 1196-203.

Sambandamurthy VK, Wang X, Chen B, Russell RG, Derrick S et al. 2002. A pantothenate auxotroph of *Mycobacterium tuberculosis* is highly attenuated and protects mice against tuberculosis. *Nat Med* **8**: 1171-4.

Sassetti CM, Boyd DH, Rubin EJ. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci* **98**: 12712-7.

Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Sassetti CM, Rubin EJ. 2003. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* **100**: 12989-94.

Savvi S, Warner DF, Kana BD, McKinney JD, Mizrahi V, et al. 2008. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in *Mycobacterium tuberculosis*: implications for propionate metabolism during growth on fatty acids. *J Bacteriol* **190**: 3886-3895.

Schilling, M.F. 1990. The Longest Run of Heads. *The College Mathematics Journal* **21**:196-207.

Schloss JV, Cleland WW. 1982. Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analogue. *Biochemistry* **21**: 4420-4427.

Schnappinger D, Ehrt S, Voskuil MI, Lui Y, Mangan JA et al. 2003. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J Exp Med* **198**: 693-704.

Selwyn Pa, Hartel D, Lewis VA, et al. 1989. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *N Engl J Med*. 320: 545-509.

Selwyn PA, Sckell BM, Alcabes P, Friedland GH, Klein RS, Schoenbaum EE. 1992. High risk of active tuberculosis in HIV-infected drug users with cutaneous anergy. *JAMA* **268**: 504-509.

Sharbati-Tehrani S, Meister B, Appel B, Lewin A. 2004. The porin MspA from *Mycobacterium smegmatis* improves growth of *Mycobacterium bovis* BCG. *Int J Med Microbiol* **294**: 235-45.

Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehrt S, Riley LW. 2003.

Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon. *Proc Natl Acad Sci USA* **100**: 15918-23.

Shin MH, Lee do Y, Liu KH, Fiehn O, Kim KH. 2010. Evaluation of sampling and extraction methodologies for the global metabolic profiling of *Saccharophagus degradans*. *Anal Chem* **82**: 6660-6.

Singh A, Crossman DK, Mai D, Guidry L, Voskuil MI, Renfrow MB, Steyn AJ. 2009. *Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. *PLoS Pathog* **5**: e1000545.

Siroy A, Mailaender C, Harder D, Koerber S, Wolschendorf F, Danilchanka O, Wang Y, et al. 2008. Rv1698 of *Mycobacterium tuberculosis* represents a new class of channel-forming outer membrane proteins. *J Biol Chem* **283**: 17827-37.

Sturgill-Koszycki S, Schlesinger PH, Chakraborty P, Haddix PL, Collins HL et al. 1994. Lack of acidification in *Mycobacterium* phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* **263**: 678-81.

Textor S, Wendisch VF, De Graaf AA, Muller U, Linder MI, et al. 1997. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Arch Microbiol* **168**: 428-436.

Timm J, Post FA, Bekker LG, Walther GB, Wainwright HC, Manganello R, Chan WT, et al. 2003. Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc Natl Acad Sci USA* **100**: 14321-6.

Turcatti G, Romieu A, Fedurco M, Tairi AP. 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* **36**: e25.

van den Berg B. 2005. The FadL family: unusual transporters for unusual substrates. *Curr Opin Struct Biol* **15**: 401-7.

Van der Geize R, Yam K, Heuser T, Wilbrink MH, Hara H, Anderton MC, Sim E, Dijkhuizen L, Davies JE, Mohn WW, et al. 2007. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proc Natl Acad Sci* **104**: 1947-52.

van Kessel JC, Hatfull GF. 2007. Recombineering in *Mycobacterium tuberculosis*. *Nat Methods* **4**: 147-152.

van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**: 767-72.

Via LE, Lin PL, Ray SM, Carrillo J, Allen SS, et al. 2008. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* **76**: 2333-2340.

Whalen C, Horsburgh CR, Jr., Hom D, Lahart C, Simberkoff M, Ellner J. 1997. Site of disease and opportunistic infection predict survival in HIV-associated tuberculosis. *AIDS* **11**: 455-460.

Wolschendorf F, Ackart D, Shrestha TB, Hascall-Dove L, Nolan S, Lamichhane G, et al. 2011. Copper resistance is essential for virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **108**: 1621-6.

World Health Organization (WHO). 2010. Global Tuberculosis Control 2010. *WHO Press* [http://www.who.int/tb/publications/global\\_report/2010/en/index.html](http://www.who.int/tb/publications/global_report/2010/en/index.html)

Wunder C, Churin Y, Winau F, Warnecke D, Vieth M, Lindner B, Zähringer U, et al. 2006. Cholesterol glucosylation promotes immune evasion by *Helicobacter pylori*. *Nat Med* **12**: 1030-8.

Yam KC, D'Angelo I, Kalscheuer R, Zhu H, Wang JX, Snieckus V, Ly LH, Converse PJ, Jacobs WR Jr, Strynadka N, Eltis LD. 2009. Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog* **5**: e1000344.

Yang X, Nesbitt NM, Dubnau E, Smith I, Samson NS. 2009. Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*. *Biochemistry* **48**: 3819-21.

Yang X, Dubnau E, Smith I, Sampson NS. 2007. Rv1106c from *Mycobacterium tuberculosis* is a 3 $\beta$ -hydroxysteroid dehydrogenase. *Biochemistry* **46**: 9058-67.

Ye J. 2007. Reliance of host cholesterol metabolic pathways for the life cycle of hepatitis C virus. *PLoS Pathog* **3**: e108.

Young TA, Delagoutte B, Endrizzi JA, Falick AM, Alber T. 2003. Structure of *Mycobacterium tuberculosis* PknB supports universal activation mechanism for Ser/Thr protein kinases. *Nat Struct Biol* **10**: 168-74.

Zimmerman MR. 1979. Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bull NY Acad Med* **55**: 604-8.

Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G, Daffé M. 2008. Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**: 5672-80.