Elucidation of the Structure and Function of the Mycosins, a family of essential subtilisin-like serine proteases of *Mycobacterium tuberculosis*

by

Zhuo Fang

Dissertation presented for the degree of Doctor of Philosophy (Molecular Biology) in the Faculty of Medicine and Health Sciences at Stellenbosch University

Supervisor: Prof. Nicolaas Claudius Gey van Pittius VI
Co-supervisor: Prof. Wolf-Dieter Schubert
Prof. Robin Mark Warren

December 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2014

Abstract

Mycobacterium tuberculosis is an ancient pathogen, which has been infecting humans for millennia. It remains globally spread, infecting one third of the world's population and causing around two million fatal cases per year. The treatment of this infectious disease remains complex, time consuming, resource demanding and costly. In the absence of adherence, good quality drugs and appropriate treatment regimens, the pathogen is highly likely to develop resistance to the antibiotics used during treatment. Searching for new drug targets and developing new drugs are constantly in progress to address these issues.

Type VII secretion system (T7SS) is a signature protein secretion system in mycobacteria, associated with virulence and nutrient acquisition. There are five copies of the T7SS in *M. tuberculosis*, namely ESX-1 to -5, with ESX-3 being essential for the bacterial growth *in vitro*. The importance of ESX-3 is supported by the fact that it influences two major iron uptake pathways in mycobacteria, namely mycobactin-mediated iron acquisition and heme acquisition. Mycosin-3 is the only membrane core component of ESX-3 that has a subtilisin-like serine protease signature and plays an essential role in ESX-3 secretion. Mycosin-3 has a unique catalytic specificity and its substrates are unknown. Elucidating the protein structure and determining the function of mycosin-3 will help the design of effective inhibitors to abolish the protease function providing for a potential therapeutic option.

In this study, the mycosin-3 gene from the *M. tuberculosis* genome was cloned and expressed in *Escherichia coli* and the protein was purified *in vitro*, with the aim of conducting structural studies. However, the amount of soluble and stable mycosin-3 was insufficient to progress to X-ray crystallography for protein structure determination. According to the literature, this technical difficulty is not uncommon for *M. tuberculosis* genes because the gene transcription and protein production machineries in mycobacteria are distinct from the conventional protein production host *E. coli*. *In vitro* expression analysis suggested that mycosin-3 possibly exerts a toxic effect on the expression hosts: *E. coli* and *Mycobacterium smegmatis*. To overcome these complexities, *M. smegmatis* was used as a model organism for functional studies; the *mycP*$_3$ gene was deleted from the genome, and the proteomes of the wild type and mycosin-3 deletion mutant were compared. No major phenotypic differences were observed between the wild type and mutant possibly because the model organism has an alternative exochelin-mediated iron acquisition pathway. Interestingly, in the absence of mycosin-3, one component of the mycobactin export system, MmpL5, was not detected in the whole cell lysate (containing both cytosolic and membrane fractions) by mass spectrometry although the *mmpL5* gene was transcribed, suggesting rapid protein degradation. We hypothesize that mycosin-3 may plays a role in maintaining the integrity of the membrane protein MmpL5 prior to its secretion and in facilitating its localization on the membrane. The direct involvement of mycosin-3 in the posttranslational modification of MmpL5 is currently under investigation. This study provides evidence that mycosin-3 may be an attractive drug target - abolishing mycosin-3 could disable mycobactin export, with ensuring toxicity from intracellular mycobactin accumulation thereby eliminating *M. tuberculosis*.

Opsomming

Mycobacterium tuberculosis is ‘n patogeen wat mense reeds eeue infekteer en tuberkulose veroorsaak. Vandag kom tuberkulose steeds wêreldwyd voor en veroorsaak ongeveer twee miljoen sterftes jaarliks. Die behandeling van hierdie aansteeklike siekte bly kompleks, tydrowend, hulpbron-veeleisend en duur. Sonder toepaslike behandeling, goeie kwaliteit antibiotika en nakoming van behandelings regimente, is dit waarskynlik dat antibiotika weerstandigheid ontwikkel. Dit is gevolglik belangrik om nuwe antibiotika teikens te vind.

Die Tipe VII sekresie sisteem (T7SS) is ‘n kenmerkende protein sekresie sisteem in mycobacteria en word met virulensie en die opname van voedingstowwe assosieer. Daar is vyf kopieë van T7SS in *M. tuberculosis* naamlik ESX-1 tot -5. ESX-3 is noodsaaklik vir bakteriële groei *in vitro*. Die belangrikheid van ESX-3 word beklemtoon deur die feit dat dit die twee hoof paaie vir ysteropname, naamlik mycobactin-bemiddelde-ysterverkryging en heme vekrygings paaie, in mycobacteria beïnvloed. Mycosin-3 is die enigste membraan kern komponent van ESX-3 wat ‘n subtilisin-agtige serien protease motief het en dit speel ‘n belangrike rol in ESX-3 sekresie. Mycosin-3 het unieke katalitiese spesifisiteit en die substrate daarvan is onbekend. Bepaling van die proteïen struktuur en funksie van mycosin-3 sal help met die ontwerp van effektiewe inhibeerders van die protease funksie. Dit kan gevolglik ‘n potensiële terapeutiese opsie verskaf.

In hierdie studie is die mycosin-3-geen van die *M. tuberculosis* genoom gekloon en uitgedruk in *Escherichia coli*. Die *in vitro* proteïen is gesuiwer om strukturele studies te doen. Die hoeveelheid oplosbare en stabiele mycosin-3 was egter onvoldoende om te vorder tot X-straalkristallografie vir proteïen struktuur bepaling. Volgens die literatuur, is hierdie tegniese probleme nie ongewoon vir *M. tuberculosis* gene, aangesien die geen transkripsie en proteïen produksie masjinerie van mycobacteria uniek is en dus van die konvensionele masjinerie in die *E. coli* gasheer verskil. Analise van *in vitro* proteïen uitdrukking eksperimente stel voor dat mycosin-3 moontlik ‘n toksiese effek op die gasheer, *E. coli* en *Mycobacterium smegmatis*, kan hê. Om hierdie probleem te oorkom, is *M. smegmatis* as ‘n model organisme gebruik om funksionele studies te doen. Die $mycP_3$ geen is uit die *M. smegmatis* genoom verwyder om ‘n delesie mutant te maak. Die proteome van die wilde-tipe en mycosin-3 geen delesie mutant is vergelyk. Daar was geen opmerklike verskille in die fenotipe van die wilde-tipe en mutant nie, moontlik omdat die model organisme ‘n alternatiewe exochelin-bemiddelde yster verkrygings pad het. Interessant genoeg, is een komponent van die mycobactin uitvoer sisteem, MmpL5, afwesig in die heel sel extrakte (van beide die sitosol en membraan fraksies) wanneer mycosin-3 afwesig is, volgens massa spektrofotometriese analise. Die *mmpL5* geen is egter getranskribeer, wat daarop dui dat die proteïen vinnig degradeer. Ons hipotese is dat mycosin-3 moontlik ‘n rol speel in die integriteit van die membraan proteïen MmpL5 voordat dit uitgeskei word en in die fasilitering van MmpL5 se posisionering in die membraan. Die direkte betrokkenheid van mycosin-3 in die posttranslasionele wysiging van MmpL5 word tans ondersoek. Hierdie studie bewys dat mycosin-3 'n aantreklike antibiotika teiken kan wees aangesien die afwesigheid van

mycosin-3 veroorsaak dat mycobactin uitvoer afgeskaf word en gevolglik tot toksiese opbou van mycobactin lei wat tot die dood van *M. tuberculosis* lei.

Acknowledgement

The author would like to address his sincere gratitude towards:

Prof. Nicolaas C. Gey van Pittius, for his guidance, support, encouragement, friendship, and financial assistance;
Prof. Wolf-Dieter Schubert, for his guidance, supervision, encouragement and friendship;
Prof. Rob Warren, for his guidance, patience, and advice;
Prof. Samamtha Sampson, for her advice and suggestions on research outputs;

Prof. Paul van Helden, for all his support and hospitality in the department;
Prof. Gerhard Walzl and Prof. Tommy Victor for their kindness and friendliness;

Dr Mae Newton-Foot, for her invaluable friendship and generous input towards all aspects of this work;
Dr Ruben van der Merwe, for his invaluable friendship, patience and generous assistance in data analysis and programming;
Dr Edukolado Mullapudi, for his invaluable friendship, generous technical support on protein production, purification and crystallization;

Dr Monique Williams, for her life-saving technical support and trouble-shoot strategies;
Dr Anzaan Dippenaar, Dr Suereta Fortuin, Dr Margaretha de Vos, Dr Natalie Bruiner, Dr Albertus Vijoen, and Dr Salome Smit for their invaluable technical support;

Miss Danicke Willemse, for her technical assistance and translating my abstract into Afrikaans;

Mr Kabongele Keith Siame, Mrs Louise Botha, Miss Nastassja Steyn, Miss Skye Fishbein for their friendship and support.

Mr Jimmy N. Warrington, Sister Yabin Mo (deceased), Mr Chaolun Wu, Dr Rachel Cooper, and Dr Laverne B. Lindenberg for their precious friendship which make me who I am now. Your friendships have had major positive impact on my life over the years. I could not thank you more.

A special thanks to my parents, Guiru Wang (mother) and Hongcai Fang (father), without you, nothing would be possible!

Table of Contents

List of Abbreviation

ABC	ATP Binding Cassette
AMP	Adenosine Monophosphate
AMS	Acyl Sulfamoyl Adenosine
ATc	Anhydrotetracycline
ATP	Adenosine triphosphate
BCG	Bacillus Calmette–Guérin
CBP	Chitin Binding Protein
CFP-10	Culture Filtrate Protein-10
CRP	cAMP Receptor Protein
DMPC	Dimyristoylphosphatidylcholine
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
Ecc	ESX Conserved Components
E. coli	*Escherichia coli*
ESAT-6	Early Secretion Antigenic Target-6
ESCRT	Endosomal Sorting Complex Required for Transport
Esp	ESX-Specific Proteins
ESX	ESAT-6 Secretion System
FAD	Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
GST	Glutathione S-Transferase
HIV	Human Immunodeficiency Virus
HO	Heme Oxygenase
IdeR	Iron Dependent Regulator
IFN	Interferon
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
KO	Knock Out
LB	Lysogeny Broth
MBP	Maltose Binding Protein
MDR	Multidrug Resistance
MMPL	Mycobacterial Membrane Protein Large
MMPS	Mycobacterial Membrane Protein Small
MPTR	Major Polymorphic Tandem Repeat
M. tuberculosis	*Mycobacterium tuberculosis*
$MycP_3$	Mycosin-3
NADH	Nicotinamide Adenine Dinucleotide
NAP	Necleoid-Associated Protein
Ni-NTA	Nickel Nitrilotriacetic Acid
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame

PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDIM	Phthiocerol dimycocerosates
PE	Proline Glutamate
PGRS	Polymorphic GC-rich Repetitive Sequence
PPE	Proline Proline Glutamate
RD	Region of Difference
RNA	Ribonucleic Acid
RT-qPCR	Real Time-quantitative PCR
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
TB	Tuberculosis
TBDB	Tuberculosis Database
TMP	Thymidine Monophosphate
T7SS	Type VII Secretion System
WHO	World Health Organization
WT	Wildtype
XDR	Extensive Drug Resistance

Chapter 1. General Introduction

1.1 General Introduction

World Health Organization (WHO) has announced tuberculosis as a global health emergency (WHO, Global tuberculosis report 2013). Tuberculosis is an infectious disease primarily affecting human lungs and leading to respiratory failure if not treated. *Mycobacterium tuberculosis*, the pathogen that causes the disease, is a gram-positive bacterium with a thick and complex cell wall that is naturally resistant to many antibiotics and toxins. *M. tuberculosis* is a slow-growing bacterium with a doubling time of 24 hours, and it may enter a dormant state when stimulated by certain molecules (e.g. CO, NO) in the host environment (Voskuil *et al*., 2003). Therefore the treatment of tuberculosis is usually prolonged and sometimes ineffective. The standard 6-month regimen consists of the first 2 months with four antibiotics, isoniazid, rifampicin, pyrazinamide and ethambutol, followed by 4 months with two antibiotics, isoniazid and rifampicin (WHO, Global tuberculosis report 2013). This regimen minimizes the risk of the bacteria developing drug resistance during the treatment. However, these drugs have many side effects, making it very difficult for the patients to adhere to the 6 month treatment. Patient's non-adherence and clinical mis-management are two major factors driving the emergence of resistance to all the first-line anti-TB drugs and some second-line drugs. Discovery of new drug targets and development of new drugs is therefore urgently required to shorten the duration of treatment and to effectively treat drug resistant TB.

A good drug target should be responsible for essential metabolic functions in *M. tuberculosis* and it should have no homology to human proteins to minimize drug toxicity to the host. The Type VII secretion system (T7SS) is a specific protein secretion system in mycobacteria. *M. tuberculosis* has five copies of T7SS which contribute to virulence and nutrient acquisition. Chapter 2 gives a thorough literature review about this recently described bacterial secretion system, delineating its substrates, its functions in *M. tuberculosis* and the host, the components of the secretion machinery, the secretion mechanism, and its influence on the pathogenicity of the mycobacteria. One of the five copies of T7SS, ESX-3, is essential for bacterial growth *in vitro*. In fact, ESX-3 enables the functioning of the two major iron acquisition mechanisms in mycobacteria, mycobactin-mediated iron acquisition and heme acquisition. Accordingly, ESX-3 was chosen for further investigation in this study. Chapter 3 provides a detailed literature review on iron and heme homeostasis in mycobacteria, including their acquisition mechanisms, regulation of the transcription of the genes encoding the machineries, the link between the two pathways, and how ESX-3 influences their functionalities.

The most appropriate drug target in ESX-3 appears to be mycosin-3. Unlike other membrane core components of the system, mycosin-3 has a unique subtilisin-like serine protease signature which is essential for ESX-3 secretion. Inhibitors could therefore be designed to

disrupt the integrity of its catalytic site. Although the substrates of mycosin-3 are unknown, it is evident that this enzyme is essential for the bacteria and it is probably able to modify certain components in the iron and heme acquisition pathways. A better understanding of the structure and substrate specificity of mycosin-3 is a prerequisite to inhibitor design. Chapter 4 describes the progress of structural genomics of *M. tuberculosis* H37Rv to date, providing a background for the structural analysis of mycosin-3, which is described in Chapter 5. In Chapter 6, *M. smegmatis* was used as a model to elucidate the function of mycosin-3. The proteomes of a mycosin-3 deletion mutant of *M. smegmatis* were compared to wild-type *M. smegmatis* to determine the downstream effects of the presence/absence of mycosin-3 on the proteome of *M. smegmatis*. The main findings of the structural and functional studies are summarized in Chapter 7, where suggestions of future research directions are made also.

The main objectives of this study are:

1. To understand the literature on Type VII Secretion System;
2. To understand the literature on iron transport and metabolism in mycobacteria;
3. To summarize the progress and experience from the past structural genomics research on *M. tuberculosis* H37Rv;
4. To attempt to study the 3D structure of mycosin-3;
5. To elucidate the possible functions of mycosin-3.

1.2 Reference

Voskuil, M.I., Schnappinger, D., Visconti, K.C., Harrell, M.I., Dolganov, G.M., Sherman, D.R., Schoolnik, G.K., 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J. Exp. Med. 198, 705–713. doi:10.1084/jem.20030205

WHO, Global tuberculosis report 2013, WHO. URL http://www.who.int/tb/publications/global_report/en/

Chapter 2. Type VII Secretion System in Mycobacteria

2.1 Introduction

Tuberculosis (TB) is one of the infectious diseases in the world. The causative agent, *Mycobacterium tuberculosis* is a successful human pathogen which manages to survive and replicate inside the hostile environment of human macrophages, spread to healthy neighbouring cells, migrate to other tissues and organs of the body, causing disease in its hosts. Its persistence and ability to avoid attacks from the host immune system are intriguing knowledge gaps. The mechanisms of its virulence remain under the spotlight of tuberculosis research.

The well-known TB vaccine, the attenuated *Mycobacterium bovis* strain BCG (Bacillus Calmette–Guérin), derived from *M. bovis* by numerous passaging in the laboratory, has several regions of difference (RD) identified by comparative genomic studies with the virulent parent strain. These RDs are believed to be responsible for the virulence of the bacteria (Mahairas *et al*., 1996). RD1 appears to be the most important region because it is directly related to the secretion of potent T-cell antigen ESAT-6 (Early Secreted Antigenic Target 6 kDa) and CFP-10 (Culture Filtrate Protein 10 kDa) (Andersen *et al*., 1991). The counterparts of the RD1 region in *M. tuberculosis* spans from Rv3871 to Rv3879c (9.5 kbp). Deletion of this region in *M. tuberculosis* led to the attenuation of the strain, similarly to the BCG attenuated phenotype, in the macrophage and mice infection models (Lewis *et al*., 2003). Re-introducing RD1 into BCG enabled the secretion of ESAT-6 and CFP-10 into the culture supernatant (Pym *et al*., 2003) and improved bacterial growth in mice, although BCG remained avirulent (Pym *et al*., 2002a).

It is believed that RD1 contains genes encoding part of a secretion system which secretes multiple virulence factors including ESAT-6 and CFP-10 (Cole *et al*., 1998; Gey Van Pittius *et al*., 2001; Tekaia *et al*., 1999). The complete gene cluster, named the ESAT-6 secretion system 1, ESX-1, contains 20 genes (Rv3864 – Rv3883c) (Brodin *et al*., 2004). There are 4 more copies of the ESAT-6 gene cluster (ESX-2, ESX-3, ESX-4 and ESX-5) in the *M. tuberculosis* genome and homologues of the ESAT-6 gene clusters exist throughout the mycobacteria and some other gram positive bacteria (Figure 1) (Gey Van Pittius *et al*., 2001). With ESX-4 as the ancestral copy and the ESX-5 as the most recently evolved copy, the evolutionary sequence for the five ESX gene clusters is ESX-4, ESX-3, ESX-1, ESX-2 and ESX-5 (Gey Van Pittius *et al*., 2001). The evolution of the ESX drove the duplication and expansion of the PE (Pro-Glu) and PPE (Pro-Glu-Glu) gene families, with the ancestral gene copies located within the ESX gene clusters (Gey van Pittius *et al*., 2006). Except for ESX-4, all of the gene clusters contain PE and PPE gene pairs (Figure 1) (Bitter *et al*., 2009). While

ESX-2 and ESX-4 are not well described, ESX-1 and ESX-5 have been associated with the virulence of pathogenic mycobacteria and ESX-3 was shown to be involved in mycobactin-mediated iron acquisition (Siegrist *et al*., 2009). Moreover, ESX-1 participates in conjugation in *M. smegmatis* (Coros *et al*., 2008) and sporulation in *Streptomyces coelicolor* (Akpe San Roman *et al*., 2010).

Besides the PE and PPE proteins, the ESX gene clusters contain genes encoding the secretion system membrane core components, EccB, EccC, EccD and EccE (ESX Conserved Components) (Houben *et al*., 2012), the secreted substrates, the Esx (ESAT-6 family proteins) and Esp (Esp stands for ESX-Specific Protein) proteins, as well as the mycosins, membrane-anchored subtilisin-like serine proteases (Brown *et al*., 2000). The ESAT-6 secretion system has a different secretion mechanism from the other six types of secretion systems discovered previously in terms of their secretion signals and two-step secretion pattern, and was thus identified as a novel type of secretion system termed the Type VII secretion systems (T7SS) (Abdallah *et al*., 2007). This review will focus on the functions of the components and substrates of T7SS, its secretion mechanism, the regulation of its gene expression, and the impact of these secretion systems on the mycobacteria and host pathogenicity.

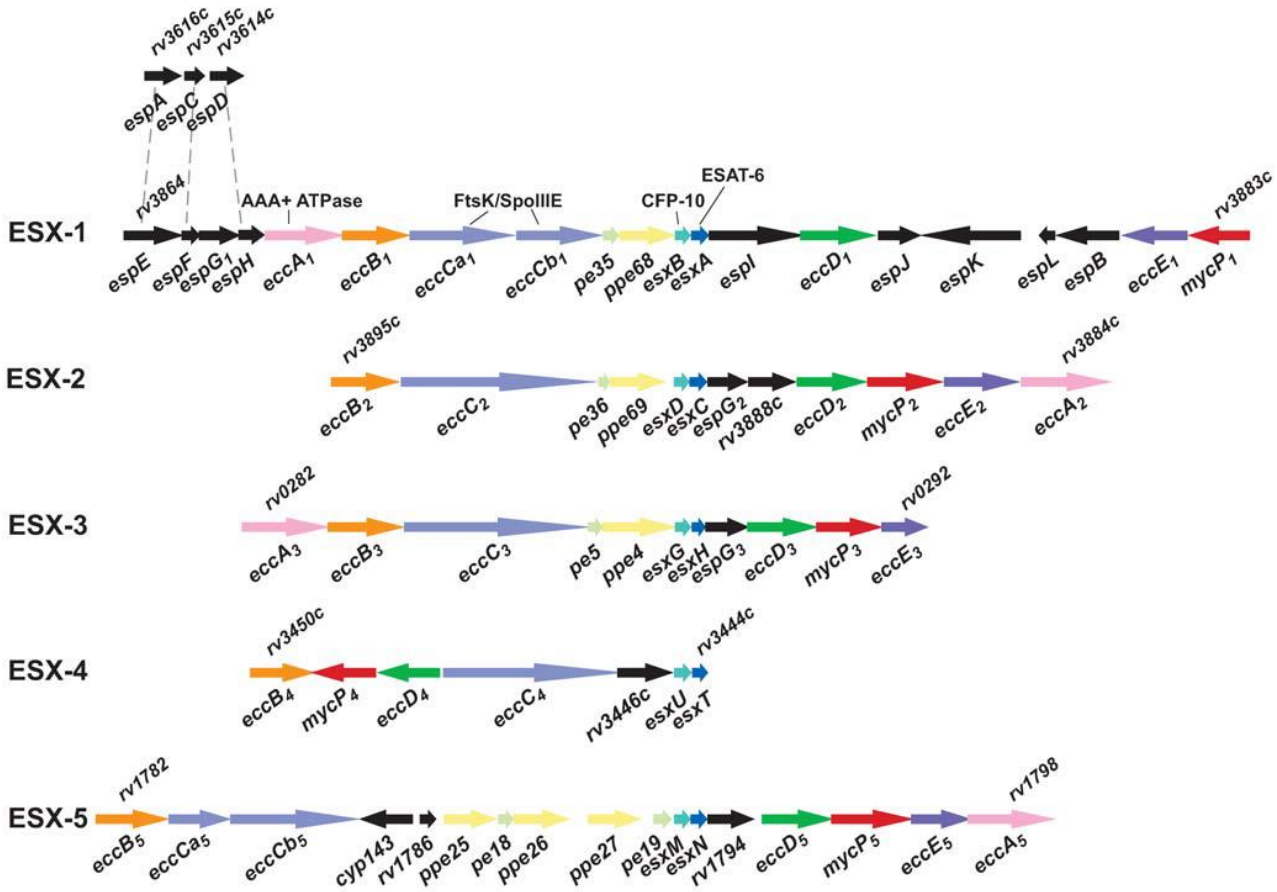


Figure 2.1. The genetic organization of the five ESX gene clusters of *Mycobacterium tuberculosis*. The abbreviations are explained in the text. The figure is adapted from Bitter, *et al*., 2009.

2.2 Type VII Secretion System Working Model

The T7SS recognizes a specific secretion signal on some of its substrates which recruit other substrates in the cytoplasm before delivering them to the ATPase-containing secretion core machinery on the plasma membrane. The ESX-1 of *M. tuberculosis* and *M. marinum* are the best studied T7SS models in terms of the substrate identities and the secretion mechanism. This section describes the ESX-1 in detail. The different ESXs are presumed to have similar secretion mechanisms although their substrates may differ.

2.2.1 The Secretion Signal

The T7SS is *sec*-independent and its substrates do not possess the N-terminal signal sequence characteristic of the *sec*-dependent secretion systems (Ligon *et al*., 2012). An ESX secretion signal sequence was identified on the C-terminus of certain substrates of the secretion system. The last 25 amino acids of CFP-10 were found to be essential for CFP-10 to interact with the membrane component EccCb to ensure its secretion. Deletion of these 25 amino acids abolishes CFP-10 secretion and secretion of other substrates which are CFP-10-dependent (Champion *et al*., 2006). The paralogues of CFP-10 (EsxB), namely ExJ, EsxK, EsxP, EsxM, and EsxW lack the CFP-10-signal sequence implying that their secretion is not ESX-1-dependent and that other secretion systems are needed for their secretion (Champion *et al*., 2006). PE25 and PPE41, which form a heterodimer before secretion, are substrates of ESX-5. Deletion of 7 or 15 amino acids from the C-terminus of PE25 completely abolished the secretion of both PE25 and PPE41 (Daleke *et al*., 2012a). More specifically, the Tyr12 and the Asp/Glu 8 from the C-terminus are the key residues determining secretion. This specific YxxxD/E motif is not only found in the flexible C terminus of PE25, but is also present on the PE domain of LipY (Daleke *et al*., 2011), PE domain of PE_PGRS33 (Cascioferro *et al*., 2011), PE35 protein, CFP-10, EspC and EspB (Daleke *et al*., 2012a). This general signal sequence, YxxxD/E, can be used to search for additional substrates of the ESX systems in the *M. tuberculosis* genome (Daleke *et al*., 2012a). However, this signal cannot differentiate between the different ESX systems. The additional signal for ESX specificity lies on the ESX secretion chaperone protein EspG which does not possess a distinct signal sequence (Daleke *et al*., 2012b).

2.2.2 The substrates and their potential functions

There are three groups of known substrates of ESX-1, the *M. tuberculosis* primary potent T-cell antigen Esx proteins including EsxA (ESAT-6) and EsxB (CFP-10), the Esp group including EspE, EspF, $EspG_1$, EspH, EspJ, EspK, EspL, EspB, EspA, EspC and EspD, and the PE and PPE families of proteins. They may perform a variety of functions after being secreted, which remain to be elucidated.

2.2.2.1 Esx proteins

EsxA (ESAT-6) and EsxB (CFP-10) are among the first discovered and studied secreted proteins from *M. tuberculosis* (Berthet *et al*., 1998; Sørensen *et al*., 1995). There are 22 known Esx genes in the *M. tuberculosis* genome (Cole *et al*., 1998) and the homologues of EsxA and EsxB are usually arranged in pairs in the same operon and are co-transcribed (Berthet *et al*., 1998; Poulsen *et al*., 2014). EsxA and EsxB form a tight heterodimeric 1:1 protein complex before secretion and are mutually dependent for secretion (Renshaw *et al*., 2005, 2002). The role of the EsxA-EsxB complex in host pathogenicity is still under investigation. Interestingly, EsxA was found to enhance the permeability of artificial membranes (Hsu *et al*., 2003). The EsxA-EsxB complex is able to interact with the host cell surface, but it does not disrupt the membrane integrity (Renshaw *et al*., 2005). EsxB showed weak and non-specific binding to the cell membrane (de Jonge *et al*., 2007). EsxA alone does not disrupt cell membrane integrity (Guinn *et al*., 2004), however, it has a high affinity to artificial liposomes, particularly when dimyristoylphosphatidylcholine (DMPC) and cholesterol are incorporated (de Jonge *et al*., 2007). EsxA may dissociate from EsxB under acidic conditions and perform a membrane-lysing activity on the liposome (de Jonge *et al*., 2007). De Jonge *et al*. (2007) proposed that the macrophage phagosome has a low pH environment which favours the dissociation of EsxA from EsxB and that EsxA is directly involved in biomembrane disruption. But this effect of EsxA has not been demonstrated in macrophage infection models (de Jonge *et al*., 2007).

2.2.2.2 Esp proteins

A unique feature about ESX-1 is that an extra genomic region *espACD* (Rv3616c-Rv3614c) is associated with its functions where the main ESX-1 substrates and the gene products of the *espACD* operon are mutually dependent for secretion (Fortune *et al*., 2005; MacGurn *et al*., 2005). In the ESX-1 gene cluster and its functionally associated *espACD* operon, there are a number of Esp genes, the encoded proteins which the functions are still largely unknown although many of them affect the stability and secretion of EsxA, EsxB and other Esps.

EspE is not secreted into the extracellular space but remains cell envelope associated (Carlsson *et al*., 2009) and treatment with the detergent Tween-80 cannot remove EspE from its location. This implies that EspE might have interactions with the host (Sani *et al*., 2010). However, deleting the *espE* gene does not affect the virulence of the bacteria (Brodin *et al*., 2006). The two *esp* genes downstream of *espE*, namely *espF* and *espG$_1$* (Figure 2.1), are co-transcribed independently from the upstream *espE* and the downstream *espH* genes (Bottai *et al*., 2011). Disruption of the *espF* and *espG$_1$* genes did not affect the secretion of EsxA and EsxB but the mutant still became severely attenuated (Bottai *et al*., 2011; Champion *et al*., 2009). EspF and EspG$_1$ probably play important roles in ESX-1 mediated pathogenicity. EspF is abundant in both the culture supernatant and the bacteria pellet of *M. marinum* but is scarce in both environments in *M. tuberculosis*. It interacts with EsxA without which EspF is not secreted (Brodin *et al*., 2006). EspG$_1$ is a cell envelope associated protein. It affects either

the gene expression or protein stability of PPE68, the PPE protein in the same ESX region (Bottai *et al*., 2011). The EspG proteins secreted by the ESXs are specific chaperone proteins for the PPE family proteins or the PE/PPE protein complexes secreted by the specific ESX. EspGs do not bind to the YxxxD/E secretion signal but to the core domain of the PE-PPE complex forming a 1:1:1 complex before secretion (Daleke *et al*., 2012b). For example, $EspG_1$ interacts with the ESX-1 substrate PPE68 but not with the ESX-5 substrate PPE41 while $EspG_5$ (Rv1794, the $EspG_1$ homologue in ESX-5) interacts with ESX-5 substrates PPE33 and PPE41 but not PPE68 (Daleke *et al*., 2012b). EspG must possess a secretion signal which directs the PPE protein or PE/PPE protein complex to its specific ESX for secretion. The presence of PE and PPE proteins which are not the substrates of a particular ESX is not affected when this ESX is disabled (Sani *et al*., 2010). Similarly to EspE, EspF and $EspG_1$, the proteins EspH, EspJ, EspK and EspB are not required for EsxA and EsxB secretion (Brodin *et al*., 2006; McLaughlin *et al*., 2007). There is very little information about the functions of EspH, EspJ and EspL but EspI was found to affect the secretion of EsxA and EsxB by abolishing EspA secretion (Fortune *et al*., 2005).

EspB is dependent on EspK but not $EspG_1$, EspH or EsxB for secretion (McLaughlin *et al*., 2007) although it was believed to form a temporary complex with the EsxA-EsxB heterodimer to facilitate secretion (Xu *et al*., 2007). EspB and EspK physically interact with one another and EspK was shown to interact with EccCb and act a chaperone that delivers EspB to the secretion machinery (McLaughlin *et al*., 2007). After secretion via ESX-1, EspK remains cell envelope associated (Sani *et al*., 2010). Interestingly, after secretion EspB is cleaved by Mycosin-1 at two sites at the C-terminus after secretion (McLaughlin *et al*., 2007; Xu *et al*., 2007), namely Pro332 and Ala392 (Ohol *et al*., 2010). The physiological or pathogenic significance of this cleavage is unknown.

The proteins encoded in the *espACD* operon, outside the ESX-1 gene cluster are substrates of ESX-1 and are inter-dependent with the ESX-1 substrates, except EspB (Chen *et al*., 2013b), for secretion (Callahan *et al*., 2010; Fortune *et al*., 2005; MacGurn *et al*., 2005). EspA can only be secreted when coupled to the EsxA-EsxB complex (Fortune *et al*., 2005), where the binding is dependent on the Trp55 and Gly57 residues in the putative WXG motif of EspA. Disruption of this motif abolishes EsxA-EsxB secretion (Chen *et al*., 2013a). The same point mutations also affect EspD stability and secretion, leading to *M. tuberculosis* attenuation. Point mutations at Phe50 and Lys62 destabilize EspA and also abolish EsxA-EsxB secretion, attenuating *M. tuberculosis*. However, if Phe5 and Lys41 are mutated, *M. tuberculosis* is not attenuated, although EspA is destabilized and EsxA-EsxB and EspD secretions are abolished (Chen *et al*., 2013a). Perhaps these two specific amino acids on EspA are responsible for the interaction of EspA with other virulence-related proteins in *M. tuberculosis*. Moreover, after secretion, EspA forms a dimer via disulfide bonds, disruption of which does not affect ESX-1 secretion but severely attenuates the ability of *M. tuberculosis* to survive and cause disease in mice and also significantly disrupts the stability of the mycobacterial cell wall (Garces *et al*., 2010). EspA was found localized within the cell envelope, suggesting that it plays a role in maintaining the bacterial cell wall integrity (Garces *et al*., 2010). These studies dissociate the

EsxA and EsxB secretion from the virulence of *M. tuberculosis*, changing the paradigm of the function of these two major T-cell antigenic proteins (Champion, 2013; Chen *et al*., 2013a).

EspC, possessing the Type VII general secretion signal (YxxxD/E), is recognized by EccA through its C-terminus. It is required for EsxA, EsxB and EspA secretion and is itself a substrate of ESX-1 (Champion *et al*., 2009). Truncating the C-terminal 7, 14, or 25 amino acids of EspC abolishes EspA secretion. EspC also affects the stability and/or the expression of EspA because the EspA level was reduced in the EspC deletion mutant (Champion *et al*., 2009).

EspD is a secreted protein, but unlike EspA and EsxA, it is not a substrate of ESX-1 and even the RD1 mutant strain of *M. tuberculosis* secretes EspD (Chen *et al*., 2012). EspD stabilizes EspA and EspC. Without EspD, the EspA and EspC cellular levels are significantly decreased and the secretion of EsxA and EsxB are disabled. However, there is no direct interaction between EspD and either EspC or EspA. Abolishing the secretion of EspD does not affect EsxA secretion. Without EspA, EspD is not produced or secreted indicating an inter-dependent mechanism between the two proteins (Chen *et al*., 2012).

2.2.2.3 PE and PPE proteins

The PE and PPE proteins are unique to the mycobacteria. They constitute almost 10% of the coding capacity of the *M. tuberculosis* genome (Cole *et al*., 1998). The PE and PPE genes in the ESX gene clusters represent the ancestral PE and PPE gene copies from which the rest of the family members were copied (Gey van Pittius *et al*., 2006). The more recently evolved PE and PPE proteins usually contain various lengths of C-terminal sequences which classify them as members of the PE-PGRS (Polymorphic GC-rich Repetitive Sequence) and PPE-MPTR (Major Polymorphic Tandem Repeat) sub-families (Gey van Pittius *et al*., 2006). In the *M. tuberculosis* genome there are 28 PE and PPE genes identified in operons, with the PE gene located directly upstream of the PPE gene and they are co-transcribed (Tundup *et al*., 2006). These PE and PPE protein pairs can only be produced solubly *in vitro* when co-expressed and co-purified (Strong *et al*., 2006; Tundup *et al*., 2006). Different PE and PPE proteins are secreted using specific ESXs when associated with a particular EspG protein. The general secretion signal was identified in some PE proteins (Daleke *et al*., 2012a). The secreted PE and PPE proteins may perform a variety of functions which are still undetermined. PPE68, a substrate of ESX-1, is localized in the cell envelope, and is able to elicit interferon (IFN)-gamma production in mice infected with *M. tuberculosis* (Sani *et al*., 2010). PPE68 may interact with EsxA (Okkels and Andersen, 2004), however, it does not interfere with the secretion and immunogenicity of EsxA and EsxB and it is not a major virulence factor of *M. tuberculosis* (Brodin *et al*., 2006; Demangel *et al*., 2004). The PE25/PPE41 complex might perform a signal function because of their resemblance to signal receptor proteins (Strong *et al*., 2006). One mycobacterial lipase, LipY, whose orthologue possesses a PE domain in *M. tuberculosis* and a PPE domain in *M. marinum* is a substrate of ESX-5, is secreted onto the surface of the bacteria, and is possibly involved in fatty acid

acquisition by hydrolysing the host lipid during infection (Daleke *et al*., 2011). PE and PPE family proteins mostly are thought to be antigenic cell surface proteins. Their diverse C-terminal protein sequences provide a large range of antigenic specificities. The expression of PE and PPE family proteins is regulated by a variety of independent mechanisms (Voskuil *et al*., 2004). With a changing microenvironment within the host which might be influenced by oxidative stress, macrophage IFN-gamma activation, nitric oxide elevation, etc., different sets of PE and PPE proteins are produced providing a dynamic antigenic profile (Voskuil *et al*., 2004). However, the significance of this phenomenon is unknown. Perhaps it exerts less pressure on genetic mutation of the bacteria and enables the mycobacteria to respond faster to the changing immunity of the host.

2.2.3 The ATPases

EccA plays a substrate recruiting role for the EsxA-EsxB complex and some Esp complexes similar to EspG for the PE and PPE proteins (Houben *et al*., 2013a). Unlike EspG, which provides a signal for the substrate delivery, EccA supplies energy for the recruitment and delivery process. EccA has two functional domains: the N-terminal domain contains 6 tetratricopeptide repeats which allow the interaction with the secreted substrates (Wagner *et al*., 2014) and the C-terminal ATPase domain hydrolyzes ATP molecules. Without EccA, EsxA and EsxB are not secreted (Brodin *et al*., 2006) and the stability of EspB is affected in its absence (Xu *et al*., 2007).

Besides facilitating substrate secretion for ESX-1, EccA seems to affect mycolic acid biosynthesis. An EccA-ATPase mutant showed a significant decrease in mycolic acid content in the cell wall, making the mutant defective in intracellular growth and reducing its virulence. EccA may be able to shuttle the enzymes involved in mycolate biosynthesis to their production sites, making the process more efficient (Joshi *et al*., 2012).

EccCb, a component of the ESX-1 membrane complex, also possesses an ATPase domain (Champion *et al*., 2009). EccCb works cooperatively with EccA (Champion *et al*., 2009). It recognizes the C-terminus of CFP-10 and brings the substrate complex to the proximity of the ESX-1 membrane complex (Abdallah *et al*., 2007; Houben *et al*., 2013a). It is also required for EspE secretion (Sani *et al*., 2010).

2.2.4 The membrane core complex

In ESX-1 and ESX-5, two proteins EccCb and EccCa comprise the membrane ATPase complex whereas in other ESX systems, EccC homologues are encoded by a single gene (Figure 2.1). ESX-5 from *M. marinum* was used as a model to elucidate the ESX membrane structure (Houben *et al*., 2012). $EccB_5$, $EccC_5$, $EccD_5$ and $EccE_5$ constitute the ESX secretion membrane core structure with 6 copies of each $EccB_5$, $EccC_5$, $EccD_5$ and 3 copies of $EccE_5$ (Houben *et al*., 2012). Although Mycosin-5 is a membrane-bound protein in the system, it cannot be isolated with the core complex, possibly indicating a weak interaction (Houben *et*

al., 2012). By performing protease digestion, the middle and the C-terminus of EccC and the N-terminus of EccE were found to contain exposed domain whereas EccB and EccD do not. It is known that the C-terminus of EccC is exposed intracellularly and is required to interact with the substrate complex prior to secretion. The cleavage of the N-terminus of EccE may suggest that this plasma membrane-localized protein extends to the outer membrane (Houben *et al*., 2012).

Extensive studies have shown that secretion is abolished through the deletion of any of the 4 membrane components of the ESX system. The $EccB_5$, $EccC_5$, and $EccD_5$ single gene deletion mutants of both *M. tuberculosis* (Di Luca *et al*., 2012) and *M. marinum* were unable to secrete EsxN, PPE41 and a number of PE proteins (Bottai *et al*., 2012; Houben *et al*., 2012); EccB, EccCa, EccCb and EccD are also essential for the secretion of EsxA and EsxB in *M. tuberculosis* (Brodin *et al*., 2006; Guinn *et al*., 2004; Stanley *et al*., 2003) and *M. smegmatis* (Converse and Cox, 2005); $EccC_3$ and $EccD_3$ are essential for the secretion of EsxG and EsxH in *M. smegmatis* (Siegrist *et al*., 2014).

2.2.5 Mycosins

The mycosins are membrane-bound subtilisin-like serine proteases (Brown *et al*., 2000) which might be shed into the extracellular matrix (Dave *et al*., 2002). They are essential for secretion through the ESX systems in *M. smegmatis* and *M. tuberculosis* (Converse and Cox, 2005; Ohol *et al*., 2010). However, disabling the protease activity of Mycosin-1 increases the secretion of ESX-1 substrates and it causes hyperactivation of ESX-1 stimulated innate signalling pathways during macrophage infection (Ohol *et al*., 2010). Mycosin-1 may have a regulatory role in ESX-1 substrate secretion where a fine tuning secretion may balance virulence and immune detection, for successful maintenance of a long-term *M. tuberculosis* infection (Ohol *et al*., 2010). Mycosin-1 is believed to have multiple substrates, but only EspB has been found to be cleaved by Mycosin-1 *in vitro* and *in vivo* (Ohol *et al*., 2010; Solomonson *et al*., 2013; Wagner *et al*., 2013).

2.3 Type VII Secretion Mechanism

The Type VII secretion mechanism has mostly been studied using ESX-1 and ESX-5 in *M. tuberculosis* and *M. marinum* as models (Figure 2.2). ESX-1 secretion seems to be more complex than ESX-5 secretion. In ESX-1 a large number of substrates have been identified, with different roles during the secretion process. The substrates, some of which contain secretion signals, and some of which are secretion chaperone proteins, form protein complexes prior to secretion thereby stabilizing one another (Houben *et al*., 2013b).

The ESX-1 substrate complex EsxA-EsxB coupled with EspA is recognized by EccCb via the secretion signal on the C-terminus of EsxB (Champion *et al*., 2009, 2006). EccCb brings this

complex in proximity to the secretion machinery which transports EsxA-EsxB out of the cell, but retains EspA in the cell envelope. The EsxA-EsxB complex may also interact with EspF, which in turn binds to EspC. This protein scaffold has C-terminus secretion signals on EsxB and EspC which are recognized by EccCb and EccA respectively (Champion *et al*., 2009). EccCb and EccA, two ATPases, work cooperatively to deliver the complex to the secretion machinery and transport EsxA, EsxB and EspF out of the cell, retaining EspC in the cell envelope (Champion *et al*., 2009). $EspG_1$ acts as an ESX-1 secretion chaperone which binds to the PE35/PPE68 complex (Bottai *et al*., 2011). The C-terminus of PE35 contains a secretion signal which is also recognized by EccCb and then secreted by the core secretion complex (Daleke *et al*., 2012a). PE35 is secreted, and PPE68 and $EspG_1$ remain cell envelope associated (Daleke *et al*., 2012b). EspK and EspB associate before secretion (McLaughlin *et al*., 2007). Although EspB contains a secretion signal at its C-terminus, EspK interacts with EccCb and brings EspB to the secretion machinery (McLaughlin *et al*., 2007). The C-terminus truncated EspB can be found in the culture supernatant while EspK remains cell envelope associated (Sani *et al*., 2010). EspD stabilizes EspA and EspC but its secretion is independent of ESX-1 via an unknown mechanism (Chen *et al*., 2012).

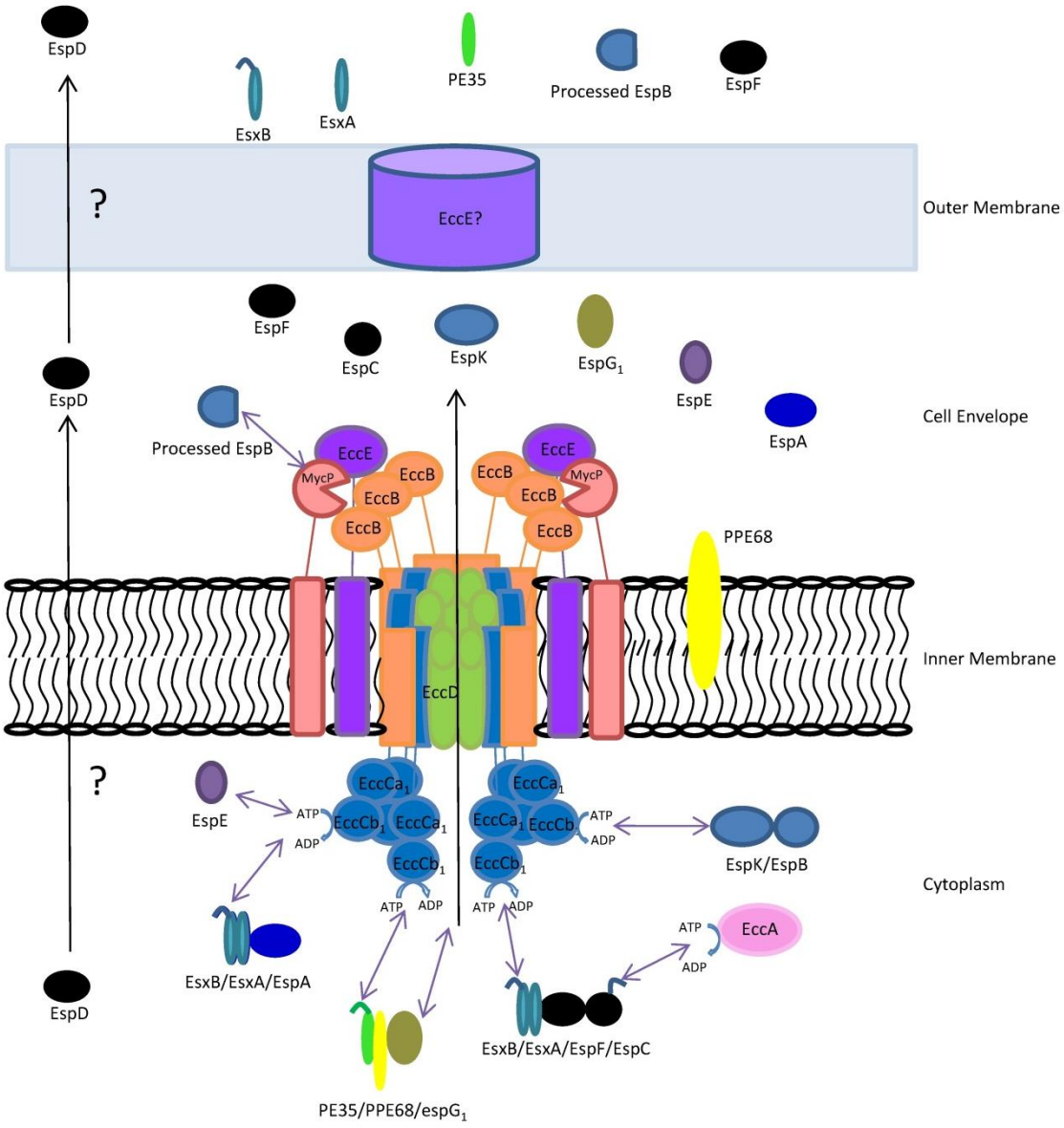


Figure 2.2. The proposed secretion mechanism of Type VII secretion system using ESX-1 from *M. tuberculosis* as a model.

2.4 ESX-1 and Its Association with the Virulence of *M. tuberculosis*

T7SSs are not virulence systems per se (Abdallah *et al*., 2007), but it has long been believed that ESX-1 is responsible for the virulence of pathogenic mycobacteria because RD1, harbouring some components of ESX-1, is absent from the attenuated *M. bovis* BCG strain (Mahairas *et al*., 1996). The characterization of the ESX-1 components and substrates confirmed that their functions are virulence-related, as deleting them leads to mycobacterial growth defects in macrophages and in mouse models. For instance, the *M. tuberculosis*

EccCa and EccCb deletion mutants showed attenuation in THP-1 macrophages (Guinn *et al*., 2004); EccD and PPE25-PE19 mutants are attenuated in both macrophages and immune-deficient mouse infection models (Bottai *et al*., 2012); and the EspB deletion mutant showed a defect in phagosome maturation (Xu *et al*., 2007). Specific EspA point mutants, in which EsxA and EsxB secretion was abolished, remained virulent showing a dissociation of EsxA-EsxB secretion from virulence in *M. tuberculosis* (Champion, 2013; Chen *et al*., 2013a). Without EsxA and EsxB, the processed form of EspB, another virulent protein, is able to bind to two phospholipids, phosphatididic acid and phosphatidylserine, possibly interfering with cell signalling in the host, facilitating virulence (Chen *et al*., 2013b). In fact, several other studies on ESX have also provided evidence to support the theory where EsxA and EsxB secretion does not necessarily confer virulence of *M. tuberculosis*. EspF and EspG deletion mutants were able to secrete EsxA and EsxB, but were attenuated (Bottai *et al*., 2011; Brodin *et al*., 2006). Non-functional EspA, where intracellular disulphide bond was disrupted, did not affect EsxA and EsxB secretion but disrupted the cell wall integrity thereby attenuating the bacteria (Garces *et al*., 2010). ESX-1 is required to prevent phagosome maturation, but none of its substrates are responsible for this virulence process (MacGurn and Cox, 2007). It has become evident that most of the ESX-1 substrates remain within the cell envelope after secretion (Figure 2.2), and serve to support the complex cell wall structure of *M. tuberculosis* rather than interacting with the host. Moreover, the observation that ESX-1 is localized at the new pole of a dividing mycobacterium where it seems involved in generating and modifying the newly synthesized cell wall, also supports this argument (Carlsson *et al*., 2009; Sani *et al*., 2010; Wirth *et al*., 2012). With such a complex barrier in *M. tuberculosis*, some ESX-1 substrates, such as EsxA, might not directly mediate virulence but target the bacteria's own cell wall to facilitate the secretion of other virulence or metabolism related factors (Garces *et al*., 2010).

2.5 ESX-5 and ESX-3 secretion systems

Compared to ESX-1, studies on ESX-5 and ESX-3 are less described. ESX-5 is responsible for secreting ESX-5-specific substrates such as EsxM-EsxN, PE25-PPE41, PPE10, PPE13, PE_PGRS45, PE-LipY (*M. tuberculosis*), PPE-LipY (*M. marinum*), and perhaps more PE, PPE, PE_PGRS and PPE-MPTR subfamily proteins (Abdallah *et al*., 2009, 2006; Bottai *et al*., 2012; Cascioferro *et al*., 2011; Daleke *et al*., 2011). Since PE and PPE proteins present strong T-cell immunogenicity (Sayes *et al*., 2012), ESX-5 could play an essential role in switching the bacterial surface structure in response to the host environment.

ESX-3 is the most conserved region among the ESXs of mycobacterial species (Kato-Maeda *et al*., 2001). It is involved in mycobactin-mediated iron acquisition and heme uptake (Serafini *et al*., 2013; Siegrist *et al*., 2014, 2009). But how this secretion system facilitates iron and heme uptake is still unknown. A detailed discussion of the functionality of ESX-3 is given in Chapter 2. EsxG and EsxH (homologues of EsxA and EsxB) are two known substrates of ESX-3. The EsxG-EsxH heterodimer structure is similar to the EsxA-EsxB

complex but contains a potential protein interaction site on the surface. There is also a specific zinc ion binding site on EsxH, implying a possible role in zinc ion acquisition. The complex does not seem to be able to bind to mycobactin (Ilghari *et al*., 2011). Unlike EsxA-EsxB which might target the lipid membrane, the EsxG-EsxH complex has been found to target and disable the host ESCRT (Endosomal Sorting Complex Required for Transport), which participates in the delivery of the *M. tuberculosis*-containing phagosomes to the lysosome to restrict bacterial growth and initiate destruction (Mehra *et al*., 2013).

2.6 Functional Regulation of the ESXs

The gene regulation of the ESX-1 gene cluster and its functionally associated *espACD* operon (Figure 2.1) has attracted much attention due to their involvement in the virulence of *M. tuberculosis*. The regulation is rather complicated because of the large number of regulatory proteins involved.

EspR (Rv3849) binds to the promoter region of the *espACD* operon (Rv3616c-Rv3614c) and represses its expression (Raghavan *et al*., 2008). EspR has an N-terminal dimerization domain and a C-terminal DNA binding domain (Blasco *et al*., 2012). It forms a dimer of a dimer with the two dimers binding at different DNA regions. Such a dimer-dimer formation is essential to enable EspR to modulate *espACD* transcriptional activity and hence pathogenesis (Blasco *et al*., 2014, 2011). The special architecture of the EspR homotetramer also allows it to perform long range interactions with other transcriptional regulators and RNA polymerase to alter the transcription of the *espACD* operon (Rosenberg *et al*., 2011). A global screen for the binding sites of EspR in the *M. tuberculosis* genome revealed 165 loci which includes the *espACD operon*, EspR itself, the ESX-2 and ESX-5 gene clusters, genes associated with the biosynthesis of cell wall lipids PDIM, and PE PPE cell surface proteins (Blasco *et al*., 2012). The fact that the binding sites of EspR are not restricted to promoter regions, together with its long range interaction with multiple genomic regions, implies that this protein acts as a global nucleoid-associated protein (NAP) in addition to its role as a gene regulator (Blasco *et al*., 2012). Deleting EspR attenuates the pathogen not only by disabling ESX-1, but also other virulence determinants such as cell wall biosynthesis machineries, which are under the regulation of EspR (Blasco *et al*., 2012).

In addition, the *espACD* operon is negatively regulated by four other transcriptional regulators. MprAB and PhoRP derepress the operon in a similar fashion when the integrity of the cell envelope is compromised (Pang *et al*., 2013, 2007; Walters *et al*., 2006). The *espACD* operon is under the regulation of one CRP (cAMP Receptor Protein), indicating that the intracellular cAMP concentration may also affect the virulence of the bacteria (Rickman *et al*., 2005). Another NAP, Lsr2, which is the part of the global regulation allowing the pathogen to enter latent state, is up-regulated and activated when *M. tuberculosis* is in a hostile host environment, thereby repressing the *espACD* operon (Gordon *et al*., 2010).

The regulations of other ESXs are not well studied. For ESX-3, the gene regulation is described in detail in Chapter 3.

2.7 ESX-1, Apoptosis and Bacterial Spread

In both *M. tuberculosis* and *M. marinum*, disrupting the functions of the ESX-1 membrane core components or some major substrates such as EsxA, EspB and EspK prevents the pathogen from migrating between host macrophages (Gao *et al*., 2004; Guinn *et al*., 2004). Intercellular migration is enabled by the induction of macrophage apoptosis which is essentially an ESX-1-dependent process (Abdallah *et al*., 2011; Aguilo *et al*., 2013; Derrick and Morris, 2007; D. Houben *et al*., 2012; Smith *et al*., 2008). Phagosomal escape and the apoptosis of macrophages were only observed with pathogenic mycobacteria, and not with saprophytic *M. smegmatis* although it also contains an ESX-1 (Aguilo *et al*., 2013; D. Houben *et al*., 2012). EsxA (ESAT-6) may be the effector protein facilitating the pore-formation process (Smith *et al*., 2008). Although none of the ESX-5 substrates is involved, both ESX-1 and ESX-5 may contribute to the necrosis of the host cells (Abdallah *et al*., 2011). The more capable the pathogenic mycobacteria are to secrete EsxA, the more likely they are to induce apoptosis in the host cells, as EsxA exerts different effects on the membrane at different concentrations. A low concentration EsxA causes pores to form, while a high concentration of 2 to 5 μg/ml may be apoptotic (Aguilo *et al*., 2013; Derrick and Morris, 2007). Apparently apoptosis has been adopted by pathogenic mycobacteria as an efficient and safe colonization mechanism, as it allows them to find a new niche without being discovered and attacked by the host immune system (Aguilo *et al*., 2013).

2.8 Conclusion

Mycobacterium tuberculosis causes tuberculosis in humans, treatment of which is prolonged and sometimes ineffective due to factors such as drug resistance and non-compliance etc.. New drug targets are needed to facilitate the treatment. Type VII secretion systems play essential and diverse roles in the virulence and physiology of *M. tuberculosis* and they provide attractive targets for drug development. However, the substrates of T7SS are still largely unidentified and the downstream possible molecular interactions between these secreted substrates and proteins from human macrophages are also not well documented. A thorough investigation of the functionality of these special secretion systems is of great significance in understanding the pathogenicity of the pathogen and developing efficient elimination strategies.

2.9 Reference

Abdallah, A.M., Bestebroer, J., Savage, N.D.L., de Punder, K., van Zon, M., Wilson, L., Korbee, C.J., van der Sar, A.M., Ottenhoff, T.H.M., van der Wel, N.N., Bitter, W., Peters, P.J., 2011. Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation. J. Immunol. Baltim. Md 1950 187, 4744–4753. doi:10.4049/jimmunol.1101457

Abdallah, A.M., Gey van Pittius, N.C., Champion, P.A.D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W., 2007. Type VII secretion--mycobacteria show the way. Nat. Rev. Microbiol. 5, 883–891. doi:10.1038/nrmicro1773

Abdallah, A.M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R.J.P., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., Luirink, J., Bitter, W., 2006. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol. Microbiol. 62, 667–679. doi:10.1111/j.1365-2958.2006.05409.x

Abdallah, A.M., Verboom, T., Weerdenburg, E.M., Gey van Pittius, N.C., Mahasha, P.W., Jiménez, C., Parra, M., Cadieux, N., Brennan, M.J., Appelmelk, B.J., Bitter, W., 2009. PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Mol. Microbiol. 73, 329–340. doi:10.1111/j.1365-2958.2009.06783.x

Aguilo, J.I., Alonso, H., Uranga, S., Marinova, D., Arbués, A., de Martino, A., Anel, A., Monzon, M., Badiola, J., Pardo, J., Brosch, R., Martin, C., 2013. ESX-1-induced apoptosis is involved in cell-to-cell spread of *Mycobacterium tuberculosis*. Cell. Microbiol. 15, 1994–2005. doi:10.1111/cmi.12169

Akpe San Roman, S., Facey, P.D., Fernandez-Martinez, L., Rodriguez, C., Vallin, C., Del Sol, R., Dyson, P., 2010. A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in Streptomyces coelicolor. Microbiol. Read. Engl. 156, 1719–1729. doi:10.1099/mic.0.037069-0

Andersen, P., Askgaard, D., Ljungqvist, L., Bentzon, M.W., Heron, I., 1991. T-cell proliferative response to antigens secreted by *Mycobacterium tuberculosis*. Infect. Immun. 59, 1558–1563.

Berthet, F.X., Rasmussen, P.B., Rosenkrands, I., Andersen, P., Gicquel, B., 1998. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). Microbiol. Read. Engl. 144 (Pt 11), 3195–3203.

Bitter, W., Houben, E.N.G., Bottai, D., Brodin, P., Brown, E.J., Cox, J.S., Derbyshire, K., Fortune, S.M., Gao, L.-Y., Liu, J., Gey van Pittius, N.C., Pym, A.S., Rubin, E.J., Sherman, D.R., Cole, S.T., Brosch, R., 2009. Systematic genetic nomenclature for type VII secretion systems. PLoS Pathog. 5, e1000507. doi:10.1371/journal.ppat.1000507

Blasco, B., Chen, J.M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., Pojer, F., Cole, S.T., 2012. Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. PLoS Pathog. 8, e1002621. doi:10.1371/journal.ppat.1002621

Blasco, B., Japaridze, A., Stenta, M., Wicky, B.I.M., Dietler, G., Dal Peraro, M., Pojer, F., Cole, S.T., 2014. Functional Dissection of Intersubunit Interactions in the EspR Virulence Regulator of *Mycobacterium tuberculosis*. J. Bacteriol. 196, 1889–1900. doi:10.1128/JB.00039-14

Blasco, B., Stenta, M., Alonso-Sarduy, L., Dietler, G., Peraro, M.D., Cole, S.T., Pojer, F., 2011. Atypical DNA recognition mechanism used by the EspR virulence regulator of

Mycobacterium tuberculosis. Mol. Microbiol. 82, 251–264. doi:10.1111/j.1365-2958.2011.07813.x

Bottai, D., Di Luca, M., Majlessi, L., Frigui, W., Simeone, R., Sayes, F., Bitter, W., Brennan, M.J., Leclerc, C., Batoni, G., Campa, M., Brosch, R., Esin, S., 2012. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol. Microbiol. 83, 1195–1209. doi:10.1111/j.1365-2958.2012.08001.x

Bottai, D., Majlessi, L., Simeone, R., Frigui, W., Laurent, C., Lenormand, P., Chen, J., Rosenkrands, I., Huerre, M., Leclerc, C., Cole, S.T., Brosch, R., 2011. ESAT-6 secretion-independent impact of ESX-1 genes espF and espG1 on virulence of *Mycobacterium tuberculosis*. J. Infect. Dis. 203, 1155–1164. doi:10.1093/infdis/jiq089

Brodin, P., Majlessi, L., Marsollier, L., de Jonge, M.I., Bottai, D., Demangel, C., Hinds, J., Neyrolles, O., Butcher, P.D., Leclerc, C., Cole, S.T., Brosch, R., 2006. Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. Infect. Immun. 74, 88–98. doi:10.1128/IAI.74.1.88-98.2006

Brodin, P., Rosenkrands, I., Andersen, P., Cole, S.T., Brosch, R., 2004. ESAT-6 proteins: protective antigens and virulence factors? Trends Microbiol. 12, 500–508. doi:10.1016/j.tim.2004.09.007

Brown, G.D., Dave, J.A., Gey van Pittius, N.C., Stevens, L., Ehlers, M.R., Beyers, A.D., 2000. The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. Gene 254, 147–155.

Callahan, B., Nguyen, K., Collins, A., Valdes, K., Caplow, M., Crossman, D.K., Steyn, A.J.C., Eisele, L., Derbyshire, K.M., 2010. Conservation of structure and protein-protein interactions mediated by the secreted mycobacterial proteins EsxA, EsxB, and EspA. J. Bacteriol. 192, 326–335. doi:10.1128/JB.01032-09

Carlsson, F., Joshi, S.A., Rangell, L., Brown, E.J., 2009. Polar localization of virulence-related Esx-1 secretion in mycobacteria. PLoS Pathog. 5, e1000285. doi:10.1371/journal.ppat.1000285

Cascioferro, A., Daleke, M.H., Ventura, M., Donà, V., Delogu, G., Palù, G., Bitter, W., Manganelli, R., 2011. Functional dissection of the PE domain responsible for translocation of PE_PGRS33 across the mycobacterial cell wall. PloS One 6, e27713. doi:10.1371/journal.pone.0027713

Champion, P.A.D., 2013. Disconnecting in vitro ESX-1 secretion from mycobacterial virulence. J. Bacteriol. 195, 5418–5420. doi:10.1128/JB.01145-13

Champion, P.A.D., Champion, M.M., Manzanillo, P., Cox, J.S., 2009. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol. Microbiol. 73, 950–962. doi:10.1111/j.1365-2958.2009.06821.x

Champion, P.A.D., Stanley, S.A., Champion, M.M., Brown, E.J., Cox, J.S., 2006. C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. Science 313, 1632–1636. doi:10.1126/science.1131167

Chen, J.M., Boy-Röttger, S., Dhar, N., Sweeney, N., Buxton, R.S., Pojer, F., Rosenkrands, I., Cole, S.T., 2012. EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis*. J. Bacteriol. 194, 884–893. doi:10.1128/JB.06417-11

Chen, J.M., Zhang, M., Rybniker, J., Basterra, L., Dhar, N., Tischler, A.D., Pojer, F., Cole, S.T., 2013a. Phenotypic profiling of *Mycobacterium tuberculosis* EspA point mutants reveals that blockage of ESAT-6 and CFP-10 secretion in vitro does not always correlate with attenuation of virulence. J. Bacteriol. 195, 5421–5430. doi:10.1128/JB.00967-13

Chen, J.M., Zhang, M., Rybniker, J., Boy-Röttger, S., Dhar, N., Pojer, F., Cole, S.T., 2013b. *Mycobacterium tuberculosis* EspB binds phospholipids and mediates EsxA-independent virulence. Mol. Microbiol. 89, 1154–1166. doi:10.1111/mmi.12336

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537–544. doi:10.1038/31159

Converse, S.E., Cox, J.S., 2005. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in Mycobacterium smegmatis. J. Bacteriol. 187, 1238–1245. doi:10.1128/JB.187.4.1238-1245.2005

Coros, A., Callahan, B., Battaglioli, E., Derbyshire, K.M., 2008. The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Mol. Microbiol. 69, 794–808. doi:10.1111/j.1365-2958.2008.06299.x

Daleke, M.H., Cascioferro, A., de Punder, K., Ummels, R., Abdallah, A.M., van der Wel, N., Peters, P.J., Luirink, J., Manganelli, R., Bitter, W., 2011. Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. J. Biol. Chem. 286, 19024–19034. doi:10.1074/jbc.M110.204966

Daleke, M.H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C.M.J.E., Luirink, J., Bitter, W., 2012a. General secretion signal for the mycobacterial type VII secretion pathway. Proc. Natl. Acad. Sci. U. S. A. 109, 11342–11347. doi:10.1073/pnas.1119453109

Daleke, M.H., van der Woude, A.D., Parret, A.H.A., Ummels, R., de Groot, A.M., Watson, D., Piersma, S.R., Jiménez, C.R., Luirink, J., Bitter, W., Houben, E.N.G., 2012b. Specific chaperones for the type VII protein secretion pathway. J. Biol. Chem. 287, 31939–31947. doi:10.1074/jbc.M112.397596

Dave, J.A., Gey van Pittius, N.C., Beyers, A.D., Ehlers, M.R.W., Brown, G.D., 2002. Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. BMC Microbiol. 2, 30.

De Jonge, M.I., Pehau-Arnaudet, G., Fretz, M.M., Romain, F., Bottai, D., Brodin, P., Honoré, N., Marchal, G., Jiskoot, W., England, P., Cole, S.T., Brosch, R., 2007. ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J. Bacteriol. 189, 6028–6034. doi:10.1128/JB.00469-07

Demangel, C., Brodin, P., Cockle, P.J., Brosch, R., Majlessi, L., Leclerc, C., Cole, S.T., 2004. Cell envelope protein PPE68 contributes to *Mycobacterium tuberculosis* RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6. Infect. Immun. 72, 2170–2176.

Derrick, S.C., Morris, S.L., 2007. The ESAT6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression. Cell. Microbiol. 9, 1547–1555. doi:10.1111/j.1462-5822.2007.00892.x

Di Luca, M., Bottai, D., Batoni, G., Orgeur, M., Aulicino, A., Counoupas, C., Campa, M., Brosch, R., Esin, S., 2012. The ESX-5 associated eccB-EccC locus is essential for

Mycobacterium tuberculosis viability. PloS One 7, e52059. doi:10.1371/journal.pone.0052059

Fortune, S.M., Jaeger, A., Sarracino, D.A., Chase, M.R., Sassetti, C.M., Sherman, D.R., Bloom, B.R., Rubin, E.J., 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. Proc. Natl. Acad. Sci. U. S. A. 102, 10676–10681. doi:10.1073/pnas.0504922102

Gao, L.-Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J.N., Brown, E.J., 2004. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. Mol. Microbiol. 53, 1677–1693. doi:10.1111/j.1365-2958.2004.04261.x

Garces, A., Atmakuri, K., Chase, M.R., Woodworth, J.S., Krastins, B., Rothchild, A.C., Ramsdell, T.L., Lopez, M.F., Behar, S.M., Sarracino, D.A., Fortune, S.M., 2010. EspA acts as a critical mediator of ESX1-dependent virulence in *Mycobacterium tuberculosis* by affecting bacterial cell wall integrity. PLoS Pathog. 6, e1000957. doi:10.1371/journal.ppat.1000957

Gey Van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. 2, RESEARCH0044.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., van Helden, P.D., Warren, R.M., 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol. Biol. 6, 95. doi:10.1186/1471-2148-6-95

Gordon, B.R.G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., Navarre, W.W., Xia, B., Liu, J., 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 107, 5154–5159. doi:10.1073/pnas.0913551107

Guinn, K.M., Hickey, M.J., Mathur, S.K., Zakel, K.L., Grotzke, J.E., Lewinsohn, D.M., Smith, S., Sherman, D.R., 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. doi:10.1046/j.1365-2958.2003.03844.x

Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeón, L., Abdallah, A.M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., van der Laan, T., Kant, A., Bossers-de Vries, R., Willemsen, P., Bitter, W., van Soolingen, D., Brosch, R., van der Wel, N., Peters, P.J., 2012. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. Cell. Microbiol. 14, 1287–1298. doi:10.1111/j.1462-5822.2012.01799.x

Houben, E.N.G., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jiménez, C.R., Ottenhoff, T.H.M., Luirink, J., Bitter, W., 2012. Composition of the type VII secretion system membrane complex. Mol. Microbiol. 86, 472–484. doi:10.1111/j.1365-2958.2012.08206.x

Houben, E.N.G., Korotkov, K.V., Bitter, W., 2013a. Take five - Type VII secretion systems of Mycobacteria. Biochim. Biophys. Acta. doi:10.1016/j.bbamcr.2013.11.003

Houben, E.N.G., Korotkov, K.V., Bitter, W., 2013b. Take five - Type VII secretion systems of Mycobacteria. Biochim. Biophys. Acta. doi:10.1016/j.bbamcr.2013.11.003

Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R.G., Derrick, S.C., Collins, F.M., Morris, S.L., King, C.H., Jacobs, W.R., Jr, 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc. Natl. Acad. Sci. U. S. A. 100, 12420–12425. doi:10.1073/pnas.1635213100

Ilghari, D., Lightbody, K.L., Veverka, V., Waters, L.C., Muskett, F.W., Renshaw, P.S., Carr, M.D., 2011. Solution structure of the *Mycobacterium tuberculosis* EsxG·EsxH complex: functional implications and comparisons with other M. tuberculosis Esx family complexes. J. Biol. Chem. 286, 29993–30002. doi:10.1074/jbc.M111.248732

Joshi, S.A., Ball, D.A., Sun, M.G., Carlsson, F., Watkins, B.Y., Aggarwal, N., McCracken, J.M., Huynh, K.K., Brown, E.J., 2012. EccA1, a component of the Mycobacterium marinum ESX-1 protein virulence factor secretion pathway, regulates mycolic acid lipid synthesis. Chem. Biol. 19, 372–380. doi:10.1016/j.chembiol.2012.01.008

Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N., Small, P.M., 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. 11, 547–554. doi:10.1101/gr166401

Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., Sherman, D.R., 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. J. Infect. Dis. 187, 117–123. doi:10.1086/345862

Ligon, L.S., Hayden, J.D., Braunstein, M., 2012. The ins and outs of *Mycobacterium tuberculosis* protein export. Tuberc. Edinb. Scotl. 92, 121–132. doi:10.1016/j.tube.2011.11.005

MacGurn, J.A., Cox, J.S., 2007. A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. Infect. Immun. 75, 2668–2678. doi:10.1128/IAI.01872-06

MacGurn, J.A., Raghavan, S., Stanley, S.A., Cox, J.S., 2005. A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. Mol. Microbiol. 57, 1653–1663. doi:10.1111/j.1365-2958.2005.04800.x

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., Stover, C.K., 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J. Bacteriol. 178, 1274–1282.

McLaughlin, B., Chon, J.S., MacGurn, J.A., Carlsson, F., Cheng, T.L., Cox, J.S., Brown, E.J., 2007. A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. PLoS Pathog. 3, e105. doi:10.1371/journal.ppat.0030105

Mehra, A., Zahra, A., Thompson, V., Sirisaengtaksin, N., Wells, A., Porto, M., Köster, S., Penberthy, K., Kubota, Y., Dricot, A., Rogan, D., Vidal, M., Hill, D.E., Bean, A.J., Philips, J.A., 2013. *Mycobacterium tuberculosis* type VII secreted effector EsxH targets host ESCRT to impair trafficking. PLoS Pathog. 9, e1003734. doi:10.1371/journal.ppat.1003734

Ohol, Y.M., Goetz, D.H., Chan, K., Shiloh, M.U., Craik, C.S., Cox, J.S., 2010. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell Host Microbe 7, 210–220. doi:10.1016/j.chom.2010.02.006

Okkels, L.M., Andersen, P., 2004. Protein-protein interactions of proteins from the ESAT-6 family of *Mycobacterium tuberculosis*. J. Bacteriol. 186, 2487–2491.

Pang, X., Samten, B., Cao, G., Wang, X., Tvinnereim, A.R., Chen, X.-L., Howard, S.T., 2013. MprAB regulates the espA operon in *Mycobacterium tuberculosis* and modulates ESX-1 function and host cytokine response. J. Bacteriol. 195, 66–75. doi:10.1128/JB.01067-12

Pang, X., Vu, P., Byrd, T.F., Ghanny, S., Soteropoulos, P., Mukamolova, G.V., Wu, S., Samten, B., Howard, S.T., 2007. Evidence for complex interactions of stress-associated regulons in an mprAB deletion mutant of *Mycobacterium tuberculosis*. Microbiol. Read. Engl. 153, 1229–1242. doi:10.1099/mic.0.29281-0

Poulsen, C., Panjikar, S., Holton, S.J., Wilmanns, M., Song, Y.-H., 2014. WXG100 Protein Superfamily Consists of Three Subfamilies and Exhibits an α-Helical C-Terminal Conserved Residue Pattern. PloS One 9, e89313. doi:10.1371/journal.pone.0089313

Pym, A.S., Brodin, P., Brosch, R., Huerre, M., Cole, S.T., 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol. Microbiol. 46, 709–717.

Pym, A.S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K.E., Marchal, G., Leclerc, C., Cole, S.T., 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat. Med. 9, 533–539. doi:10.1038/nm859

Raghavan, S., Manzanillo, P., Chan, K., Dovey, C., Cox, J.S., 2008. Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. Nature 454, 717–721. doi:10.1038/nature07219

Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norman, J., Williamson, R.A., Carr, M.D., 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J. 24, 2491–2498. doi:10.1038/sj.emboj.7600732

Renshaw, P.S., Panagiotidou, P., Whelan, A., Gordon, S.V., Hewinson, R.G., Williamson, R.A., Carr, M.D., 2002. Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. J. Biol. Chem. 277, 21598–21603. doi:10.1074/jbc.M201625200

Rickman, L., Scott, C., Hunt, D.M., Hutchinson, T., Menéndez, M.C., Whalan, R., Hinds, J., Colston, M.J., Green, J., Buxton, R.S., 2005. A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the rpfA gene coding for a resuscitation promoting factor. Mol. Microbiol. 56, 1274–1286. doi:10.1111/j.1365-2958.2005.04609.x

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., 2002. ideR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. 70, 3371–3381.

Rosenberg, O.S., Dovey, C., Tempesta, M., Robbins, R.A., Finer-Moore, J.S., Stroud, R.M., Cox, J.S., 2011. EspR, a key regulator of *Mycobacterium tuberculosis* virulence, adopts a unique dimeric structure among helix-turn-helix proteins. Proc. Natl. Acad. Sci. U. S. A. 108, 13450–13455. doi:10.1073/pnas.1110242108

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jiménez, C.R., Daffé, M., Appelmelk, B.J., Bitter, W., van der Wel, N., Peters, P.J., 2010. Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. PLoS Pathog. 6, e1000794. doi:10.1371/journal.ppat.1000794

Sayes, F., Sun, L., Di Luca, M., Simeone, R., Degaiffier, N., Fiette, L., Esin, S., Brosch, R., Bottai, D., Leclerc, C., Majlessi, L., 2012. Strong immunogenicity and cross-reactivity of *Mycobacterium tuberculosis* ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. Cell Host Microbe 11, 352–363. doi:10.1016/j.chom.2012.03.003

Serafini, A., Pisu, D., Palù, G., Rodriguez, G.M., Manganelli, R., 2013. The ESX-3 Secretion System Is Necessary for Iron and Zinc Homeostasis in *Mycobacterium tuberculosis*. PLoS ONE 8, e78351. doi:10.1371/journal.pone.0078351

Siegrist, M.S., Steigedal, M., Ahmad, R., Mehra, A., Dragset, M.S., Schuster, B.M., Philips, J.A., Carr, S.A., Rubin, E.J., 2014. Mycobacterial esx-3 requires multiple components for iron acquisition. mBio 5. doi:10.1128/mBio.01073-14

Siegrist, M.S., Unnikrishnan, M., McConnell, M.J., Borowsky, M., Cheng, T.-Y., Siddiqi, N., Fortune, S.M., Moody, D.B., Rubin, E.J., 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc. Natl. Acad. Sci. U. S. A. 106, 18792–18797. doi:10.1073/pnas.0900589106

Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J., Liu, J., McDonald, K.L., Szyk, A., LaRonde-LeBlanc, N., Gao, L.-Y., 2008. Evidence for Pore Formation in Host Cell Membranes by ESX-1-Secreted ESAT-6 and Its Role in Mycobacterium marinum Escape from the Vacuole. Infect. Immun. 76, 5478–5487. doi:10.1128/IAI.00614-08

Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., Strynadka, N.C.J., 2013. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. J. Biol. Chem. 288, 17782–17790. doi:10.1074/jbc.M113.462036

Sørensen, A.L., Nagai, S., Houen, G., Andersen, P., Andersen, A.B., 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infect. Immun. 63, 1710–1717.

Stanley, S.A., Raghavan, S., Hwang, W.W., Cox, J.S., 2003. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. Proc. Natl. Acad. Sci. U. S. A. 100, 13001–13006. doi:10.1073/pnas.2235593100

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D., 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 103, 8060–8065. doi:10.1073/pnas.0602606103

Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell, B.G., Cole, S.T., 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tuber. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. 79, 329–342. doi:10.1054/tuld.1999.0220

Tundup, S., Akhter, Y., Thiagarajan, D., Hasnain, S.E., 2006. Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. FEBS Lett. 580, 1285–1293. doi:10.1016/j.febslet.2006.01.042

Voskuil, M.I., Schnappinger, D., Rutherford, R., Liu, Y., Schoolnik, G.K., 2004. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. Tuberc. Edinb. Scotl. 84, 256–262. doi:10.1016/j.tube.2003.12.014

Wagner, J.M., Evans, T.J., Chen, J., Zhu, H., Houben, E.N.G., Bitter, W., Korotkov, K.V., 2013. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. J. Struct. Biol. 184, 115–128. doi:10.1016/j.jsb.2013.09.022

Wagner, J.M., Evans, T.J., Korotkov, K.V., 2014. Crystal structure of the N-terminal domain of $EccA_1$ ATPase from the ESX-1 secretion system of *Mycobacterium tuberculosis*. Proteins 82, 159–163. doi:10.1002/prot.24351

Walters, S.B., Dubnau, E., Kolesnikova, I., Laval, F., Daffe, M., Smith, I., 2006. The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol. Microbiol. 60, 312–330. doi:10.1111/j.1365-2958.2006.05102.x

Wirth, S.E., Krywy, J.A., Aldridge, B.B., Fortune, S.M., Fernandez-Suarez, M., Gray, T.A., Derbyshire, K.M., 2012. Polar assembly and scaffolding proteins of the virulence-associated ESX-1 secretory apparatus in mycobacteria. Mol. Microbiol. 83, 654–664. doi:10.1111/j.1365-2958.2011.07958.x

Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S.J., Edwards, N., Zhu, X., Fenselau, C., Gao, L.-Y., 2007. A unique Mycobacterium ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. Mol. Microbiol. 66, 787–800. doi:10.1111/j.1365-2958.2007.05959.x

Chapter 3. Mycobacterial Strategies for Acquiring Scarce Iron

3.1 Introduction

Mycobacteria, like most other living organisms, require iron for many vital functions. Iron forms the essential catalytic centre of the active site of various enzymes enabling enzymatic reactions. The ability of iron to receive and donate an electron thereby oscillating between the ferrous (Fe^{2+}) and ferric (Fe^{3+}) oxidative states enables the enzyme to catalyse redox reactions. For this reason, iron is often associated with cytochromes responsible for oxidative phosphorylation and energy production (Banerjee *et al*., 2011). Iron-sulfur clusters are essential cofactors of many enzymes involved in amino acid and pyrimidine biogenesis, the tricarboxylic acid cycle as well as electron transport (Saini *et al*., 2012). Iron is one of the most abundant elements on earth, however, under the earth's oxidizing environment at physiological pH, iron exists predominantly as insoluble ferric salts such as iron oxide, iron hydroxide and iron phosphate which cannot be assimilated by bacteria. Free iron ions are therefore scarce. Iron acquisition is even more challenging for pathogenic bacteria because iron ions are bound to host iron-binding proteins, such as transferrin and lactoferrin which serve as host iron transporters, the iron storage protein ferritin and protoporphyrins in heme and hemoproteins (Weinberg, 1984). During infection, the host restricts the amount of circulating transferrin-bound iron in the body and reduces the uptake of dietary iron utilising iron-deprivation as a host anti-microbial defense mechanism (Frederick *et al*., 2009; Kontoghiorghes and Weinberg, 1995; Ratledge, 2004). Pathogenic mycobacteria are, however, able to cause disease despite the severely iron limited host environment. To overcome iron-deprivation mycobacterial pathogens have evolved iron acquiring machineries which are more efficient than those of their vertebrate hosts. In this review, the strategies mycobacteria use to access iron and the iron-dependent gene regulation responsible for the maintenance of iron homeostasis in mycobacteria are discussed. Additionally, a brief opinion is given regarding the selection of iron acquisition machineries as drug targets.

3.2 Siderophore-mediated Iron Acquisition

A major mechanism employed by mycobacteria to compete for the limited available iron is the use of high affinity iron chelators, siderophores, which are predominantly produced during iron deprivation (De Voss *et al*., 1999). There are three types of siderophores, mycobactin, carboxymycobactin and exochelin, where mycobactin and carboxymycobactin share a core structure which is distinct from exochelin. Mycobactin is cell envelope-associated and facilitates the transport of iron through the cell envelope into the cytoplasm while carboxymycobactin and exochelin are secreted iron-chelators which acquire iron in the

extracellular milieu and transport iron into the cytoplasm of the bacteria (De Voss *et al*., 1999).

3.2.1 The siderophores: Mycobactin, carboxymycobactin and exochelin

Mycobactins can be isolated from almost all mycobacterial species except *M. microti* (Barclay and Ratledge, 1988), *M. paratuberculosis*, and *M. vaccae* (Messenger *et al*., 1986). They have a core structure comprising 5 amino acids with a characteristic phenyloxazolidine ring derived from salicylate with saturated or unsaturated alkyl side chains of various lengths on the hydroxylated lysine residue at the centre of the molecule (Barclay *et al*., 1985) (Figure 3.1 and Figure 3.2a). The long alkyl chain confers the lipophilic property of mycobactin and thereby locates it within the cell envelope (De Voss *et al*., 1999).

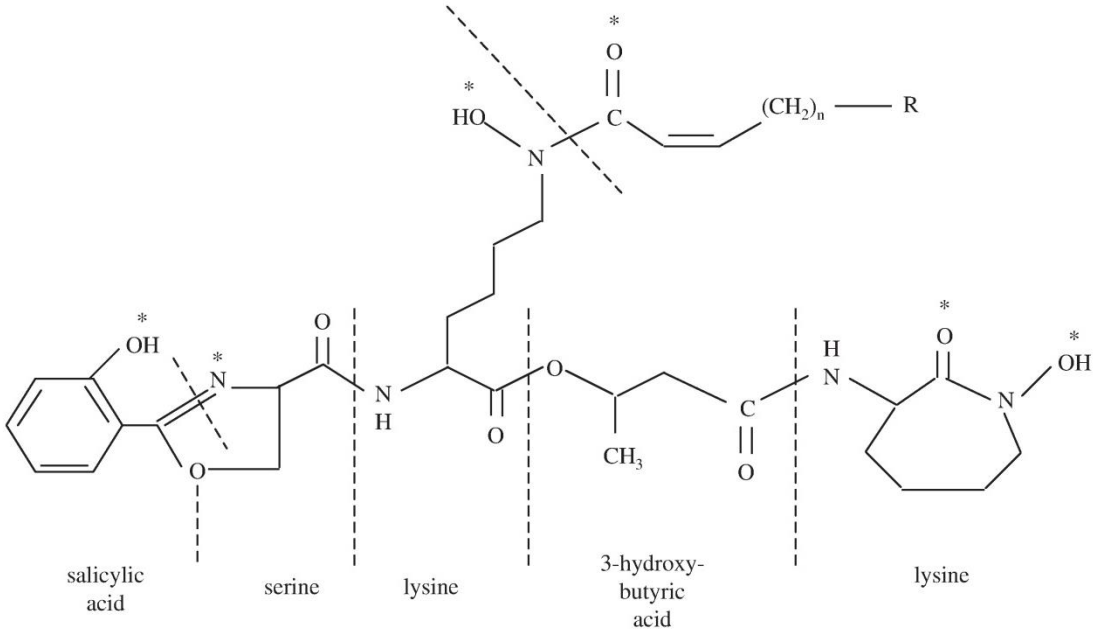
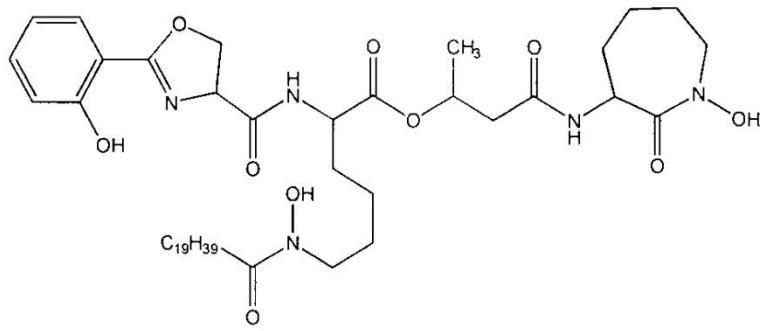


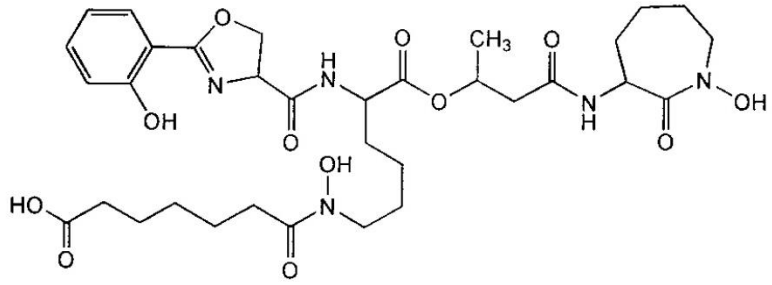
Figure 3.1. The general structure of mycobactin from *M. tuberculosis*. The structure consists of 5 amino acids including one salicylic acid, one serine, two lysines, and one 3-hydroxybutyric acid. A long alkyl chain extends from the side chain of the middle lysine residue where the length of the R-group may vary amongst the mycobacteria species. The asterisk (*) indicates the chelating group for the ferric ion binding. (Ratledge, 2004)

Early studies created confusion by using the term "exochelin" to describe all extracellular siderophores isolated from both saprophytic and pathogenic mycobacteria. This became inappropriate when "exochelin" from pathogenic mycobacteria was found to have a different core structure from that of saprophytic mycobacteria. The former has a core structure similar to that of mycobactin, but with shorter alkyl side chains that terminate with either a carboxyl group or a methyl ester (Figure 3.2b) (Gobin *et al*., 1995; Wong *et al*., 1996). This feature

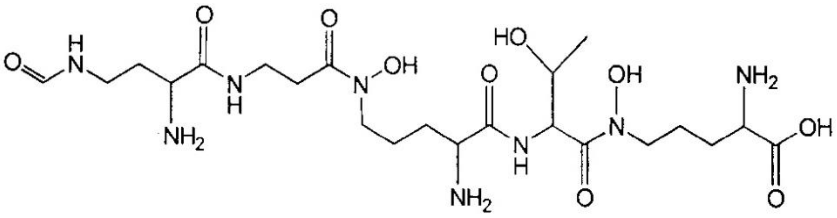
makes these "exochelin" molecules more hydrophilic than the lipophilic mycobactins. These molecules were renamed carboxymycobactin (Ratledge and Ewing, 1996), although less frequently "exomycobactin" is used.



(a)



(b)



(c)

Figure 3.2. Examples of mycobacterial siderophores. *M. tuberculosis* mycobactin (a) and carboxymycobactin (exomycobactin) structures contain a typical phenyloxazolidine ring (b); The difference between the two molecules is that the *N*-acyl chain on the hydroxylated lysine is a 19-carbon alkyl group (a) while that of carboxymycobactin is a 6-carbon alkyl group coupled with a carboxylic acid (b) (De Voss *et al*., 2000). *M. smegmatis* exochelin structure containing 5 amino acids connected with 2 unusual peptide bonds (c). (De Voss *et al*., 1999)

The extracellular siderophores from saprophytic mycobacteria retained the name exochelin. This water-soluble molecule has a different core structure to mycobactin/carboxymycobactin. consisting of a formylated pentapeptide: *N*-(δ-*N*-formayl, δ-*N*-hydroxy-R-ornithyl)-βalaninyl-δ-*N*-hydroxy-R-ornithinyl-*R*-allo-threoninyl-δ-*N*-hydroxy-S-ornithine (Figure 3.2c) (Sharman *et al*., 1995a). The two peptide bonds within the molecule are atypical making them resistant to peptidase hydrolysis. The structure of exochelin varies in some mycobacterial species, e.g.

in *M. neoaurum* exochelin is a hexapeptide with an unusual β-hydroxyhistidine residue (Sharman *et al*., 1995b).

Carboxymycobactin was thought to be exclusive to pathogenic mycobacteria until it was found to comprise about 5% of the total siderophore content in *M. smegmatis* (Ratledge and Ewing, 1996). The coexistence of carboxymycobactin and exochelin in *M. smegmatis* further assisted in resolving the historical confusion between the two siderophores. The difference in core structure of mycobactin/carboxymycobactin and exochelin necessitates distinctive biosynthetic pathways and transport mechanisms for the two unrelated siderophores. The identification of the two biosynthetic pathways further clarified the distinction between the two types of siderophores. The machinery for mycobactin biosynthesis is encoded in two gene clusters, *mbtA-J* (Quadri *et al*., 1998) and *mbtK-N* (Krithika *et al*., 2006) in both *M. tuberculosis* and *M. smegmatis*. The enzymes responsible for exochelin biosynthesis are encoded by *fxbA*, *fxbB* and *fxbC* genes in *M. smegmatis* (Fiss *et al*., 1994; Yu *et al*., 1998; Zhu *et al*., 1998). The biosynthetic pathways of the mycobactin (Madigan *et al*., 2012; McMahon *et al*., 2012) and exochelin (Zhu *et al*., 1998) are described in detail elsewhere.

3.2.2 Siderophore transport

The import and export of the secreted siderophores, carboxymycobactin and exochelin utilises unique mechanisms and machineries. Carboxymycobactin transport is dominant in pathogenic mycobacteria while exochelin transport is more important in saprophytic species.

3.2.2.1 Carboxymycobactin transport

Hydrophilic deferri-carboxymycobactin is secreted into the extracellular milieu to acquire iron from the environment, and presumably the host transferrin (Gobin *et al*., 1995) and at a slower rate, ferritin (Gobin and Horwitz, 1996). The ferri-carboxymycobactin is transported back through the mycobacterial outer membrane via the Msp family of porins (Jones and Niederweis, 2010). The Msp porins are abundant membrane proteins which allow a variety of nutrients to enter the cell by non-specific diffusion (Jones and Niederweis, 2010; Ojha and Hatfull, 2007).

In the mycobacterial periplasmic space, ferri-carboxymycobactin either transfers its iron to the cell-wall associated mycobactin or delivers it to the plasma membrane-bound protein complex, IrtAB (iron-regulated transporter A and B, Rv1348 and Rv1349) (Rodriguez and Smith, 2006). IrtA and IrtB are ABC transporters, which are homologues of the YbtPQ system involved in iron transport in *Yersinia pestis* (Fetherston *et al*., 1999). Inactivation of this gene pair impairs the utilization of ferri-carboxymycobactin and affects the optimal growth of *M. tuberculosis* under low iron conditions but does not abolish the production and secretion of the siderophores (Rodriguez and Smith, 2006). It has been proposed that IrtA possesses the dual function of iron import and ferric iron reduction. It is thought to release the

bound iron in the ferrous form, possibly due to the reductase role of its FAD-binding domain (Ryndak *et al*., 2010). However, this activity was not obtained *in vitro*, possibly because other components are also required to assist in this role (Ryndak *et al*., 2010). A possible candidate is ViuB (Rv2895c), whose homologue in *Vibrio cholera* encodes a cytosolic protein required for the utilization of ferric-vibriobactin (Butterton and Calderwood, 1994). ViuB forms a complex with IrtB (Farhana *et al*., 2008), but is not essential for iron acquisition (Santhanagopalan and Rodriguez, 2012). Whether ViuB is involved in siderophore import remains unclear.

After releasing the bound iron, deferri-carboxymycobactin may either be degraded or secreted again to acquire more iron. Siderophore catabolism has been relatively neglected in the literature whereas siderophore secretion has been well described. The carboxymycobactin export system was shown to be the membrane complex of MmpL5/MmpS5 and MmpL4/MmpS4 (Wells *et al*., 2013). The MmpS4/S5 double mutant showed an iron starvation phenotype under high iron conditions in which the uptake of carboxymycobactin was not affected but the siderophore biosynthesis and secretion were strongly impaired indicating that MmpS4/S5 is directly involved in mycobactin/carboxymycobactin export.

3.2.2.2 Exochelin transport

Exochelin transport has not been well characterized in mycobacteria (Figure 3.3). The deferri-exochelin may be exported by an ABC transporter protein, ExiT, encoded in the same operon as the exochelin synthetase genes *fxbB* and *fxbC* (Yu *et al*., 1998; Zhu *et al*., 1998). ExiT appears to be involved in both the biosynthesis and export of exochelin, directly coupling these processes (Zhu *et al*., 1998). After binding extracellular iron, ferri-exochelin is recognized by some unknown receptors on the outer membrane and imported into the periplasmic space (Jones and Niederweis, 2010). There, ferri-exochelin may associate with FxuD, the ferri-exochelin carrier, and either transfer the iron to mycobactin or to the ferric-exochelin importer membrane complex formed by FxuA, FxuB, and FxuC. FxuA, FxuB and FuxC are homologues of FepG, FepC and FepD which form the membrane-bound permease which is responsible for the uptake of the siderophore ferrienterobactin in *Escherichia coli* (Fiss *et al*., 1994). In the cytoplasm, exochelin-bound ferric iron is reduced to ferrous iron by an unidentified reductase and released.

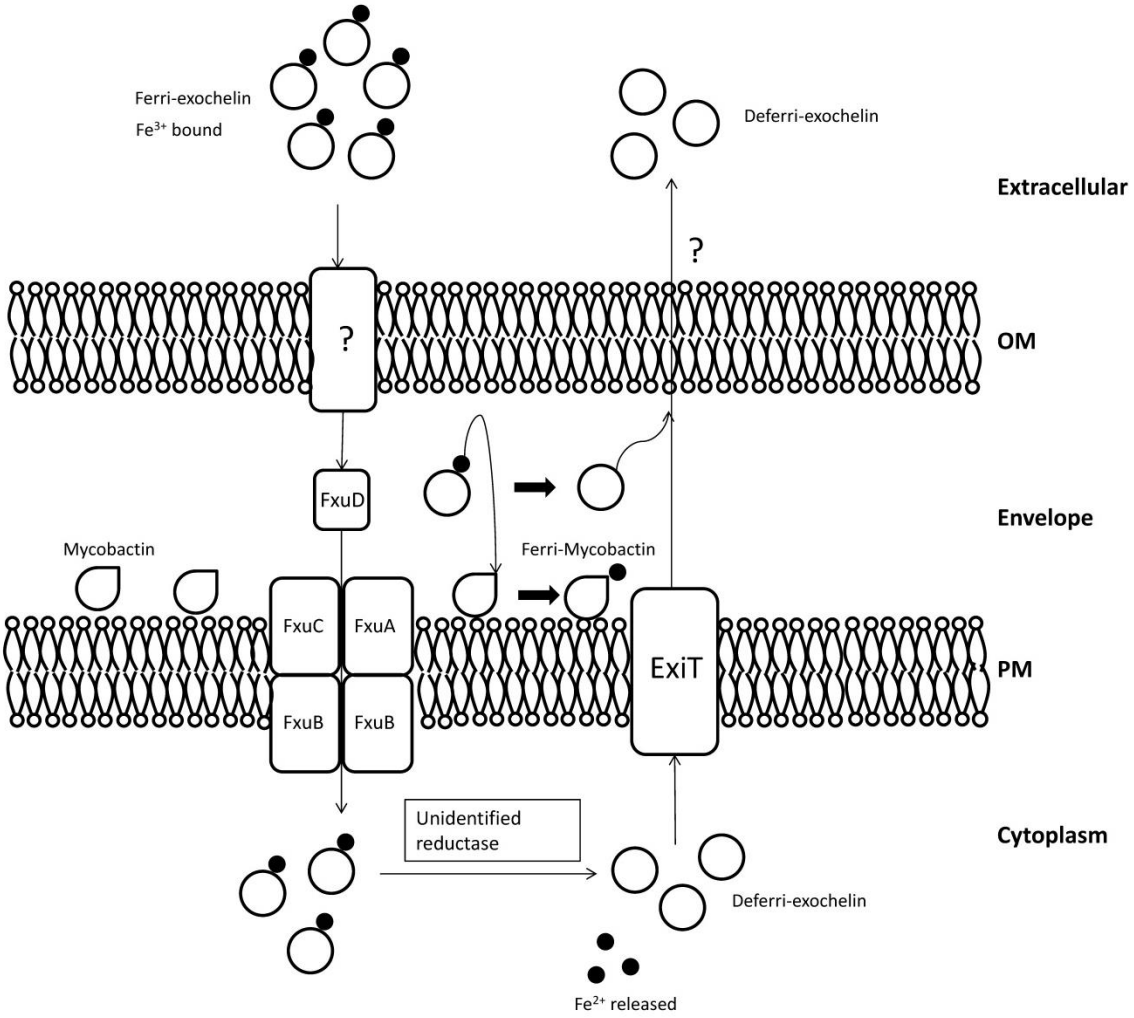


Figure 3.3. Exochelin import and export in *M. smegmatis* (adapted from Ratledge, 2004). The ferri-exochelin gets recognized by an unknown protein (indicated with a "?") on the outer membrane and delivered to FxuA-B-C complex on the plasma membrane while coupled with FxuD. While in the cell envelope, ferri-exochelin may transfer the iron to mycobactin on the membrane. Once ferri-exochelin are imported into the cytoplasm, the ferric iron gets reduced by unidentified reductase and released for utilization. The deferri-exochelin is transported out of the cell via ExiT. (Fiss *et al*., 1994; Yu *et al*., 1998; Zhu *et al*., 1998)

3.3 Heme Transport and Its Metabolism in Mycobacteria

Wells and colleagues used heme as the alternative iron source to successfully rescue their *M. tuberculosis* mutants in which the mycobactin biosynthetic pathway was disabled under low iron conditions (Wells *et al*., 2013). Heme is the iron-containing prosthetic group in the hemoglobin molecules which are abundant in erythrocytes. Iron binds to the oxygen molecule allowing erythrocytes to carry and supply oxygen for cellular respiration in the vertebrate's body where two thirds of iron is in the heme-bound state. Heme is therefore a reservoir of

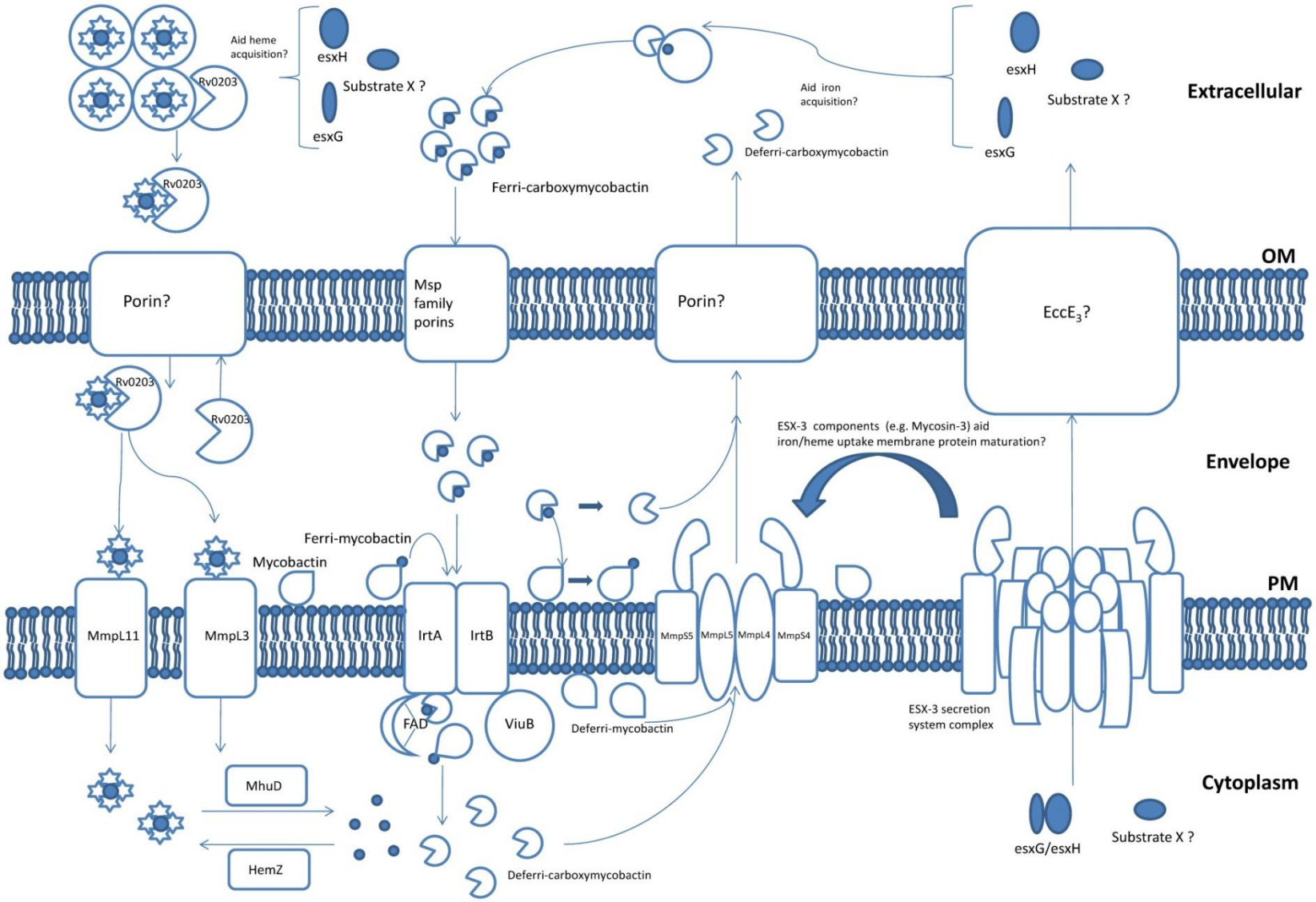
iron for many pathogenic bacteria and their heme acquisition systems have been well described (Cescau *et al*., 2007; Wandersman and Delepelaire, 2004; Wilks and Burkhard, 2007). *M. tuberculosis* may encounter heme or heme-containing protein both intracellularly and extracellularly during infection. Erythrocytes are degraded in the phagosomes of macrophages resulting in heme release (Ganz, 2012; Soe-Lin *et al*., 2009), exposing *M. tuberculosis* residing in the phagosome directly to the heme. Extracellular *M. tuberculosis* may access hemoglobin in the bloodstream where hemolysins produced by *M. tuberculosis* facilitate the release of hemoglobin from the red blood cells (Deshpande *et al*., 1997; Rahman *et al*., 2010). Many studies have demonstrated the ability of *M. tuberculosis* to utilize heme as an iron source in low iron culture media (Jones and Niederweis, 2011; Raghu *et al*., 1993; Tullius *et al*., 2011).

The gene cluster *Rv0201c-Rv0207c* was shown to be responsible for heme acquisition in *M. tuberculosis*. Rv0203 is a secreted hemophore which has similar a heme binding affinity to its orthologues in some gram-negative bacteria (Owens *et al*., 2012; Tullius *et al*., 2011). Rv0202c (MmpL11) and Rv0206c (MmpL3) are transmembrane heme importers (Figure 3.4). Rv0203 is secreted via an unidentified mechanism and acquires heme from hemoglobin, transports it across the outer membrane and delivers to the extracellular domain (E1) of either MmpL11 and MmpL3 (Owens *et al*., 2012). Inside *M. tuberculosis*, the heme can either be degraded to release iron to be incorporated into other metabolic enzymes or utilized directly for the synthesis of heme-containing enzymes. Heme degradation is catalysed by heme oxygenase (HO) in a well characterized process in which heme is broken down to biliverdin, releasing carbon monoxide (CO) and iron ions (Wandersman and Stojiljkovic, 2000). The HO orthologue in *M. tuberculosis*, MhuD (Mycobacterial heme utilization Degrader, Rv3592), however, degrades heme to mycobilin without generating CO. As CO induces *M. tuberculosis* to enter dormant state (Voskuil *et al*., 2003), this particular heme degradation mechanism does not interfere with the CO sensing of the bacteria in the host environment (Nambu *et al*., 2013). As MhuD is able to bind to two heme molecules in its inactive state, it may serve as an additional heme storage molecule (Chim *et al*., 2010).

Although *M. tuberculosis* favours heme as an iron source, it expresses a heme biosynthetic enzyme, ferrochelatase, encoded by *hemZ* (Rv1485). This protein contains a [2Fe-2S] cluster and catalyzes the last step of heme biosynthesis in which ferrous iron is inserted into protoporthyrin IX to form protoheme (Dailey and Dailey, 2002). This gene is essential *in vitro* implying that heme is not only used as an iron source by *M. tuberculosis* but is also an essential cofactor for some metabolic enzymes (Parish *et al*., 2005). For example, the catalase-peroxidase (KatG) and the DosS/DosT redox sensing two component system proteins contain heme in their active sites (Svistunenko, 2005). *M. tuberculosis* and other mycobacteria may utilise iron imported via the siderophore-mediated iron acquisition system to produce heme to synthesize these enzymes.

It is interesting that addition of heme to the low iron media could not rescue the MmpS4/S5 double mutant, in which the mycobactin export was destroyed, unless the siderophore biosynthesis pathway was also disabled (Wells *et al*., 2013). This implies that the intracellular

accumulation of carboxymycobactin/mycobactin exerts a toxic effect on the bacteria perhaps by removing irons from the bacteria's own iron-containing proteins or from imported heme.



(Figure 3.4)

Figure 3.4. Relationships between iron acquisition, heme acquisition, siderophore export and the ESX-3 secretion system. It is currently thought that mycobacteria possess transport systems that may acquire iron from two types of iron source, one from transferrin and the other one from heme. Carboxymycobactin acquires iron from the host transferrin and gets recognized by Msp family porins on the outer membrane and imported to the periplasmic space where they either deliver the iron to the cell envelope associated mycobactin or to the importer protein complex IrtA and IrtB. IrtA contains an intracellularly-located FAD domain which might perform reductase function releasing ferrous iron from carboxymycobactin. The deferri-carboxymycobactin gets exported through MmpL5/4 MmpS5/4 siderophore export complex and acquires more iron if necessary. Mycobacteria release heme-binding protein Rv0203 which acquires heme from hemoproteins in the host and delivers the heme to the heme importers, MmpL11 and MmpL3. Once inside the bacteria, heme either gets degraded to release iron by MhuD or utilized by bacterial heme-binding proteins such as katG, DosT and DosS. Heme can also be synthesized by mycobacteria via the ferrochelatase HemZ, which catalyzes the addition of iron to protophorphyrin IX to form protoheme. The iron and heme acquisition processes are enabled by the ESX-3 components or their substrates. The detailed roles played by ESX-3 on these processes are to be investigated.

3.4 The role of ESX-3 in siderophore and heme transport

M. tuberculosis contains five copies of the Type VII protein secretion system, namely ESX-1- to -5 (Abdallah *et al*., 2007; Gey van Pittius *et al*., 2001). ESX-1 and ESX-5 are involved in virulence in pathogenic mycobacteria (Abdallah *et al*., 2008, p. -5; Banu *et al*., 2002; Hsu *et al*., 2003; Lewis *et al*., 2003; Pym *et al*., 2002b). ESX-3 has shown to be involved in the acquisition of iron from the mycobactin-mediated iron acquisition pathway, but not the exochelin pathway in *M. smegmatis* (Siegrist *et al*., 2009). In *M. tuberculosis*, ESX-3 is essential for *in vitro* growth (Griffin *et al*., 2011; Sassetti *et al*., 2003). The growth defect of the *M. tuberculosis* conditional ESX-3 knock-out strain in 7H9 was only rescued by the addition of extremely high concentrations of iron and zinc (Serafini *et al*., 2009). This implies that alternative iron acquisition systems exist in *M. tuberculosis* and sustain the bacteria with a basal iron supply. The transcriptional profile of the *M. tuberculosis* ESX-3 conditional mutant resembles that of iron starvation even under these high iron conditions (Serafini *et al*., 2013). Repression of ESX-3 expression has a negative effect on iron import, but ESX-3 is not directly responsible for iron uptake. This suggests that ESX-3 may participate in iron uptake by allowing the modification and secretion of unidentified factors which assist in iron acquisition, as suggested by the survival of the ESX-3 conditional mutant in the culture supernatant of wildtype *M. tuberculosis* (Serafini *et al*., 2009). Alternatively ESX-3 may modify the *M. tuberculosis* cell membrane surface composition, thereby facilitating the import of ferri-carboxymycobactin (Serafini *et al*., 2009) (Figure 3.4).

ESX-3 also appears to influence the heme uptake pathway in *M. tuberculosis*. Heme supplementation could not rescue the *M. tuberculosis* ESX-3 conditional mutant under low iron conditions (Serafini *et al*., 2013). As the heme uptake pathway has been well characterized (Tullius *et al*., 2011), ESX-3 is not directly involved in heme uptake, but may play a similar supportive role as observed for mycobactin-mediated iron acquisition where ESX-3 may modify and secrete unknown factors which are essential for iron and heme acquisition in *M. tuberculosis*. Further characterisation of the role of ESX-3 in both iron and heme acquisition is of great significance due to the essential nature of these processes.

3.5 Iron-dependent gene regulation

Mycobacteria utilise elaborate molecular mechanisms to maintain iron homeostasis. Iron acquisition mechanisms are activated in response to low intracellular iron levels to acquire iron from the environment. When there is sufficient intracellular iron, excess iron is stored to prevent it from participating in the Fenton reaction, which would generate harmful hydroxyl radicals (Rodriguez, 2006). Mycobacteria have developed complex gene regulation mechanisms which respond to different iron levels to maintain iron homeostasis.

The iron dependent regulator, IdeR, of *M. tuberculosis* was identified based on homology to the Diphtheria Toxin Repressor (DtxR), an iron-dependent gene expression regulator in *Corynebacterium diphtheria*. IdeR is essential in *M. tuberculosis in vitro* (Pandey and Rodriguez, 2014; Rodriguez *et al*., 2002). Abolishing the iron-responsive regulator leads to attenuation of *M. tuberculosis* in macrophages and the mouse model (Pandey and Rodriguez, 2014). IdeR binding sites have been identified throughout the *M. tuberculosis* H37Rv genome (Gold *et al*., 2001; Rodriguez *et al*., 1999). The expression of 153 *M. tuberculosis* genes is regulated in response to iron concentration and 44 of these genes are regulated directly by IdeR (See examples in Table 3.1 and Table 3.2).

Of the 40 genes that are repressed by IdeR in high iron conditions, 32 encode proteins involved in three iron acquisition related processes: 1. Iron acquisition including the ESX-3 secretion system and the siderophore import transmembrane protein pair IrtA/IrtB; 2. Siderophore biosynthesis; and 3. Siderophore export, MmpL4/S4 (Rodriguez *et al*., 2002) (Table 3.1). The expression of four genes is induced by IdeR under high iron levels, two of which encode the iron storage bacterioferritin proteins (*bfr*A and *bfr*B). The same transcriptional response was observed in *M. tuberculosis* infected macrophages implying that the macrophage presents an iron-deprived condition (Gold *et al*., 2001). Similarly in *M. smegmatis*, under high iron conditions, IdeR negatively regulates the expression of *fxb*A, which encodes a formyltransferase which is essential in the exochelin biosynthetic pathway (Dussurget *et al*., 1999). Mycobacteria therefore respond to high iron levels by inhibiting the expression of the iron acquisition and transport machineries while increasing the production of iron storage proteins. These activities show that mycobacteria maintain iron homeostasis by preventing excessive iron uptake which could cause oxidative damage and storing excess intracellular iron for future use.

Many *M. tuberculosis* genes were found to be regulated by iron independently of IdeR. Expression of 26 genes was repressed, and 23 induced, by a high iron concentration. These genes are involved in a variety of cellular functions, including carbohydrate metabolism, lipid metabolism, energy metabolism, cofactor biosynthesis, aerobic growth, transcriptional regulation, iron transport, and toxin-antitoxin systems (Table 3.2). Of the genes involved in iron acquisition related processes, the expression of *mmpL5* and *mmpS5* is repressed under high iron conditions, consistent with the other components of the siderophore export system, MmpL4 and MmpS4. Interestingly these four genes are not within the same IdeR regulon which may indicate that *M. tuberculosis* also utilises alternative iron regulators.

One substrate of ESX-1, EspD, is also repressed under high iron condition. EspD might play a role in facilitating the secretion of other ESX-1 substrates thereby enhancing the virulence of *M. tuberculosis* (Chen *et al*., 2012), so the iron-deprived environment in macrophages possibly enables the functionality of ESX-1. The gene, *ethA*, encoding a monooxygenase which activates the anti-TB drug ethionamide, was also induced under low-iron condition. The effectiveness of this drug may correlate with different iron concentrations. Sufficient iron in the culture media allows the bacteria to synthesize iron-sulfur cluster-containing metabolic

enzymes including NADH dehydrogenase (Vinogradov, 2001), so the transcription of this gene cluster is induced (n*uoA-D*, *nuoH-I*, and *nuoK-N*).

Table 3.1 Genes that are regulated (repressed/induced) by IdeR under high iron condition, adapted from Rodriguez, *et al*. 2002, quantitative data were shown in the original publication.

Rv No.	Gene Name	Regulation	Gene product	Function
0116c	*ldtA*	Repressed	Probable L,D-transpeptidase LdtA	It catalyzes the formation of 3->3 crosslinks between peptidoglycan subunits in peptidoglycan biosynthesis
0282	*eccA*$_3$		AAA+ ATPase	These genes encode ESX-3, a TypeVII secretion system, which might be involved in mycobactin-mediated iron acquisition (Siegrist *et al*., 2009), where eccB3, eccC3, eccD3 and eccE3 form the membrane-bound secretion machinery (Houben *et al*., 2012); esxG, esxH, PE5 and espG$_3$ might be the secretion substrates for this secretion system. EccA is the ATPase and recruiter for the substrates before secretion (Joshi *et al*., 2012)
0283	*eccB*$_3$		Transmembrane protein (1 TM)	
0284	*eccC*$_3$		FtsK/SpoIIIE-like transmembrane protein (1-3 TMs)	
0285	*pe5*		PE subfamily protein	
0286	*ppe4*		PPE subfamily protein	
0287	*esxG*		esx family protein, CFP-10 like	
0288	*esxH*		esx family protein, ESAT-6 like	
0289	*espG*$_3$		ESX secretion protein 3	
0290	*eccD*$_3$		Transmembrane protein (10-11 TMs)	
0291	*mycP3*		Mycosin-3, subtilisin-like serine protease	
0292	*eccE*$_3$		Transmembrane protein (2 TMs)	
0766c	*cyp123*		Probable cytochrome P450 123, Cyp123	Heme-thiolate monooxygenases. It may oxidize a variety of structurally unrelated compounds, including steroids, fatty acids and xenobiotics.
0450c	*mmpL4*		Probable conserved transmembrane transport protein MmpL4	Together with MmpL5 and MmpS5, the four protein constitute a membrane-bound protein complex which is responsible for siderophore export, maintaining iron homeostasis within the bacteria (Wells *et al*., 2013)
0451c	*mmpS4*		Probable conserved transmembrane transport protein	

			MmpS4	
0587	*yrbE2A*		Conserved hypothetical integral membrane protein	Unknown
Rv No.	Gene Name	Regulation	Gene Product	Function
1343c	*lprD*	Repressed	Probable conserved lipoprotein LprD	Unknown
1344	*mbtL*		Acyl carrier protein MbtL	Involved in the mycobactin biogenesis (Huang *et al*., 2006)
1345	*mbtM (fadD33)*		Probable fatty acyl-AMP ligase MbtM	Involved in the mycobactin biogenesis (LaMarca *et al*., 2004) It activates the acyl chain and transfer it to holo-MbtL to form covalently acylated MbtL (Vergnolle *et al*., 2013)
1346	*mbtN (fadE14)*		Acyl-CoA dehydrogenase	It catalyzes the double bond formation in the fatty –acyl moiety of mycobactin (Krithika *et al*., 2006; Quadri *et al*., 1998)
1347c	*mbtK*		Lysine N-acetyltransferase MbtK	Involved in the mycobactin biogenesis (Card *et al*., 2005) where it catalyzes the transfer of a variety of aliphatic chains from mbtL onto the ε-amino group of lysine residue on mycobactin skeleton (Krithika *et al*., 2006). It prefers long acyl chain-CoA substrates (Frankel and Blanchard, 2008)
1348	*IrtA*		Iron-regulated transporter IrtA	Together, IrtA and IrtB form complex on the plasma membrane, involved in the transport of Fe-carboxymycobactin (Rodriguez and Smith, 2006)
1349	*IrtB*		Iron-regulated transporter IrtB	
1519			Conserved hypothetical protein	Function unknown
2122c	*hisE*		Phosphoribosyl-AMP pyrophosphatase	Involved in histidine biosynthesis
2123	*ppe37*		PPE family protein PPE37	Function unknown
2377c	*mbtH*		Putative conserved protein mbtH	It stimulates amino acid adenylation steps catalysed by the non-ribosomal peptide synthetase (Felnagle *et al*., 2010)
2378c	*mbtG*		Lysine 6-monooxygenase (lysine N6-hydroxylase)	Required for N6-hydroxylation of the two lysine residues during mycobactin assembly, cell membrane associated indicating a connection

				between siderophore biosynthesis and export (Wells *et al*., 2013)
2379c	*mbtF*		Peptide synthetase MbtF	Possibly activate the two lysine residues before incorporated into mycobactin
2380c	*mbtE*		Peptide synthetase	
Rv No.	Gene Name	Regulation	Gene Product	Function
2381c	*mbtD*	Repressed	Polyketide synthetase MbtD	
2382c	*mbtC*		Polyketide synthetase MbtC	
2383c	*mbtB*		Phenyloxazoline synthase MbtB	It catalyzes the amide bond formation between the carboxylic acid of salicylate and the α-amino group of serine in the mycobactin biosynthesis
2384	*mbtA*		Bifunctional enzyme MbtA: salicyl-AMP ligase and salicyl-S-ArCP synthetase	It catalyzes the initiation step of mycobactin synthesis where it activates the salicylic acid as acyladenylate and transfers it to the MbtB ArCP as a thioester.
2385	*mbtJ (lipK)*		Putative acetyl hydrolase MbtJ	Required for N-hydroxylation of the two lysine residues during mycobactin assembly
2386c	*mbtI*		Salicylate synthase MbtI	It catalyzes the initial transformation in mycobactin biosynthesis by converting chorismate to salicylate (Harrison *et al*., 2006)(Zwahlen *et al*., 2007)
3402c			Conserved hypothetical protein	Function unknown
3403c			Hypothetical protein	Function unknown
3839			Conserved hypothetical protein	Function unknown
3840			Possible transcriptional regulatory protein	Involved in transcriptional mechanism
0009	*ppiA*	Induced	Probable iron-regulated peptidyl-prolyl cis-trans isomerase A	It catalyzes the folding of the proteins, in particular cis-trans isomerization of proline imidic peptide bonds in oligopeptides
0338c			Probable iron sulfur binding reductase	Function unknown
1876	*bfrA*		Probable bacterioferritin BfrA	Involved in iron storage
3841	*bftB*		Bacterioferritin BfrB	

Table 3.2. Genes that are regulated (repressed/induced) independent from IdeR under high iron condition, adapted from Rodriguez, *et al*. 2002, quantitative data were shown in the original publication.

Rv No.	Gene Name	Regulation	Gene Product	Function
0279c	*PE_PGRS4*	Repressed	PE_PGRS family protein PE_PGRS4	Unknown
0464c			Conserved hypothetical protein	Unknown
0465c			Probable transcriptional regulatory protein	Possibly involved in transcriptional mechanism
0467	*icl1*		Isocitrate lyase	Involved in glyoxylate bypass, an alternative to the TCA cycle
0676c	*mmpL5*		Probable conserved transmembrane transport protein MmpL5	Together with MmpL4 and MmpS4, the four protein constitute a membrane-bound protein complex which is responsible for siderophore export, maintaining iron homeostasis within the bacteria (Wells *et al*., 2013)
0677c	*mmpS5*		Probable conserved transmembrane transport protein MmpS5	
0692			Conserved hypothetical protein	Unknown
0693	*pqqE*		Probable coenzyme PQQ (pyrroloquinoline quinone) synthesis protein E, PqqE	Required for coenzyme PQQ biosynthesis
0694	*lldD1*		Possible L-lactate dehydrogenase, LldD1	Catalyze conversion of lactate into pyruvate in cellular respiration
1169c	*lipX*		PE family protein, possible lipase, LipX	Unknown
1184c			Possible exported protein	Unknown
1195			PE family protein, PE13	Unknown
1393c			Probable monoxygenase	Unknown
1461			Conserved hypothetical protein	Unknown
1462			Conserved hypothetical protein	Unknown
1463			Probable conserved ATP-binding protein ABC transporter	Involved in active transport across the membrane. Responsible for energy coupling to the transport system

Rv No.	Gene Name	Regulation	Gene Product	Function
1464	*csd*	Repressed	Probable cysteine desulfurase, Csd	Catalyze the removal of elemental sulfur and selenium atoms from L-cysteine, L-cystine, L-selenocysteine, and L-selenocystine to produce L-alanine
1465			Possible nitrogen fixation related protein	Unknown
1466			Conserved hypothetical protein	Unknown
1520			Probable sugar transferase	Unknown
2621c			Possible transcriptional regulatory protein	Involved in transcriptional mechanism
2794c	*pptT*		Phosphopantetheinyl transferase PptT	Involved in biosynthesis of fatty acid and lipids. It transfers the 4'-phosphopantetheine moiety from coenzyme A to a SER of acyl-carrier protein. It catalyzes the formation of holo-ACP, which mediates transfer of acyl fatty-acid intermediates during the biosynthesis of fatty acids and lipids.
3229c	*desA3*		Possible linoleoyl-CoA desaturase	Involved in lipid metabolism
3230c			Hypothetical oxidoreductase	Unknown
3614c	*espD*		ESX-1 secreted protein EspD	Stabilise EspA and EspC for their secretion but itself is not secreted by ESX-1 (Chen *et al*., 2012)
3854c	*ethA*		Monooxygenase, EthA	It activates the pro-drug ethionamide (ETH); it induces ETH sensitivity when overexpressed in *M. tuberculosis*
0706	*rplV*	Induced	50S ribosomal protein L22, RplV	It binds to 23S rRNA, important during early stages of 50S reconsititution
1252c	*lprE*		Probable lipoprotein, LprE	Unknown
1305	*atpE*		Probable ATP synthase C chain	It is one of the three chains of the nonenzymatic component of the ATPase complex
1908c	*katG*		Catalase-peroxidase-peroxynitritase T,	It exhibits both a catalase, a broad-spectrum peroxidase, and

			KatG	a peroxynitritase activities.
Rv No.	Gene Name	Regulation	Gene Product	Function
1943c	mazE5	Induced	Possible antitoxin MazE5	Signal programmed cell death pathway (Engelberg-Kulka *et al*., 2005)
2526	vapB17		Possible antitoxin VapB17	Unknown
2549c	vapC20		Possible toxin, VapC20	Unknown
2550c	vapB20		Possible antitoxin, VapB20	Unknown
2741	PE-PGRS47		PE_PGRS family protein PE_PGRS47	Unknown
2927c			Conserved hypothetical protein	Unknown
3075c			Conserved hypothetical protein	Unknown
3145	nuoA		Probable NADH dehydrogenase I, Chain A	Involved in aerobic/anaerobic respiration, reaction: NADH + ubiquinone = NAD(+) + ubiquinol. Those enzymes contain iron-sulfur clusters (Vinogradov, 2001)
3146	nuoB		Probable NADH dehydrogenase I, Chain B	
3147	nuoC		Probable NADH dehydrogenase I, Chain C	
3148	nuoD		Probable NADH dehydrogenase I, Chain D	
3152	nuoH		Probable NADH dehydrogenase I, Chain H	
3153	nuoI		Probable NADH dehydrogenase I, Chain I	
3155	nuoK		Probable NADH dehydrogenase I, Chain K	
3156	nuoL		Probable NADH dehydrogenase I, Chain L	
3157	nuoK		Probable NADH dehydrogenase I, Chain M	
3158	nuoN		Probable NADH dehydrogenase I, Chain N	
3246	mtrA		Two component	Transcriptional activator part of

			sensory transduction transcriptional regulatory protein MtrA	a two component regulatory system
3394			Conserved hypothetical protein	Unknown

3.6 Iron and heme acquisition-centred drug target discovery

Mycobacteria cannot survive without iron; a malfunctioning iron acquisition pathway causes a growth defect *in vitro* and in macrophages, and attenuates *M. tuberculosis* in mouse model (Rodriguez and Smith, 2006; Serafini *et al*., 2013; Siegrist *et al*., 2009; Wells *et al*., 2013). Intervention strategies targeting siderophore-mediated iron acquisition could be effective because this pathway is not only essential in *M. tuberculosis* but also is absent from the human host. Acyl sulfamoyl adenosine (acyl-AMS), an inhibitor to one mycobactin biosynthesis enzyme, MbtA, has been synthesized and proven to be effective in abolishing the pathway under low iron conditions *in vitro* (Ferreras *et al*., 2005). However, as mycobacteria are also able to acquire iron from heme, an abundant iron source in the host, this may compensate the loss of the siderophore-mediated iron acquisition pathway. This approach may therefore not be successful *in vivo* unless heme acquisition is targeted at the same time. Drugs which target or disable both pathways would be potentially therapeutic. MmpS5/S4 is a promising candidate, as disrupting this complex would not only abolish the siderophore-mediated iron acquisition and recycling of the siderophores, but would also cause intracellular siderophore accumulation, disrupting heme utilization as previously described where heme could not rescue the MmpS5/S4 mutant (Jones *et al*., 2014; Wells *et al*., 2013). MmpL3 is another candidate which is essential for cell viability *in vitro* (Domenech *et al*., 2005). Abolishing MmpL3 may impair trehalose monomycolate export preventing proper assembly of the mycobacterial cell wall (Tahlan *et al*., 2012) and certainly heme uptake (Tullius *et al*., 2011). Two anti-TB drugs, SQ109 (Tahlan *et al*., 2012) and BM212 (La Rosa *et al*., 2012), have shown an inhibitory effect on MmpL3. Drug target exploration and characterization regarding the heme uptake pathway is well described elsewhere (Owens *et al*., 2013). Moreover, ESX-3 also influences both the iron and heme acquisition pathways (Serafini *et al*., 2013). The substrates of ESX-3 may play a variety of roles in fulfilling the metabolic needs of mycobacteria, although these roles remain to be determined. Drugs targeting ESX-3 may have potent therapeutic effects. The iron-dependent regulator, IdeR, although not directly involved in iron or heme uptake, regulates the global iron-responsive metabolism. The *M. tuberculosis* IdeR mutant has a growth defect *in vitro*, *in vivo* and in a mouse model (Pandey and Rodriguez, 2014) making it a promising drug target, although delivery of the drug to the bacterial cytosol may be challenging.

3.7 Conclusion

M. tuberculosis utilises elaborate iron-responsive gene regulation and various iron acquisition strategies to maintain intracellular iron homeostasis and overcome the challenges of accessing scarce iron in their natural and host environments. Numerous studies have demonstrated the essentiality of these mechanisms for the survival and virulence of mycobacteria. Therapeutic strategies targeting those pathways offer promising medical solutions for relieving the global burden of tuberculosis.

3.8 Reference

Abdallah, A.M., Gey van Pittius, N.C., Champion, P.A.D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W., 2007. Type VII secretion--mycobacteria show the way. Nat. Rev. Microbiol. 5, 883–891. doi:10.1038/nrmicro1773

Abdallah, A.M., Savage, N.D.L., van Zon, M., Wilson, L., Vandenbroucke-Grauls, C.M.J.E., van der Wel, N.N., Ottenhoff, T.H.M., Bitter, W., 2008. The ESX-5 secretion system of *Mycobacterium marinum* modulates the macrophage response. J. Immunol. Baltim. Md 1950 181, 7166–7175.

Banerjee, S., Farhana, A., Ehtesham, N.Z., Hasnain, S.E., 2011. Iron acquisition, assimilation and regulation in mycobacteria. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 11, 825–838. doi:10.1016/j.meegid.2011.02.016

Banu, S., Honoré, N., Saint-Joanis, B., Philpott, D., Prévost, M.-C., Cole, S.T., 2002. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol. Microbiol. 44, 9–19.

Barclay, R., Ewing, D.F., Ratledge, C., 1985. Isolation, identification, and structural analysis of the mycobactins of *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium scrofulaceum*, and *Mycobacterium paratuberculosis*. J. Bacteriol. 164, 896–903.

Barclay, R., Ratledge, C., 1988. Mycobactins and exochelins of *Mycobacterium tuberculosis*, *M. bovis*, *M. africanum* and other related species. J. Gen. Microbiol. 134, 771–776.

Butterton, J.R., Calderwood, S.B., 1994. Identification, cloning, and sequencing of a gene required for ferric vibriobactin utilization by *Vibrio cholerae*. J. Bacteriol. 176, 5631–5638.

Cescau, S., Cwerman, H., Létoffé, S., Delepelaire, P., Wandersman, C., Biville, F., 2007. Heme acquisition by hemophores. Biometals Int. J. Role Met. Ions Biol. Biochem. Med. 20, 603–613. doi:10.1007/s10534-006-9050-y

Chen, J.M., Boy-Röttger, S., Dhar, N., Sweeney, N., Buxton, R.S., Pojer, F., Rosenkrands, I., Cole, S.T., 2012. EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis*. J. Bacteriol. 194, 884–893. doi:10.1128/JB.06417-11

Chim, N., Iniguez, A., Nguyen, T.Q., Goulding, C.W., 2010. Unusual diheme conformation of the heme-degrading protein from *Mycobacterium tuberculosis*. J. Mol. Biol. 395, 595–608. doi:10.1016/j.jmb.2009.11.025

Dailey, T.A., Dailey, H.A., 2002. Identification of [2Fe-2S] clusters in microbial ferrochelatases. J. Bacteriol. 184, 2460–2464.

De Voss, J.J., Rutter, K., Schroeder, B.G., Barry, C.E., 3rd, 1999. Iron acquisition and metabolism by mycobacteria. J. Bacteriol. 181, 4443–4451.

Deshpande, R.G., Khan, M.B., Bhat, D.A., Navalkar, R.G., 1997. Isolation of a contact-dependent haemolysin from *Mycobacterium tuberculosis*. J. Med. Microbiol. 46, 233–238.

Domenech, P., Reed, M.B., Barry, C.E., 3rd, 2005. Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect. Immun. 73, 3492–3501. doi:10.1128/IAI.73.6.3492-3501.2005

Dussurget, O., Timm, J., Gomez, M., Gold, B., Yu, S., Sabol, S.Z., Holmes, R.K., Jacobs, W.R., Jr, Smith, I., 1999. Transcriptional control of the iron-responsive *fxbA* gene by the mycobacterial regulator IdeR. J. Bacteriol. 181, 3402–3408.

Farhana, A., Kumar, S., Rathore, S.S., Ghosh, P.C., Ehtesham, N.Z., Tyagi, A.K., Hasnain, S.E., 2008. Mechanistic insights into a novel exporter-importer system of

Mycobacterium tuberculosis unravel its role in trafficking of iron. PloS One 3, e2087. doi:10.1371/journal.pone.0002087

Ferreras, J.A., Ryu, J.-S., Di Lello, F., Tan, D.S., Quadri, L.E.N., 2005. Small-molecule inhibition of siderophore biosynthesis in *Mycobacterium tuberculosis* and *Yersinia pestis*. Nat. Chem. Biol. 1, 29–32. doi:10.1038/nchembio706

Fetherston, J.D., Bertolino, V.J., Perry, R.D., 1999. YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. Mol. Microbiol. 32, 289–299.

Fiss, E.H., Yu, S., Jacobs, W.R., Jr, 1994. Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways. Mol. Microbiol. 14, 557–569.

Frederick, R.E., Mayfield, J.A., DuBois, J.L., 2009. Iron trafficking as an antimicrobial target. Biometals Int. J. Role Met. Ions Biol. Biochem. Med. 22, 583–593. doi:10.1007/s10534-009-9236-1

Ganz, T., 2012. Macrophages and systemic iron homeostasis. J. Innate Immun. 4, 446–453. doi:10.1159/000336423

Gey van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. 2, research0044.1–research0044.18.

Gobin, J., Horwitz, M.A., 1996. Exochelins of *Mycobacterium tuberculosis* remove iron from human iron-binding proteins and donate iron to mycobactins in the *Mycobacterium tuberculosis* cell wall. J. Exp. Med. 183, 1527–1532.

Gobin, J., Moore, C.H., Reeve, J.R., Jr, Wong, D.K., Gibson, B.W., Horwitz, M.A., 1995. Iron acquisition by *Mycobacterium tuberculosis*: isolation and characterization of a family of iron-binding exochelins. Proc. Natl. Acad. Sci. U. S. A. 92, 5189–5193.

Gold, B., Rodriguez, G.M., Marras, S.A., Pentecost, M., Smith, I., 2001. The *Mycobacterium tuberculosis* IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. Mol. Microbiol. 42, 851–865.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 7, e1002251. doi:10.1371/journal.ppat.1002251

Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R.G., Derrick, S.C., Collins, F.M., Morris, S.L., King, C.H., Jacobs, W.R., Jr, 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc. Natl. Acad. Sci. U. S. A. 100, 12420–12425. doi:10.1073/pnas.1635213100

Jones, C.M., Niederweis, M., 2010. Role of porins in iron uptake by *Mycobacterium smegmatis*. J. Bacteriol. 192, 6411–6417. doi:10.1128/JB.00986-10

Jones, C.M., Niederweis, M., 2011. *Mycobacterium tuberculosis* can utilize heme as an iron source. J. Bacteriol. 193, 1767–1770. doi:10.1128/JB.01312-10

Jones, C.M., Wells, R.M., Madduri, A.V.R., Renfrow, M.B., Ratledge, C., Moody, D.B., Niederweis, M., 2014. Self-poisoning of *Mycobacterium tuberculosis* by interrupting siderophore recycling. Proc. Natl. Acad. Sci. U. S. A. 111, 1945–1950. doi:10.1073/pnas.1311402111

Kontoghiorghes, G.J., Weinberg, E.D., 1995. Iron: mammalian defense systems, mechanisms of disease, and chelation therapy approaches. Blood Rev. 9, 33–45.

Krithika, R., Marathe, U., Saxena, P., Ansari, M.Z., Mohanty, D., Gokhale, R.S., 2006. A genetic locus required for iron acquisition in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 103, 2069–2074. doi:10.1073/pnas.0507924103

La Rosa, V., Poce, G., Canseco, J.O., Buroni, S., Pasca, M.R., Biava, M., Raju, R.M., Porretta, G.C., Alfonso, S., Battilocchio, C., Javid, B., Sorrentino, F., Ioerger, T.R., Sacchettini, J.C., Manetti, F., Botta, M., De Logu, A., Rubin, E.J., De Rossi, E., 2012. MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. Antimicrob. Agents Chemother. 56, 324–331. doi:10.1128/AAC.05270-11

Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., Sherman, D.R., 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. J. Infect. Dis. 187, 117–123. doi:10.1086/345862

Madigan, C.A., Cheng, T.-Y., Layre, E., Young, D.C., McConnell, M.J., Debono, C.A., Murry, J.P., Wei, J.-R., Barry, C.E., 3rd, Rodriguez, G.M., Matsunaga, I., Rubin, E.J., Moody, D.B., 2012. Lipidomic discovery of deoxysiderophores reveals a revised mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. 109, 1257–1262. doi:10.1073/pnas.1109958109

McMahon, M.D., Rush, J.S., Thomas, M.G., 2012. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. J. Bacteriol. 194, 2809–2818. doi:10.1128/JB.00088-12

Messenger, A.J., Hall, R.M., Ratledge, C., 1986. Iron uptake processes in *Mycobacterium vaccae* R877R, a mycobacterium lacking mycobactin. J. Gen. Microbiol. 132, 845–852.

Nambu, S., Matsui, T., Goulding, C.W., Takahashi, S., Ikeda-Saito, M., 2013. A new way to degrade heme: the *Mycobacterium tuberculosis* enzyme MhuD catalyzes heme degradation without generating CO. J. Biol. Chem. 288, 10101–10109. doi:10.1074/jbc.M112.448399

Ojha, A., Hatfull, G.F., 2007. The role of iron in *Mycobacterium smegmatis* biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonic growth. Mol. Microbiol. 66, 468–483. doi:10.1111/j.1365-2958.2007.05935.x

Owens, C.P., Chim, N., Goulding, C.W., 2013. Insights on how the *Mycobacterium tuberculosis* heme uptake pathway can be used as a drug target. Future Med. Chem. 5, 1391–1403. doi:10.4155/fmc.13.109

Owens, C.P., Du, J., Dawson, J.H., Goulding, C.W., 2012. Characterization of heme ligation properties of Rv0203, a secreted heme binding protein involved in *Mycobacterium tuberculosis* heme uptake. Biochemistry (Mosc.) 51, 1518–1531. doi:10.1021/bi2018305

Pandey, R., Rodriguez, G.M., 2014. IdeR is required for iron homeostasis and virulence in *Mycobacterium tuberculosis*. Mol. Microbiol. 91, 98–109. doi:10.1111/mmi.12441

Parish, T., Schaeffer, M., Roberts, G., Duncan, K., 2005. HemZ is essential for heme biosynthesis in *Mycobacterium tuberculosis*. Tuberc. Edinb. Scotl. 85, 197–204. doi:10.1016/j.tube.2005.01.002

Pym, A.S., Brodin, P., Brosch, R., Huerre, M., Cole, S.T., 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. Mol. Microbiol. 46, 709–717.

Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., Walsh, C.T., 1998. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem. Biol. 5, 631–645.

Raghu, B., Sarma, G.R., Venkatesan, P., 1993. Effect of iron on the growth and siderophore production of mycobacteria. Biochem. Mol. Biol. Int. 31, 341–348.

Rahman, A., Srivastava, S.S., Sneh, A., Ahmed, N., Krishnasastry, M.V., 2010. Molecular characterization of tlyA gene product, Rv1694 of *Mycobacterium tuberculosis*: a non-conventional hemolysin and a ribosomal RNA methyl transferase. BMC Biochem. 11, 35. doi:10.1186/1471-2091-11-35

Ratledge, C., 2004. Iron, mycobacteria and tuberculosis. Tuberc. Edinb. Scotl. 84, 110–130.

Ratledge, C., Ewing, M., 1996. The occurrence of carboxymycobactin, the siderophore of pathogenic mycobacteria, as a second extracellular siderophore in *Mycobacterium smegmatis*. Microbiology 142, 2207–2212. doi:10.1099/13500872-142-8-2207

Rodriguez, G.M., 2006. Control of iron metabolism in *Mycobacterium tuberculosis*. Trends Microbiol. 14, 320–327. doi:10.1016/j.tim.2006.05.006

Rodriguez, G.M., Gold, B., Gomez, M., Dussurget, O., Smith, I., 1999. Identification and characterization of two divergently transcribed iron regulated genes in *Mycobacterium tuberculosis*. Tuber. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. 79, 287–298.

Rodriguez, G.M., Smith, I., 2006. Identification of an ABC transporter required for iron acquisition and virulence in *Mycobacterium tuberculosis*. J. Bacteriol. 188, 424–430. doi:10.1128/JB.188.2.424-430.2006

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., 2002. ideR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. 70, 3371–3381.

Ryndak, M.B., Wang, S., Smith, I., Rodriguez, G.M., 2010. The *Mycobacterium tuberculosis* high-affinity iron importer, IrtA, contains an FAD-binding domain. J. Bacteriol. 192, 861–869. doi:10.1128/JB.00223-09

Saini, V., Farhana, A., Glasgow, J.N., Steyn, A.J.C., 2012. Iron sulfur cluster proteins and microbial regulation: implications for understanding tuberculosis. Curr. Opin. Chem. Biol. 16, 45–53. doi:10.1016/j.cbpa.2012.03.004

Santhanagopalan, S.M., Rodriguez, G.M., 2012. Examining the role of Rv2895c (ViuB) in iron acquisition in *Mycobacterium tuberculosis*. Tuberc. Edinb. Scotl. 92, 60–62. doi:10.1016/j.tube.2011.09.010

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

Serafini, A., Boldrin, F., Palù, G., Manganelli, R., 2009. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. J. Bacteriol. 191, 6340–6344. doi:10.1128/JB.00756-09

Serafini, A., Pisu, D., Palù, G., Rodriguez, G.M., Manganelli, R., 2013. The ESX-3 Secretion System Is Necessary for Iron and Zinc Homeostasis in *Mycobacterium tuberculosis*. PLoS ONE 8, e78351. doi:10.1371/journal.pone.0078351

Sharman, G.J., Williams, D.H., Ewing, D.F., Ratledge, C., 1995a. Isolation, purification and structure of exochelin MS, the extracellular siderophore from *Mycobacterium smegmatis*. Biochem. J. 305 (Pt 1), 187–196.

Sharman, G.J., Williams, D.H., Ewing, D.F., Ratledge, C., 1995b. Determination of the structure of exochelin MN, the extracellular siderophore from *Mycobacterium neoaurum*. Chem. Biol. 2, 553–561.

Siegrist, M.S., Unnikrishnan, M., McConnell, M.J., Borowsky, M., Cheng, T.-Y., Siddiqi, N., Fortune, S.M., Moody, D.B., Rubin, E.J., 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc. Natl. Acad. Sci. U. S. A. 106, 18792–18797. doi:10.1073/pnas.0900589106

Soe-Lin, S., Apte, S.S., Andriopoulos, B., Jr, Andrews, M.C., Schranzhofer, M., Kahawita, T., Garcia-Santos, D., Ponka, P., 2009. Nramp1 promotes efficient macrophage recycling of iron following erythrophagocytosis in vivo. Proc. Natl. Acad. Sci. U. S. A. 106, 5960–5965. doi:10.1073/pnas.0900808106

Svistunenko, D.A., 2005. Reaction of haem containing proteins and enzymes with hydroperoxides: the radical view. Biochim. Biophys. Acta 1707, 127–155. doi:10.1016/j.bbabio.2005.01.004

Tahlan, K., Wilson, R., Kastrinsky, D.B., Arora, K., Nair, V., Fischer, E., Barnes, S.W., Walker, J.R., Alland, D., Barry, C.E., 3rd, Boshoff, H.I., 2012. SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. 56, 1797–1809. doi:10.1128/AAC.05708-11

Tullius, M.V., Harmston, C.A., Owens, C.P., Chim, N., Morse, R.P., McMath, L.M., Iniguez, A., Kimmey, J.M., Sawaya, M.R., Whitelegge, J.P., Horwitz, M.A., Goulding, C.W., 2011. Discovery and characterization of a unique mycobacterial heme acquisition system. Proc. Natl. Acad. Sci. U. S. A. 108, 5051–5056. doi:10.1073/pnas.1009516108

Vinogradov, A.D., 2001. Respiratory complex I: structure, redox components, and possible mechanisms of energy transduction. Biochem. Biokhimiia 66, 1086–1097.

Voskuil, M.I., Schnappinger, D., Visconti, K.C., Harrell, M.I., Dolganov, G.M., Sherman, D.R., Schoolnik, G.K., 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J. Exp. Med. 198, 705–713. doi:10.1084/jem.20030205

Wandersman, C., Delepelaire, P., 2004. Bacterial iron sources: from siderophores to hemophores. Annu. Rev. Microbiol. 58, 611–647. doi:10.1146/annurev.micro.58.030603.123811

Wandersman, C., Stojiljkovic, I., 2000. Bacterial heme sources: the role of heme, hemoprotein receptors and hemophores. Curr. Opin. Microbiol. 3, 215–220.

Weinberg, E.D., 1984. Iron withholding: a defense against infection and neoplasia. Physiol. Rev. 64, 65–102.

Wells, R.M., Jones, C.M., Xi, Z., Speer, A., Danilchanka, O., Doornbos, K.S., Sun, P., Wu, F., Tian, C., Niederweis, M., 2013. Discovery of a siderophore export system essential for virulence of *Mycobacterium tuberculosis*. PLoS Pathog. 9, e1003120. doi:10.1371/journal.ppat.1003120

Wilks, A., Burkhard, K.A., 2007. Heme and virulence: how bacterial pathogens regulate, transport and utilize heme. Nat. Prod. Rep. 24, 511–522. doi:10.1039/b604193k

Wong, D.K., Gobin, J., Horwitz, M.A., Gibson, B.W., 1996. Characterization of exochelins of *Mycobacterium avium*: evidence for saturated and unsaturated and for acid and ester forms. J. Bacteriol. 178, 6394–6398.

Yu, S., Fiss, E., Jacobs, W.R., Jr, 1998. Analysis of the exochelin locus in *Mycobacterium smegmatis*: biosynthesis genes have homology with genes of the peptide synthetase family. J. Bacteriol. 180, 4676–4685.

Zhu, W., Arceneaux, J.E., Beggs, M.L., Byers, B.R., Eisenach, K.D., Lundrigan, M.D., 1998. Exochelin genes in *Mycobacterium smegmatis*: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. Mol. Microbiol. 29, 629–639.

Chapter 4. Assessing the Progress of *Mycobacterium tuberculosis* H37Rv Structural Genomics

4.1. Introduction

With one third of the world population affected, tuberculosis (TB) remains one of the most wide-spread infectious diseases (*Guideline*, 2013). Its etiological agent, *Mycobacterium tuberculosis* (*M. tuberculosis*), is estimated to have contributed to ~1.3 million deaths in 2012 *inter alia* (WHO, Global tuberculosis report 2013) through the emergence of multidrug (MDR) and extensive drug resistance (XDR) *M. tuberculosis* strains, as well as the complex co-infection with human immune-deficient virus (HIV) (WHO, Global tuberculosis report 2013). The importance of identifying new drug targets and developing new drugs can thus hardly be overstated. Structural information on *M. tuberculosis* proteins can help to characterize known drug targets including *inter alia* arabinosyltransferase and DNA gyrase, respective targets for ethambutol (Alderwick *et al*., 2011) and moxifloxacin (Tretter *et al*., 2010), and to identify novel molecular intervention strategies (Arora and Banerjee, 2012; Chim *et al*., 2012, 2009; Xiong *et al*., 2013). Understanding the structural effects of resistance mutations can further support the identification of compensating modifications to drugs to stay ahead in the evolutionary race. A comprehensive structural database of *M. tuberculosis* proteins is a treasure trove of potential drug targets. This review aims to summarize the current status on structural data for *M. tuberculosis* strain H37Rv indicating the progress of *M. tuberculosis* structural genomics, providing clues on producing "difficult-to-express" *M. tuberculosis* proteins, encouraging international research group collaboration and the exploration on the structures of proteins unique to mycobacteria.

4.2. Current Status of Structural Data for *Mycobacterium tuberculosis* H37Rv

The first *M. tuberculosis* protein structure was deposited with the Protein Data Bank (PDB) in 1994 (Wang *et al*., 1994). Since then much effort has been invested in structurally elucidating mycobacterial proteins (Arora *et al*., 2011; Ehebauer and Wilmanns, 2011). By the end of June of 2014, the PDB contained 1612 mycobacterial protein entries including 1045 without specified strain, 542 for H37Rv, 20 for H37Ra, four for CDC1551 and one for F11.

The 542 PDB entries assigned to *M. tuberculosis* H37Rv describe 199 distinct proteins and are associated with 245 publications. However, this number dramatically underestimates the true availability of structural data for *M. tuberculosis* H37Rv both because many *M. tuberculosis* structures in the PDB (potentially including many H37Rv proteins) are not explicitly assigned to any strains and because a close evolutionary relationship of H37Rv to other *M. tuberculosis* strains or a more distant kinship to other bacteria means that structural data for H37Rv proteins can be directly inferred. Mapping all 4031 *M. tuberculosis* open reading frame (ORF) sequences from Tuberculist (March 2013 update) (Lew *et al*., 2011) to all PDB entries using BLAST (Altschul *et al*., 1990) identifies excellent structural matches for 324 H37Rv proteins (Table S4.1) using a cut-off of 95 in the "structural match score", calculated as the product of sequence identity (%) for the best match and the sequence coverage (%) divided by 100. If the cut-off is lowered to 70 (Figure 4.1), 402 or ~10% of all *M. tuberculosis* ORFs may be described as structurally "well defined". A further 287 ORFs (7.1%) with match values between 30 and 70 provide intermediate to "solid structural data". For this group, the overall fold or a partial model of good quality can mostly be derived. Overall, good structural data is thus available for ~689 *M. tuberculosis* H37Rv ORFs (17.1%). For scores below 30, structural data is limited or non-existent implying that structures need to be solved experimentally or that more sophisticated modelling tools may be required.

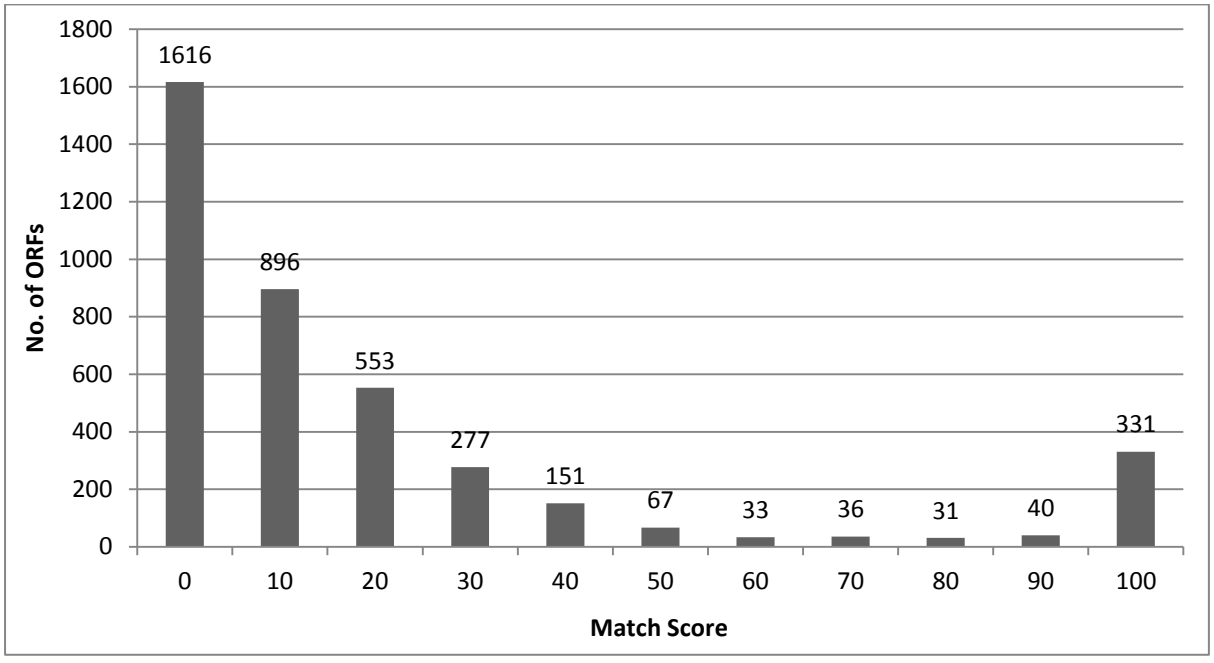


Figure 4.1. Distribution of 4031 *M. tuberculosis* H37Rv ORFs by structural match score (sequence identity (%) x coverage (%) / 100) for the best match in the PDB. Most *M. tuberculosis* H37Rv ORFs have no or a low scoring match (≤ 30%)

To avoid ambiguities, the following more detailed analyses will concentrate on those PDB entries explicitly assigned to the *M. tuberculosis* strain H37Rv. This decision is made based on the fact that this laboratory reference strain has been studied and annotated in most detail,

that set of proteins is broadly representative of all H37Rv entries, and because the PDB provides a convenient platform for analysis.

Structural characterization of *M. tuberculosis* H37Rv proteins mostly involve the elucidation of functional mechanisms *inter alia* by comparing apo-protein structures with complexes involving ions, inhibitors, coenzymes or protein partners bound to the protein of interest. This frequently results in multiple crystal structures for individual proteins and a range of separate publications. While publication output for entries assigned to H37Rv initially increased rapidly from 1998 to 2006, it clearly plateaued thereafter (Figure 4.2). Apart from characterizing new drug targets, the structural elucidation has also helped to correct mis-annotation of potential protein functions (Brown *et al*., 2007; Layre *et al*., 2014; Sinha *et al*., 2005), to assign functions to previously uncharacterized or "hypothetical" proteins (17–19), and to expand data on *M. tuberculosis* physiology and pathogenicity. Both X-ray crystallography (524 structures, or 97%) and Nuclear Magnetic Resonance Spectroscopy (18 structures, or 3%) were used in solving this set of protein structures.

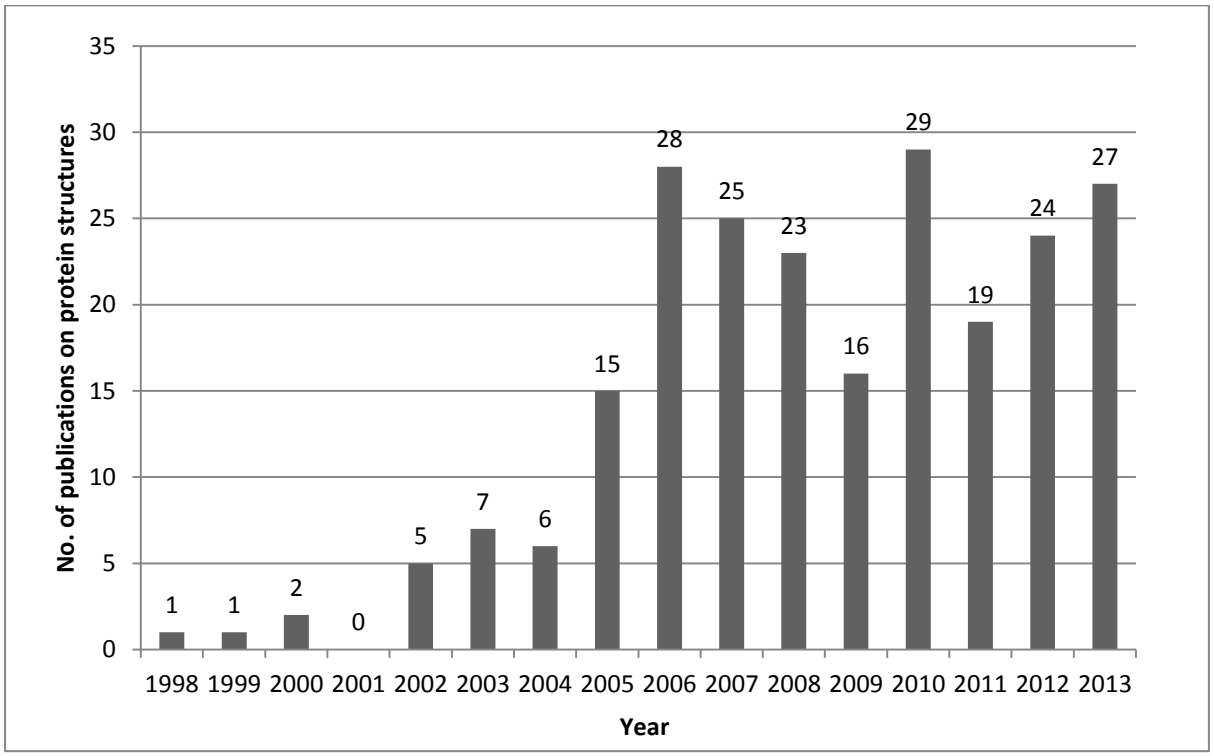


Figure 4.2. Number of publications on *M. tuberculosis* H37Rv protein structures by year.

4.3. Functional Categorization of *M. tuberculosis* H37Rv Crystal Structures

Ideal anti-TB drug targets either involve enzymes catalysing essential biological reactions or signal transduction components mediating nutrient acquisition or metabolic regulation. Targets should further be unrelated or should share low homology to human orthologs to limit drug toxicity in patients. The 199 structurally characterized proteins from *M. tuberculosis* H37Rv were mostly selected according to these criteria and can be assigned to 5 functional groups: 126 (64%) to "house-keeping metabolism" and "resistance/survival mechanism", 31 (15%) to "genetic information processing", 14 (7%) to "cellular information processing and substrate transport", 11 (6%) "virulence factors", while 17 (8%) are of unknown function mostly annotated as "hypothetical". Note that house-keeping metabolism processes dominate amongst studied proteins while only a few virulence associated proteins have been characterized. Structures of previously uncharacterized "hypothetical" proteins were overwhelmingly elucidated by structural genomics consortia. With some exceptions, the structurally solved proteins in H37Rv are mostly similar in function to their recognized orthologs in other organisms and share structural features with other members in the same protein family.

4.3.1 House-keeping metabolism and resistance/survival mechanisms

Metabolic and environmental stress associated enzymes constitute 64% of available *M. tuberculosis* H37Rv protein structures reflecting the ability of this bacterium to synthesize amino acids, vitamins and cofactors, as well as cell wall components such as peptidoglycan, sugars and mycolic acids. As humans do not code for the latter proteins, these are particularly attractive as drug targets. Enzymes of carbohydrate metabolism, redox homeostasis, nutrient acquisition, oxidative stress relief, antibiotic and harmful chemical elimination share only minimal homology to their human counterparts again making for good potential drug targets. The 126 proteins in this group include nine in carbohydrate metabolism (Table S4.2), 28 in amino acid metabolism (Table S4.3), 39 in lipid and polyketide metabolism (Table S4.4), 15 in nucleotide metabolism (Table S4.5), eight in glycan biosynthesis (Table S4.6), nine in cofactor and vitamin metabolism (Table S4.7), and 18 in self-defense and detoxification, (Table S4.8).

4.3.2 Genetic information processing

Genetic integrity is vital for survival. Gene transcription and protein production are hence critical metabolic processes. Targeting components in these pathways will kill or limit the growth of *M. tuberculosis*. Protein structures assigned to this category include 17 proteins in transcriptional regulation (Table S4.9), five in translation/protein folding (Table S4.10) and nine in DNA replication/repair (Table S4.11).

4.3.3 Cellular information processing and substrate transport

Cellular signalling pathways allow bacteria to sense and respond to environmental stimuli such chemical gradients including PO_4^{3-}, O_2, NO, Fe^{2+}, N_2, cAMP. Proteins in this group mostly belong to two-component systems and serine/threonine kinases (Pkns) though their exact function often remains unclear. Fourteen protein structures fall into this category (Table S12).

4.3.4 Virulence factors and hypothetical proteins

Despite the molecular basis for *M. tuberculosis* pathogenicity being of critical importance, knowledge in this area remains quite limited. *M. tuberculosis* virulence factors appear to mostly be directed towards ensuring mycobacterial survival in human macrophages. Only 11 of 195 or 6% of *M. tuberculosis* H37Rv structurally elucidated proteins belong to this group (Table S13). Identifying human targets of virulence factors as well as crystallizing the associated molecular complexes remains challenging with the result that data in this regard is currently not available.

Structural analysis of 17 proteins of unknown function (Table S14) has led to novel fold discovery and the assignment of possible functions to some proteins. Thus, Rv2714 was found to be a representative of a new nucleoside phosphorylase-like family of actinobacterial proteins (20); Rv1980c represents a new family of secreted β-grasp proteins (21); Rv2827c has a unique DNA-binding fold although the exact function remains unknown (22); Rv1155 resembles flavin mononucleotide (FMN)-binding proteins but has no affinity towards FMN (23). Rv0802, resembling members of the GCN5-related N-acetyltransferase family, utilizes succinyl-CoA as substrate rather than acetyl-CoA, the first in this family (24). In contrast to the above, no function could be identified for many structurally-elucidated hypothetical proteins. Examples include Rv0020c, a substrate of Ser/Thr protein kinase PknB (25), and Rv2140c with marked homology to phosphatidylethanolamine-binding proteins indicating a potential role in lipid metabolism (26) (Table S13). These highlight the important contribution structural biology can make to our understanding of mycobacterial biology.

4.4. Technical Aspects of *M. tuberculosis* Structural Analysis

Structural elucidation of *M. tuberculosis* proteins remains challenging mainly due to difficulties with gene expression, protein solubility, protein purification and crystallization. Unexpectedly, codon-optimization of *M. tuberculosis* genes has not been frequently applied. With costs for DNA synthesis steadily decreasing this modification may become more

accessible in future. Overall, affinity tag fusion proteins remain popular: 168 of 199 proteins or 84% were produced with a tag, though tagless proteins also feature (31 or 16%). N-terminal His_6-tags dominate (56%), followed by C-terminal His_6 (23%). Larger tags such as glutathione S-transferase (GST, 26 kDa), maltose-binding protein (MBP, 42.5 kDa) and intein/chitin-binding protein (CBP, 55 kDa) which may help to increase solubility were used less frequently. The GST-tag was correspondingly used for six proteins (3%), MBP for three (1%), while CBP and streptavidin tag were used once each (0.5%).

Proteins used to derive the deposited *M. tuberculosis* H37Rv structures were with very few exceptions all heterologously produced in *Escherichia coli* (525 of 542 PDB entries or 196 of 199 unique proteins, See Table 4.1). Established protocols and the superior efficiency of this host compared to *M. tuberculosis* account for this landslide. The *M. tuberculosis* genome is more G/C rich (>65%) and its codon usage differs from that of *E. coli*. It incorporates a higher proportion of glycine, alanine, proline and arginine and *M. tuberculosis* post-translational modification machinery is lacking in *E. coli* (27,28). *M. smegmatis*, a fast-growing saprophytic relative of *M. tuberculosis*, is a potential expression host for *M. tuberculosis* genes overcoming many of the restrictions of *E. coli* (28–31). Unexpectedly, however, only three proteins of H37Rv represented by 16 structures were produced in *M. smegmatis* (29,32,33). Clearly the faster growth and production rates of *E. coli* ensure that it remains the most popular expression host. Also strikingly, BL21 (DE3) was used most frequently to express *M. tuberculosis* genes despite the presence of rare codons. Only a small proportion of *M. tuberculosis* proteins were produced in Rosetta or CodonPlus strains that code for tRNAs of these rare codons. Again the observed bias towards BL21 (DE3) may largely be practical with researchers simply using the most widely available strain. In future, other expression strains may come to the help for non-standard proteins such as membrane proteins.

Table 4.1. Expression host for H37Rv protein production

Expression Host	Number
E. coli BL21(DE3) and its derivative strains	157
E. coli Rosetta or BL21 CodonPlus strain	21
E. coli B834	7
E. coli C41(DE3)	4
E. coli DH5α	3
E. coli Origami	2
Pseudomonas putida KT4224	2
M. smegmatis mc^2 4517	2
M. smegmatis mc^2 155	1

Production conditions were optimized for most proteins. Overwhelmingly *lac* promoter constructs (T7 and *tac*) were used for isopropyl β-D-1-thiogalactopyranoside (IPTG) induction in Lysogeny Broth (179 of 199 proteins). Only for three proteins was Terrific Broth (TB) used, auto-inducing media for 11, and arabinose induction for six proteins. Plotting IPTG concentration against production temperature (Figure 4.3) indicates that IPTG concentration between 0.1 and 0.5 mM and production temperatures between 16 and 25ºC were most common, with additional spikes at 37°C and 30°C. IPTG concentrations between 0.01 and 0.1 mM are known to minimize protein insolubility, growth inhibition and cell lysis (34). Similarly low temperatures (16-25ºC) limit heat shock protein production, proteolytic degradation, protein aggregation (inclusion bodies) and improve protein stability (34). Though only rarely reported, *E. coli* strains such as Arctic Express and expression vectors such as pCOLD, co-produce cold shock chaperone proteins which help to overcome production hurdles for *M. tuberculosis* proteins. Arylamine N-acetyltransferase (TBNAT), for example, was produced insolubly in *E. coli* BL21 (DE3) pLysS but solubly in ArcticExpress (27).

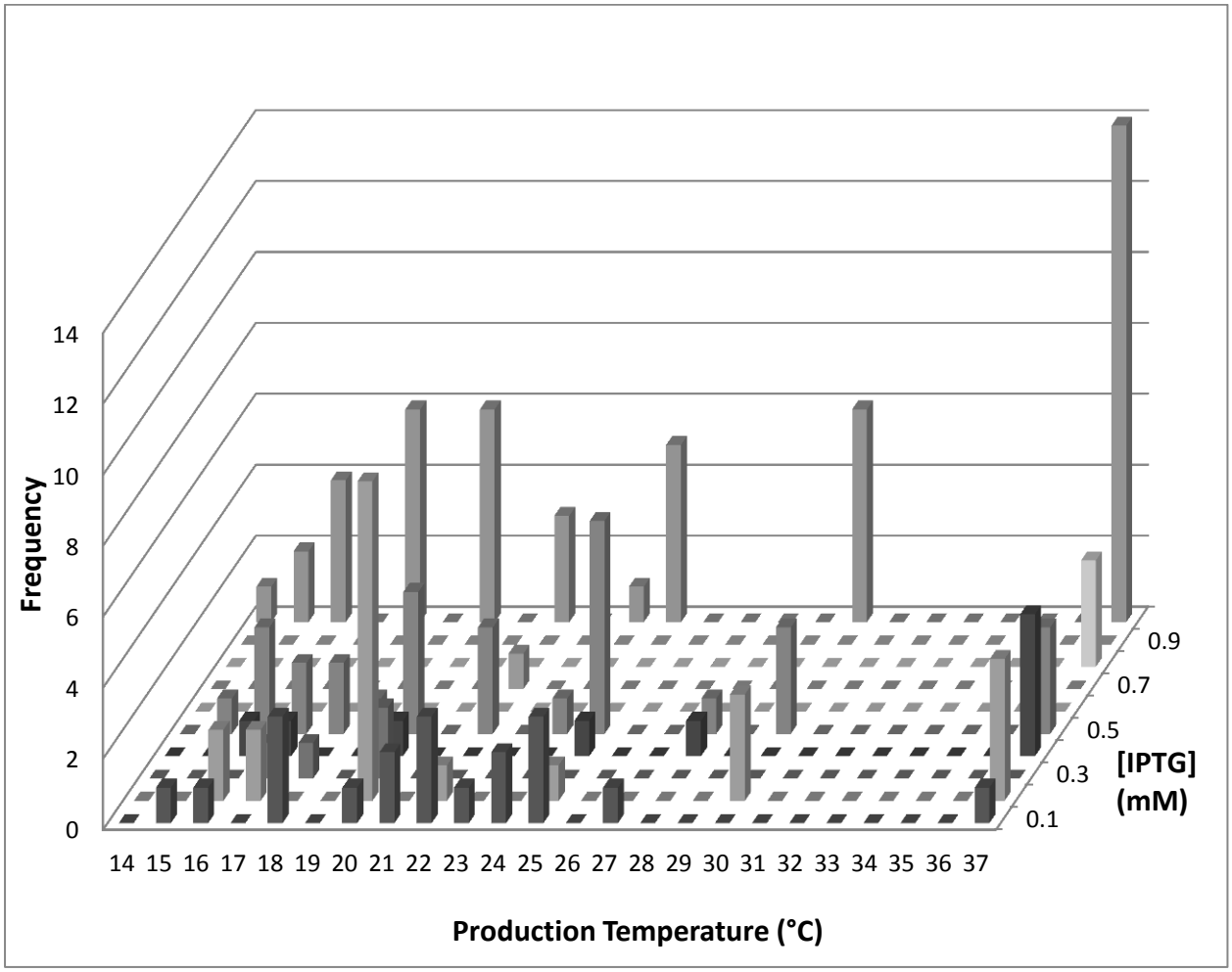


Figure 4.3. *M. tuberculosis* H37Rv strain protein production in *E. coli*; expression temperature is plotted against IPTG concentration. Most proteins were expressed at moderately low temperatures (16-25ºC) and IPTG concentrations (0.1-0.5 mM).

Protein constructs resulting in successful structures are mostly small to medium in size (10-60 kDa) with neutral pI (Figure 4.4). A majority of proteins are further inherently soluble and stable enzymes or signal transduction components. These features clearly enhance successful production in *E. coli* but indicate that future progress will presumably be much slower. Membrane protein structures are correspondingly rare with structural studies mostly being limited to soluble domain such as in MmpS4 (35) MmpL3 and MmpL11 (36).

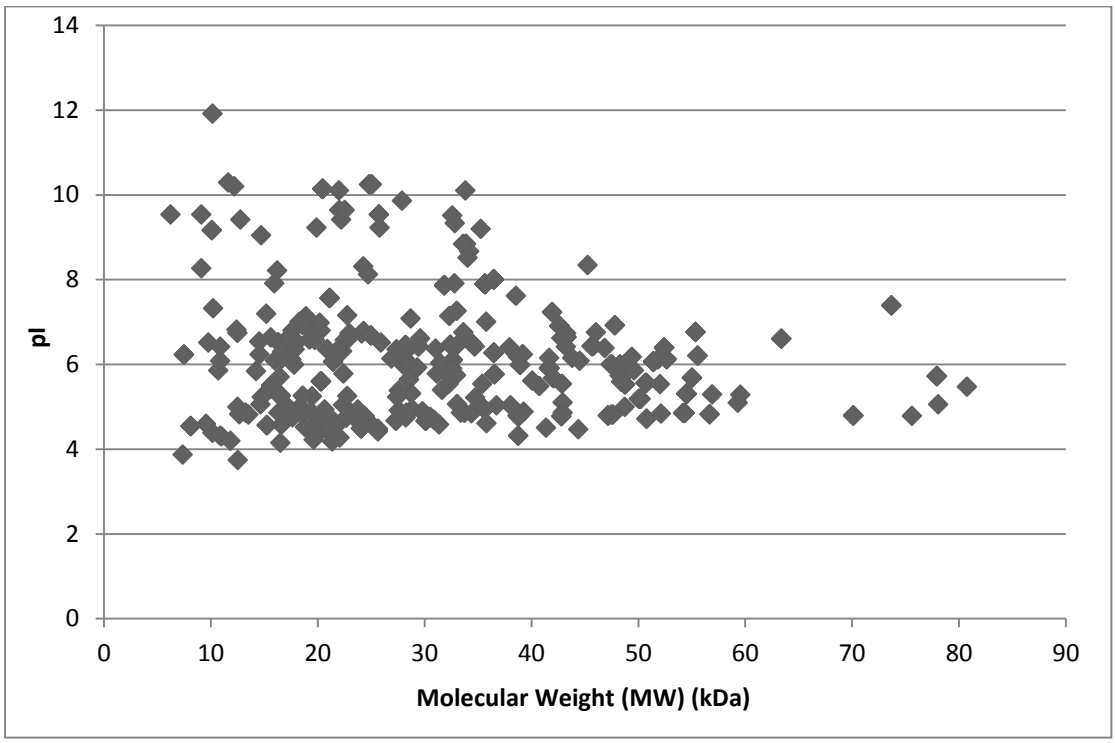


Figure 4.4. A plot of pI versus molecular weight of all *M. tuberculosis* H37Rv protein structures in the PDB. A clustering around a MW of 10-60 kDa and a pI of 4-7 is apparent.

Insolubility of target *M. tuberculosis* proteins remains a major hurdle for structural analysis. The solubility of secreted target proteins may be improved by co-expressing the soluble domain of the membrane-anchored disulphide bond-forming (Dsb) protein (Rv2969c) (37,38). Similarly, GlgB (Rv1326c), Cyp125A1 (Rv3545c), and tuberculosinyl transferase (Rv3378c) were converted to soluble proteins by co-producing *E. coli* chaperone proteins (15,39,40). Corresponding *M. tuberculosis* chaperones such as Cpn60.1, an ortholog of *E. coli* GroEL, significantly improved the solubility of DprE1, the target for the drug benzothiazinone (41). Protein solubility may also be optimized by co-expressing other proteins of the same metabolic pathway. Siroheme-dependent sulfite reductase (Rv2391) was hence solubly produced by co-producing uroporphyrinogen III C-methyl-transferase (Rv2847c) (42). Similarly, non-ribosomal peptide synthetases, MbtE and MbtF of mycobactin biosynthesis, were solubly produced only by co-producing 8 kDa MbtH with a previously unknown role in the pathway (43,44). Yet another strategy of obtaining structural data for difficult or insoluble *M. tuberculosis* proteins is to replace them with mycobacterial orthologues. Thus, Mycosin-1

and Mycosin-3 proteases from ESX-1 and ESX-3 Type VII secretion systems were structurally characterized from *M. smegmatis* in place of their *M. tuberculosis* counterparts. Amino acid sequence identities of 72% and 60%, respectively ensure that structural features may directly be inferred for the latter (45–47). Underlying causes for difficulties in producing *M. tuberculosis* proteins mostly remain unresolved as projects are generally discontinued if problems prove intractable.

4.5. Scope of future structural genomics studies

In choosing targets for structural analysis, proteins unrelated to existing PDB entries may be of most interest as they may represent new protein families or new functions. Rv2179c, for example, discovered to have RNase activity but without structural similarity to other RNases, was identified as the founding member of a large family of bacterial RNases (17). Many proteins without structural counterparts in the PDB, however, are membrane proteins limiting rapid progress in structural analysis. Furthermore some *M. tuberculosis H37Rv* ORFs code for highly unusual proteins. Rv3512 for example contains 598 glycines (55%) and 140 alanines (13%) within 1079 amino acids, identifying this protein as potentially inherently unfolded.

4.6. Structural Biology Research Group World Wide

Table S4.15 lists the principle investigators, the institutions, the countries and the category of proteins analysed. Overall, 113 principle investigators contributed to the 245 publications for *M. tuberculosis* H37Rv protein structures in the PDB. Of these, 58 involved with more than one publication are listed. Groups are mostly based in Europe and Northern America but complemented by groups from India, New Zealand, China and South Korea.

4.7. Reference

Abendroth, J., Ollodart, A., Andrews, E.S.V., Myler, P.J., Staker, B.L., Edwards, T.E., Arcus, V.L., Grundner, C., 2013. Mycobacterium tuberculosis Rv2179c establishes a new exoribonuclease family with broad phylogenetic distribution. J. Biol. Chem. doi:10.1074/jbc.M113.525683

Abuhammad, A., Lack, N., Schweichler, J., Staunton, D., Sim, R.B., Sim, E., 2011. Improvement of the expression and purification of Mycobacterium tuberculosis arylamine N-acetyltransferase (TBNAT) a potential target for novel anti-tubercular agents. Protein Expr. Purif. 80, 246–252. doi:10.1016/j.pep.2011.06.021

Alderwick, L.J., Lloyd, G.S., Ghadbane, H., May, J.W., Bhatt, A., Eggeling, L., Fütterer, K., Besra, G.S., 2011. The C-terminal domain of the Arabinosyltransferase Mycobacterium tuberculosis EmbC is a lectin-like carbohydrate binding module. PLoS Pathog. 7, e1001299. doi:10.1371/journal.ppat.1001299

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Arora, A., Chandra, N.R., Das, A., Gopal, B., Mande, S.C., Prakash, B., Ramachandran, R., Sankaranarayanan, R., Sekar, K., Suguna, K., Tyagi, A.K., Vijayan, M., 2011. Structural biology of Mycobacterium tuberculosis proteins: the Indian efforts. Tuberc. Edinb. Scotl. 91, 456–468. doi:10.1016/j.tube.2011.03.004

Arora, N., Banerjee, A.K., 2012. Targeting tuberculosis: a glimpse of promising drug targets. Mini Rev. Med. Chem. 12, 187–201.

Bashiri, G., Squire, C.J., Baker, E.N., Moreland, N.J., 2007. Expression, purification and crystallization of native and selenomethionine labeled Mycobacterium tuberculosis FGD1 (Rv0407) using a Mycobacterium smegmatis expression system. Protein Expr. Purif. 54, 38–44. doi:10.1016/j.pep.2007.01.014

Batt, S.M., Jabeen, T., Bhowruth, V., Quill, L., Lund, P.A., Eggeling, L., Alderwick, L.J., Fütterer, K., Besra, G.S., 2012. Structural basis of inhibition of Mycobacterium tuberculosis DprE1 by benzothiazinone inhibitors. Proc. Natl. Acad. Sci. U. S. A. 109, 11354–11359. doi:10.1073/pnas.1205735109

Brown, A.K., Meng, G., Ghadbane, H., Scott, D.J., Dover, L.G., Nigou, J., Besra, G.S., Fütterer, K., 2007. Dimerization of inositol monophosphatase Mycobacterium tuberculosis SuhB is not constitutive, but induced by binding of the activator Mg2+. BMC Struct. Biol. 7, 55. doi:10.1186/1472-6807-7-55

Canaan, S., Sulzenbacher, G., Roig-Zamboni, V., Scappuccini-Calvo, L., Frassinetti, F., Maurin, D., Cambillau, C., Bourne, Y., 2005. Crystal structure of the conserved hypothetical protein Rv1155 from Mycobacterium tuberculosis. FEBS Lett. 579, 215–221. doi:10.1016/j.febslet.2004.11.069

Chim, N., Harmston, C.A., Guzman, D.J., Goulding, C.W., 2013. Structural and biochemical characterization of the essential DsbA-like disulfide bond forming protein from Mycobacterium tuberculosis. BMC Struct. Biol. 13, 23. doi:10.1186/1472-6807-13-23

Chim, N., McMath, L.M., Beeby, M., Goulding, C.W., 2009. Advances in Mycobacterium tuberculosis structural genomics: investigating potential chinks in the armor of a deadly pathogen. Infect. Disord. Drug Targets 9, 475–492.

Chim, N., Owens, C.P., Contreras, H., Goulding, C.W., 2012. Advances In Mycobacterium Tuberculosis Therapeutics Discovery Utlizing Structural Biology. Infect. Disord. Drug Targets.

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R., Kaufmann, S.H.E., 2003. The RD1 proteins of Mycobacterium tuberculosis: expression in

Mycobacterium smegmatis and biochemical characterization. Microbes Infect. Inst. Pasteur 5, 1082–1095.

Donovan, R.S., Robinson, C.W., Glick, B.R., 1996. Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. J. Ind. Microbiol. 16, 145–154.

Ehebauer, M.T., Wilmanns, M., 2011. The progress made in determining the Mycobacterium tuberculosis structural proteome. Proteomics 11, 3128–3133. doi:10.1002/pmic.201000787

Eulenburg, G., Higman, V.A., Diehl, A., Wilmanns, M., Holton, S.J., 2013. Structural and biochemical characterization of Rv2140c, a phosphatidylethanolamine-binding protein from Mycobacterium tuberculosis. FEBS Lett. 587, 2936–2942. doi:10.1016/j.febslet.2013.07.038

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., Thomas, M.G., 2010. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. Biochemistry (Mosc.) 49, 8815–8817. doi:10.1021/bi1012854

Garg, S.K., Alam, M.S., Kishan, K.V.R., Agrawal, P., 2007. Expression and characterization of alpha-(1,4)-glucan branching enzyme Rv1326c of Mycobacterium tuberculosis H37Rv. Protein Expr. Purif. 51, 198–208. doi:10.1016/j.pep.2006.08.005

Goldstone, R.M., Moreland, N.J., Bashiri, G., Baker, E.N., Shaun Lott, J., 2008. A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. Protein Expr. Purif. 57, 81–87. doi:10.1016/j.pep.2007.08.015

Graña, M., Bellinzoni, M., Miras, I., Fiez-Vandal, C., Haouz, A., Shepard, W., Buschiazzo, A., Alzari, P.M., 2009. Structure of Mycobacterium tuberculosis Rv2714, a representative of a duplicated gene family in Actinobacteria. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 65, 972–977. doi:10.1107/S1744309109035027

Guideline: Nutritional Care and Support for Patients with Tuberculosis, 2013. , WHO Guidelines Approved by the Guidelines Review Committee. World Health Organization, Geneva.

Hegde, S.S., Vetting, M.W., Roderick, S.L., Mitchenall, L.A., Maxwell, A., Takiff, H.E., Blanchard, J.S., 2005. A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. Science 308, 1480–1483. doi:10.1126/science.1110699

Janowski, R., Panjikar, S., Eddine, A.N., Kaufmann, S.H.E., Weiss, M.S., 2009. Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from Mycobacterium tuberculosis. J. Struct. Funct. Genomics 10, 137–150. doi:10.1007/s10969-009-9060-4

Lagautriere, T., Bashiri, G., Paterson, N.G., Berney, M., Cook, G.M., Baker, E.N., 2014. Characterization of the proline-utilization pathway in Mycobacterium tuberculosis through structural and functional studies. Acta Crystallogr. D Biol. Crystallogr. 70, 968–980. doi:10.1107/S1399004713034391

Layre, E., Lee, H.J., Young, D.C., Jezek Martinot, A., Buter, J., Minnaard, A.J., Annand, J.W., Fortune, S.M., Snider, B.B., Matsunaga, I., Rubin, E.J., Alber, T., Moody, D.B., 2014. Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. Proc. Natl. Acad. Sci. U. S. A. 111, 2978–2983. doi:10.1073/pnas.1315883111

Lew, J.M., Kapopoulou, A., Jones, L.M., Cole, S.T., 2011. TubercuList--10 years after. Tuberc. Edinb. Scotl. 91, 1–7. doi:10.1016/j.tube.2010.09.008

McMahon, M.D., Rush, J.S., Thomas, M.G., 2012. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in Mycobacterium tuberculosis. J. Bacteriol. 194, 2809–2818. doi:10.1128/JB.00088-12

Noens, E.E., Williams, C., Anandhakrishnan, M., Poulsen, C., Ehebauer, M.T., Wilmanns, M., 2011. Improved mycobacterial protein production using a Mycobacterium smegmatis groEL1ΔC expression strain. BMC Biotechnol. 11, 27. doi:10.1186/1472-6750-11-27

Ohol, Y.M., Goetz, D.H., Chan, K., Shiloh, M.U., Craik, C.S., Cox, J.S., 2010. Mycobacterium tuberculosis MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell Host Microbe 7, 210–220. doi:10.1016/j.chom.2010.02.006

Ouellet, H., Guan, S., Johnston, J.B., Chow, E.D., Kells, P.M., Burlingame, A.L., Cox, J.S., Podust, L.M., de Montellano, P.R.O., 2010. Mycobacterium tuberculosis CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. Mol. Microbiol. 77, 730–742. doi:10.1111/j.1365-2958.2010.07243.x

Prigozhin, D.M., Mavrici, D., Huizar, J.P., Vansell, H.J., Alber, T., 2013. Structural and biochemical analyses of Mycobacterium tuberculosis N-acetylmuramyl-L-alanine amidase Rv3717 point to a role in peptidoglycan fragment recycling. J. Biol. Chem. 288, 31549–31555. doi:10.1074/jbc.M113.510792

Roumestand, C., Leiba, J., Galophe, N., Margeat, E., Padilla, A., Bessin, Y., Barthe, P., Molle, V., Cohen-Gonsaud, M., 2011. Structural insight into the Mycobacterium tuberculosis Rv0020c protein and its interaction with the PknB kinase. Struct. Lond. Engl. 1993 19, 1525–1534. doi:10.1016/j.str.2011.07.011

Schiebel, J., Kapilashrami, K., Fekete, A., Bommineni, G.R., Schaefer, C.M., Mueller, M.J., Tonge, P.J., Kisker, C., 2013. Structural Basis for the Recognition of Mycolic Acid Precursors by KasA, a Condensing Enzyme and Drug Target from Mycobacterium Tuberculosis. J. Biol. Chem. 288, 34190–34204. doi:10.1074/jbc.M113.511436

Schnell, R., Sandalova, T., Hellman, U., Lindqvist, Y., Schneider, G., 2005. Siroheme- and [Fe4-S4]-dependent NirA from Mycobacterium tuberculosis is a sulfite reductase with a covalent Cys-Tyr bond in the active site. J. Biol. Chem. 280, 27319–27328. doi:10.1074/jbc.M502560200

Sinha, S.C., Wetterer, M., Sprang, S.R., Schultz, J.E., Linder, J.U., 2005. Origin of asymmetry in adenylyl cyclases: structures of Mycobacterium tuberculosis Rv1900c. EMBO J. 24, 663–673. doi:10.1038/sj.emboj.7600573

Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., Strynadka, N.C.J., 2013. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. J. Biol. Chem. 288, 17782–17790. doi:10.1074/jbc.M113.462036

Tretter, E.M., Schoeffler, A.J., Weisfield, S.R., Berger, J.M., 2010. Crystal structure of the DNA gyrase GyrA N-terminal domain from Mycobacterium tuberculosis. Proteins Struct. Funct. Bioinforma. 78, 492–495. doi:10.1002/prot.22600

Tullius, M.V., Harmston, C.A., Owens, C.P., Chim, N., Morse, R.P., McMath, L.M., Iniguez, A., Kimmey, J.M., Sawaya, M.R., Whitelegge, J.P., Horwitz, M.A., Goulding, C.W., 2011. Discovery and characterization of a unique mycobacterial heme acquisition system. Proc. Natl. Acad. Sci. U. S. A. 108, 5051–5056. doi:10.1073/pnas.1009516108

Vetting, M.W., Errey, J.C., Blanchard, J.S., 2008. Rv0802c from Mycobacterium tuberculosis: the first structure of a succinyltransferase with the GNAT fold. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 64, 978–985. doi:10.1107/S1744309108031679

Wagner, J.M., Evans, T.J., Chen, J., Zhu, H., Houben, E.N.G., Bitter, W., Korotkov, K.V., 2013. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. J. Struct. Biol. 184, 115–128. doi:10.1016/j.jsb.2013.09.022

Wang, L., Li, J., Wang, X., Liu, W., Zhang, X.C., Li, X., Rao, Z., 2013. Structure analysis of the extracellular domain reveals disulfide bond forming-protein properties of Mycobacterium tuberculosis Rv2969c. Protein Cell 4, 628–640. doi:10.1007/s13238-013-3033-x

Wang, Z., Choudhary, A., Ledvina, P.S., Quiocho, F.A., 1994. Fine tuning the specificity of the periplasmic phosphate transport receptor. Site-directed mutagenesis, ligand binding, and crystallographic studies. J. Biol. Chem. 269, 25091–25094.

Wang, Z., Potter, B.M., Gray, A.M., Sacksteder, K.A., Geisbrecht, B.V., Laity, J.H., 2007. The solution structure of antigen MPT64 from Mycobacterium tuberculosis defines a new family of beta-grasp proteins. J. Mol. Biol. 366, 375–381. doi:10.1016/j.jmb.2006.11.039

Wells, R.M., Jones, C.M., Xi, Z., Speer, A., Danilchanka, O., Doornbos, K.S., Sun, P., Wu, F., Tian, C., Niederweis, M., 2013. Discovery of a siderophore export system essential for virulence of Mycobacterium tuberculosis. PLoS Pathog. 9, e1003120. doi:10.1371/journal.ppat.1003120

WHO, Global tuberculosis report 2013, WHO. URL http://www.who.int/tb/publications/global_report/en/

Xiong, X., Xu, Z., Yang, Z., Liu, Y., Wang, D., Dong, M., Parker, E.J., Zhu, W., 2013. Key targets and relevant inhibitors for the drug discovery of tuberculosis. Curr. Drug Targets 14, 676–699.

Chapter 5. Optimization of Cloning and Expression of Mycosin-3 from *Mycobacterium tuberculosis* H37Rv

5.1 Introduction

Pathogenic bacteria utilise different types of secretion systems to transport virulence factors into host cells to manipulate their immune mechanisms (Finlay and Falkow, 1997). The Type VII Secretion System (T7SS) is the most recently described bacterial secretion system (Abdallah *et al*., 2007) and it is found mostly in mycobacteria and some other high GC gram-positive bacteria (Gey van Pittius *et al*., 2001). *Mycobacterium tuberculosis* (*M. tuberculosis*), the etiological agent of tuberculosis, has five copies of T7SS, namely ESX-1 – ESX-5 evolved via gene duplication (Gey van Pittius *et al*., 2006). ESX-1 and ESX-5 are involved in the virulence of virulent mycobacteria (Houben *et al*., 2013a) whereas ESX-3 participates in mycobactin-mediated iron acquisition (Siegrist *et al*., 2014, 2009). Owing to the close associations with these fundamental processes, T7SS has been extensively studied, with ESX-1 and ESX-5 from *M. tuberculosis* and *M. marinum*, being the most highly studied (Houben *et al*., 2013b). The working mechanisms, involvement in pathogenicity and gene regulation of the T7SS are discussed in detail in Chapter 2.

Previous studies have shown that the mycosins, as one essential component of the T7SS, could not be isolated with the core machinery (EccB, EccC, EccD and EccE), indicating a weak interaction on the membrane (Houben *et al*., 2012). Mycosins have enzymatic signatures similar to the subtilisin-like serine proteases and they share the conserved catalytic traid (Asp-His-Ser) (Brown *et al*., 2000). However, the initial screening experiments did not reveal their substrates and it was believed that the substrates are more specific than other subtilisin family members (Brown *et al*., 2000) hence their functions remained unknown. In-depth studies on Mycosin-1 revealed that EspB is one substrate where it was cleaved twice upon its secretion (Ohol *et al*., 2010). Although the reason for the processing of EspB by Mycosin-1 is unknown, it is speculated that it could facilitate the maturation of EspB which become a functional effector to interact with the host targets. Mycosin-1 is essential for the secretion of ESX-1 substrates but abolishing its protease activity up-regulates the substrate secretion (Ohol *et al*., 2010). Possibly Mycosin-1 plays a regulatory role on maintaining the balance between immune detection and virulence to achieve the long term persistence for *M. tuberculosis* (Ohol *et al*., 2010). EspB is unlikely to be the only mycosin substrate, as the *espB* gene is not present in the other 4 ESX regions in *M. tuberculosis*.

Mycosins contain, from N-terminus to C-terminus, a signal sequence, a "pro-peptide", the catalytic domain, the proline-rich linker region and the hydrophobic transmembrane region (Figure 5.1) and it was believed that the mycosins would only be active when the "pro-

peptide" is removed (Brown *et al*., 2000). The crystal structure of Mycosin-1 ($MycP_1$) from both *M. smegmatis* and *M. thermoresistabile* and Mycosin-3 ($MycP_3$) from *M. smegmatis* suggest that the "pro-peptide" does not have such a regulatory role on the activities of mycosins; instead it wraps around the catalytic domain and stabilizes it. Retaining the "pro-peptide" does not affect the protease activity (Solomonson *et al*., 2013; Wagner *et al*., 2013). Thus the "pro-peptide" was renamed as the "N-terminal extension region" (Solomonson *et al*., 2013; Wagner *et al*., 2013). Moreover, Solomonson and colleagues found that their *M. smegmatis* $MycP_1$ orthologue was inefficient at cleaving EspB *in vitro* and they argued the reason is that the $MycP_1$ was separated from the rest of the secretion system in the experiments. It was not fully activated but showed leaky activity (Solomonson *et al*., 2013). Although the Mycosin-1 orthologues in *M. smegmatis* and *M. thermoresistabile* have more than 70% amino acid sequence identity with that in *M. tuberculosis*, their substrate specificities may differ considering the Mycosin-1 orthologues in those saprophytic species may not be involved in virulence related processes. Both research groups did not manage to produce the recombinant *M. tuberculosis* orthologue of Mycosin-1 or Mycosin-3 (Solomonson and Korotkov, personal communication).

```
1-   MIRAAFACLAATVVVAGWWTPPAWAIGPPVVDAAAQPPSGDPGPVAPMEQRGACSVSGVI
61-  PGTDPGVPTPSQTMLNLPAAWQFSRGEGQLVAIIDTGVQPGPRLPNVDAGGDFVESTDGL
121- TDCDGHGTLVAGIVAGQPGNDGFSGVAPAARLLSIRAMSTKFSPRTSGGDPQLAQATLDV
181- AVLAGAIVHAADLGAKVINVSTITCLPADRMVDQAALGAAIRYAAVDKDAVIVAAAGNTG
241- ASGSVSASCDSNPLTDLSRPDDPRNWAGVTSVSIPSWWQPYVLSVASLTSAGQPSKFSMP
301- GPWVGIAAPGENIASVSNSGDGALANGLPDAHQKLVALSGTSYAAGYVSGVAALVRSRYP
361- GLNATEVVRRLTATAHRGARESSNIVGAGNLDAVAALTWQLPAEPGGGAAPAKPVADPPV
421- PAPKDTTPRNVAFAGAAALSVLVGLTAATVAIARRRREPTE
```

Figure 5.1. The primary sequence of $MycP_3$ from *M. tuberculosis*. Red: signal peptide; Blue: "pro-peptide"; Brown: proline-rich linker; Purple: hydrophobic transmembrane region; Green (underlined): catalytic triad, Glu^{95} – His^{126} – Ser^{342}, Oxyanion hole: Asn^{237}.

In this study, the *mycP*$_3$ gene from *M. tuberculosis* was cloned and expressed. Although the role of $MycP_3$ remains enigmatic, it is essential to *M. tuberculosis* survival *in vitro* (Griffin *et al*., 2011; Sassetti *et al*., 2003) making it a potential anti-TB drug target (Frasinyuk *et al*., 2014; Wagner *et al*., 2013). Extensive efforts were undertaken to optimize the production of $MycP_3$ to increase protein yield, solubility and stability. While small amounts of protein were produced, this could not be easily handled limiting its use in downstream analysis such as substrate library screening and X-ray crystallography. The description below outlines the many efforts undertaken to clone *mycP*$_3$. This may prove useful in future if a related project is attempted. The section also advises possible alternative approaches that may be conducted in future.

5.2 Methods and Materials

5.2.1 Media, Plasmids and Bacteria Strains

Lysogeny broth (LB) was used to culture all *Escherichia coli* strains including XL-1 blue (Promega), BL21 (DE3) pLysS (Promega), Arctic Express (Agilent Technologies), and Origami (Novagen). Middlebrook 7H9 broth and 7H11 agar, each supplemented with 0.2% Tween 80, 0.5% glucose and 0.5% glycerol, were used to culture *M. smegmatis* MC^2 155 strain in liquid and solid media respectively. *E. coli* expression vector pET-28a (Novagen) (Figure S5.2), pGEX-6P-1 (EMBL) (Figure S5.3), pCOLD (Takara) (Figure S5.4) were used to produce C-terminal His_6-tagged, N-terminal glutathione *S*-transferase (GST)-tagged, and N-terminal His_6-tagged fusion proteins respectively while pET-21a (Novagen) was used for *in vitro* tagless construct expression.

5.2.2 Cloning, Expression and Purification

A range of truncated *mycP*$_3$ constructs were generated by Phusion DNA polymerase (ThermoScientific) using specific primer pairs (Table 5.1) in polymerase chain reactions (PCR). The PCR experiments were conducted with the following procedure: 1. initial denaturation at 98°C for 15 min; 2. 35 cycles of amplification including denaturation step at 98°C for 30 sec, annealing step at 56°C for 30 sec, and extension step at 72°C for 1 min; 3. extension at 72°C for 8 min; and 4. cooling at 4°C for 10 min. The PCR products were ligated into pJET2.1 cloning vector (Thermo Scientific) (Figure S5.1) using T4 DNA ligase (Promega). The recombinant pJET2.1 vector was restriction digested and the insert ligated into the expression vectors. Deletion constructs of Mycosin-3 were generated using QuickChange Site-Directed Mutagenesis Kit (Agilent Technologies) as per manufacturer's instructions. The recombinant or mutated expression vectors were electroporated into the expression host cells, where 1 µl of the vector is mixed with 50 µl of electro-compotent *E. coli* cells and the mixture was then transferred into an electroporation cuvette and the electroporation was conducted at 2.4 kV and the resistance was set at 4 kΩ using a GenePulser (BioRad, USA).

For the expression in *E. coli*, an overnight 50 ml starter culture was prepared from a single transformant colony. An aliquot of the starter culture was re-inoculated into 1 liter LB media, allowed to grow to mid-log phase (OD_{600nm} = 0.6 to 0.8) at 37°C and induced with 0.1 to 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16, 25 and 30°C for 18 hours for test expression. *E. coli* Arctic Express strain and pCOLD *E. coli* transformants were induced with 0.1 mM IPTG at 13°C and expression continued for 40 hours.

For the expression in *M. smegmatis*, a 1 ml starter culture in 7H9 medium was prepared from a single transformant colony. It was re-inoculated to a 10 ml and then to a 500 ml enriched 7H9 medium. The expression construct p19Kpro (Ashbridge *et al.*, 1989) and pDMN1

(Mailaender *et al*., 2004) did not require inducer while 50 ng/ml anhydrotetracycline (ATc) (Sigma) was used to induce $MycP_3$ expression in pSE100/pMC5m Tet-on expression system where ATc binds to TetR which is coded in pMC5m causing it to dissociate from tet operon located in pSE100 thereby allowing the transcription of the downstream gene cloned in the same vector (Ehrt *et al*., 2005). The expression was carried out for 16 hours at 37°C.

For the GST-tagged fusion protein purification, cultures were harvested by centrifugation at 4000 x g at 4°C for 10 min. The pellet was resuspended in phosphate buffered saline (PBS) and stored at -80°C overnight. The frozen resuspension was thawed and sonicated using a probe sonicator (MSE) at an amplitude of 20 µm for 5 cycles of 30 s with 30 s incubation on ice. The soluble and insoluble fractions of the cell lysate were separated by centrifugation at 14000 x g at 4°C for 45 min. The supernatant was mixed and incubated with PBS-equilibrated glutathione agarose beads (ABT) at 4°C on a roller mixer for 1 hour for efficient coupling. The mixture was poured into an empty drip column and the flowthrough was allowed to elute under gravity. The column was washed with 20 column volumes (CV) of PBS. The fusion protein was either digested on column by PreScission protease (Roche) or eluted with elution buffer (25 mM Tris-HCl, pH 7.4, 150 NaCl, 15 mM reduced glutathione) followed by fusion protein digestion in solution. The elution fractions containing $MycP_3$ were pooled and concentrated using a centrifugal filter device (Millipore) and applied to Superdex 75 size exclusion column (GE HealthCare) for further purification. The desired fraction was collected and pooled.

For the purification of The His_6-tagged fusion proteins, the *M. smegmatis* cell pellet was resuspended in lysis buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10 mM imidazole) and lysed by a tissue and cell homogenizer FastPrep-24 (MP Biomedical) at a speed of 6 M/s for 5 cycles of 30 s with 30 s incubation on ice in between. The soluble fraction was separated as described above and the target protein was coupled to Ni-NTA beads (Qiagen). The beads were loaded into an empty drip column and washed with 20 CV wash buffer (25 mM Tris-HCl, pH 7.4, 150 NaCl, 20 mM imidazole) and eluted with elution buffer (25 mM Tris-HCl, pH 7.4, 150 NaCl, 250 mM imidazole).

$MycP_3$ was also produced using PURExpress *In Vitro* Protein Synthesis Kit (New England Biolab) as per manufacturer's instructions. The quantity of the protein production and the presence of $MycP_3$ during purification were monitored using SDS-PAGE (Laemmli, 1970) and mass spectrometer (Central Analytical Facility, Stellenbosch University).

Table 5.1. Primers used to generate the starting *M. tuberculosis mycP*$_3$ construct (Construct A, not codon-optimized), Eleven codon-optimized *mycP*$_3$ constructs (Constructs B to L) including 3 deletion constructs (Constructs I to K), as well as three *M. smegmatis* expression constructs based on the Construct A (Constructs M to O). Expression hosts and vectors are as listed.

Construct Name	Encoded Amino Acid Sequence	Expression Host	Expression Vector	Primer Sequences and Their Restriction Sites
Construct A	Ile^{26}-Asn^{430}	*E. coli* BL21 (DE3) pLysS	pET-28a	forward: 5'-CCATGGCGATCGGGCCGCCGG-3' (*Nco*I)
				reverse: 5'-CTCGAGGTTGCGCGGTGTGGTG-3' (*Xho*I)
			pGEX-6P-1	Not required (cloned from synthetic gene)
Construct B	Arg^{51}-Asn^{430}		pET-28a	forward: 5'-CCATGGAACGCGGTGCGTGCAG-3' (*Nco*I)
				Construct A reverse primer
			pGEX-6P-1	forward: 5'-GGATCCCGCGGTGCGTGCAG-3' (*BamH*I)
				Construct A reverse primer
Construct C	Gly^{52}-Leu^{401}		pGEX-6P-1	forward: 5'-GGATCCGGTGCATGTAGCG-3' (*BamH*I)
				reverse: 5'-CTCGAGTCACAGCTGCCAGGTC-3' (*Xho*I)
Construct D	Ser^{57}-Leu^{401}		pGEX-6P-1	forward: 5'-GGATCCGGTGTTATTCCGG-3' (*BamH*I)
				Construct C reverse primer
Construct E	Gly^{62}-Leu^{401}		pGEX-6P-1	forward: 5'-GGATCCGGTACAGATCCGG-3' (*BamH*I)
				Construct C reverse primer
Construct F	Val^{67}-Leu^{401}		pGEX-6P-1	forward: 5'-GGATCCGTTCCGACCCCGAG-3' (*BamH*I)
				Construct C reverse primer
Construct G	Ser^{71}-Leu^{401}		pGEX-6P-1	forward: 5'-GGATCCCAGACCATGCTG-3' (*BamH*I)
				Construct C reverse primer
		E. coli Arctic Express and BL21 (DE3) pLysS	pCOLD	forward: 5'-CATATGCAGACCATGCTGAATC-3' (*Nde*I)
				Construct C reverse primer
		Cell Free	pET-21a	As per primer set for pCOLD
Construct H	Leu^{77}-Leu^{401}	*E. coli* BL21	pGEX-6P-1	forward: 5'-GGATCCCTGCCAGCAGCATG-3' (*BamH*I)

		(DE3) pLysS		Construct C reverse primer
Construct I	Construct G ΔPro^{164}-Asp^{179}		pGEX-6P-1	forward: 5'-GAGGTTGCAGTTCTGGCAGG-3'
				reverse: 5'-GAGTGAAAATTTGGTGCTCATTG-3'
Construct J	Construct G ΔPro^{253}-Ser^{284}		pGEX-6P-1	forward: 5'-GAGCTGAGCGTTGCAAGC-3'
				reverse: 5'-GAGCGGATTGCTATCACAG-3
Construct K	Construct G ΔGly^{320}-Lys^{334}		pGEX-6P-1	forward: 5'-TCCCAGAAACTGGTTGCACTG-3'
				reverse: 5'-TCCATCACCGCTATTGCTAAC-3'
Construct L	Ile^{24}-Leu^{401}	*E. coli* Origami	pGEX-6P-1	forward: 5'-GGATCCATTGGTCCGCCTGTTG-3' (*BamH*I)
				Construct C reverse primer
Construct M	Gly^{52}-Leu^{401}	*M. smegmatis* MC^2 155	p19Kpro	forward: 5'-GGATCCAGGCGGAGAACAAATGGGTGCGTGCAGCGTCTC-3' (*BamH*I)
				reverse: 5'-ATCGATTTAGTGGTGGTGGTGGTGGTGCAGTTGCCAGGTCAGGGC-3' (*Cla*I)
Construct N	Gly^{52}-Leu^{401}		pDMN1	forward: 5'-AAGCTTTTCGTGCGTGCAGCGTCTCCGGTGTTATCCCGGGC-3' (*Hind*III)
				reverse: 5'-GTTAACTCAGTGGTGGTGGTGGTGGTGCAGTTGCCAGGTCAGGGC-3' *Hpa*I
Construct O	Ser^{71}-Leu^{401}		pSE100	forward: 5'-GGATCCAGGCGGACTGCAGATGCAAACGATGCTGAATCTGCCTGCG-3' (*BamH*I)
				reverse: 5'-AAGCTTTCAGTGGTGGTGGTGGTGGTGCAGTTGCCAGGTCAGGGCCGCC-3' (*Hind*III)

*The underlined sequence reflects the sequence of restriction site the name of which is indicated in the brackets following the sequence

5.3 Results

5.3.1 Protein Production in *E. coli*

The role that the putative "pro-peptide" (Ile^{26}-Gln^{50}) plays in $MycP_3$ remains unknown. As it is common for signal sequences and pro-peptide regions of proteases to be cleaved off during the maturation of the protein, two truncated constructs were generated for protein production, including (i) Ile^{26}-Asn^{430} (Construct A), eliminating the signal sequence (Met^{1}-Ala^{25}) and transmembrane region (Val^{431}-Glu^{461}); and (ii) Arg^{51}-Asn^{430} (Construct B), eliminating the signal sequence, "pro-peptide" and transmembrane region. Both constructs A and B were cloned into the expression vector pET-28a. However, neither pET-*mycP*$_3$ constructs produced sufficient amounts of MycP3-His_6 fusion protein to be visible by SDS-PAGE.

E. coli codon usage differs significantly from that of mycobacteria (Scapoli *et al*., 2009). A synthetic *mycP*$_3$ gene (encoding the amino acid sequence from Ile^{26} to Asn^{430}), codon-optimized for expression in *E. coli* (GeneArt), was cloned into pGEX-6P-1 expression vector to code for $MycP_3$ fused to an N-terminal GST-tag. GST-$MycP_3$ (Ile^{26}-Asn^{430}) production in *E. coli* BL21 (DE3) at 37ºC is little (Figure 5.2) and mostly insoluble (Figure 5.3a). The soluble fraction of the fusion protein was unstable and prone to rapid proteolysis and precipitation (Figure 5.4 and 5.5). It co-purified with *E. coli* chaperonin DnaK (69.1 kDa) and heat shock protein (60 kDa) identified by mass spectrometry (Figure 5.4). Adding ATP and detergent Triton X-100 allowed chaperone removal (Schönfeld *et al*., 1995) while increasing the rate of precipitation and degradation (Figure 5.5) hampering purification. Expression at lower temperatures (16, 25, and 30ºC) and a range of IPTG concentrations for long periods did not improve the solubility or stability of the fusion protein (Figure 5.3b).

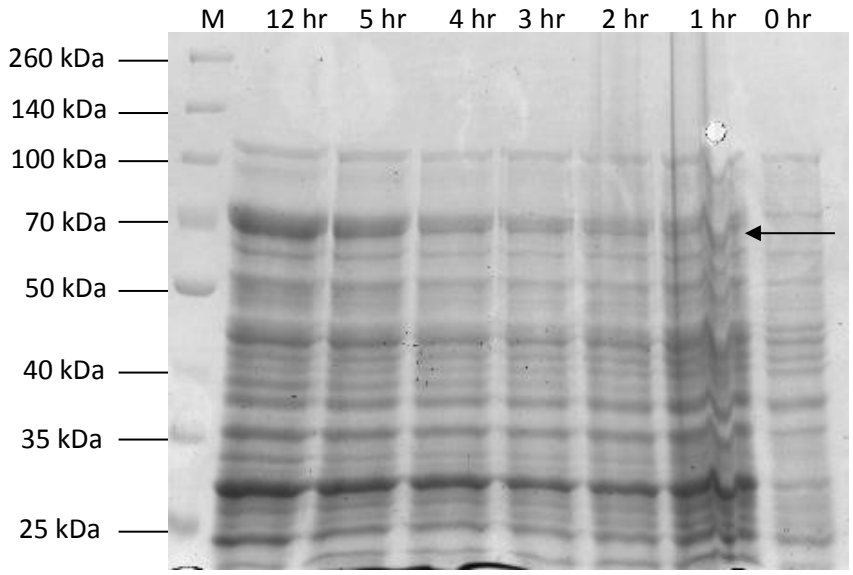
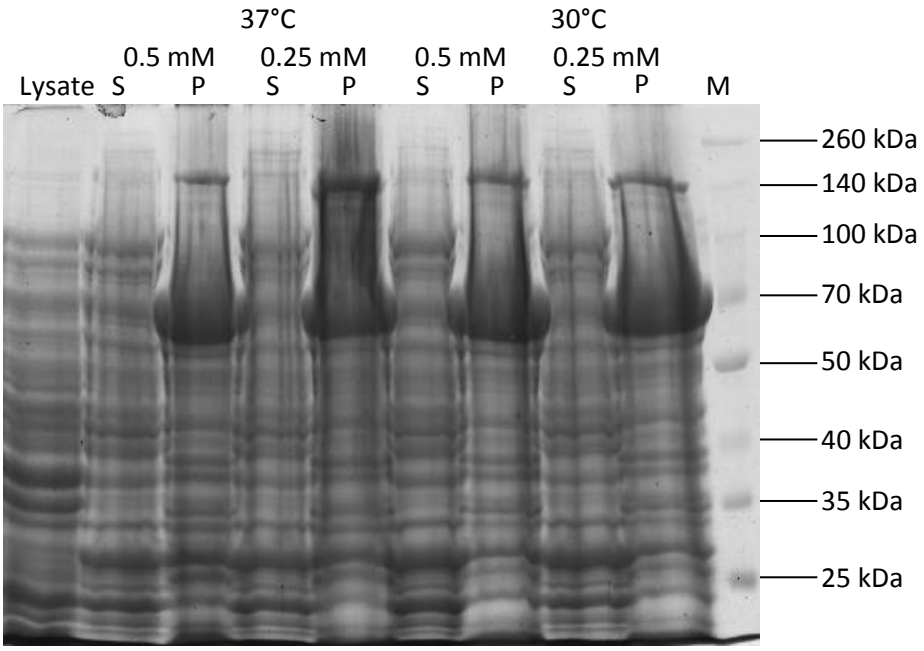
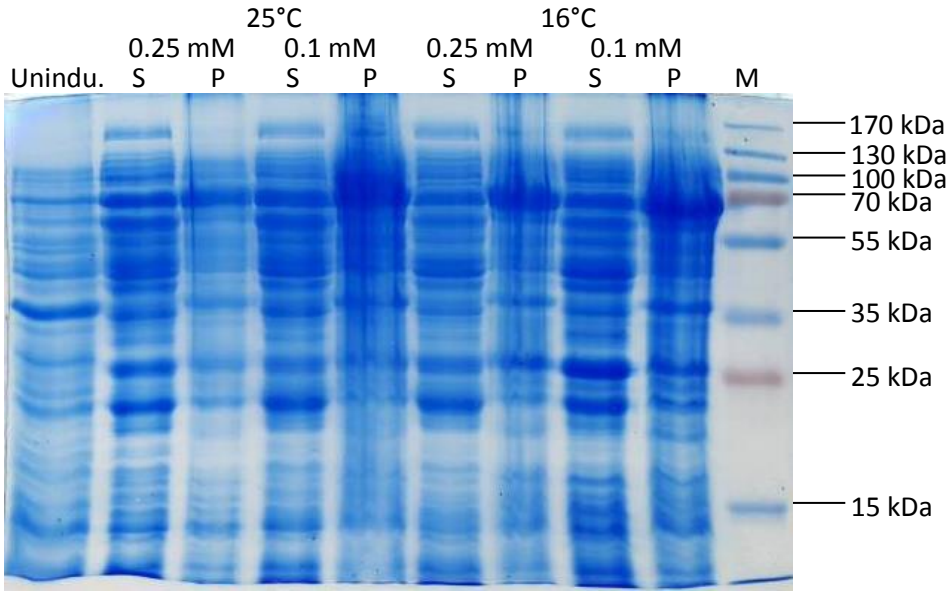


Figure 5.2. Test expression profile over time for GST-tagged $MycP_3$ (Ile^{26}-Asn^{430}) (Construct A) in *E. coli* BL21 (DE3) pLysS strain. The black arrow indicates the increasing amount of fusion protein produced over time. The fusion protein has a size of about 67 kDa.



(a)



(b)

Figure 5.3. Solubility test on Construct A produced from *E. coli* BL21 (DE3) pLysS strain under a range of IPTG concentrations and temperatures with (a) for 30 and 37°C and (b) for 16 and 25°C. The fusion protein was produced mostly as inclusion bodies under all conditions. S: Cell lysate supernatant; P: Cell lysate pellet; M: protein ladder.

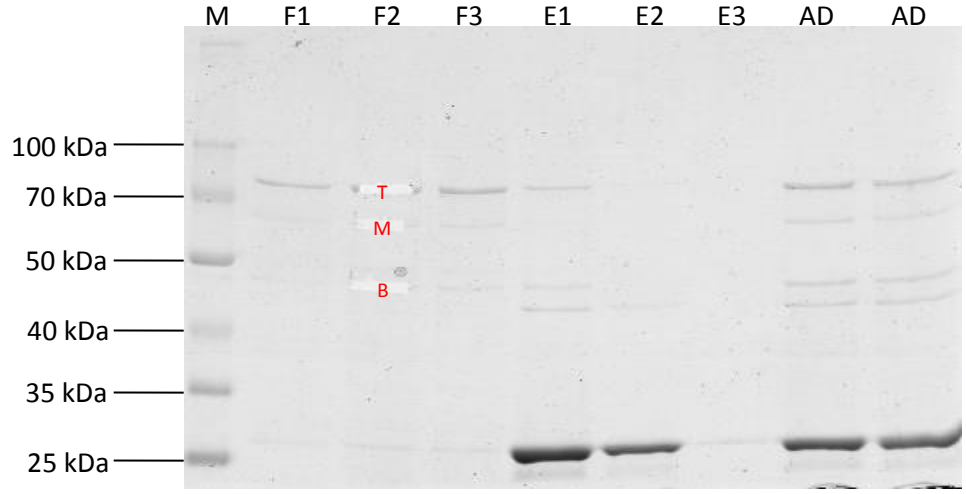


Figure 5.4. Construct A purification profile with on-column PreScission Protease digestion. After on-column digestion, the flowthrough (F) contained $MycP_3$ and the eluent (E) contained undigested fusion protein and GST. AD lanes were loaded with digested fusion protein before purification. Band T was identified as Chaperonin DnaK (MW = 69.1 kDa) from *E. coli*. Band M was identified as the mixture of heat shock chaperonin (60 kDa) from *E. coli* and $MycP_3$ from *M. tuberculosis*. Band B was identified as $MycP_3$ from *M. tuberculosis*. F: flowthrough; E: eluent; M: protein ladder.

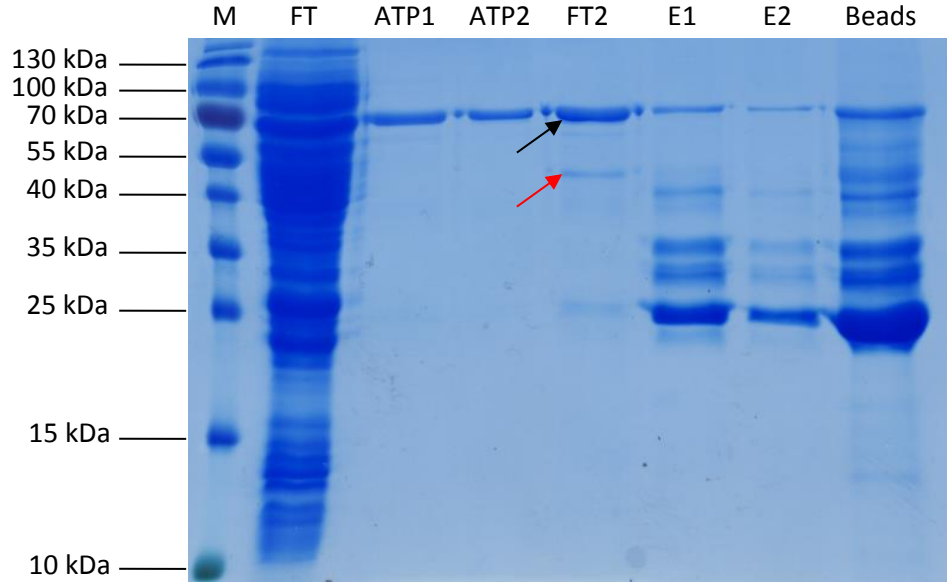
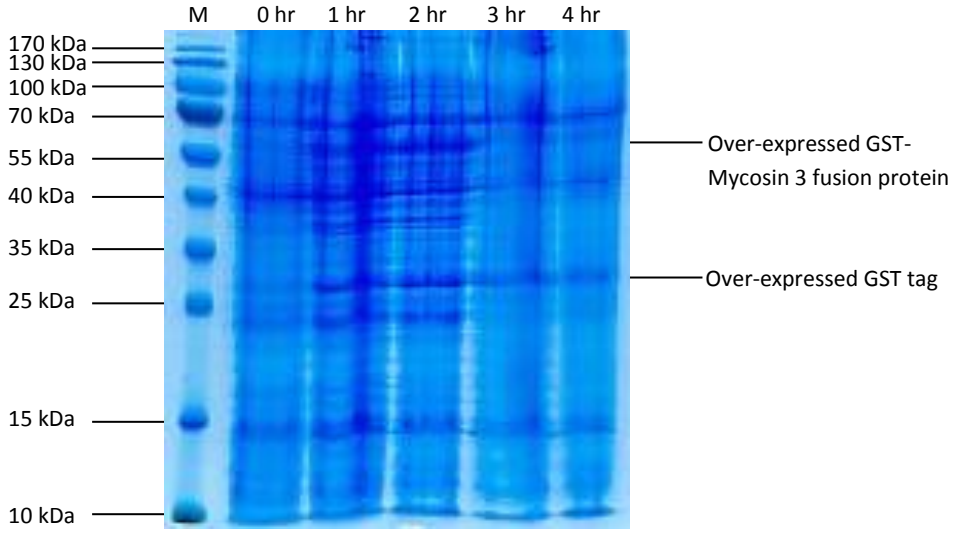
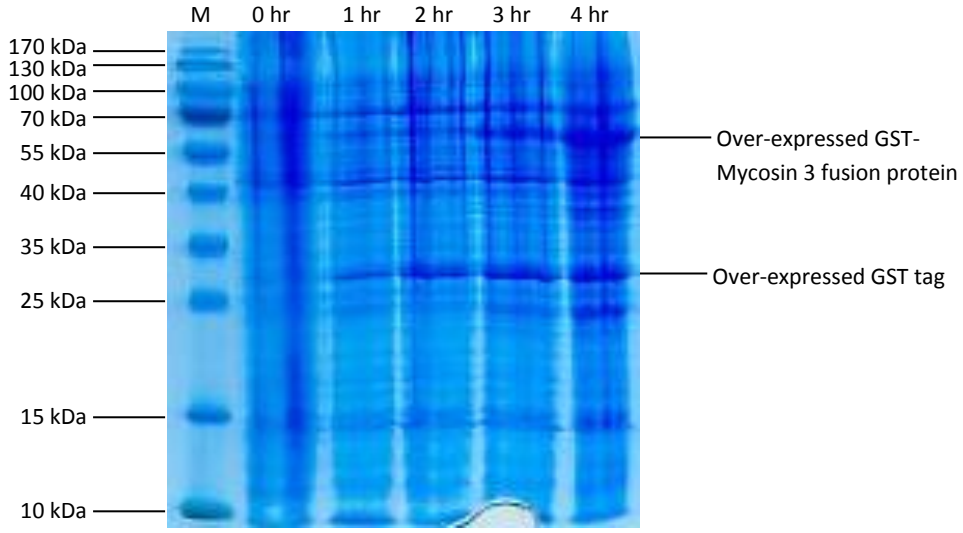


Figure 5.5. Construct A purification profile with on-column PreScission Protease digestion where ATP and Triton X-100 were added in all buffers (Black arrow: DnaK; Red arrow: $MycP_3$). DnaK were dissociated from DnaK-$MycP_3$ complex. However DnaK elimination was not achieved and $MycP_3$ became ready to precipitate and susceptible to proteolysis after DnaK detachment. M: Protein Marker; FT: Flow Through; ATP: eluent fraction containing ATP; E: elution with elution buffer; Beads: an aliquot of column matrix after elution was loaded.

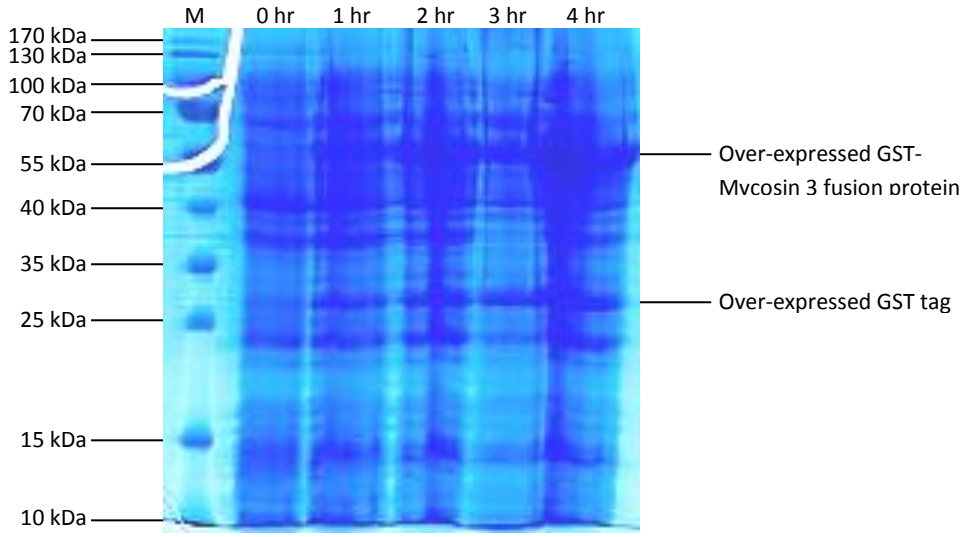
At the time of undertaking the studies outlined here, structural information for $MycP_3$ from *M. tuberculosis* H37Rv was not available. It is not clear what the factors are that made it difficult for *E. coli* to express stable $MycP_3$ (Ile^{26}-Asn^{430}). Analysing $MycP_3$ for inherently disordered regions (IDR) using PrDOS (Ishida and Kinoshita, 2007) identified Met^{1}-Pro^{70} and Pro^{402}-Glu^{461} as disordered. Secondary structure was predicted using a range of different software (http://www.expasy.org/tools/). Aligning the IDRs to the secondary structure predictions indicates that IDRs correspond to regions without defined secondary structures. Pro^{402}-Glu^{461} was excluded from analysis as it encompasses the proline-rich linker (Pro^{402}-Pro^{423}) and the C-terminal transmembrane α-helix (Lys^{424}-Glu^{461}) (Brown *et al*., 2000). The predicted IDR Met^{1}-Pro^{70} includes the entire "pro-peptide" (Ile^{26}-Gln^{50}) together with the next 20 amino acids. A series of *mycP*$_3$ constructs removing residues Ile^{26}-Gln^{50} at intervals of 5-6 amino acids was generated encoding for Gly^{52}-Leu^{401} (Construct C), Ser^{57}-Leu^{401} (Construct D), Gly^{62}-Leu^{401} (Construct E), Val^{67}-Leu^{401} (Construct F), Ser^{71}-Leu^{401} (Construct G), and Leu^{77}-Leu^{401} (Construct H). These six constructs were cloned into the expression vector pGEX-6P-1 as the resulting N-terminal GST-tagged fusion proteins were produced at higher levels than the fusion proteins with C-terminal His_6-tag from pET-28a. Production of six $MycP_3$ protein variants was first attempted at 37ºC inducing with 0.1 mM IPTG. Protein production levels of constructs C and D proved low while those for constructs E, F, G and H were higher (Figure 5.6). However, the GST-$MycP_3$ fusion proteins were invariably incorporated into insoluble inclusion bodies (Figure 5.7). Expression at reduced temperatures (16, 25 and 30ºC) did not improve the solubility of the resulting protein significantly. Construct G (for Ser^{71}-Leu^{401}, i.e. excluding the IDR) was selected for gene expression optimization.



(a)



(b)



(c)

(Figure 5.6)

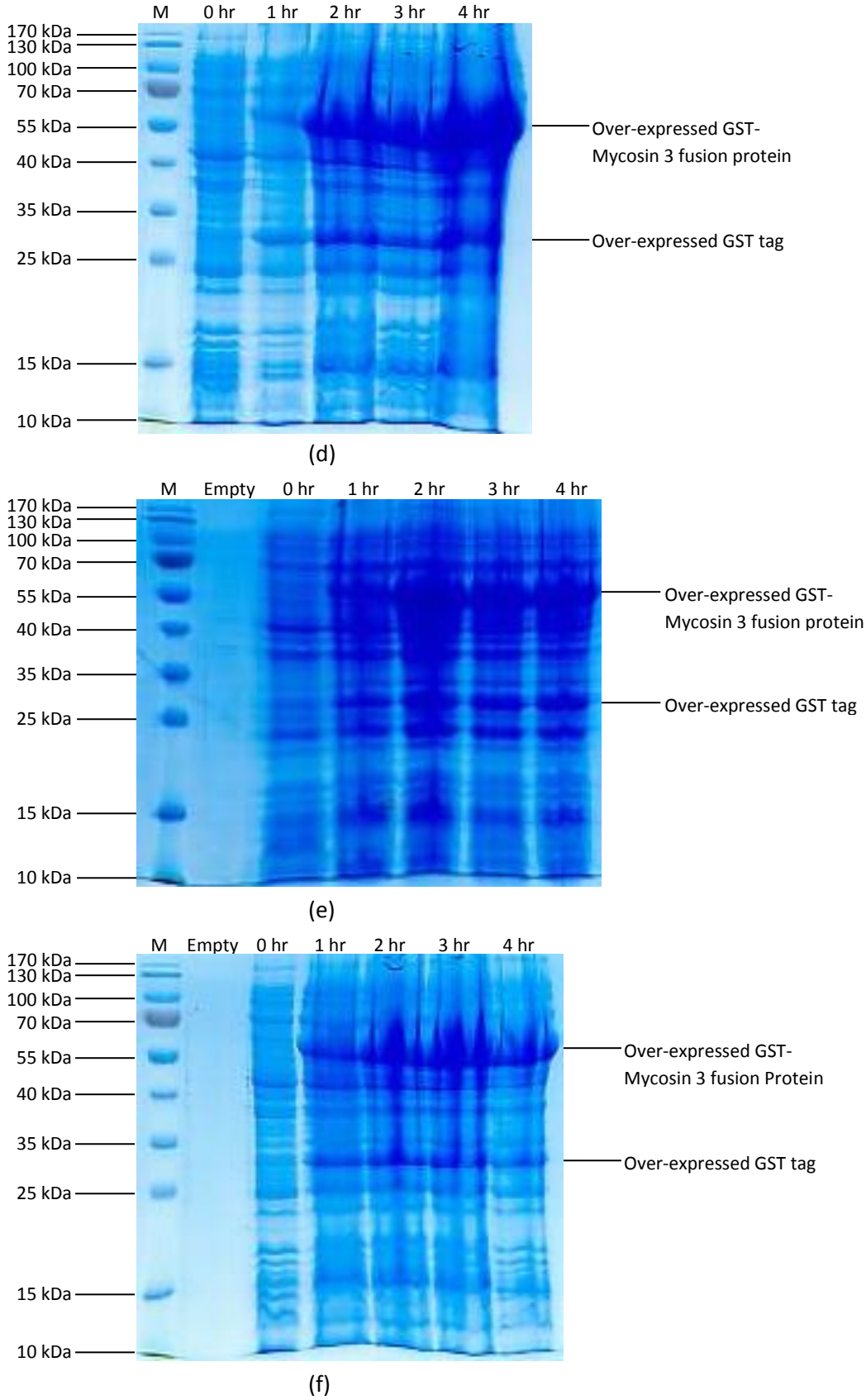


Figure 5.6. Test expression of 6 truncated constructs of $MycP_3$ (Constructs C to H). The protein production levels of construct C (a) and D (b) were low while those of constructs E (c), F (d), G (e) and H (f) were high. M: Protein ladder.

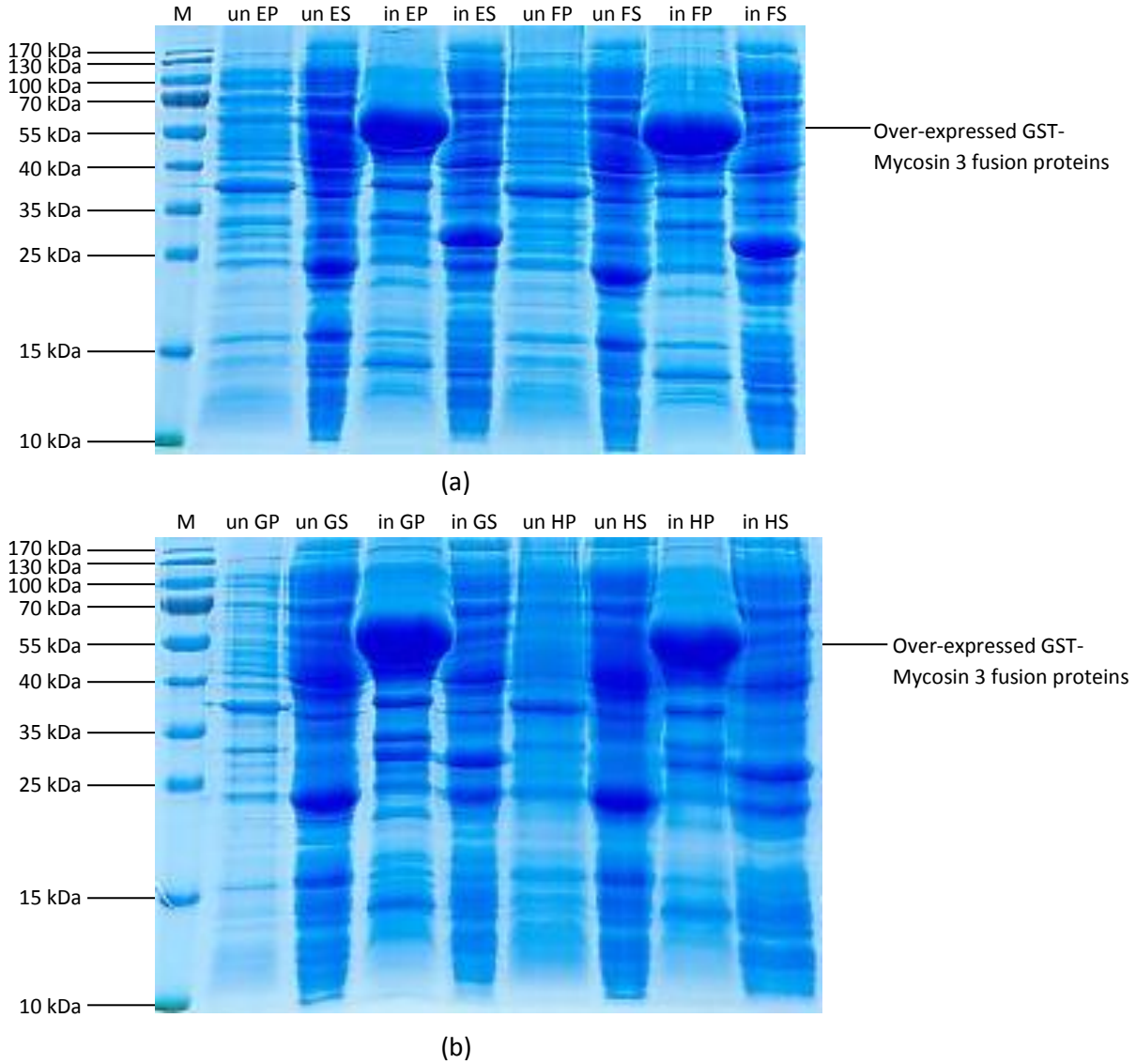
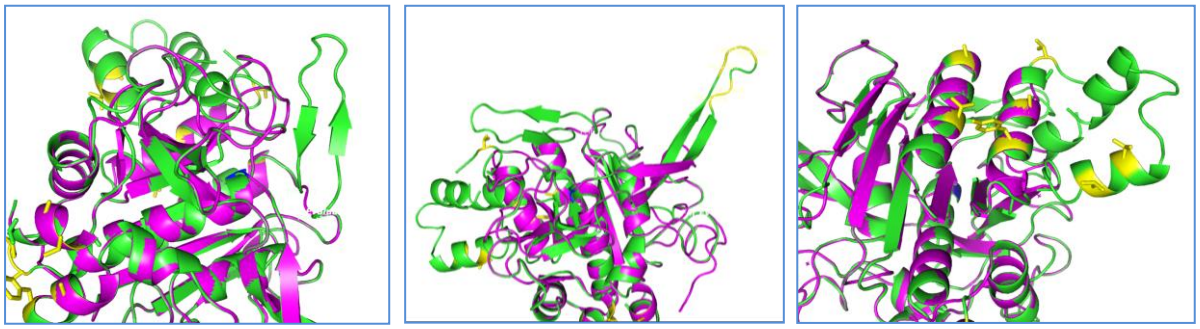


Figure 5.7. Solubility test for construct E and F (a), G and H (b) because they showed high expression levels. All four constructs were expressed insolubly at 0.1 mM IPTG under 37°C. The lanes are labelled according to the construct used (e.g. EP: Construct E pellet; ES: Construct E supernatant). M: Protein Marker; "un": the culture was un-induced; "in": the culture was induced with IPTG.

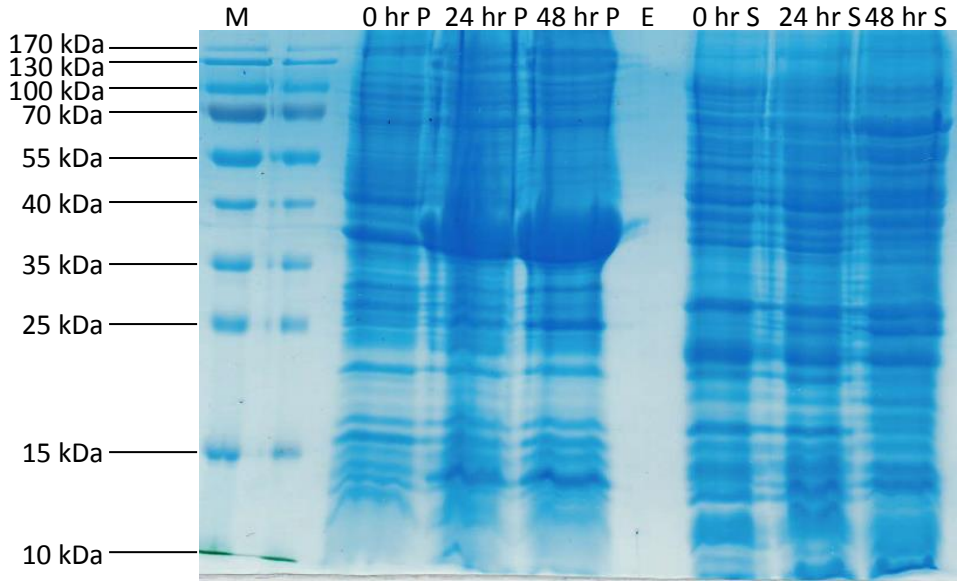
Exposed hydrophobic regions of a partly (mis)folded protein may increase intermolecular interaction and hence precipitation. Removing exposed hydrophobic regions of $MycP_3$ could improve its solubility in *E. coli*. Threading the amino acid sequence of $MycP_3$ onto the best structural model thermitase, by Swiss Model (Arnold *et al*., 2006) revealed three probable surface exposed hydrophobic regions (Figure 5.8). Using Construct G as a template the three regions were individually removed by site-directed mutagenesis generating Construct I (minus Pro^{164}-Asp^{179}), Construct J (minus Pro^{253}-Ser^{284}) and Construct K (minus Gly^{320}-Lys^{334}). However, none of the three constructs showed improved solubility of $MycP_3$ under a range of incubation temperatures and IPTG concentrations. The high expression level was however maintained (results not shown).



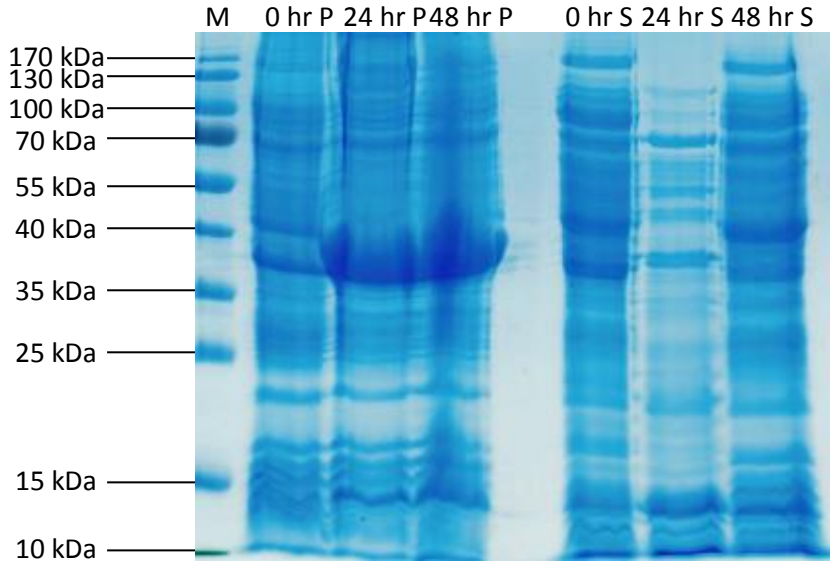
(a) (b) (c)

Figure 5.8. Identification of three possible surface hydrophobic regions in $MycP_3$ following threading onto thermitase with Swiss Model. The 3 regions include (a) Pro^{164}-Asp^{179}, (b) Pro^{253}-Ser^{284}, and (c) Gly^{320}-Lys^{334}.

Some *E. coli* expression vectors, such as pCOLD (Takara), and certain strains, such as Arctic Express (Agilent Technologies), are designed specifically to increase the solubility of produced proteins (Hayashi and Kojima, 2008). Construct G was thus cloned into pCOLD vector resulting in an N-terminal His_6-tagged *mycP*$_3$ recombinant gene. Even with the co-expression of cold-shock chaperone proteins in the pCOLD vector, $MycP_3$ was still produced insolubly in both BL21 (DE3) and Arctic Express strains (Figure 5.9). However, expression of Construct G in pGEX-6P-1 vector in Arctic Express was successful and resulted in the increased production of soluble GST-tagged $MycP_3$ fusion protein (Figure 5.10).



(a)



(b)

Figure 5.9. Test expression for construct G in pCOLD expression vector in BL21(DE3) strain (a) and Arctic Express strain (b). The N-terminal His_6-tagged fusion protein was produced as inclusion bodies. M: Protein Marker; P: Cell Lysate Pellet; E: Empty; S: Cell Lysate Supernatant.

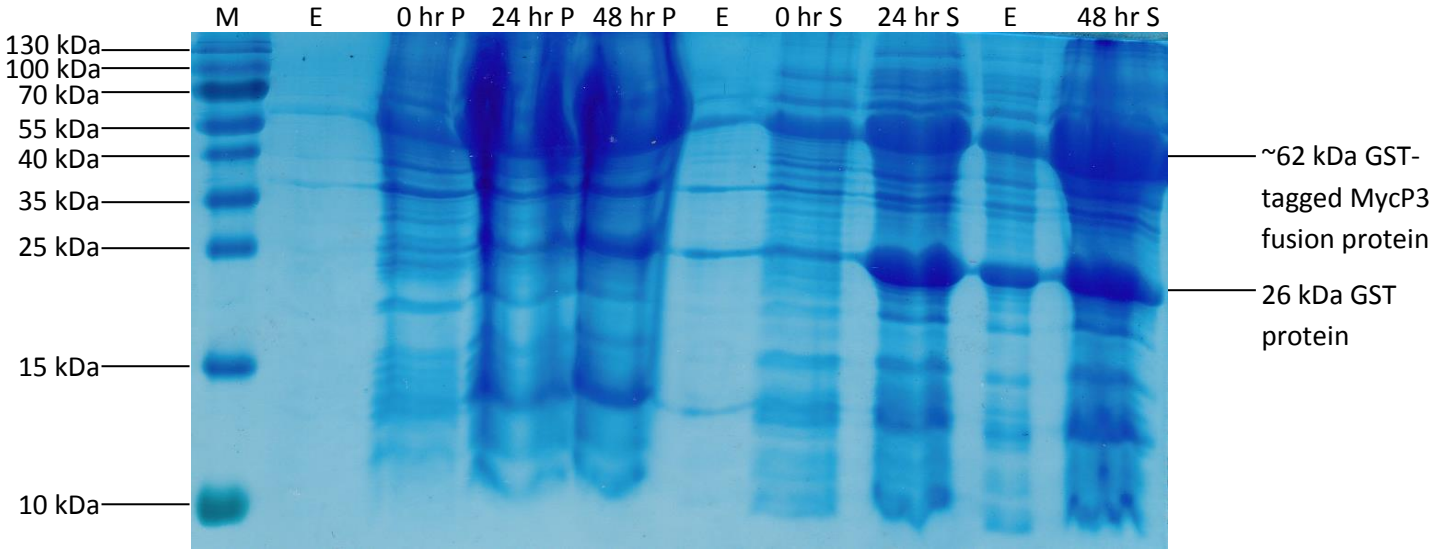


Figure 5.10. Test expression and solubility test for Construct G expressed in *E. coli* Arctic Express strain. High expression level can be observed in both soluble and insoluble fraction of the cell lysate. M: Protein Marker; E: Empty (the proteins in those lanes were from adjacent lanes by overflow); P: Cell Lysate Pellet; S: Cell Lysate Supernatant.

Further problems were encountered when it was found that the soluble GST-tagged $MycP_3$ fusion protein bound only weakly to glutathione agarose (GA) beads while conducting affinity chromatography (Figure 5.11). The purified fusion protein and unbound fusion protein in the flow through could both be cleaved by PreScission protease (Figure 5.12). However, removal of the GST-tag destabilized $MycP_3$ resulting in rapid precipitation (Figure 5.13). Addition of detergents (glycerol and Triton X-100) and reducing agents (β-mercaptoethanol and DTT) at various concentrations did not improve the stability of $MycP_3$ (results not shown). Omitting reducing agent from purification buffers resulted in"smears" above proteins bands in SDS-PAGE indicating intermolecular disulfide bonds (Kim and Robinson, 2006) implying that the $MycP_3$ was not folded properly. During this study, the structures of the $MycP_1$ ortholog (MycP1, PDB ID: 4J94 & 4KPG) in *M. smegmatis* and in *M. thermoresistibile* ATCC 19527 (PDB ID: 4HVL), and the $MycP_3$ ortholog in *M. smegmatis* (PDB ID: 4KG7) were solved (Solomonson *et al*., 2013; Wagner *et al*., 2013). These orthologs have at least 70% sequence identity with $MycP_3$ in *M. tuberculosis*. In these studies, the "pro-peptide" region was included in the constructs while the signal peptide, proline-rich linker region and transmembrane regions were excluded. The "pro-peptide" was re-annotated as an "N-terminal extension region", because it does not possess pro-peptide function but seems to maintain the stability of the subtilisin domain in these organisms. Cysteine[51] residue in *M. smegmatis* $MycP_1$, Cysteine[46] residue in *M. thermoresistibile* $MycP_1$ and cysteine[54] residue in *M. smegmatis* $MycP_3$ were all essential for the formation of disulfide bonds with one of the other 3 cysteine residues, contributing to the stability of the overall structure. As an ortholog to those mycosins, cysteine[54] in *M. tuberculosis* $MycP_3$ should also play an essential role in protein stability and functionality. Following this

information, Construct L, Ile^{24}-Leu^{401}, which is homologous to the deposited constructs described above, was generated and cloned into pGEX-6P-1 vector. This recombinant expression construct was expressed in disulfide bond promoting *E. coli* Origami strain (Novagen). The protein production was conducted at a range of temperatures (16, 25, 30, and 37ºC) and IPTG concentrations (0.1, 0.2, 0.5, and 1 mM). However, the production of GST-tagged $MycP_3$ fusion protein remained insignificant (results not shown). At 16ºC, the production of GST-tag was observed, but nothing was observed for the fusion protein (results not shown).

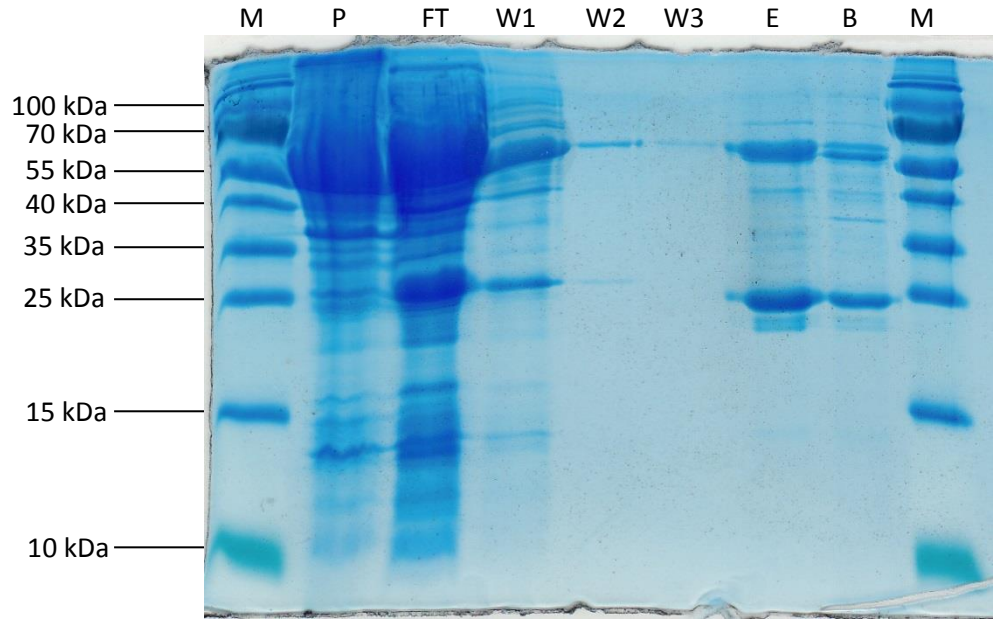


Figure 5.11. GST affinity chromatography purification profile. Besides a significant amount of fusion protein was produced as inclusion bodies (P), majority of fusion proteins came out in the flowthrough (FT) and some was washed off during extensive washing (W). A small quantity of the fusion proteins together with GST came off the column during elution (E). The beads (B) (column matrix after elution) bound to the residual fusion proteins and GSTs. M: Protein Marker.

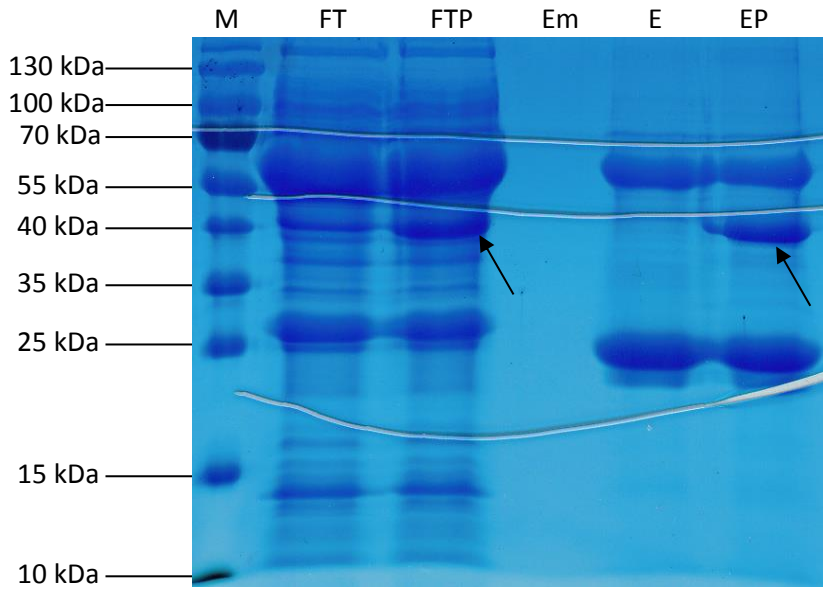
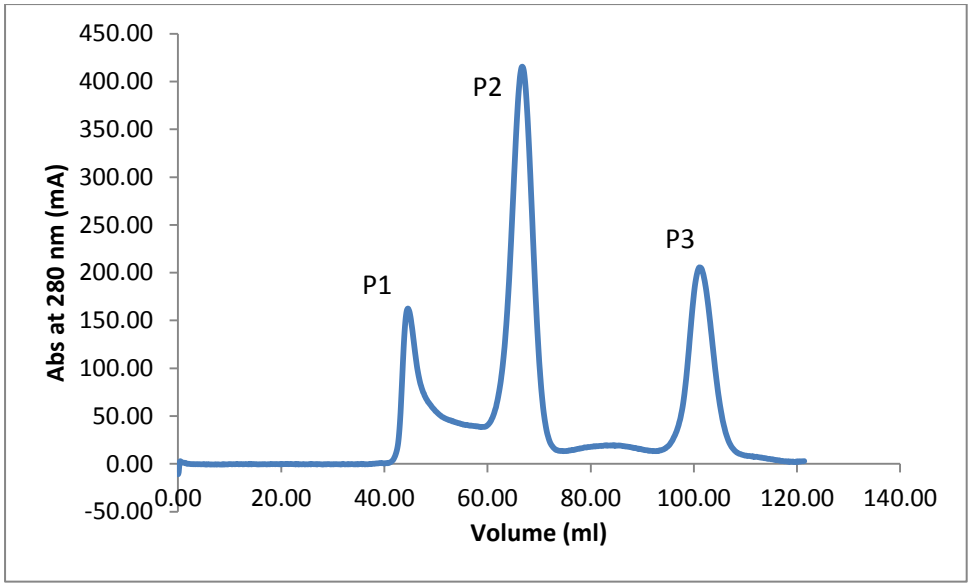
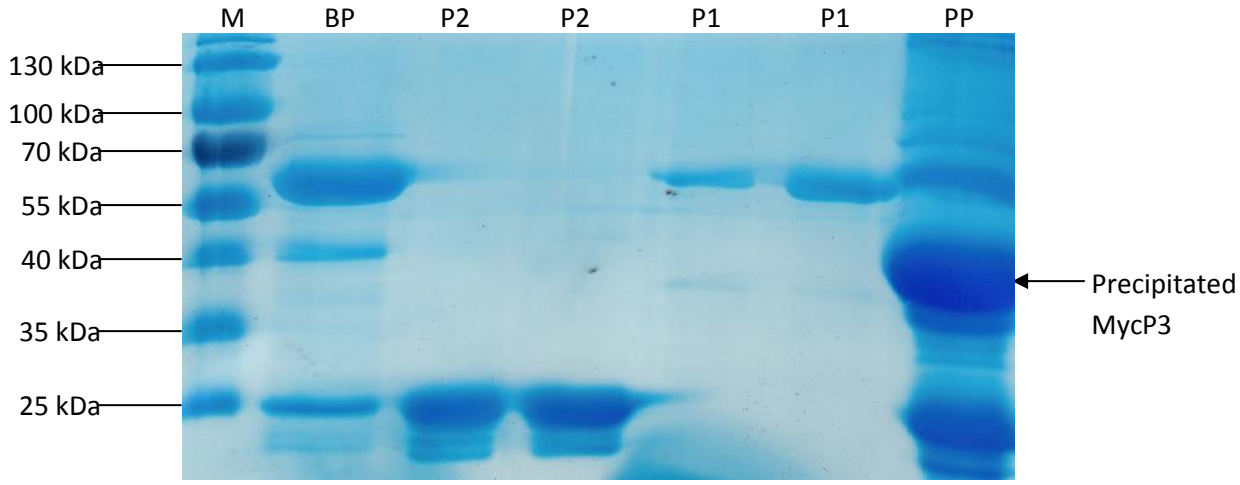


Figure 5.12. The purified GST-tagged MycP3 fusion protein and the unbound fusion protein in the flowthrough can both be cleaved by PreScission protease (black arrows show MycP3 without GST-tag), further confirming the true identity of this over-expressed fusion protein. M: Protein Marker; FT: FlowThrough; FTP: Protease treated FlowThrough; Em: Empty; E: Eluent; EP: Protease treated Eluent



(a)



(b)

Figure 5.13. Protein purification profile of size exclusion chromatography following GST affinity chromatography. There are 3 peaks on the chromatogram (a), where the first peak (P1) is the remaining un-cleaved fusion protein (b), the second peak (P2) is GST (b), and the 3rd peak was identified as reduced glutathione in the solution. The cleaved MycP3 was supposed to eluted between the 1st peak and the 2nd peak. However, it was very unstable and precipitated (PP) before loading onto the gel filtration column (b).

5.3.2 Protein production in *M. smegmatis* and *in vitro*

M. smegmatis has often been regarded as a better *M. tuberculosis* protein production host compared to *E. coli* due to its similar cellular environment as *M. tuberculosis* (Daugelat *et al.*, 2003). $MycP_3$ construct A and B were cloned into p19Kpro (Ashbridge *et al.*, 1989) and

pDMN1 (Mailaender *et al*., 2004) two mycobacterial non-inducible expression vectors both containing C-terminal His_6 tags. The transformed cultures were grown at 37ºC for 40 hours resulting in insignificant fusion protein production (results not shown). The growth of *M. smegmatis* transformants harbouring pDMN1-$MycP_3$ was severely inhibited due to the possible toxicity of over-expressed $MycP_3$ (results not shown) compared to the wild type strain transformed with empty pDMN1 vector. An inducible mycobacterial gene expression system (pSE100 and pMC5m) with a tet_{on} promoter (Ehrt *et al*., 2005) was used in an attempt to eliminate the growth inhibition. However, although the growth of this transformant was normal, the fusion protein production was insignificant (results not shown).

To eliminate all possible host cell inhibitory factors to exogenous protein production, $MycP_3$ construct G was cloned into pET-21a and expressed taglessly *in vitro*. The expression quality and conditions were carefully controlled but no $MycP_3$ protein was produced while a couple of translational machinery components such as translation initiation factor IF-2 and elongation factor Tu1 and Tu2 were degraded (Figure 5.14).

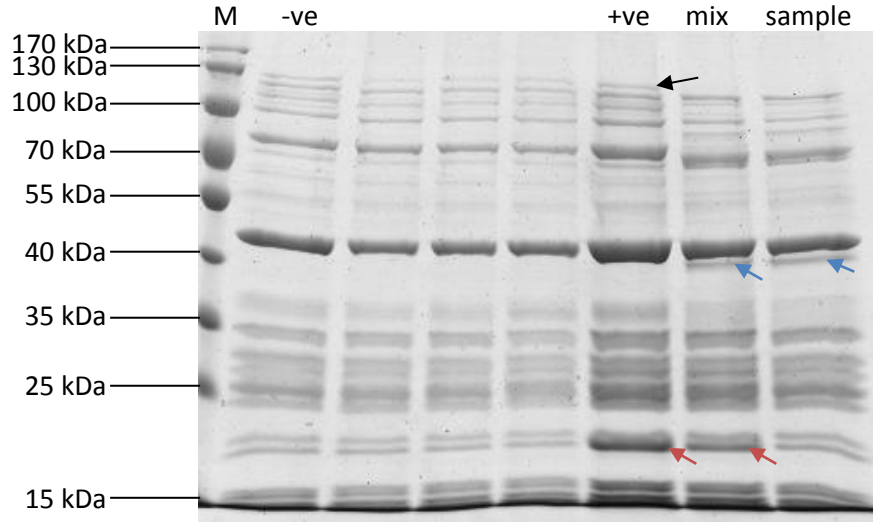


Figure 5.14. Cell free expression profile for pET-$MycP_3$ recombinant expression construct. Brown arrow: positive control expression; blue arrow: elongation factor Tu1 and Tu2, 43.3 kDa; black arrow: translation initiation factor IF-2, 97.3 kDa. M: protein marker; -ve: negative control; +ve: positive control; mix: using mixed positive control DNA with pET-MycP3 recombinant DNA; Sample: *in vitro* test expression for pET-$MycP_3$.

5.4 Discussion and Conclusion

Optimization of $MycP_3$ protein production constitutes the main body of this work. Two *mycP*$_3$ constructs originated from *M. tuberculosis* and 12 different codon-optimized truncated *mycP*$_3$ constructs were expressed with 7 expression vectors in 3 *E. coli* expression strains, 1 *M. smegmatis* strain as well as *in vitro* (Table 5.1). *In silico* sequence analysis and N-terminal/C-terminal truncation were instrumental to determine which construct was the highest expression construct (Construct G). The *E. coli* Arctic Express strain enabled this construct to produce soluble fusion protein. However, the soluble GST-tagged truncated $MycP_3$ fusion protein was only partially folded, making it unstable and possibly non-functional. The reason for this is not clear.

During the time that this study was conducted, two structural biology groups solved the 3-D structures of three other mycosin orthologues from environmental mycobacterial species with high homology to their *M. tuberculosis* counterparts (Figure 5.15). *M. smegmatis* $MycP_1$ has an identity score of 71.97 to *M. tuberculosis* MycP1 while that of *M. thermoresistibile* has a score of 71.11 to *M. tuberculosis* $MycP_1$ (Table 5.2) (Larkin *et al*., 2007). The sequence identity score between two $MycP_1$ orthologues of the two environmental species is the best, 78.78 (Table 5.2). The analysis showed that the best match for *M. tuberculosis* $MycP_3$ is *M. smemgatis* $MycP_3$ with a score of 59.04. Swiss Model also used $MycP_3$ from *M. smegmatis* as the model structure for *M. tuberculosis* $MycP_3$ (Figure 5.16) with the QMEANZ-score of -1.45 which indicates a relatively poor model (Benkert *et al*., 2011a) although the overlay structures do not show any distinct differences. Moreover, PROSO II, a sequence-based protein solubility evaluator, shows that $MycP_3$ construct Ile^{24}-Leu^{401} from *M. tuberculosis*, just like its orthologues from *M. smegmatis* and *M. thermoresistibile*, can be produced as a soluble protein (http://mips.helmholtz-muenchen.de/prosoII/prosoII.seam).

```
Msmeg_0083|MycP1    --MQRVAVMVLAVLLALFSAPPAWAIDPPVIDAGAVPP-DETGPDQPTEQRKICATPTVM 57
KEK_05522|MycP1     -------MLITALLLAVVTAPPAGAIEPPVIDPAAVPP-DETGPDNPMEQRRVCAAPTVY 52
Rv3883c|MycP1       ----VHRIFLITVALALLTASPASAITPPPIDPGALPP-DVTGPDQPTEQRVLCASPTTL 55
Rv0291|MycP3        MIRAAFACLAATVVVAGWWTPPAWAIGPPVVDAAAQPPSGDPGPVAPMEQRGACSVSGVI 60
Msmeg_0624|MycP3    MIHKSLGVVATAGLVLLIGCPSAGAVSPPQVDPQIAPPPGTAGPAQPMMQRSECITTSVL 60
                             .  :  :     ..* *: ** :*.   ** . .**  *  **  *  . .

Msmeg_0083|MycP1    PNSNFADRPWANDYLRIQEAQKFATGAGVTVAVIDTGVNGSPRVP-AEPGGDFVDAAGNG 116
KEK_05522|MycP1     PDSNFADRPWASDYLRLTEAHKFATGAGITVAVIDTGVNGSPRVP-AEPGGDFVDAAGNG 111
Rv3883c|MycP1       PGSGFHDPPWSNTYLGVADAHKFATGAGVTVAVIDTGVDASPRVP-AEPGGDFVDQAGNG 114
Rv0291|MycP3        PGTDPGVPTPSQTMLNLPAAWQFSRGEGQLVAIIDTGVQPGPRLPNVDAGGDFVEST-DG 119
Msmeg_0624|MycP3    PGTDPGAVSPNQLALNLSGAWQHSRGAGQTVAVIDTGVQPGPRLPNVEAGGDYIEST-DG 119
                    *.:.    .  .  * :  * :.: * *  **:*****: .**:* .:.***::: : :*

Msmeg_0083|MycP1    MSDCDAHGTMTAAIIGGRPSPTDGFVGMAPDVRLLSLRQTSVAFQPKGARQDPNDPNTTQ 176
KEK_05522|MycP1     MSDCDAHGTLTASVIAGRGAPTDGFIGVAPDARILSLRHTSAAFQPVGARTDPNNPNTTQ 171
Rv3883c|MycP1       LSDCDAHGTLTASIIAGRPAPTDGFVGVAPDARLLSLRQTSEAFEPVGSQANPNDPNATP 174
Rv0291|MycP3        LTDCDGHGTLVAGIVAGQPGN-DGFSGVAPAARLLSIRAMSTKFSP---RTSGGDPQLAQ 175
Msmeg_0624|MycP3    LTDCDGHGTSVAGLIAGQPGP-DGFSGVAPEARLISIRQNSPRFAP---RTPGADSEATR 175
                    ::***.*** .*.::.*: .  *** *:** .*::*:*  *  * *   :    :.: :

Msmeg_0083|MycP1    TAGSIRSLARSVVHAANLGAQVINISEAACYKVTRRIDETSLGAAINYAVNVKGAVIVVA 236
KEK_05522|MycP1     TAGSLRSLARAIVHAANLGAQVINISEAACYKVTRPIDETGVGAAVNYAVHVKNAVVIAA 231
Rv3883c|MycP1       AAGSIRSLARAVVHAANLGVGVINISEAACYKVSRPIDETSLGASIDYAVNVKGVVVVVA 234
Rv0291|MycP3        ATLDVAVLAGAIVHAADLGAKVINVSTITCLPADRMVDQAALGAAIRYAAVDKDAVIVAA 235
Msmeg_0624|MycP3    AASDAETLARAVVRAADMGARVINISLVTCLPADRTIDQSVLGAALRYAALEKDAVIVAA 235
                    :: .   ** ::*:**::*. ***:*  :*  . * :*:: :**:: **.  *..*::.*

Msmeg_0083|MycP1    AGN------TGQDCSQNPPPDPSVPSDPRGWREVQTIVSPAWYDPLVLTVGSIGQNGQPS 290
KEK_05522|MycP1     AGN------TGQDCTQNPPPDPAVPSDPRGWQQVQTIVSPAWYSPLVLTVGGIGPTGQPS 285
Rv3883c|MycP1       AGN------TGGDCVQNPAPDPSTPGDPRGWNNVQTVVTPAWYAPLVLSVGGIGQTGMPS 288
Rv0291|MycP3        AGNTGASGSVSASCDSNPLTDLSRPDDPRNWAGVTSVSIPSWWQPYVLSVASLTSAGQPS 295
Msmeg_0624|MycP3    AGNNRGGVSTGAACESNPLPSGT-PGDPRNWNGVTSVSIPSWWQPYVLSVGAVDSTGQPS 294
                    ***      ..  * .** .. : *.***.*  * ::  *:*: * **:*..:   * **

Msmeg_0083|MycP1    NFSMSGPWVGAAAPGENLTSLGYD--GQPVNATPG-EDGPVPLNGTSFSAAYVSGLAALV 347
KEK_05522|MycP1     NFSMSGPWVGAAAPAENITALGYG--GEPVNALQG-QDGLVPVAGTSFAAAYVSGLAALI 342
Rv3883c|MycP1       SFSMHGPWVDVAAPAENIVALGDT--GEPVNALQG-REGPVPIAGTSFAAAYVSGLAALL 345
Rv0291|MycP3        KFSMPGPWVGIAAPGENIASVSNSGDGALANGLPDAHQKLVALSGTSYAAGYVSGVAALV 355
Msmeg_0624|MycP3    SFTMAGPWVGIAAPGENIVSVSNAPDGGLSNALPSERDRLVPLTGTSYAAAYVSGVAALV 354
                    .*:* ****. ***.**:.::.    *   *.  . .:  *.: ***::*.****:***:

Msmeg_0083|MycP1    KQRFPDLTPAQIINRITATARHPGGGVDNYVGAGVIDPVAALTWEIPDGPEKAPFRV--K 405
KEK_05522|MycP1     RQRYPDLTPAQVINRITATARHPGGGVDNYVGAGVIDPVAALTWDVPEGPETAPYRA--K 400
Rv3883c|MycP1       RQRFPDLTPAQIIHRITATARHPGGGVDDLVGAGVIDAVAALTWDIPPGPASAPYNV--R 403
Rv0291|MycP3        RSRYPGLNATEVVRRLTATAHRGARESSNIVGAGNLDAVAALTWQLPAEPGGG-AAP-AK 413
Msmeg_0624|Mycp3    RSKFPDLTARQVVHRLTTTAQGAARSPSNLIGAGMVDPVAALTWDVADVPLDGPAAPEGR 414
                    :.::*.*.. :::.*:*:**:  .   .: :*** :*.******::.  *  .      :

Msmeg_0083|MycP1    EVPPPVYIPPPDRGPITAVVIAGATLAFALGIGALARRALRRKQ---- 449
KEK_05522|MycP1     EIPEPEFIPPPDRGPITWVVVSSAAVALALGIGALTRRALRRR----- 443
Rv3883c|MycP1       RLPPPVVEPGPDRRPITAVALVAVGLTLALGLGALARRALSRR----- 446
Rv0291|MycP3        PVADPPVPAPKDTTPRNVAFAGAAALSVLVGLTAATVAIARRRREPTE 461
Msmeg_0624|MycP3    PIAAPPEPEPRDNTPRIIAFVGTGVLAAAVG--AAFYATYRRKDKTT- 459
                     :. *      *  *   .      ::  :*  *       *:
```

Figure 5.15. The global sequence alignment between $MycP1_{sm}$ (Msmeg), $MycP1_{th}$ (KEK), $MycP1_{tb}$ (Rv), $MycP_{3tb}$, and $MycP_{3sm}$. The rectangles indicate the conserved cysteine residues.

Table 5.2. Sequence identity between mycosin orthologues from *M. tuberculosis*, *M. smegmatis* and *M. thermoresistibile*

SeqA Name	Length	SeqB Name	Length	Alignment Score*
Msmeg_0083|$MycP_1$	449	M.therm KEK_05522|$MycP_1$	443	78.78
M. tuberculosis Rv3883c|$MycP_1$	446	Msmeg_0083|$MycP_1$	449	71.97
M. tuberculosis Rv3883c|$MycP_1$	446	M. therm KEK_05522|$MycP_1$	443	71.11
M. tuberculosis Rv0291|$MycP_3$	461	Msmeg_0624|$MycP_3$	459	59.04
Msmeg_0624|$MycP_3$	459	M. therm KEK_05522|$MycP_1$	443	44.02
M. tuberculosis Rv3883c|$MycP_1$	446	Msmeg_0624|$MycP_3$	459	43.95
M. tuberculosis Rv0291|$MycP_3$	461	M. therm KEK_05522|$MycP_1$	443	43.79
Msmeg_0624|$MycP_3$	459	Msmeg_0083|$MycP_1$	449	43.21
M. tuberculosis Rv0291|$MycP_3$	461	Msmeg_0083|$MycP_1$	449	42.76
M. tuberculosis Rv0291|$MycP_3$	461	*M. tuberculosis* Rv3883c|$MycP_1$	446	42.38

*Alignment score is defined as the number of identities between the two sequences divided by the length of the alignment (Larkin et al., 2007)

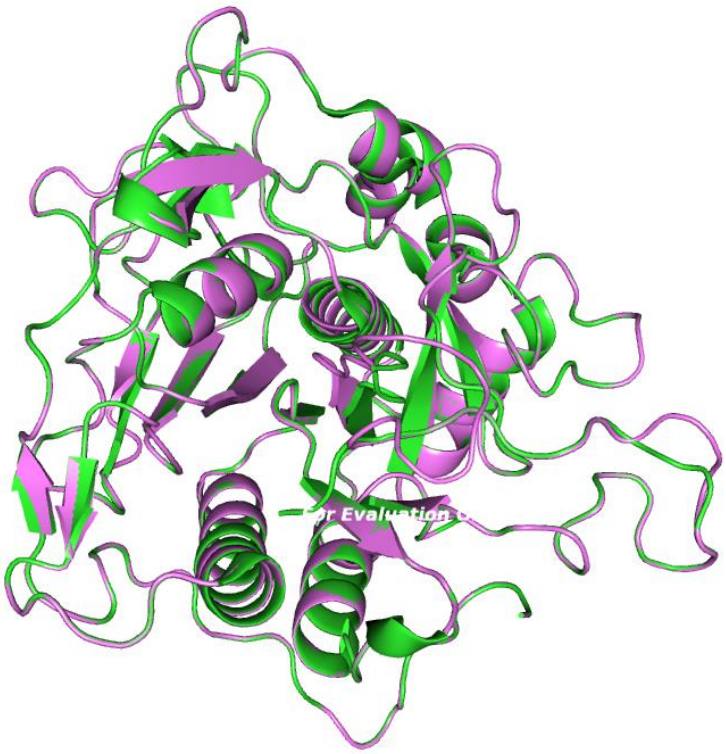


Figure 5.16. The overlay model of $MycP_{3tb}$ structure and $MycP_{3sm}$ structure. Few distinct differences can be observed, however, the model has a low confidence with a QMEAN Z-score of -1.45 (Benkert *et al*., 2011b).

From the results of the *in silico* analysis on $MycP_{3tb}$, it was possible to construct an *E. coli* expression strain that is capable of producing soluble and stable $MycP_{3tb}$ and its orthologues $MycP_{3sm}$, $MycP_{1sm}$ and $MycP_{1th}$. However, some unknown mechanism inhibited the production of $MycP_{3tb}$ but not the mycosins from environmental species. Over-expression of *M. tuberculosis* $MycP_3$ in *M. smegmatis* inhibited the growth of the expression host. Cell-free $MycP_3$ expression was also unsuccessful because some *E. coli* translational machinery components (translation initiation factor IF-2 and elongation factor Tu1 and Tu2) were degraded by unidentified protease activity. The cleavage sites on these two proteins were not identified due to the difficulty in isolating cleaved products in the reaction milieu. The results from all the expression experiments suggest that $MycP_3$ exerted a strongly toxic effect on the expression hosts, including the mycobacterial host, in a way that it might harm the translation processes of the host cells. Perhaps, in the *M. tuberculosis* cytoplasmic environment, $MycP_3$ remains inactive via an unknown inhibitory mechanism until it is transported out of the cell. To test such a hypothesis, one can delete the signal sequence M^1-A^{25} and find out whether such a mutant $MycP_3$ may inhibit the growth of *M. tuberculosis*; or abolish the function of $MycP_3$ via single point mutation S342A and produce inactive $MycP_3$ mutant *in vitro* and *in vivo*. This would enable one to see whether stable and soluble $MycP_3$ can be produced. These experiments are the subject of future investigations.

ESX or T7S systems are widely found in both pathogenic and environmental mycobacterial species. The function of *M. tuberculosis* mycosins possibly no longer has the same substrate specificity as its orthologues in environmental mycobacterium species, considering ESX-1 in *M. tuberculosis* is associated with virulence, while in *M. smegmatis* it is essential for DNA transfer (Coros *et al*., 2008; Flint *et al*., 2004). They probably have evolved to possess virulence towards host cells and even to themselves if not transported correctly out of the cell. To differentiate enzyme specificity between mycosins is vital to further understand the functions of different ESXs. Furthermore, the mycosins have evolved distantly from other substilisins and they serve as attractive drug target candidates due to the fact that $MycP_1$ seems to regulate the secretion, and process the secreted antigens, from ESX-1, and $MycP_3$ is involved with the iron acquisition pathway in mycobacteria. These are both essential processes for infection with the organism. Disabling $MycP_1$ will attenuate the bacteria and abolishing the function of $MycP_3$ may disrupt the iron acquisition thereby inhibiting the cell growth. The use of mycosins as drug targets will be the subject of future studies.

5.5 Reference

Abdallah, A.M., Gey van Pittius, N.C., Champion, P.A.D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W., 2007. Type VII secretion--mycobacteria show the way. Nat. Rev. Microbiol. 5, 883–891. doi:10.1038/nrmicro1773

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinforma. Oxf. Engl. 22, 195–201. doi:10.1093/bioinformatics/bti770

Ashbridge, K.R., Booth, R.J., Watson, J.D., Lathigra, R.B., 1989. Nucleotide sequence of the 19 kDa antigen gene from *Mycobacterium tuberculosis*. Nucleic Acids Res. 17, 1249.

Benkert, P., Biasini, M., Schwede, T., 2011a. Toward the estimation of the absolute quality of individual protein structure models. Bioinforma. Oxf. Engl. 27, 343–350. doi:10.1093/bioinformatics/btq662

Benkert, P., Biasini, M., Schwede, T., 2011b. Toward the estimation of the absolute quality of individual protein structure models. Bioinforma. Oxf. Engl. 27, 343–350. doi:10.1093/bioinformatics/btq662

Brown, G.D., Dave, J.A., Gey van Pittius, N.C., Stevens, L., Ehlers, M.R., Beyers, A.D., 2000. The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. Gene 254, 147–155.

Coros, A., Callahan, B., Battaglioli, E., Derbyshire, K.M., 2008. The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. Mol. Microbiol. 69, 794–808. doi:10.1111/j.1365-2958.2008.06299.x

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R., Kaufmann, S.H.E., 2003. The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. Microbes Infect. Inst. Pasteur 5, 1082–1095.

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W., Schnappinger, D., 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res. 33, e21. doi:10.1093/nar/gni013

Finlay, B.B., Falkow, S., 1997. Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. MMBR 61, 136–169.

Flint, J.L., Kowalski, J.C., Karnati, P.K., Derbyshire, K.M., 2004. The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. Proc. Natl. Acad. Sci. U. S. A. 101, 12598–12603. doi:10.1073/pnas.0404892101

Frasinyuk, M.S., Kwiatkowski, S., Wagner, J.M., Evans, T.J., Reed, R.W., Korotkov, K.V., Watt, D.S., 2014. Pentapeptide boronic acid inhibitors of *Mycobacterium tuberculosis* MycP1 protease. Bioorg. Med. Chem. Lett. doi:10.1016/j.bmcl.2014.05.056

Gey Van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. 2, RESEARCH0044.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., van Helden, P.D., Warren, R.M., 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol. Biol. 6, 95. doi:10.1186/1471-2148-6-95

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 7, e1002251. doi:10.1371/journal.ppat.1002251

Hayashi, K., Kojima, C., 2008. pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. Protein Expr. Purif. 62, 120–127. doi:10.1016/j.pep.2008.07.007

Houben, E.N.G., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jiménez, C.R., Ottenhoff, T.H.M., Luirink, J., Bitter, W., 2012. Composition of the type VII secretion system membrane complex. Mol. Microbiol. 86, 472–484. doi:10.1111/j.1365-2958.2012.08206.x

Houben, E.N.G., Korotkov, K.V., Bitter, W., 2013a. Take five - Type VII secretion systems of Mycobacteria. Biochim. Biophys. Acta. doi:10.1016/j.bbamcr.2013.11.003

Houben, E.N.G., Korotkov, K.V., Bitter, W., 2013b. Take five - Type VII secretion systems of Mycobacteria. Biochim. Biophys. Acta. doi:10.1016/j.bbamcr.2013.11.003

Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 35, W460–464. doi:10.1093/nar/gkm363

Kim, J., Robinson, A.S., 2006. Dissociation of intermolecular disulfide bonds in P22 tailspike protein intermediates in the presence of SDS. Protein Sci. Publ. Protein Soc. 15, 1791–1793. doi:10.1110/ps.062197206

Laemmli, U. K., 1970, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227: 680-684

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinforma. Oxf. Engl. 23, 2947–2948. doi:10.1093/bioinformatics/btm404

Mailaender, C., Reiling, N., Engelhardt, H., Bossmann, S., Ehlers, S., Niederweis, M., 2004. The MspA porin promotes growth and increases antibiotic susceptibility of both *Mycobacterium bovis* BCG and Mycobacterium tuberculosis. Microbiol. Read. Engl. 150, 853–864.

Ohol, Y.M., Goetz, D.H., Chan, K., Shiloh, M.U., Craik, C.S., Cox, J.S., 2010. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell Host Microbe 7, 210–220. doi:10.1016/j.chom.2010.02.006

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

Scapoli, C., Bartolomei, E., De Lorenzi, S., Carrieri, A., Salvatorelli, G., Rodriguez-Larralde, A., Barrai, I., 2009. Codon and aminoacid usage patterns in mycobacteria. J. Mol. Microbiol. Biotechnol. 17, 53–60. doi:10.1159/000195674

Schönfeld, H.J., Schmidt, D., Schröder, H., Bukau, B., 1995. The DnaK chaperone system of *Escherichia coli*: quaternary structures and interactions of the DnaK and GrpE components. J. Biol. Chem. 270, 2183–2189.

Siegrist, M.S., Steigedal, M., Ahmad, R., Mehra, A., Dragset, M.S., Schuster, B.M., Philips, J.A., Carr, S.A., Rubin, E.J., 2014. Mycobacterial esx-3 requires multiple components for iron acquisition. mBio 5. doi:10.1128/mBio.01073-14

Siegrist, M.S., Unnikrishnan, M., McConnell, M.J., Borowsky, M., Cheng, T.-Y., Siddiqi, N., Fortune, S.M., Moody, D.B., Rubin, E.J., 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc. Natl. Acad. Sci. U. S. A. 106, 18792–18797. doi:10.1073/pnas.0900589106

Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., Strynadka, N.C.J., 2013. Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. J. Biol. Chem. 288, 17782–17790. doi:10.1074/jbc.M113.462036

Wagner, J.M., Evans, T.J., Chen, J., Zhu, H., Houben, E.N.G., Bitter, W., Korotkov, K.V., 2013. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. J. Struct. Biol. 184, 115–128. doi:10.1016/j.jsb.2013.09.022

Chapter 6. Functional Study of Mycosin-3, a proteomics approach

6.1 Introduction

Mycobacterium tuberculosis (*M. tuberculosis*) has acquired various combinations of resistance to all the drugs to which it has been exposed to during treatment, leading to mono-, multi- and even extensively drug resistant strains of *M. tuberculosis* circulating worldwide. This necessitates the continuous discovery of novel drug targets to combat tuberculosis. The iron acquisition pathway in *M. tuberculosis* has several attractive drug targets (Owens *et al*., 2013). *M. tuberculosis* cannot survive without iron as it is required for numerous enzymes responsible for many cellular processes such as electron transport, amino acid and pyrimidine biosynthesis and the tricarboxylic acid cycle. Although free iron is scarce in the host, *M. tuberculosis* has evolved effective iron acquisition machineries to survive inside the host. There are two sources of iron in the vertebrate hosts, the non-heme iron which is mostly bound in the iron transport proteins, transferrin and lactotransferrin, and the iron storage protein, ferritin, and the heme iron which constitutes the prosthetic group in hemoglobin inside erythrocytes circulating in the bloodstream. *M. tuberculosis* has access to both iron sources while residing in macrophages. Transferrin can be engulfed into macrophage endosomes which fuse with *M. tuberculosis*-containing phagosomes (Clemens and Horwitz, 1996), and macrophages can degrade red blood cells releasing heme intracellularly (Frederick *et al*., 2009; Kontoghiorghes and Weinberg, 1995). Additionally, hemoglobin may be degraded by hemolysins secreted by *M. tuberculosis* in the bloodstream (Wren *et al*., 1998).

The acquisition systems for the two iron sources are distinct. The non-heme iron is imported via a siderophore-mediated pathway. Siderophores are small iron chelators with high affinity for ferric iron. There are two types of siderophores which are distinguished by their core structures, biosynthetic pathways and transport machineries. Exochelin is a secreted, peptide-based, water soluble molecule which is only produced by environmental mycobacteria. The genes responsible for exochelin assembly include *fxbA*, *fxbB* and *fxbC* (Fiss *et al*., 1994; Yu *et al*., 1998). The exochelin importers are FxuA, FxuB, FxuC and FuxD. One exochelin exporter, ExiT, has been identified (Zhu *et al*., 1998). The other siderophores are the salicylate-derived, chloroform extractable mycobactin and carboxymycobactin which are produced by most mycobacteria (Barclay and Ratledge, 1988; Messenger *et al*., 1986). The lipophilic mycobactin is cell envelope associated whereas the hydrophilic carboxymycobactin is secreted into the extracellular space. The two are synthesized via the same biosynthetic pathway encoded by two gene clusters, *mbt*A-J and *mbt*K-N (Krithika *et al*., 2006; Quadri *et al*., 1998). Carboxymycobactin chelates iron from transferrin and is taken up by the Msp family porin in the mycobacterial outer membrane (Jones and Niederweis, 2010). In the cell envelope, ferri-carboxymycobactin either transfers the iron to mycobactin or delivers it to the iron importer complex IrtA-B (Rodriguez and Smith, 2006). The ferri-carboxymycobactin

releases the iron upon reduction by an unknown reductase (Ryndak *et al*., 2010). The deferri-carboxymycobactin is recycled and exported by the membrane complex MmpL5/MmpS5 and MmpL4/MmpS4 to acquire more iron (Wells *et al*., 2013) or prevent siderophore-related self-poisoning (Jones *et al*., 2014).

Heme is acquired from hemoglobin by a mycobacterial hemophore Rv0203 (Tullius *et al*., 2011). Rv0203 carries heme to the heme importer protein MmpL11 or MmpL3 on the plasma menbrane. The imported heme is either degraded by MhuD to release iron (Chim *et al*., 2010) or utilized for the biosynthesis of heme containing proteins such as KatG, catalase-peroxidase, and DosT/S, the redox sensing two component system (Svistunenko, 2005).

The expression of the genes encoding the mycobactin biosynthesis and transport machineries is regulated by the global iron-dependent regulator, IdeR. When the iron level is high, iron binds to IdeR which in turn binds to its target promoter regions and suppresses transcription (Rodriguez *et al*., 2002). The regulation of heme uptake has not been elucidated, however, ESX-3, one of the type VII secretion systems, seems to influence both siderophore and hemophore-mediated iron acquisition in mycobacteria (Serafini *et al*., 2013, 2009; Siegrist *et al*., 2014, 2009). In *M. tuberculosis*, ESX-3 is essential for growth in iron deplete 7H9 media (Griffin *et al*., 2011; Sassetti *et al*., 2003). The gene expression profile of the *M. tuberculosis* ESX-3 conditional mutant resembles that of *M. tuberculosis* during iron starvation and this mutant cannot be rescued by heme supplementation (Serafini *et al*., 2013). Both the non-heme and heme acquisition systems have been well characterized, indicating that ESX-3 is unlikely to be directly involved in iron uptake. ESX-3 may, however, modify and secrete unidentified factors which facilitate iron and heme acquisition, assist in the post-translational modification of certain membrane proteins to enable siderophore and hemophore export and import, or increase the permeability of the membrane to iron. ESX-3 membrane core structure is composed of 5 proteins, $EccB_3$, $EccC_3$, $EccD_3$, $EccE_3$ and mycosin-3 ($MycP_3$), where $EccB_3$, $EccD_3$ and $EccE_3$ are proposed to form the transmembrane pore of the secretion system on the plasma membrane, $EccC_3$ contains an ATPase domain providing energy for the active export process, and $MycP_3$ has a serine protease active site (Houben *et al*., 2012).

$MycP_3$ is the only protease in the ESX-3 system. Although its function is not well characterized (Abdallah *et al*., 2007; Houben *et al*., 2012; Wagner *et al*., 2013), it may play a role in modifying the secreted proteins and other membrane proteins. In this study, a *Mycobacterium smegmatis* $MycP_3$ knock out (KO) mutant ($\Delta MycP_{3ms}$) was generated and used to determine the impact of the absence or presence of $MycP_3$ on the proteome of *M. smegmatis*, under high iron conditions *in vitro* where $MycP_3$ is probably at its basal level. The proteins with significantly different abundances were further investigated by RT-qPCR. The $MycP_3$ KO mutant was complemented with $MycP_3$ expressed episomally from both ESX-3 promoters. The intracellular iron levels of *M. smegmatis* WT, $MycP_3$ KO and complemented strains were assayed under high iron and low iron conditions, and after iron rescue.

6.2 Materials and Methods

6.2.1 Bacterial Strains, Culture media and Plasmid DNA

Escherichia coli XL-1 blue (Promega, USA) was used for electro-competent cell preparation, manipulating and generating plasmid DNA constructs. *Mycobacterium smegmatis* mc^2 155 was used as the wildtype (WT) parent strain from which the $MycP_3$ mutant was derived. *E. coli* was cultured in either Lysogeny Broth (LB) liquid media (1% tryptone [Merck, Germany], 0.5% yeast extract [Merck, Germany], 1% sodium chloride [Sigma, Germany], or LB solid media (LB supplemented with 1.5 % bacterial agar [Merck, Germany]). *M. smegmatis* was cultured in either 7H9 liquid media (ammonium sulphate, 0.5 g/l; *L*-glutamic acid, 0.5 g/l; sodium citrate, 0.1 g/l; pyridoxine, 0.001 g/l; biotin, 0.0005 g/l; disodium phosphate, 2.5 g/l; monopotassium phosphate, 1 g/l; ferric ammonium citrate, 0.04 g/l; magnesium sulphate, 0.05 g/l; calcium chloride, 0.0005 g/l; zinc sulphate, 0.0001 g/l; copper sulphate, 0.001 g/l; pH 6.6 ± 0.2 at 25°C, BD Biosciences, Germany), or Difco 7H11 solid media (enzymatic digest of casein, 1 g/l; disodium phosphate, 1.5 g/l; monopotassium phosphate, 1.5 g/l; ammonium sulphate, 0.5 g/l; monosodium glutamate, 0.5 g/l; sodium citrate, 0.4 g/l; ferric ammonium citrate, 0.04 g/l; magnesium sulphate, 0.05 g/l; copper sulphate, 0.001 g/l; pyridoxine, 0.001 g/l; zinc sulphate, 0.001 g/l, biotin, 0.0005 g/l; malachite green, 0.00025 g/l; agar, 13.5 g/l, pH 6.6 ± 0.2 at 25°C, BD Biosciences, Germany), both supplemented with 0.05% Tween 80 (Sigma, Germany), 0.5% glucose (Kimix, South Africa), and 0.5% glycerol (Sigma, Germany).

M. smegmatis growth was monitored in 7H9 liquid media and Fe-free Sauton's media (3.5 mM KH_2PO_4 [Merck, Germany], 25 mM L-asparagine [Merck, Germany], 10 mM citric acid [Merck, Germany], 4 mM $MgSO_4 \cdot 6H_2O$ [Merck, Germany], 5% glycerol (v/v), and 0.05% Tween-80 [Sigma, Germany]). The strains were cultured in 7H9 for gene expression analysis. For intracellular iron quantitation, *M. smegmatis* was grown in (i) 7H9, (ii) iron-deprived 7H9 from which ferric ammonium citrate was omitted, (iii) extremely low iron 7H9 medium in which the iron-deprived 7H9 was treated with 1% Chelex® 100 resin (Bio-Rad, USA) at room temperature for 48 hours and filter-sterilized and all divalent cation salts except ferric ammonium citrate were added after filter sterilisation, or (iv) iron-rescued 7H9 which was initially iron-deprived 7H9 but ferric ammonium citrate was added to the level as per 7H9 medium. Tween-80, glucose and glycerol were supplemented to all types of 7H9 media before use (7H9 medium mentioned in this chapter refers to the 7H9 with the Tween-80, glucose and glyercol supplementation).

The CloneJet1.2 vector (Fermentas, USA) was used for target DNA replication and preservation before cloning into the target vectors. The p2Nil suicide vector (Parish and Stoker, 2000) (Figure S6.1) and the pGoal17 selection gene cassette (Figure S6.2) were used to generate $\Delta MycP_{3ms}$, and pMV306 (Stover *et al.*, 1991) was used for $MycP_3$ complementation (Figure S6.3).

6.2.2 Generation of *M. smegmatis* $MycP_3$ gene KO strain

The principle and the method of generating a gene KO mycobacterial strain using homologous recombination was described previously (Parish and Stoker, 2000). One thousand basepair (bp) fragments upstream (UP1000) and downstream (DOWN1000) of $MycP_3$ (MSMEG_0624) were amplified by Polymerase Chain Reaction (PCR) with Phusion® DNA polymerase (Fermentas, USA) using two pairs of primers (Table 6.1). The UP1000 and DOWN1000 PCR fragments were cloned into pJet1.2 vector according to the manufacturer's instructions. The UP1000 and DOWN1000 inserts were digested out of the recombinant DNA using the appropriate restriction enzymes. Both inserts were ligated into p2Nil via three-way cloning using T4 DNA ligase (Promega, USA).

The selection gene cassette (P_{Ag85}-*lac*Z P_{hsp60}-*sac*B) from pGOAL17 was inserted into the recombinant p2Nil-UP1000-DOWN1000 plasmid at the PacI restriction site. The recombinant p2Nil-UP1000-DOWN1000-*lac*Z-*sac*B construct was electroporated into electro-competent *M. smegmatis* cells and blue single-crossover colonies were selected on LB agar supplemented with 50 µg/ml kanamycin and 0.2% X-gal (Roche, Switzerland). The colonies were picked and passaged in LB media in the absence of kanamycin to induce a second crossover event. Double crossover colonies were selected on LB agar supplemented with 5% sucrose and X-gal. White colonies were further screened by colony PCR using screening primers (Table 6.1) to distinguish between WT and $\Delta MycP_{3ms}$ strains. The WT PCR product was approximately 1600 bp while that of the $\Delta MycP_{3ms}$ was approximately 200 bp.

6.2.3 Proteome Analysis

M. smegmatis WT and $\Delta MycP_{3ms}$ strains were cultured in 100 ml 7H9 medium with supplementation to OD_{600nm} of 0.7~0.9 (mid-log phase) and harvested by centrifugation at 3000 xg at 4°C for 10 min in a desktop swing-bucket centrifuge (Eppendorf, USA). The cell pellets were washed twice with cold Tris-HCl buffer (5 mM tris, pH 7.6, 0.005% Tween 80) with protease inhibitors and 20 units per milliliter DNase (Roche, Germany). The pellets were mixed with an equal volume of 0.1 mm diameter glass beads (Biospec, USA) and resuspended in 500 µl sterile cold Tris-HCl buffer. The mixture was ribolyzed using a FastPrep®-24 ribolyzer (MP Biomedicals, USA) at 6.0 m/s for 30 s repeated 3 times with 30 s incubation on ice between each repeat. The whole cell lysate was cleared up by centrifugation at 12000 xg at 4°C for 30 min. The supernatant was filtered through a 0.22 micron filter device (Millipore, US). The protein concentrations of the samples were assayed using the RC-DC protein assay kit (Bio-Rad, USA). Equal amounts of the WT and $\Delta MycP_{3ms}$ protein extracts (60 µg) were loaded in triplicate onto a Criterion™ XT 4-12% Precast SDS-PAGE GEL (Bio-Rad, USA) and electrophoretically fractionated. The in-gel proteins were treated and cleaned up as previously described (de Souza *et al*., 2010; Rappsilber *et al*., 2007). Mass spectrometry was done at the Proteomics Unit of the Central Analytical Facility, Stellenbosch University using the Thermofisher LTQ™ Orbitrap Velos mass spectrometer (Thermo

Scientific, Bremen, Germany). The identified peptides were used to identify the proteins using an automated database searching software (Mascot, Matrix Science, London, UK, and Sequest) against the *M. smegmatis* protein database in TBDB (Tuberculosis Database) (Reddy *et al*., 2009). The proteins were considered positively identified when they were identified with at least two unique tryptic peptide per protein. Proteins were searched using the Adromeda function from Maxquant 1.2.2.5. The relative protein abundance was determined as previously described (Vogel and Marcotte, 2008) and compared between the WT and $\Delta MycP_{3ms}$ strains.

Table 6.1. All the primers used for ΔMycP$_{3ms}$ generation, MycP$_3$ complementation and RT-qPCR assay.

Experiment	Gene Related	Primer Name	Primer Sequence	Restriction Site (underlined)
ΔMycP$_{3ms}$ Generation	*mycP$_3$*	UP1000 forward	5'-AAGCTTTCCCACGCACATCG-3'	*HindIII*
		UP1000 reverse	5'-CTCG AGATCACCTGTCGAGCACG-3'	*XhoI*
		DOWN1000 forward	5'-CTCGAGATGACCGCCCGGATAGC-3	*XhoI*
		DOWN1000 reverse	5'-GGATCCCCGGTCTCGGTGAC-3'	*BamHI*
ΔMycP$_{3ms}$ Construct verification	*mycP$_3$*	*mycP$_3$* screening forward	5'-GCTCAACCCGAAGATC GCCTC-3'	N/A
		mycP$_3$ screening reverse	5'-AGGAACATGCCTTTCCACCAGG-3'	N/A
MycP$_3$ Complementa-tion	*mycP$_3$*	pr1 forward	5'-CCATGGGACGCTGAACGAGTGTTTAC-3'	*NcoI*
		pr1 reverse	5'-GACGCCCAGACTCTTGTGGATCACATCGCGGTCGACCCGGGGCG-3'	N/A
		mycP$_3$-pr1 forward	5'-CGCCCCGGGTCGACCGCGATGTGATCCACAAGAGTCTGGGCGTC-3'	N/A
		mycP$_3$ reverse	5'-AAGCTTTCATGTGGTCTTGTCCTTCC-3'	*HindIII*
		pr2 forward	5'-CCATGGACGTGGGACGGCGACGA GAATC-3'	*NcoI*
		pr2 reverse	5'-GACGCCCAGACTCTTGTGGATCACGACTGTTTCC TTTCGAAGGTGGTG-3'	N/A
		mycP$_3$-pr2 forward	5'-CACCACCTTCGAAAGGAAAC AGTCGTGATCCACAAGAGTCTGGGCGTC-3'	N/A
RT-qPCR assay	*SigA*	*SigA* forward	5'-GGGCGTGATGTCCATCTGCT-3'	N/A
		SigA reverse	5'-GTATCCCGGTGCATGGTC-3'	

(Table 6.1 continued)

Experiment	Gene Related	Primer Name	Primer Sequence	Restriction Site
RT-qPCR assay	*mycP*$_3$	*mycP*$_3$ forward	5'-GGATCATCGCGTTCGTGGGTAC-3'	N/A
		mycP$_3$ reverse	5'-GTCTTGTCCTTCCGACGGTAGG-3'	
	eccE$_3$	*eccE*$_3$ forward	5'-GAGCCGTTGTTGACGGTTTG-3'	
		eccE$_3$ reverse	5'-GTTCGGTCGACAACGGGTTC-3'	
	eccC$_3$	*eccC*$_3$ forward	5'-ACACAAAGGCGAGGGCTTCC-3'	
		eccC$_3$ reverse	5'-AATCGACTTCTCGGCGCGTG-3'	
	mmpL5	*mmpL5* forward	5'-GATCAAGCTGGCCGCCAAAG-3'	
		mmpL5 reverse	5'-GTTGCCTTCCTGCATGTCCTTG-3'	
	mmpL4	*mmpL4* forward	5'-AAGCGGCCATGGAGAACCAG-3'	
		mmpL4 reverse	5'-CGCGCTTGAAGTCCGGATTC-3'	

6.2.4 MycP$_3$ Complementation

The genomic region of the ESX-3 gene cluster contains two promoters, namely pr1 and pr2 which are upstream of the MSMEG_0615 and the MSMEG_0620 genes, respectively (Figure 6.1) (Maciag *et al*., 2009). Both promoters were used to make complementation constructs separately.

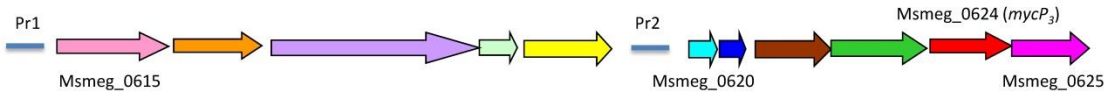


Figure 6.1. Genetic organization of ESAT-6 cluster 3 in *Mycobacterium smegmatis*. The positions of the promoters, Pr1 and Pr2, are indicated (Maciag *et al*., 2009).

The two pairs of primers for making the joined DNA fragments of pr1 and MycP$_3$ in the first complementation construct and the two pairs of primers for making the joined DNA fragments of pr2 and MycP$_3$ in the second complementation construct are given in Table 6.1. The reverse primer sequences for the two promoters contain a region which is complementary to a region of the sense primers for *myc*P$_3$ to facilitate single-joint PCR connecting pr1/pr2 and MycP$_3$. The single-joint PCR method was described previously (Yu *et al*., 2004). The final PCR products, named pr1-mycP$_3$ and pr2-mycP$_3$ were cloned into pMV306 integrative vector. The recombinant pMV306-pr1-*mycP$_3$* and pMV306-pr2-*mycP$_3$* plasmids were electroporated into the *M. smegmatis* ΔMycP$_{3ms}$ mutant strain to generate two MycP$_3$ complementation strains, designated ΔMycP$_{3ms}$::pr1*myc*P$_{3ms}$ and ΔMycP$_{3ms}$::pr2*myc*P$_{3ms}$.

6.2.5 Growth curves under iron-rich and iron-deprived conditions

M. smegmatis WT, ΔMycP$_{3ms}$, ΔMycP$_{3ms}$::pr1MycP$_{3ms}$ and ΔMycP$_{3ms}$::pr2MycP$_{3ms}$ strains were cultured in 7H9 broth and iron-free Sauton's media from a starting OD$_{600nm}$ of 0.01. They were incubated at 37°C with a rotating rate of 200 rpm. The OD$_{600nm}$ reading was taken every 3 hours for 48 hours. The growth curves were generated by plotting OD$_{600nm}$ readings versus time.

6.2.6 Intracellular iron quantitation

The intracellular iron quantitation technique was used as previously described (Riemer *et al*., 2004). Fifty millilitres of cultures at mid-log phase (OD$_{600nm}$ of 0.7~0.9) were harvested by centrifugation. The cell pellet was resuspended in 500 µl 50 mM NaOH and ribolyzed as previously stated. The whole cell lysate was cleared by centrifugation at 12 000 xg at 4°C for 30 min. One hundred microliters of the whole cell lysate was used for the iron quantification assay while another 25 µl was used for protein quantitation using the RC-DC protein assay. The 100 µl whole cell lysate was transferred into one well of a 96-well microtiter plate and

mixed with 100 µl of 10 mM HCl and then 100 µl iron-releasing reagent (a freshly made solution of equal volumes of 1.4 M HCl and 4.5%, $KMnO_4$ [Merck, Germany] in water). This mixture was incubated at 60°C for 2 hours, cooled to room temperature, and 30 µl of iron-detecting agent (6.5 mM ferrozine [Sigma, Germany], 6.5 mM neocuproine [Sigma, Germany], 2.5 M ammonium acetate [Merck, Germany], and 1 M ascorbic acid [Sigma, Germany]) was added to the well and incubated for 30 min at room temperature. The absorbance was read at 550 nm on a photospectrometer. Ferric chloride was used as iron standards at the concentration of 10 µM, 20 µM, 30 µM, 40 µM, 50 µM, 60 µM, 70 µM, and 80 µM, in 50 mM NaOH. The iron concentration was normalized against the protein content in that well, resulting in the unit of nmol (of iron) per mg (of protein).

6.2.7 RT-qPCR

Fifteen millilitres of each *M. smegmatis* strain was harvested at mid-log phase from each culturing condition (7H9, Fe-free 7H9, and Fe rescued 7H9) by centrifugation. The supernatant was discarded and the pellet was resuspended with 1 ml FastRNA® Blue solution (MP Biomedicals, USA) and ribolyzed as previously described. The whole cell lysate was cleared by centrifugation at 12 000 xg at 4°C for 30 min, and 700 µl of the supernatant was transferred into a new 1.5 ml tube and thoroughly mixed with 300 µl chloroform. The mixture was centrifuged at 12 000 xg at 4°C for 10 min. The top aqueous layer was transferred to a new 1.5 ml tube and mixed with 500 µl pre-chilled 100% ethanol (Sigma, Germany). The mixture was transferred onto the RNA purification column from NucleoSpin® RNA isolation kit (Macherey-Nagel, Germany) and further total RNA purification was done according to manufacturer's instructions. The quality of the total RNA was assayed using a Bioanalyzer (Agilent Technologies, USA) at the Central Analytical Facility (Stellenbosch University, South Africa).

Five micrograms of total RNA was treated with Turbo DNase (Ambion, USA) according to the manufacturer's instructions. One microgram of Turbo DNase-treated total RNA was used for cDNA synthesis using the PrimeScript™ 1st strand cDNA synthesis kit (TaKaRa, USA) using the appropriate reverse primers (Table 6.1). Quantitative PCR was conducted using SYBR *Premix Ex Taq*™ (TaKaRa, USA) on the Bio-Rad CFX96™ Real-Time PCR Detection System (Bio-Rad, USA) with the following cycling conditions: initial denaturation at 95°C for 30 sec; 39 cycles of amplification at 95°C for 5 sec followed by 30 sec at 60°C. The melt curve followed with a denaturation step at 95°C for 10 sec and then a melting step from 65°C to 95°C with 5 sec staying at each 0.5°C interval. *SigA* was selected as the reference gene due to its constitutive expression. The expression of all genes of interest was normalized against that of *SigA* in the same RNA sample.

6.2.8 ESX-3 promoter activity in response to iron levels

The two complementation strains were generated using the two ESX-3 promoters (see section 6.2.4). The promoter activity in response to iron levels were indirectly assayed using RT-qPCR on the gene expression levels of $MycP_3$ in WT, $\Delta MycP3_{ms}$, $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ and $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ strains under normal 7H9, iron deprived 7H9 and iron rescued 7H9 media.

6.3 Results

6.3.1 MycP3 gene KO in *M. smegmatis* and genetic complementation

The *Mycobacterium smegmatis* ΔMycP3 strain was generated via homologous recombination. Two MycP3 complementation strains $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ and $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ were obtained by electroporating recombinant pMV306-Pr1-MycP3 and pMV306-Pr2-MycP3 plasmid DNA into the $\Delta MycP3_{ms}$ strain respectively. The WT_{ms}, $\Delta MycP3_{ms}$, $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ and $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ strains were confirmed by colony PCR (Figure 6.2).

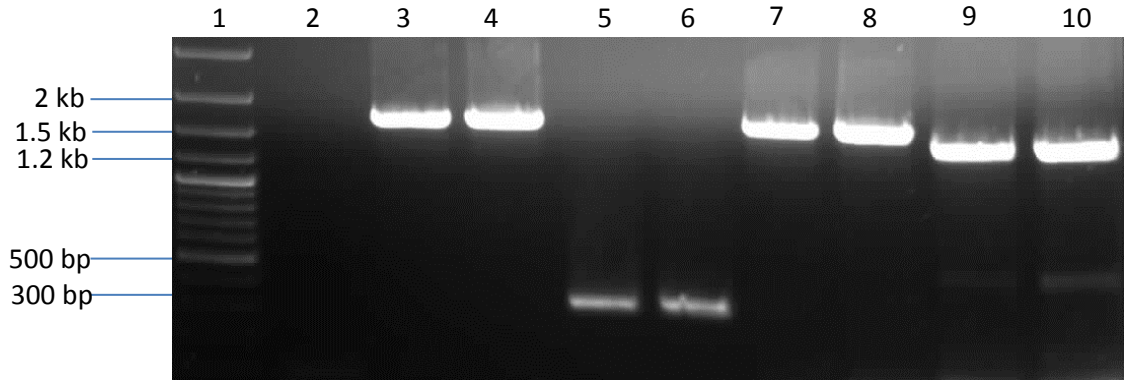
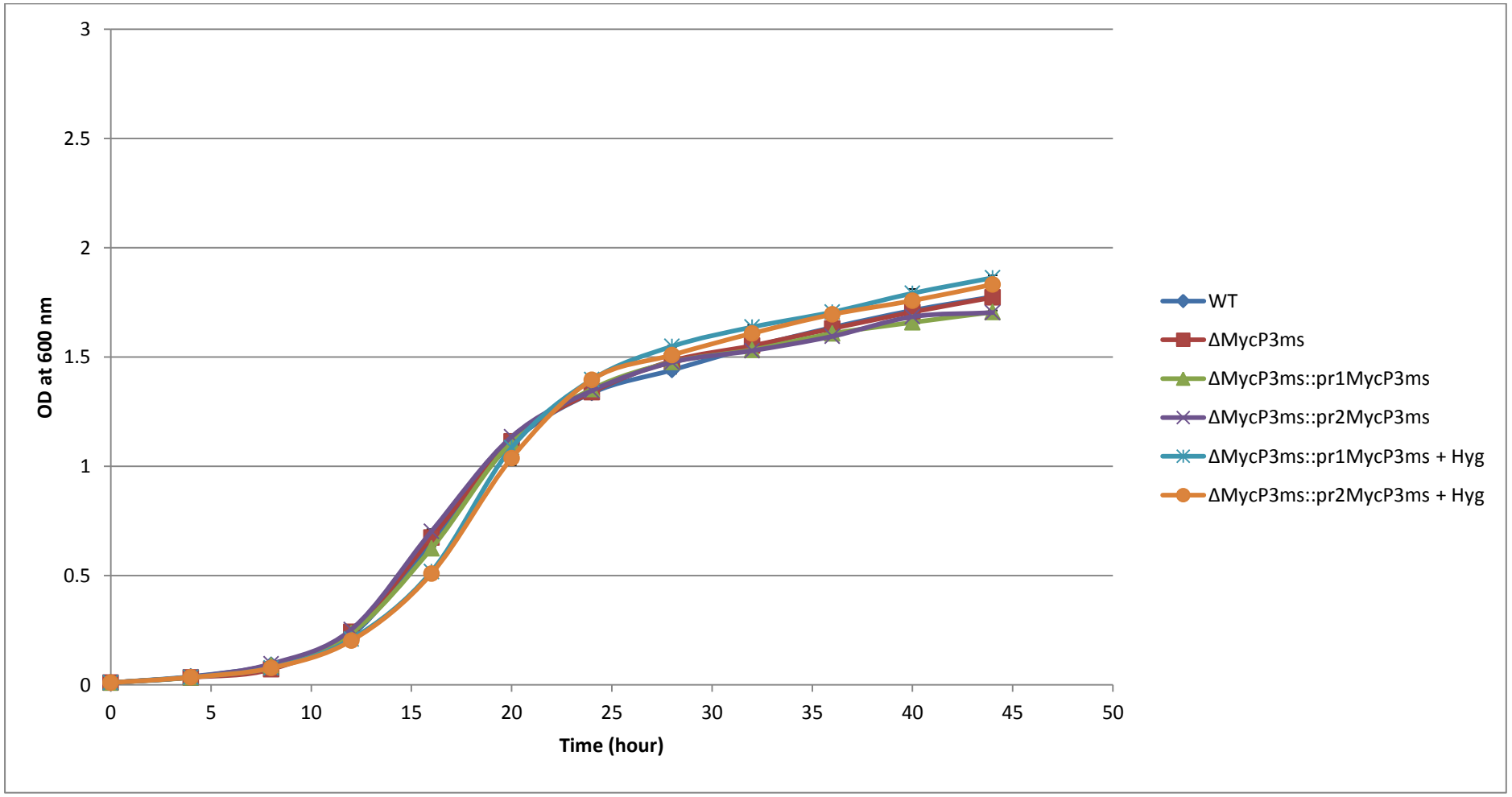


Figure 6.2 Colony PCR to validate the strains where $\Delta MycP_{3ms}$ screening primers (Table 6.1) were used to distinguish between WT and $\Delta MycP_{3ms}$, and complementation strain generating primers were used to distinguish between $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ and $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ (Table 6.1). Lane 1: DNA Marker [1 kb DNA ladder Plus (Fermentas, USA)], 2: no template control, 3 and 4: WT_{ms} (1674 bp), 5 and 6: $\Delta MycP3_{ms}$ (251 bp), Lane 7 and 8, $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ (1684 bp), Lane 9 and 10: $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ (1509 bp).

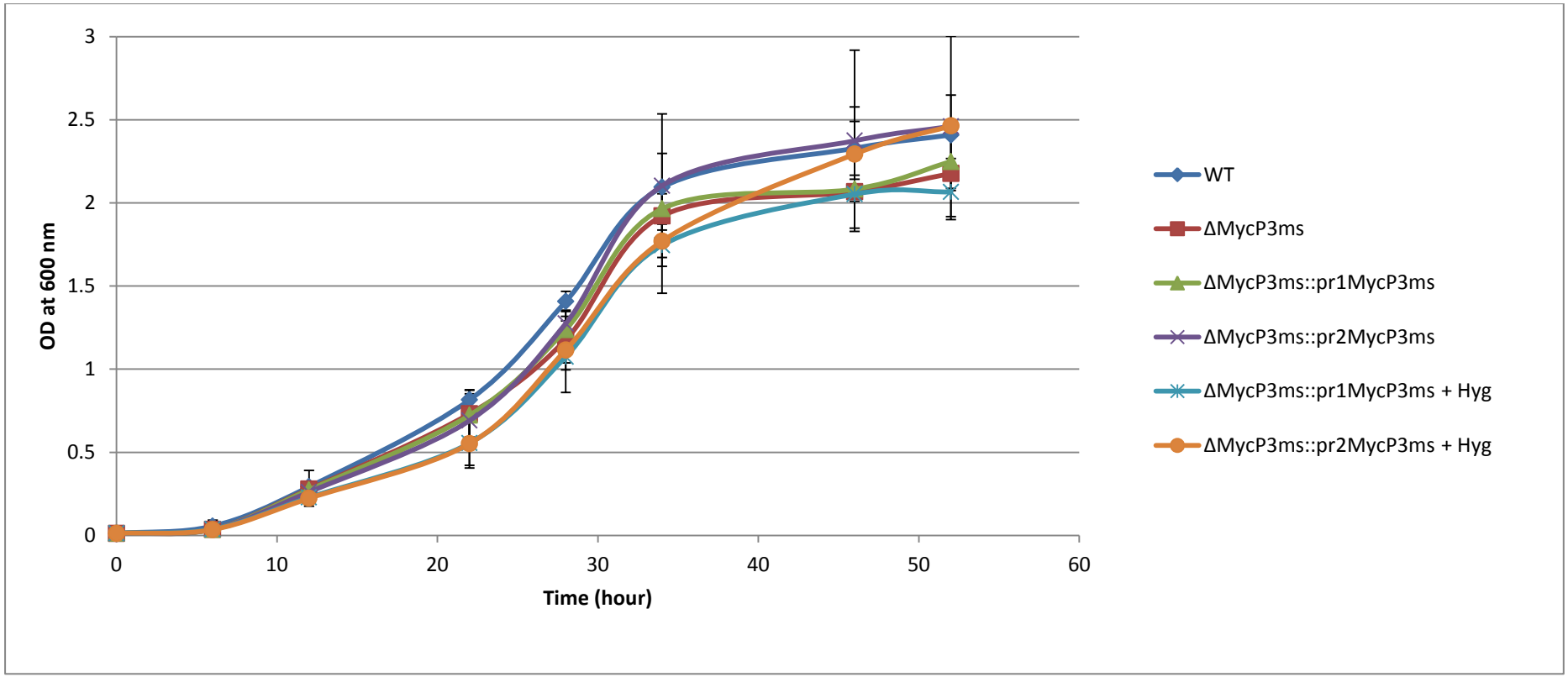
6.3.2 Deletion of MycP3 does not impact the growth of M. *smegmatis*

To determine whether deletion of MycP3 would directly exert a negative effect on the growth of *M. smegmatis*, growth curves were done in 7H9 and Fe-free Sauton's medium (Figure 6.3). *M. smegmatis* WT, $\Delta MycP3_{ms}$, $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$, and $\Delta MycP3_{ms}$::pr2$MycP3_{ms}$ showed sigmoidal growth without significant differences in iron-rich 7H9 and in iron-free Sauton's medium respectively. The addition of hygromycin to the media did not influence the growth of the complementation strains. The standard deviation values for the growth in iron-free Sauton's medium are larger than those in 7H9 medium. In addition, RT-qPCR analysis demonstrated that the expression level of *eccE$_3$*, the gene immediately downstream of *mycP$_3$*, was the same in both WT and $\Delta MycP3_{ms}$ strains, indicating that the deletion of *mycP$_3$* does not have a polar effect on the *eccE$_3$* gene.



(a)

(Figure 6.3)



(b)

Figure 6.3 Growth curves of *M. smegmatis* WT and ΔMycP3$_{ms}$ strains in 7H9 (a) and Fe-free Sauton's media (b) respectively. Hygromycin was either added or omitted in the medium where the ΔMycP3$_{ms}$::pr1MycP3$_{ms}$ and ΔMycP3$_{ms}$::pr2MycP3$_{ms}$ were cultured in triplicates. The error bars show the standard deviation of the triplicate OD$_{600nm}$ readings.

6.3.3 Proteomic analysis on *M. smegmatis* WT and ΔMycP3$_{ms}$ strains

Deletion of MycP$_3$ did not have a significantly negative impact on the growth of *M. smegmatis* in general, however, that assay does not measure its impact on the proteome. To assess the impact of MycP$_3$ KO on the *M. smegmatis* proteome, the WT and ΔMycP3$_{ms}$ strains were cultured in 7H9 media and subjected to proteomic analysis. The experiment was done in biological duplicate (strains were cultured at different time) and technical duplicate (each biological replicate was analysed twice separately). There were 2421 proteins identified from the entire set of experiments. Thirty-two proteins with significantly different abundances (student T-test result P value smaller than 0.05) between the two strains were identified (Table 6.2, Table 6.3, Table 6.4 and Table 6.5) and they represent 1.3% of the overall expressed proteome (32 out of 2421). These proteins were found to be involved in a large variety of cellular processes including iron acquisition, copper transport, glycine catabolism, drug resistance, the maintenance of cellular redox homeostasis and membrane potential, siderophore export, DNA replication, repair and protection, dTMP biosynthesis, cell cycle and cell division regulation, carbohydrate metabolism, energy metabolism, RNA and DNA catabolism, transcriptional control on multidrug efflux pumps, as well as a variety of metabolism proteins responding to environmental stress (See Table 6.2-6.5).

Table 6.2. Proteins identified to be more abundant in the proteome of the ΔMycP3$_{ms}$ strain compared to the WT strain. Where available, the *M. tuberculosis* orthologs are indicated.

Gene Name	**Gene Product**	**Function**	**Up-regulation Level (fold)**	**P-value**
MSMEG_0617 (Rv0284)	Ftsk-spoIIIe family protein ESX-3 conserved component EccC	Iron acquisition pathway (Siegrist *et al*., 2009)	2.2	0.0128
MSMEG_1017	Glutaredoxin	Unknown function, its orthologs are involved in cell redox homeostasis utilizing the reducing power of glutathione to maintain and regulate the cellular redox state and redox-dependent signalling pathways (Lillig *et al*., 2008)	2.2	0.0456
MSMEG_2771 (Rv2691)	TrkA protein K^+ ion uptake regulator	Unknown function, it involves in controlling membrane potential, pH homeostasis and multidrug susceptibility in *M. smegmatis*. (Castañeda-García *et al*., 2011)	2.3	0.0471
MSMEG_3515	3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase	Unknown function, its orthologs are involved in redox reactions. (Ghosh *et al*., 1991)	2.3	0.0023
MSMEG_4324 (Rv2242)	Hypothetical protein	Unknown function, conserved hypothetical protein	3.0	0.0098
MSMEG_4495 (Rv2366c)	CBS domain protein	Unknown function, orthologs are involved in multimerization and sorting of proteins, channel gating, and ligand binding, possibly function as sensors of intracellular metabolites. (Ignoul and Eggermont, 2005)	3.3	0.0239
MSMEG_5014 (Rv0969)	Copper-translocating P-type ATPase	Unknown function, its orthologs have copper-exporting ATPase activity, binding to ATP and conduct copper ion transport. (Hassani *et al*., 2010)	2.2	0.0478
MSMEG_5796 (Rv0811c)	Glycine cleavage T-protein	Glycine catabolism, it breaks down excess glycine molecules. (Kikuchi *et al*., 2008)	4.3	0.0073
MSMEG_6575 (Rv3775)	Beta-lactamase	It provides resistance to β-lactam antibiotics like penicillins, cephamycins, and carbapenems which inhibit the formation of	3.1	0.0037

		peptidoglycan cross-links in the bacterial cell wall.		

Table 6.3 Proteins only detected in ΔMycP3$_{ms}$ (not detected in the WT strain)

Gene Name	**Gene Product**	**Function**	**P-value**
MSMEG_0922 (Rv0478)	Deoxyribose-phosphate aldolase	Carbohydrate degradation, it involves in deoxyribonucleotide catabolic process	0.0085
MSMEG_3688	Hypothetical protein	Unknown function, conserved hypothetical protein	0.0020
MSMEG_4383 (Rv0676c)	MmpL5 protein	One component of siderophore export system. Iron acquisition pathway (Wells *et al*., 2013)	0.0034

Table 6.4 Proteins detected in lower abundance in ΔMycP3$_{ms}$ than in the WT strain

Gene Name	**Gene Product**	**Functional Category**	**Down regulation level (fold)**	**P-value**
MSMEG_1131	Tryptophan-rich sensory protein	Unknown function, its ortholog is involved in binding of intermediates in the tetrapyrrole biosynthetic pathway	2.2	0.0371
MSMEG_1633	DNA polymerase III, alpha subunit	DNA replication, DNA synthesis and repair	2.1	0.0463
MSMEG_1944	Membrane protein	Unknown function	3.1	0.0334
MSMEG_2683 (Rv2754c)	Thymidylate synthase, flavin-dependent	dTMP biosynthetic process, nucleotide metabolism	3.3	0.0202
MSMEG_2736 (Rv2725c)	GTP-binding protein	Signal transduction	7.3	0.0002
MSMEG_3026	Hypothetical protein	Unknown function, conserved hypothetical protein	6.1	0.0487
MSMEG_3496	MmpL4 protein	One component of siderophore export system. Iron acquisition pathway (Wells *et al*., 2013)	3.5	0.0296
MSMEG_3741 (Rv1710)	Transcriptional regulator	Unknown function, orthologs are involved in regulation in cell cycle and cell division	3.6	0.0162
MSMEG_4305	Phosphoglycerate mutase	Catalyze step 8 of glycolysis. Carbohydrate metabolism	6.5	0.0148

MSMEG_5307	TetR-family protein transcriptional regulator	Unknown function, orthologs are involved in the transcriptional control of multidrug efflux pumps, pathways for the biosynthesis of antibiotics, response to osmotic stress and toxic chemicals, control of catabolic pathways, differentiation processes and pathogenicity (Ramos *et al*., 2005)	2.3	0.0003
MSMEG_5479 (Rv0991c)	Type I antifreeze protein	Permit bacterial survival in subzero environments	4.6	0.0367
MSMEG_5926	Putative secreted protein	Unknown function	4.7	0.0130
MSMEG_6401	Prenyltransferase, UbiA family protein	Unknown function, orthologs catalyse the transfer of an isoprenyl pyrophosphate to a protein or a peptide	6.1	0.0194
MSMEG_6467	Starvation-induced DNA protecting protein	Unknown function, orthologs are anti-stress iron proteins preserving bacteria from oxidative damage, also responsible for cellular iron ion homeostasis (Wei *et al*., 2007)	2.6	0.0491
MSMEG_6687	Aldehyde dehydrogenase, thermostable	Unknown function, oxidize aldehyde to carboxylic acid	2.0	0.0032

Table 6.5 Proteins only detected in the WT strain (not detected in the ΔMycP3$_{ms}$ strain)

Gene Name	**Gene Product**	**Functional Category**	**P-value**
MSMEG_1966	Hypothetical protein	Unknown function, conserved hypothetical protein	0.0477
MSMEG_3237	ATP-binding protein	Unknown function, orthologs are involved in providing energy to carry out biological processes including translocation of various substrates across membranes and non-transport-related processes such as translation of RNA and DNA repair	0.0136
MSMEG_4559	ABC transporter, membrane spanning protein	Unknown function, involved in substrate transport across the membrane	0.0137
MSMEG_6029	Hypothetical protein	Unknown function, conserved hypothetical protein	0.0091
MSMEG_6945	Ribonuclease P protein component	Unknown function, orthologs are responsible for RNA molecule cleavage. Ususally it is involved in efficient production of tRNA and other small non-coding RNA genes	0.0199

6.3.4 RT-qPCR of EccC$_3$, MmpL5, MmpL4

ESX-3 was previously found to be associated with mycobactin-mediated iron acquisition (Siegrist *et al*., 2009). We were therefore interested in further exploring iron transport related proteins with significant difference in quantity between the WT and ΔMycP3$_{ms}$ strains cultured in 7H9 (high Fe). The proteomic results identified EccC$_3$ (Ftsk-SpoIIIE family protein, ESX-3 conserved component (Gey van Pittius *et al*., 2001)), MmpL5 and MmpL4 (both are related to siderophore export (Wells *et al*., 2013)) involved in Fe sequestration. RT-qPCR was conducted on the genes encoding these three proteins to determine whether the gene expression levels correlated with the proteomics results (Figure 6.4).

The expression level of *eccC$_3$* in ΔMycP3$_{ms}$ was higher than the WT and two complementation strains indicating an up-regulation in the ΔMycP3$_{ms}$ strain which correlates with the proteomic finding. MmpL4 expression was down-regulated on transcriptional level in ΔMycP3$_{ms}$, also correlating with the proteomic finding. MmpL5 expression was down-regulated at the transcriptional level in ΔMycP3$_{ms}$ but the MmpL5 protein was absent in the WT proteome. The complementation strains did not show the same expression level of *eccC$_3$* as the wild type while those of *mmpL5* and *mmpL4* were restored.

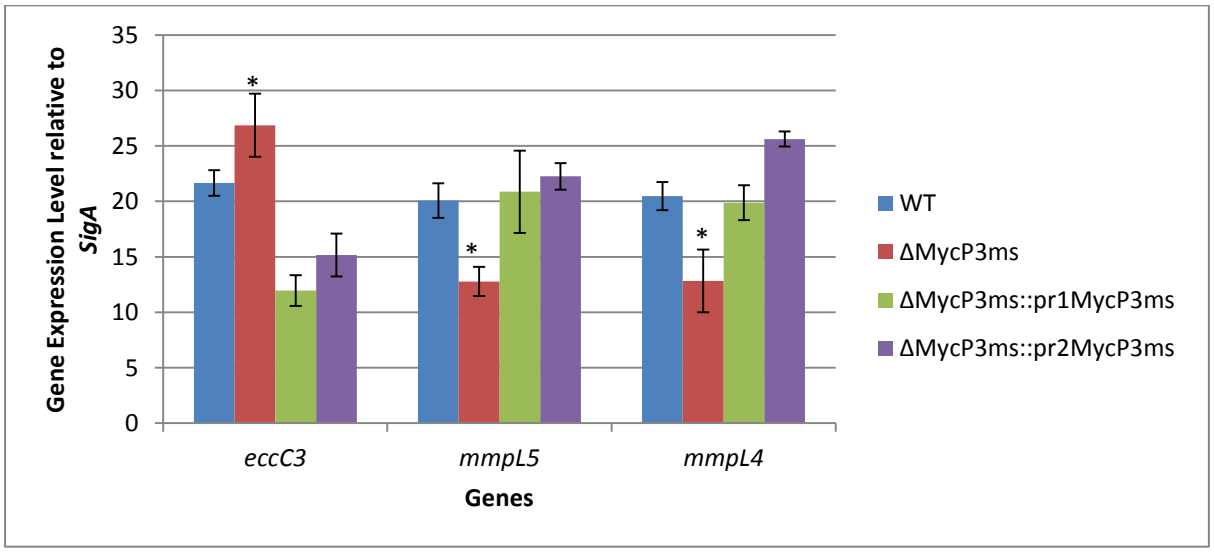


Figure 6.4. Gene expression levels of *ecc*C$_3$, *mmpL*5 and *mmpL*4 relative to that of *sigA* when strains were cultured in 7H9 media. The experiment was done in triplicates. The gene expression level relative to *sigA* is plotted against the gene name. Error bars show standard deviation. Complementation strains were able to restore the expression levels of the target genes, thus the significance of the difference between the wildtype and ΔMycP$_{3ms}$ only are calculated which, for three genes, *eccC3*, *mmpL5* and *mmpL4*, are p = 0.012, p = 0.000066, p = 0.0019 respectively (indicated with *).

6.3.5 Intracellular iron quantitation

To investigate whether the altered expression of *eccC*$_3$, *mmpL*5 and *mmpL*4 resulting from the deletion of MycP$_3$ in the *M. smegmatis* genome might influence the intracellular iron level in *M. smegmatis*, the intracellular iron content of the four strains was quantified. However, no significant differences in the intracellular iron levels were detected between the strains cultured under the same conditions were detected (Figure 6.5). As expected, iron levels dropped dramatically for all 4 strains when they were sub-cultured in Fe-free 7H9, showing an approximately 4-fold decrease. In contrast, intracellular iron level rose to a much higher level than in normal 7H9, increasing by approximately 2-fold, when iron was added to the Fe-deprived 7H9.

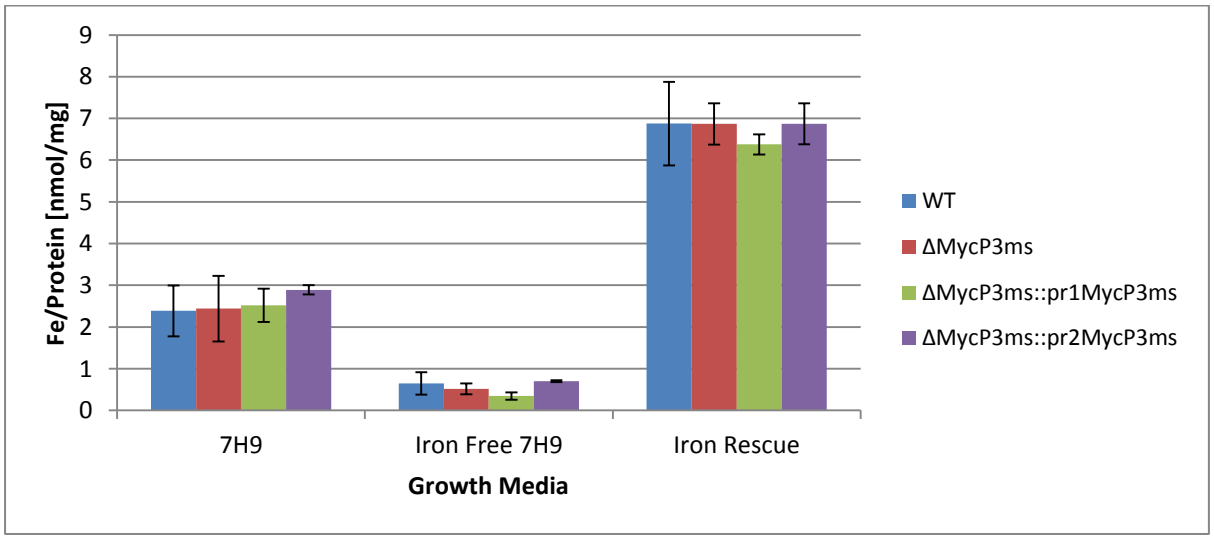


Figure 6.5. The comparison of the intracellular iron levels of WT, ΔMycP3$_{ms}$, ΔMycP3$_{ms}$::pr1MycP3$_{ms}$ and ΔMycP3$_{ms}$::pr2MycP3$_{ms}$ strains under 7H9, Fe-free 7H9 and Fe rescued 7H9 media. The experiment was done in triplicates. The error bars show standard deviations.

6.3.6 Functional analysis of the two promoters in the ESX-3 gene cluster

To determine how the ESX-3 promoters respond to different iron conditions, the MycP$_3$ expression levels from the two complementation strains were compared with those in the WT and KO using RT-qPCR on RNA recovered from bacteria grown in 7H9, Fe-deprived 7H9 and Fe-rescued 7H9 media.

In 7H9 media, the expression levels of *mycP*$_3$ were similar in the WT and ΔMycP$_{3ms}$::pr1MycP$_{3ms}$. However, ΔMycP$_{3ms}$::pr2MycP$_{3ms}$ demonstrated a *mycP*$_3$ expression

level about 6-fold higher than in WT_{ms} and $\Delta MycP_{3ms}$::pr1$MycP_{3ms}$ (Figure 6.6). When grown in iron-deprived media, the situation reversed with a similar *mycP*$_3$ expression in WT_{ms} and $\Delta MycP_{3ms}$::pr2$MycP_{3ms}$ but 7-fold higher *mycP*$_3$ expression in $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ (Figure 6.6). Predictably, when the Fe-deprived culture was rescued with Fe, the MycP3 expression level resembled that in normal 7H9 media, where the expression in $\Delta MycP_{3ms}$::pr2$MycP_{3ms}$ was 6-fold higher than WT and $\Delta MycP_{3ms}$::pr1$MycP_{3ms}$ (Figure 6.6).

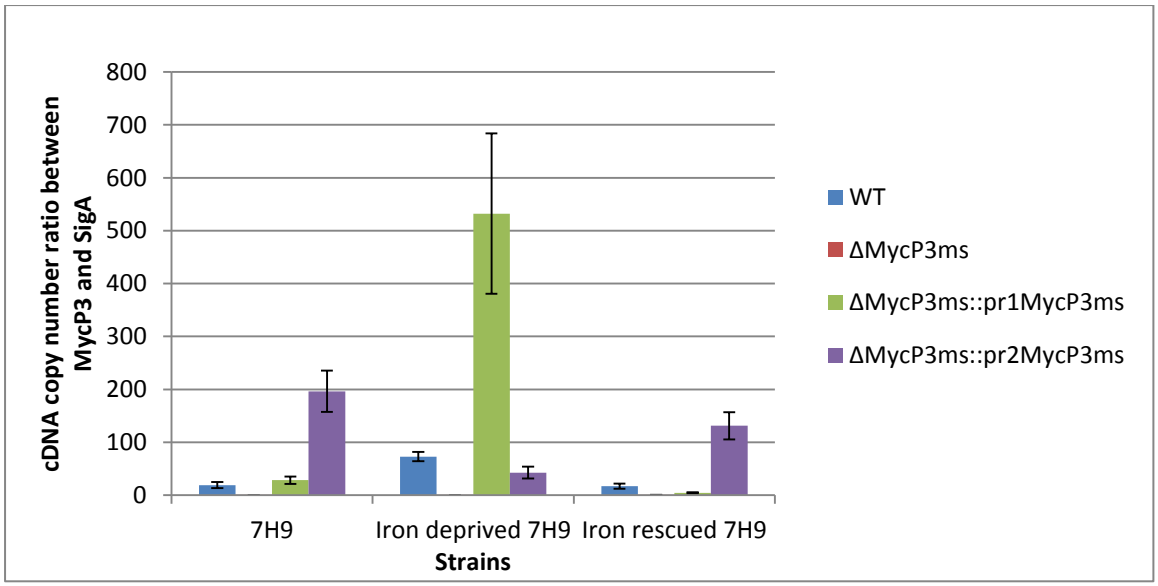


Figure 6.6 Comparison of the expression levels of *mycP*$_3$ in the WT, $\Delta MycP_{3ms}$ and two $MycP_3$ complementation strains under 7H9, Fe-free 7H9 and Fe rescued 7H9 media, where the experiments were done in triplicates. The error bars show standard deviations. Seeing that the *mycP*$_3$ expression level of wildtype is always similar to that of one of the complementation strains, and those of the two complementation strains have distinct difference on *mycP*$_3$ expression levels, the significance of difference among the strains can be represented by comparing the two complementation strains. In 7H9, iron deprived 7H9 and iron rescued 7H9 media, the *p* value for the significance of different $mycP_3$ expression levels between the two complementation strains are 0.0073, 0.00048 and 0.013 respectively.

6.4 Discussion

ESX-3 is involved in the mycobactin-mediated iron acquisition pathway (Siegrist *et al*., 2014, 2009). The growth of the *Mycobacterium smegmatis* Δ*mbt*DΔ*esx*-3 double mutant was not affected unless the exochelin pathway was also disabled (Siegrist *et al*., 2009). Possessing the exochelin-mediated iron acquisition pathway renders *M. smegmatis* tolerant to the loss of the

mycobactin pathway. It is not known to what extent the loss of MycP3 will influence the functionality of the ESX-3 system with regards to mycobactin-mediated iron acquisition.

To better understand the contribution of $MycP_3$ to iron acquisition in *M. smegmatis*, we generated and characterised *mycP3* knockout and complemented strains. Although $\Delta MycP3_{ms}$ did not show any significant *in vitro* growth defect compared to the WT_{ms} (Figure 6.3), some molecular pathways may be altered. To address this possibility, we exploited high resolution mass spectrometry to analyse the proteomes of WT_{ms} and $\Delta MycP3_{ms}$ under normal nutrient sufficient conditions. These were compared to determine the effect of the absence of $MycP_3$ on overall metabolism in *M. smegmatis*. This revealed that, although deletion of MycP3 did not have a major impact on the overall proteome of *M. smegmatis* (1.3%), it led to the differential abundance of iron homeostasis-related proteins including $EccC_3$, MmpL5 and MmpL4 between WT_{ms} and the $\Delta MycP3_{ms}$.

To investigate whether the differential abundance of iron-homeostasis-related proteins had impact on proteins played a role in iron acquisition/homeostasis, we compared intracellular iron levels in the WT, mutant and complemented strains. This demonstrated that the intracellular iron levels were consistent across all strains (Figure 6.5) suggesting that the $\Delta MycP3_{ms}$ mutant did not show a defect in iron acquisition. This is likely because exochelin is able to compensate for the loss of carboxymycobactin as carboxymycobactin constitutes less than 5% of the total extracellular iron-binding capacity in *M. smegmatis* (Ratledge and Ewing, 1996). Such a loss of carboxymycobactin might exert a negative effect on the iron acquisition efficiency. The intracellular iron levels for all strains decreased when sub-cultured to iron-deprived media from iron rich media but increased to a level even higher than that in the original iron rich media when rescued with iron. This observation may be explained by the hypothesis that cell envelope-associated mycobactins serves to temporarily store iron (Ratledge, 2004). Therefore we reason that the production of mycobactin was suppressed in the original iron rich conditions, but derepressed during iron deprivation to produce a large amount of mycobactin which was transported into the cell envelope (Figure 6.7). When the iron was added to rescue the culture, the normal iron uptake was restored. The increased iron is possibly from the iron-bound mycobactin retained in the cell envelope (Figure 6.7).

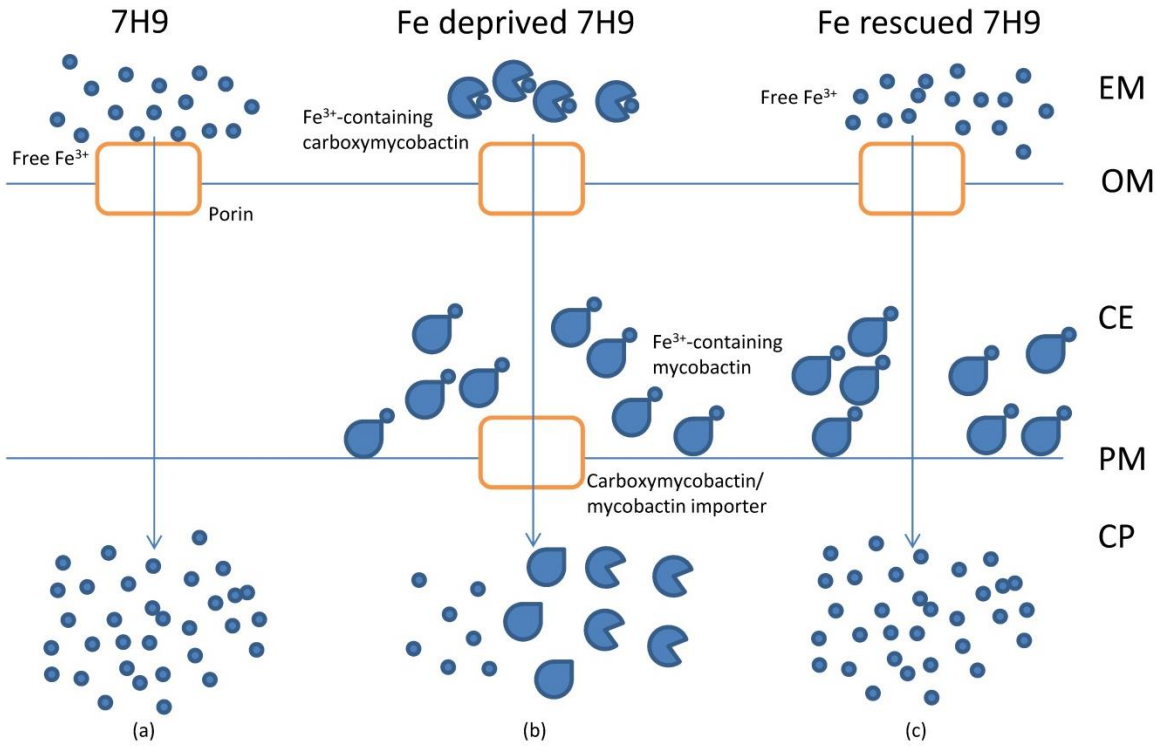


Figure 6.7 An illustration of the change in intracellular iron levels of *M. smegmatis* cultured in 7H9 medium along iron deprivation and iron rescue processes. When iron is sufficient, free Fe^{3+} can be imported freely through porin (a). When Fe^{3+} is depleted, carboxymycobactin and mycobactin are produced to assist in iron acquisition (b). When the Fe deprived culture is rescued with Fe^{3+}, the Fe^{3+} can be imported freely again meanwhile there are still significant amount of Fe^{3+}-containing mycobactin in the cell envelope playing an Fe^{3+} storage role thus elevating the intracellular iron content (c).

Proteomic analysis demonstrated that the ESX-3 component, $EccC_3$, was more abundant in $\Delta MycP3_{ms}$ than in WT. This was confirmed by RT-qPCR, which also demonstrated that *ecc*C_3 expression was reduced upon complementation. It appears that an unknown mechanism, possibly feedback from the downstream effectors of ESX-3, compensates for the loss of the functional ESX-3 by up-regulating the production of the entire system (Figure 6.8). The other components of ESX-3 were not detected by mass spectrometry, probably due to the protein extraction protocol which did not enrich for membrane proteins and the low expression of ESX-3 under iron sufficient conditions. Evidence in the literature shows that ESX-3 not only affects the siderophore-mediated iron acquisition but also heme uptake (Serafini *et al*., 2013). It is thus interesting to speculate that ESX-3 influences the secretion and maturation of multiple growth factors, as well as post-translational modification on membrane-bound proteins. The higher abundance of ESX-3 might result in more growth factors being secreted and facilitating nutrient acquisition, including Fe^{3+}, resulting in excess intracellular quantity (although this was not observed in the assay).

The proteomic results also showed differential abundance of many other proteins playing various roles in the metabolism of the bacteria. The changes in the abundance of these

proteins are probably the downstream effects of the excess intracellular Fe^{3+} ions imported via the exochelin pathway, and it is probably not affected by ESX-3. It is hypothesized that all those downstream effects are linked (Figure 6.8). After being transported into the cytoplasm, the ferric iron bound to exochelin is reduced to ferrous iron upon release by an unknown reductase (Zhu *et al*., 1998) which in turn needs to be reduced for recycling. The extra pressure on redox homeostasis maintenance might be relieved by producing more redox regulators such as glutaredoxin as observed (Table 6.2) (Lillig *et al*., 2008). Another downstream effect is the disruption of the existing membrane potential of the bacteria due to the excess intracellular positive charges exerted by the Fe ions. The higher abundance of the TrkA K^+ uptake regulator (Table 6.2) may remedy the situation to maintain a normal membrane potential (Castañeda-García *et al*., 2011). This could also be reinforced by the more abundant copper exporters (copper translocating P-type ATPase, Table 6.2) which may render the bacteria more resistant to copper toxicity (Hodgkinson and Petris, 2012). The ESX-3 is a multi-component membrane transporter complex consisting of the $EccB_3$, $EccC_3$, $EccD_3$ and $EccE_3$ transmembrane partners, with $EccE_3$ possibly stretching into the outer membrane (Houben *et al*., 2012). This machinery may play a role in maintaining the cell wall integrity, disruption of which by deleting $MycP_3$ may have resulted in the increased production of β-lactamase as seen by the increased levels of MSMEG_6575 (Table 6.2). This may confer resistance to many antibiotics, including penicillin, cephamycin and carbapenems which inhibit peptidoglycan cross-link formation. These proteomic findings can be further investigated on transcriptional level by RT-qPCR. Alternatively, the downstream products of the enzymes involved in those cellular processes can be used as markers to assess and compare their activities in the WT_{ms} and $\Delta MycP_{3ms}$ strains.

The proteomics data also provided evidence that the $MycP_3$ mutant had slowed down cellular metabolism on a micro-scale (not observed from the growth curves) with many house-keeping enzymes present in lower abundance (Figure 6.8). These include the DNA replication and repair enzyme (DNA polymerase III), nucleotide biosynthetic enzyme (thymidylate synthase), cell cycle and cell division regulator (MSMEG_3741), carbohydrate metabolic enzyme (phosphoglycerate mutase), as well as some components in bacterial survival mechanisms against physiological stress (MSMEG_5307, MSMEG_5479 and MSMEG_6467). A few proteins were not identified in the mutant including an ATP-binding protein (MSMEG_3237) possibly providing energy to various substrate transport systems across the membrane as well as RNA translation and DNA repair systems, and a ribonuclease P protein component which is involved in tRNA production (Guerrier-Takada *et al*., 1983). Although technical limitations might have limited sensitivity, the failure to identify these proteins suggests that the mutant was not at its optimal growth where full scale production of cellular components and the survival mechanism are not required.

An interesting observation is that MmpL5 was only detected in $\Delta MycP3_{ms}$ mutant whereas MmpL4 was less abundant compared to WT_{ms} (Figure 6.8). MmpL5 and MmpL4, together with MmpS5 and MmpS4, are components of a hetero-tetramer complex on the plasma membrane which is involved in/responsible for siderophore export (Wells *et al*., 2013). Although the genes encoding MmpL4 and MmpL5 are located in different operons

(Rodriguez *et al*., 2002), the expression of both is regulated by iron and they should show similar levels of transcription and translation. Since deletion of $MycP_3$ is likely to compromise ESX-3 function, mycobactin mediated iron uptake may be impaired making mycobactin export less important in the $\Delta MycP3_{ms}$ mutant compared to WT_{ms}. Expression of this export system may therefore be down-regulated. This is supported by RT-qPCR results (Figure 6.3) which showed that both genes were down-regulated in the ΔMycP3 mutant and the effect was reversed by MycP3 complementation. Down-regulation of transcription corresponded to less MmpL4 protein in the whole cell lysate of $\Delta MycP3_{ms}$, although similar was not observed for MmpL5. This study did not enrich for membrane proteins and MmpL5 was abundant in the soluble fraction, suggesting that the maturation and export of this protein onto plasma membrane may be impaired, and that $MycP_3$ may play a direct or indirect role in the post-translational modification on MmpL5.

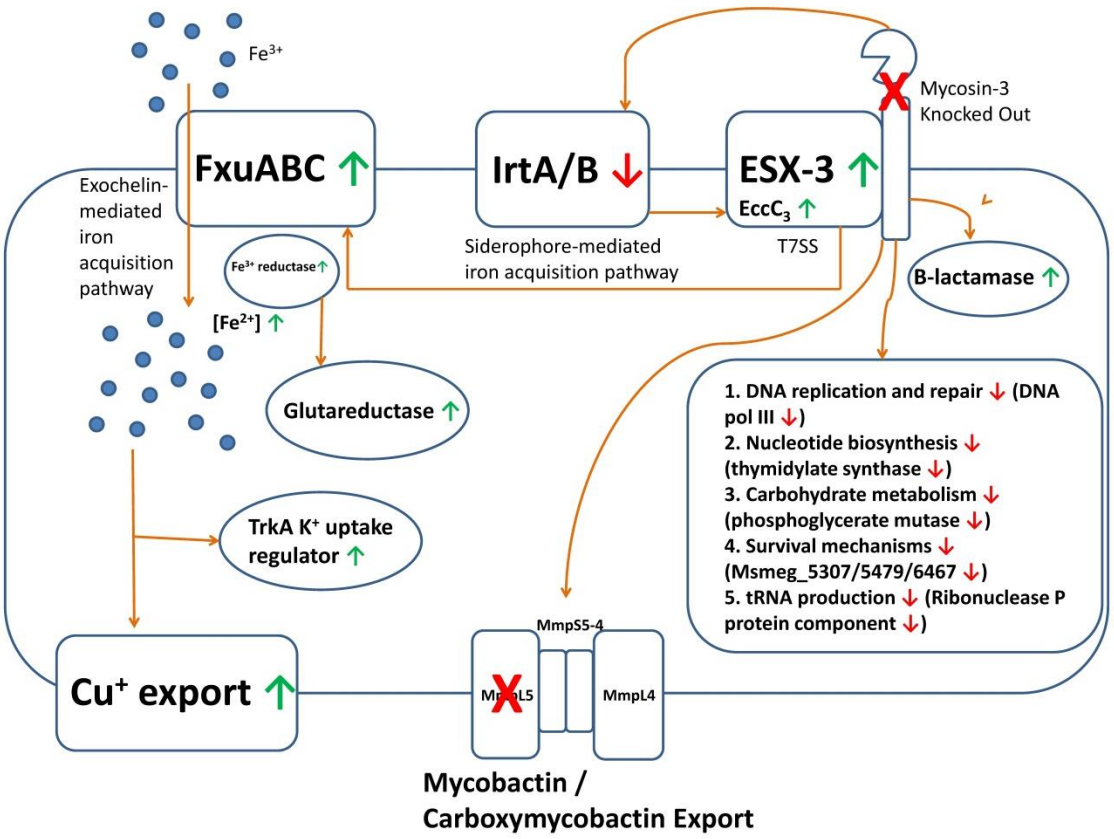


Figure 6.8 The proposed model of the downstream effects of $MycP_3$ knock-out in *M. smegmatis* under high iron conditions. Loss of $MycP_3$ had a negative impact on the siderophore-mediated iron acquisition pathway and this effect signalled the bacteria to produce more ESX-3 which in turn allowed more ferric ion to be imported via exochelin-mediated iron acquisition pathway. The increased import of Fe^{3+} resulted in more oxidized form of Fe^{3+} reductases which were possibly reduced by glutareductase to maintain the redox homeostasis in the bacteria. Meanwhile the increased amount of Fe^{2+} elevated the intracellular positive charge; to maintain the cell membrane potential, TrkA K^+ uptake regulator and Cu^+ exporters were more abundant to secrete excess positive charges. A number of cellular metabolisms were slowed down, including DNA replication and repair, nucleotide biosynthesis, carbohydrate metabolism, tRNA production, and some survival mechanisms. Interestingly, MmpL5 was not observed from $\Delta MycP_{3ms}$ possibly implying that $MycP_3$ may regulate the stability and export of MmpL5.

The *M. smegmatis* ESX-3 is expressed from two promoters as mentioned previously (Maciag *et al*., 2009). The efficiency of complementation using both promoters in the ESX-3 gene cluster was investigated. Previous studies did not find major differences in expression of the genes in the cluster downstream of these two promoters under various culturing conditions except acid stress (Maciag *et al*., 2009). The transcription of $MycP_3$ gene from the promoter immediately upstream of the cluster (pr1) responds to different iron levels in the same fashion as WT_{ms} (Figure 6.6) implying that the transcription of *mycP*$_3$ gene is controlled by the first promoter (pr1) in the gene cluster even though the gene is downstream of the 2nd promoter (pr2) (Figure 6.1) (Maciag *et al*., 2009). However, the transcription level of MycP3 under the control of the first promoter is almost 5 fold higher than in the WT_{ms}. In the integrative complementation vector, the *myc*P_3 gene was positioned directly downstream of the promoter, rather than being 9 genes downstream as in the WT_{ms} genome. Such a gene re-arrangement brings the gene closer to the transcription start site increasing the rate of transcription, hence a higher number of transcripts (Lim *et al*., 2011). The second promoter is not regulated in the same manner as the first, indicating its independence from the IdeR regulation. It was proposed that the first promoter in ESX-3 gene cluster is responsible for the transcription of the entire operon while the second promoter only influences the transcription of the 6 genes downstream of it (Maciag *et al*., 2009). Although iron was deprived, the expression level of $MycP_3$ by the second promoter still reached the same level as in WT_{ms}, possibly to match with those of other ESX-3 components. When iron was sufficient, the second promoter was up-regulated by an unknown mechanism to express the downstream PE protein (MSMEG_0620), esx-like antigen 7 (MSMEG_0621) and secreted protein $EspG_3$ (MSMEG_0622) which may be required in other metabolic pathways. The two promoters in the gene cluster are not redundant as they respond to the different iron levels in opposite ways which was not observed previously. The regulation of the second promoter is of interest for further investigation.

The influence of $MycP_3$ on the proteome under iron-deprived conditions may reveal more clues as to the protein's function. The proteomes of both $MycP_3$ complementation strains can be analysed together with the WT_{ms} and the $\Delta MycP_{3\,ms}$ mutant as the two promoters are non-redundant. This work is currently under investigation. For further validation of the modification role of $MycP_3$ on MmpL5, both proteins need to be produced *in vitro* and their binding affinities or possible cleavage activity may then be analysed. The cellular localization of MmpL5 in the WT and $\Delta MycP_{3ms}$ mutant can be determined and compared using MmpL5 antibodies and Western blotting of the membrane and cytosolic fractions.

6.5 Conclusion

This study confirmed the direct relationship between siderophore-mediated iron acquisition and ESX-3 secretion system. Loss of one component of this secretion system, Mycosin-3, leads to various downstream effects on divalent cation transport, redox homeostasis, membrane potential, drug resistance, and many other house-keeping processes. Evidence has started to reveal and support the pleiotropic effect of substrate secretion and modification by ESX-3. Abolishing the ESX-3 secretion system may impair many cellular functions to stop the bacterial growth of *M. tuberculosis*, attenuating the pathogen in the host, hence offering excellent targets for anti-TB drug development.

6.6 Reference

Abdallah, A.M., Gey van Pittius, N.C., Champion, P.A.D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M.J.E., Appelmelk, B.J., Bitter, W., 2007. Type VII secretion--mycobacteria show the way. Nat. Rev. Microbiol. 5, 883–891. doi:10.1038/nrmicro1773

Barclay, R., Ratledge, C., 1988. Mycobactins and exochelins of Mycobacterium tuberculosis, M. bovis, M. africanum and other related species. J. Gen. Microbiol. 134, 771–776.

Castañeda-García, A., Do, T.T., Blázquez, J., 2011. The K+ uptake regulator TrkA controls membrane potential, pH homeostasis and multidrug susceptibility in Mycobacterium smegmatis. J. Antimicrob. Chemother. 66, 1489–1498. doi:10.1093/jac/dkr165

Chim, N., Iniguez, A., Nguyen, T.Q., Goulding, C.W., 2010. Unusual diheme conformation of the heme-degrading protein from Mycobacterium tuberculosis. J. Mol. Biol. 395, 595–608. doi:10.1016/j.jmb.2009.11.025

Clemens, D.L., Horwitz, M.A., 1996. The Mycobacterium tuberculosis phagosome interacts with early endosomes and is accessible to exogenously administered transferrin. J. Exp. Med. 184, 1349–1355.

De Souza, G.A., Fortuin, S., Aguilar, D., Pando, R.H., McEvoy, C.R.E., van Helden, P.D., Koehler, C.J., Thiede, B., Warren, R.M., Wiker, H.G., 2010. Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo- and hypervirulent clinical Mycobacterium tuberculosis Beijing isolates. Mol. Cell. Proteomics MCP 9, 2414–2423. doi:10.1074/mcp.M900422-MCP200

Fiss, E.H., Yu, S., Jacobs, W.R., Jr, 1994. Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways. Mol. Microbiol. 14, 557–569.

Frederick, R.E., Mayfield, J.A., DuBois, J.L., 2009. Iron trafficking as an antimicrobial target. Biometals Int. J. Role Met. Ions Biol. Biochem. Med. 22, 583–593. doi:10.1007/s10534-009-9236-1

Gey Van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria. Genome Biol. 2, RESEARCH0044.

Ghosh, D., Weeks, C.M., Grochulski, P., Duax, W.L., Erman, M., Rimsay, R.L., Orr, J.C., 1991. Three-dimensional structure of holo 3 alpha,20 beta-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. Proc. Natl. Acad. Sci. U. S. A. 88, 10064–10068.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 7, e1002251. doi:10.1371/journal.ppat.1002251

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S., 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35, 849–857. doi:10.1016/0092-8674(83)90117-4

Hassani, B.K., Astier, C., Nitschke, W., Ouchane, S., 2010. CtpA, a copper-translocating P-type ATPase involved in the biogenesis of multiple copper-requiring enzymes. J. Biol. Chem. 285, 19330–19337. doi:10.1074/jbc.M110.116020

Hodgkinson, V., Petris, M.J., 2012. Copper Homeostasis at the Host-Pathogen Interface. J. Biol. Chem. 287, 13549–13555. doi:10.1074/jbc.R111.316406

Houben, E.N.G., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jiménez, C.R., Ottenhoff, T.H.M., Luirink, J., Bitter, W., 2012. Composition of the type VII

secretion system membrane complex. Mol. Microbiol. 86, 472–484. doi:10.1111/j.1365-2958.2012.08206.x

Ignoul, S., Eggermont, J., 2005. CBS domains: structure, function, and pathology in human proteins. Am. J. Physiol. Cell Physiol. 289, C1369–1378. doi:10.1152/ajpcell.00282.2005

Jones, C.M., Niederweis, M., 2010. Role of porins in iron uptake by Mycobacterium smegmatis. J. Bacteriol. 192, 6411–6417. doi:10.1128/JB.00986-10

Jones, C.M., Wells, R.M., Madduri, A.V.R., Renfrow, M.B., Ratledge, C., Moody, D.B., Niederweis, M., 2014. Self-poisoning of Mycobacterium tuberculosis by interrupting siderophore recycling. Proc. Natl. Acad. Sci. U. S. A. 111, 1945–1950. doi:10.1073/pnas.1311402111

Kikuchi, G., Motokawa, Y., Yoshida, T., Hiraga, K., 2008. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 84, 246–263.

Kontoghiorghes, G.J., Weinberg, E.D., 1995. Iron: mammalian defense systems, mechanisms of disease, and chelation therapy approaches. Blood Rev. 9, 33–45.

Krithika, R., Marathe, U., Saxena, P., Ansari, M.Z., Mohanty, D., Gokhale, R.S., 2006. A genetic locus required for iron acquisition in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 103, 2069–2074. doi:10.1073/pnas.0507924103

Lillig, C.H., Berndt, C., Holmgren, A., 2008. Glutaredoxin systems. Biochim. Biophys. Acta 1780, 1304–1317. doi:10.1016/j.bbagen.2008.06.003

Lim, H.N., Lee, Y., Hussein, R., 2011. Fundamental relationship between operon organization and gene expression. Proc. Natl. Acad. Sci. U. S. A. 108, 10626–10631. doi:10.1073/pnas.1105692108

Maciag, A., Piazza, A., Riccardi, G., Milano, A., 2009. Transcriptional analysis of ESAT-6 cluster 3 in Mycobacterium smegmatis. BMC Microbiol. 9, 48. doi:10.1186/1471-2180-9-48

Messenger, A.J., Hall, R.M., Ratledge, C., 1986. Iron uptake processes in Mycobacterium vaccae R877R, a mycobacterium lacking mycobactin. J. Gen. Microbiol. 132, 845–852.

Owens, C.P., Chim, N., Goulding, C.W., 2013. Insights on how the Mycobacterium tuberculosis heme uptake pathway can be used as a drug target. Future Med. Chem. 5, 1391–1403. doi:10.4155/fmc.13.109

Parish, T., Stoker, N.G., 2000. Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. Microbiol. Read. Engl. 146 (Pt 8), 1969–1975.

Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., Walsh, C.T., 1998. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem. Biol. 5, 631–645.

Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., Tobes, R., 2005. The TetR family of transcriptional repressors. Microbiol. Mol. Biol. Rev. MMBR 69, 326–356. doi:10.1128/MMBR.69.2.326-356.2005

Rappsilber, J., Mann, M., Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906. doi:10.1038/nprot.2007.261

Ratledge, C., 2004. Iron, mycobacteria and tuberculosis. Tuberc. Edinb. Scotl. 84, 110–130.

Ratledge, C., Ewing, M., 1996. The occurrence of carboxymycobactin, the siderophore of pathogenic mycobacteria, as a second extracellular siderophore in Mycobacterium smegmatis. Microbiology 142, 2207–2212. doi:10.1099/13500872-142-8-2207

Reddy, T.B.K., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H., Koehrsen, M., Larson, L., Mao, M., Nitzberg, M., Sisk, P., Stolte, C., Weiner, B., White, J., Zachariah, Z.K., Sherlock, G., Galagan, J.E., Ball, C.A., Schoolnik, G.K., 2009. TB database: an integrated platform for tuberculosis research. Nucleic Acids Res. 37, D499–508. doi:10.1093/nar/gkn652

Riemer, J., Hoepken, H.H., Czerwinska, H., Robinson, S.R., Dringen, R., 2004. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. Anal. Biochem. 331, 370–375. doi:10.1016/j.ab.2004.03.049

Rodriguez, G.M., Smith, I., 2006. Identification of an ABC transporter required for iron acquisition and virulence in Mycobacterium tuberculosis. J. Bacteriol. 188, 424–430. doi:10.1128/JB.188.2.424-430.2006

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., 2002. ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. 70, 3371–3381.

Ryndak, M.B., Wang, S., Smith, I., Rodriguez, G.M., 2010. The Mycobacterium tuberculosis high-affinity iron importer, IrtA, contains an FAD-binding domain. J. Bacteriol. 192, 861–869. doi:10.1128/JB.00223-09

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

Serafini, A., Boldrin, F., Palù, G., Manganelli, R., 2009. Characterization of a Mycobacterium tuberculosis ESX-3 conditional mutant: essentiality and rescue by iron and zinc. J. Bacteriol. 191, 6340–6344. doi:10.1128/JB.00756-09

Serafini, A., Pisu, D., Palù, G., Rodriguez, G.M., Manganelli, R., 2013. The ESX-3 Secretion System Is Necessary for Iron and Zinc Homeostasis in Mycobacterium tuberculosis. PLoS ONE 8, e78351. doi:10.1371/journal.pone.0078351

Siegrist, M.S., Steigedal, M., Ahmad, R., Mehra, A., Dragset, M.S., Schuster, B.M., Philips, J.A., Carr, S.A., Rubin, E.J., 2014. Mycobacterial esx-3 requires multiple components for iron acquisition. mBio 5. doi:10.1128/mBio.01073-14

Siegrist, M.S., Unnikrishnan, M., McConnell, M.J., Borowsky, M., Cheng, T.-Y., Siddiqi, N., Fortune, S.M., Moody, D.B., Rubin, E.J., 2009. Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc. Natl. Acad. Sci. U. S. A. 106, 18792–18797. doi:10.1073/pnas.0900589106

Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F., 1991. New use of BCG for recombinant vaccines. Nature 351, 456–460. doi:10.1038/351456a0

Svistunenko, D.A., 2005. Reaction of haem containing proteins and enzymes with hydroperoxides: the radical view. Biochim. Biophys. Acta 1707, 127–155. doi:10.1016/j.bbabio.2005.01.004

Tullius, M.V., Harmston, C.A., Owens, C.P., Chim, N., Morse, R.P., McMath, L.M., Iniguez, A., Kimmey, J.M., Sawaya, M.R., Whitelegge, J.P., Horwitz, M.A., Goulding, C.W., 2011. Discovery and characterization of a unique mycobacterial heme acquisition system. Proc. Natl. Acad. Sci. U. S. A. 108, 5051–5056. doi:10.1073/pnas.1009516108

Vogel, C., Marcotte, E.M., 2008. Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. Nat. Protoc. 3, 1444–1451. doi:10.1038/nport.2008.132

Wagner, J.M., Evans, T.J., Chen, J., Zhu, H., Houben, E.N.G., Bitter, W., Korotkov, K.V., 2013. Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. J. Struct. Biol. 184, 115–128. doi:10.1016/j.jsb.2013.09.022

Wei, X., Mingjia, H., Xiufeng, L., Yang, G., Qingyu, W., 2007. Identification and biochemical properties of Dps (starvation-induced DNA binding protein) from cyanobacterium Anabaena sp. PCC 7120. IUBMB Life 59, 675–681. doi:10.1080/15216540701606926

Wells, R.M., Jones, C.M., Xi, Z., Speer, A., Danilchanka, O., Doornbos, K.S., Sun, P., Wu, F., Tian, C., Niederweis, M., 2013. Discovery of a siderophore export system essential for virulence of Mycobacterium tuberculosis. PLoS Pathog. 9, e1003120. doi:10.1371/journal.ppat.1003120

Wren, B.W., Stabler, R.A., Das, S.S., Butcher, P.D., Mangan, J.A., Clarke, J.D., Casali, N., Parish, T., Stoker, N.G., 1998. Characterization of a haemolysin from Mycobacterium tuberculosis with homology to a virulence factor of Serpulina hyodysenteriae. Microbiol. Read. Engl. 144 (Pt 5), 1205–1211.

Yu, J.-H., Hamari, Z., Han, K.-H., Seo, J.-A., Reyes-Domínguez, Y., Scazzocchio, C., 2004. Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi. Fungal Genet. Biol. FG B 41, 973–981. doi:10.1016/j.fgb.2004.08.001

Yu, S., Fiss, E., Jacobs, W.R., Jr, 1998. Analysis of the exochelin locus in Mycobacterium smegmatis: biosynthesis genes have homology with genes of the peptide synthetase family. J. Bacteriol. 180, 4676–4685.

Zhu, W., Arceneaux, J.E., Beggs, M.L., Byers, B.R., Eisenach, K.D., Lundrigan, M.D., 1998. Exochelin genes in Mycobacterium smegmatis: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. Mol. Microbiol. 29, 629–639.

Chapter 7. Conclusion and Future Directions

7.1 Conclusion

Type VII secretion systems are essential for the survival and virulence of *Mycobacterium tuberculosis*. Much is still unknown about the five copies of the T7SS in this pathogen, including the identity of all substrates, the secretion mechanism, and the role in pathogenicity. It is evident that a large number of substrates of ESX-1 and ESX-5, such as the Esp and PPE proteins, are responsible for maintaining the cell wall integrity of the bacteria and conferring specific antigenic features. It seems the well-known T-cell antigens, ESAT-6 and CFP-10, are more likely to target and lyse the bacterial membranes than the host membranes (de Jonge *et al*., 2007; Garces *et al*., 2010). We postulate that that the substrates of ESX-1 and ESX-5 appear not to be directly related to the virulence of the pathogen. ESX-3 influences the acquisition of both iron and heme, mediated by siderophores and hemophores respectively. Possibly the substrates or the membrane complex of ESX-3 facilitate either the export of siderophores and hemophores or the iron and heme acquisition processes extracellularly.

The subtilisin-like serine protease, mycosin-3, one of the membrane component of ESX-3, was unable to be produced either by expression hosts or *in vitro*, suggesting that it may exert a toxic effect on the host if not exported.

The connection between ESX-3 and iron/heme acquisition has been observed previously by a few international research groups. However, how ESX-3 influences the iron/heme uptake has not been studied in detail. Our proteomics and RT-qPCR data on the wild type and mycosin-3 deletion mutant of *M. smegmatis* suggests that ESX-3, especially mycosin-3, may enable the functionality of siderophore export system by facilitating the localization of MmpL5, without which siderophore can no longer get recycled and exported to acquire iron and intracellular accumulation of siderophores is toxic to mycobacteria possibly through intracellular iron chelating. Although this result needs to be validated, primarily mycosin-3 can be regarded as an important anti-TB drug target, the abolishment of which may lead to iron deprivation of the bacteria which cannot be rescued by heme. Without iron, the pathogen may not survive and cause disease in the host.

7.2 Future Directions

Many protein production methods were used to investigate the production of recombinant $MycP_3$, however, despite many attempts with different methods, it was not possible with all the tools available to us. Two other international groups also did not manage to clone and produce *M. tuberculosis* orthologues of $MycP_1$ and $MycP_3$ either (Solomonson *et al*., 2013;

Wagner *et al*., 2013). Their results agree with our findings that mycosins from *M. tuberculosis* may be toxic to the expression host and not possible to be expressed at this stage.

To test this hypothesis, the following approaches can be employed in future studies:

1. The signal peptide of $MycP_3$ could be deleted in the *M. tuberculosis* genome to determine whether the un-exported $MycP_3$ exerts a toxic effect on the *M. tuberculosis* mutant, and/or

2. The catalytic triad of $MycP_3$ can be deactivated by generating a point mutation at the serine residue to establish whether *E. coli* is able to produce inactive $MycP_3$.

The functional studies of $MycP_3$ have provided new insights into the regulation of iron homeostasis in the mycobacteria. To validate the possible impact of $MycP_3$ on MmpL5, the following methods can be adopted:

1. The soluble domains of MmpL5 can be cloned and produced and anti-MmpL5 antibodies generated against them. The antibodies may be used to determine the cellular localization of MmpL5 in the wild type and $\Delta MycP_3$ mutant of *M. smegmatis* to assess whether $MycP_3$ affects the export of MmpL5.

2. An enzyme assay can be done using the soluble domains of MmpL5 and $MycP_3$ *in vitro* to determine whether $MycP_3$ has catalytic activity against MmpL5. The cleavage results can be visualized using SDS-PAGE.

In addition, to reveal novel substrates of ESX-3, the secretomes of WT_{ms}, $\Delta MycP_{3\ ms}$ and $\Delta MycP3_{ms}$::pr1$MycP3_{ms}$ can be compared. The proteins which are absent from, or dramatically less abundant in, $\Delta MycP_{3ms}$ may be substrates of ESX-3.

Addemdum

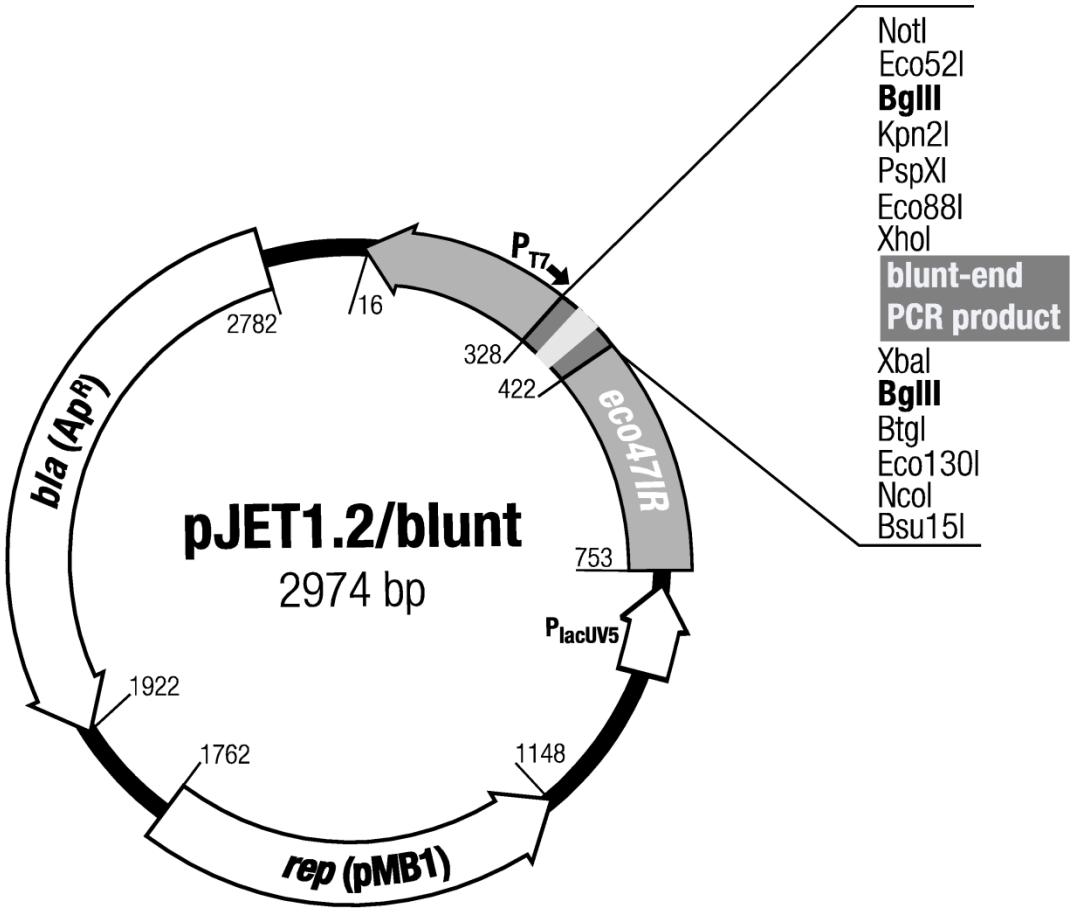


Figure S5.1 Plasmid DNA map of pJET1.2 cloning vector.

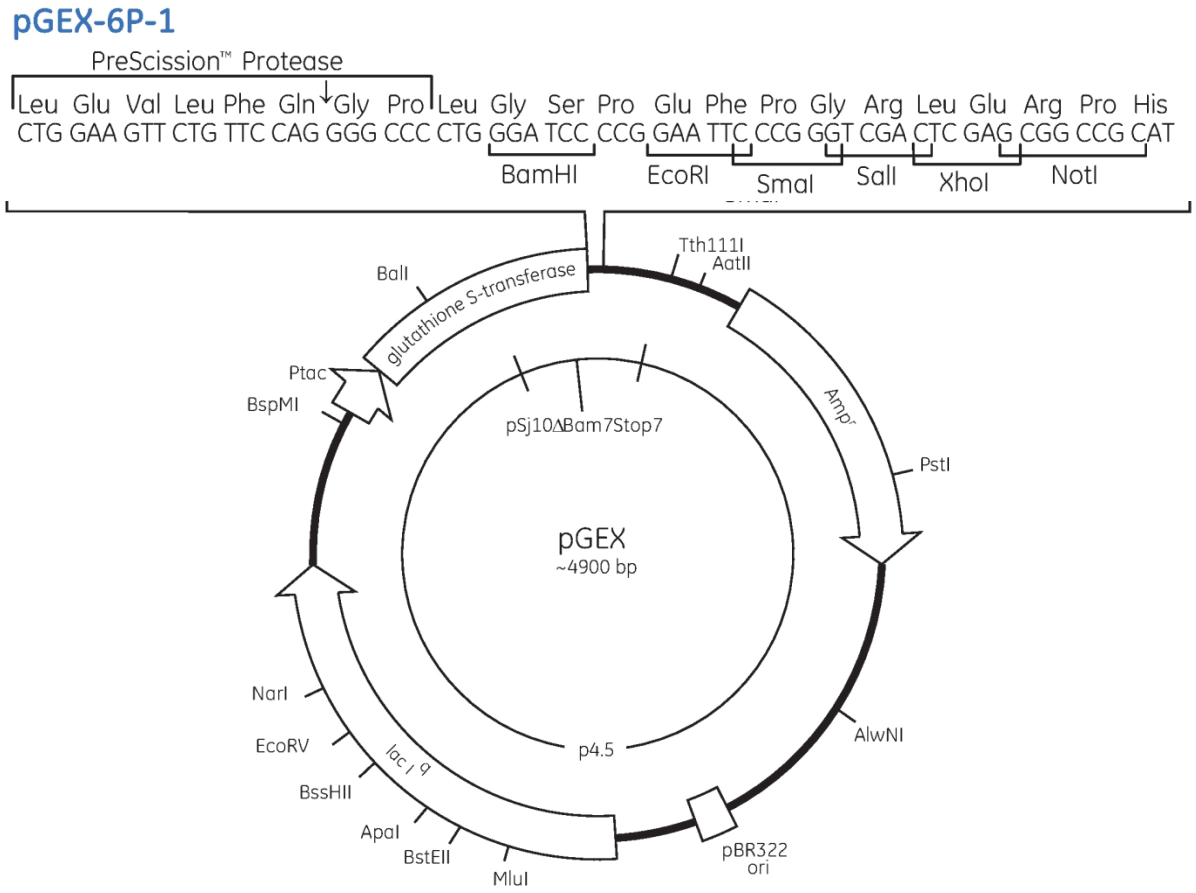


Figure S5.2 Plasmid DNA map of pGEX-6P-1 expression vector.

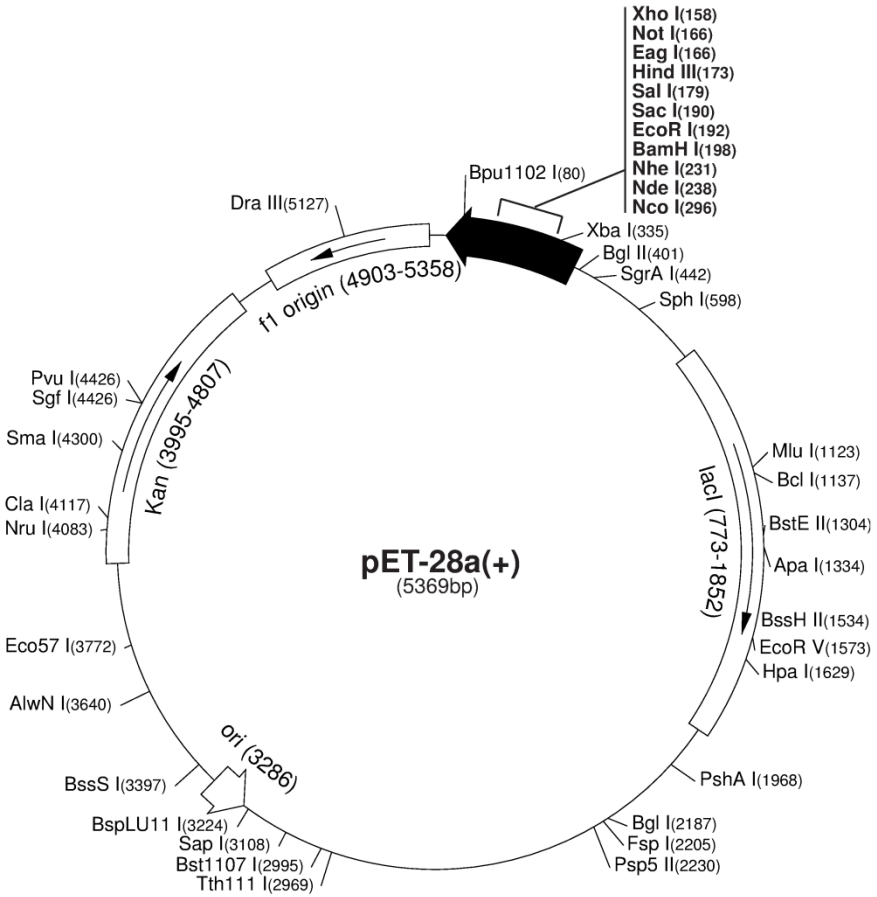


Figure S5.3 Plasmid DNA map of pET-28a expression vector.

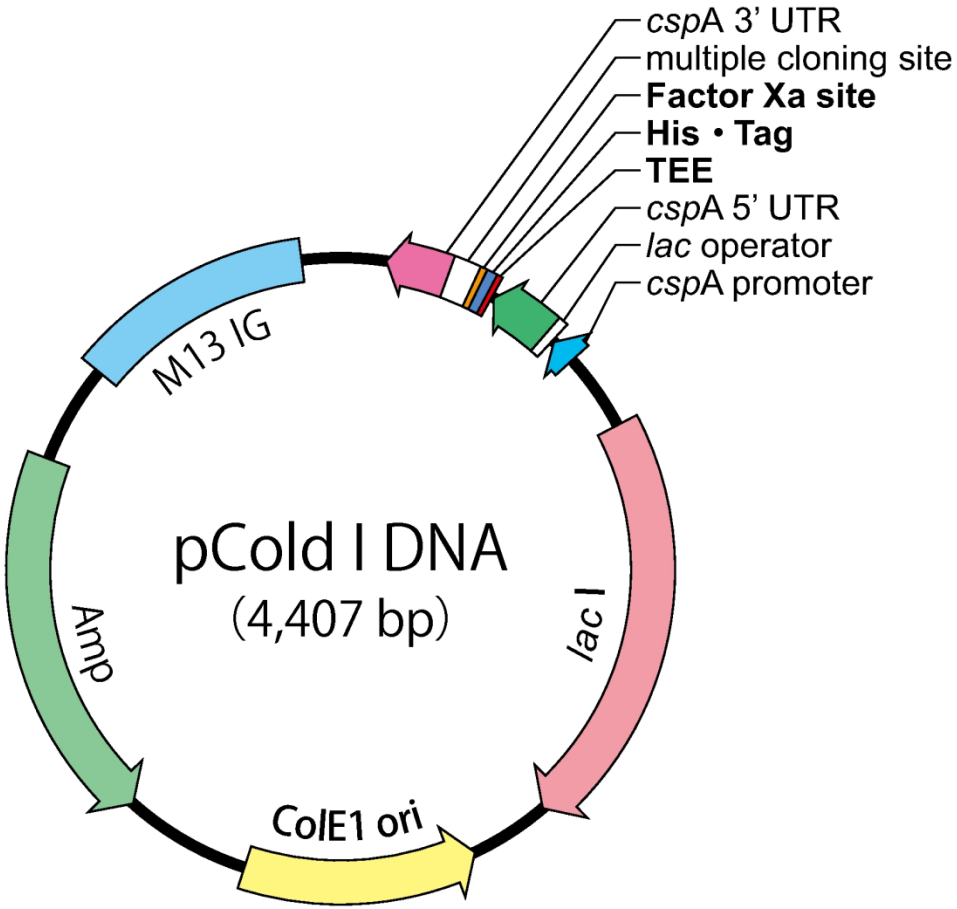


Figure S5.4 Plasmid DNA map of pCOLD expression vector.

Table S4.1 can be found in the CD attached.

Table S4.2. Proteins in carbohydrate metabolism and its regulation

Accession Number	Gene	Protein	Function	Reference info
Rv0062	*cel*A1	Cellulase	Endoglucanase for cellulose.	(Varrot et al. 2005)
Rv0126	*tre*S	Trehalose synthase	Glucan synthesis: isomerize trehalose to maltose	(Roy *et al*., 2013)
Rv0363c	*fba*	Fructose-1,6-biphosphate aldolase	Reversible cleavage of fructose 1,6-bisphosphate to dihydroxyacetone phosphate and glyceroldehyde 3-phosphate in glycolysis/gluconeogenesis.	(de la Paz Santangelo et al. 2011)
Rv0946c	*pgi*	Phosphoglucose isomerase	2nd enzyme of glycolysis, converting glucose-6-phosphate and fructose-6-phosphate	(Mathur et al. 2007) (Anand et al. 2010)
Rv0998-Rv1151c	Not assigned	Protein lysine acetyltransferase (PAT) and NAD^+-dependent deacetylase	Carbon metabolism regulator. NAD^+-dependent deacetylase to reverse PAT-acetylation.	(Lee et al. 2012)
Rv1098c	*fum*	Fumarase	Fumarate to L-malate hydration in tricarboxylic acid cycle	(Mechaly et al. 2012)
Rv1326c	*glg*B	α-(1,4)-glucan branching enzyme	Transfers oligosaccharide from α-(1,4)-glucosidic link to new α-(1,6)-branch (glucogen biosynthesis)	(Garg et al. 2007)
Rv1438	*tpi*A	Triosephosphate isomerase	Converts DHAP to D-GAP in gluconeogenesis.	(Connor, et al., 2011)

Table S4.3. Proteins in amino acid metabolism

Accession	Gene	Protein Name	Function	Reference

Number	Name			
Rv0858c	*dap*C	N-succinyldiaminopimelate aminotransferase	6th enzyme in lysine biosynthesis from L-aspartate	(Weyand, Kefala & Weiss 2006) (Weyand, Kefala & Weiss 2007)
Rv0884c	*Ser*C	Phosphoserine aminotransferase	2nd enzyme in phosphorylated serine biosynthesis	(Coulibaly, et al., 2012)
Rv0948c	Not assigned	Intracellular chorismate mutase	Chorismate to prephenate conversion in Shikimate aromatic amino acid pathway	(Kim et al. 2008) (Sasso et al. 2009)
Rv1885c	Not assigned	Secreted chorismate mutase		(Qamra et al. 2005) (Qamra et al. 2006) (Kim et al. 2006)
Rv1187	*pru*A	NAD^{+}-dependent Δ-pyrroline-5-carboxylic dehydrogenase	2nd enzyme in the conversion of proline to glutamate in proline utilizatiom system	(Lagautriere *et al*., 2014)
Rv1201c	*dap*D	Tetrahydrodipicolinate N-succin-yltransferase	diaminopimelate (DAP) pathway to L-lysine: 5th step	(Schuldt et al. 2009) (Schuldt et al. 2008)
Rv1237/ Rv1293	*lys*A	Diaminopimelate decarboxylase	L-lysine biosynthesis: Last step, duplicated gene	(Gokulan et al. 2003) (Kefala, Perry & Weiss 2005) (Weyand et al. 2009)
Rv1335-Rv1336	*cys*O-*cys*M	Sulfur carrier protein complex OM	Cysteine biosynthesis: CysM replaces O-acetylserine acetyl by Cys-O thiocarboxylate. Resulting cysteine released proteolytically.	(Jurgenson et al. 2008)
Rv1603	*his*A	Histidine/Tryptophan biosynthesis isomerase	Histidine/tryptophan biosynthesis: Metabolite isomerization	(Due et al. 2011)
Rv1652	*arg*C	N-acetyl-γ-glutamyl-phosphate reductase (putative)	Arginine biosynthesis: 3rd step	(Cherney et al. 2007) (Moradian et al. 2006)
Rv1656	*arg*F	Ornithine carbamoyltransferase	Arginine biosynthesis: 6th step	(Moradian et al. 2006) (Sankaranarayanan et al. 2008)
Rv1826	*gcv*H	Glycine cleavage system protein H	Glycine cleavage system: intermediate and electron shuttling.	(Abendroth et al. 2010)
Rv2121	*his*G	N-1-(5'-phosphoribosyl)-ATP transferase	Histidine biosynthesis: 1st step	(Cho, Sharma & Sacchettini 2003)
Rv2122c	*his*E	Phosphoribosyl-ATP pyrophos-phohydrolase	Histidine biosynthesis: 2nd step	(Javid-Majd et al. 2008)

Rv2178c	*aro*G	3-Deoxy-D-arabino-heptulo-sonate-7-phosphate synthase	Shikimate pathway of aromatic amino acids biosynthesis: 1st step	(Webby et al. 2005) (Webby et al. 2010) (Jiao et al. 2012)
Rv2192c	*trp*D	Anthranilate phosphoribosyltransferase	Tryptophan biosynthesis: 2nd step	(Lee et al. 2006)
Rv2220	*gln*A1	Glutamine synthetase	Binds ammonia to glutamate to produce glutamine	(Krajewski, Jones & Mowbray 2005)
Rv2391	*nir*A	Siroheme-dependent sulphite reductase	Reduces sulphite to sulphide in sulphur-containing amino acids and cofactors.	(Schnell et al. 2005)
Rv2537c	*aro*D	3-dehydroquinate dehydratase	Shikimate/quinate pathways: 3-dehydroquinate dehydration to 3-dehydroshikimate	(Evans et al. 2002)
Rv2539c	*aro*K	Shikimate kinase	Chorismate biosynthesis: 5th step	(Hartmann et al. 2006)
Rv2753	*dap*A	Dihydrodipicolinate synthase	L-lysine biosynthesis: 1st step	(Kefala & Weiss 2006)
Rv2773c	*dap*B	Dihydrodipicolinate reductase	L-lysine biosynthesis: 2nd step	(Kefala et al. 2005)
Rv2780	*ald*	L-alanine dehydrogenase	L-alanine to pyruvate reversible oxidative deamination	(Tripathi & Ramachandran 2008)
Rv2995c	*leu*B	3-isopropylmalate dehydrogenase	Leucine biosynthesis: 3rd step	(Singh et al. 2005)
Rv3290c	*lat*	Lysine ε-aminotransferase	L-lysine ε-amino transfer to α-ketoglutarate to yield L-glutamate and α-aminoadipate-δ-semi-aldehyde	(Tripathi & Ramachandran 2006)
Rv3710	*leu*A	α-isopropylmalate synthase	Leucine biosynthesis: 1st step	(Huisman et al. 2012)

Table S4.4. Proteins in lipid and polyketide metabolism

Accession	Gene	Name of Protein	Function	Reference

Number	Name			
Rv0099-Rv0100	*fad*D10-Rv0100	Fatty acyl-AMP ligase and acyl carrier protein	Lipopeptide biosynthesis: transfer of fatty-acyl chains to Rv0100.	(Liu et al. 2013)
Rv0242c	*fab*G4	β-ketoacyl coA reductase	Disaggregated fatty acid synthesis acetoacyl-CoA reduction to 3-hydroxyacyl-CoA	(Dutta et al. 2011) (Dutta et al. 2013)
Rv0642c	*hma*	S-adenosylmethionine-dependent methyl-transferase	Keto- and methoxymycolate production by methylation.	(Boissier et al. 2006)
Rv0764c	*cyp*51	Cytochrome P450 51	Sterol biosynthesis in cell membrane	(Podust, Poulos & Waterman 2001)
Rv0824c-Rv1094	*des*A1-*des*A2	Acyl-ACP desaturase	Mycolic acid precursor synthesis: variable double bonds for diverse mycolic acids	(Dyer et al. 2005)
Rv0855	*far*	Fatty acid-CoA racemase	Fatty acid metabolism: (2R)-branched-chain fatty acid-CoA to (2S)-stereoisomer conversion	(Rhee et al. 2005)
Rv0859-R0860	*fad*A-*fad*B	Acyl-CoA thiolase fatty acid oxidation protein	β-Oxidation lipid degradation: trifunctional enzyme	(Venkatesan & Wierenga 2013)
Rv1143	*mcr*	α-methylacyl-coA racemase	Interconversion of (2R)- and (2S)-methylacyl-CoA.	(Savolainen et al. 2005) (Bhaumik et al. 2007)
Rv1208	*gpg*S	Glucosyl-3-phosphoglycerate synthase	Methylglucose lipopolysaccharide biosynthesis: 1st glucosylation	(Gest et al. 2008) (Urresti et al. 2012)
Rv1256c	*cyp*130	Cytochrome P450 CYP130	Unknown function. Binds primary arylamines	(Ouellet, Podust & Ortiz de Montellano 2008)
Rv1284-Rv3588c	*can*A-*can*B	Carbonic anhydrases	Fatty acid biosynthesis: hydration of CO_2 to bicarbonate (reversible)	(Covarrubias et al. 2005) (Covarrubias et al. 2006)β
Rv1372	*pks*18	Type III polyketide synthase 18	Aliphatic long-chain acyl-CoA starter unit conversion to tri- and tetraketide pyrone precursors for cell wall lipid synthesis by fatty acid synthases.	(Sankaranarayanan, et al., 2004)
Rv1411c	*lpr*G	Conserved lipoprotein LprG	Binds glycolipid to enhance Toll-Like Receptor 2 recognition of triacylated glycolipids. Potential support of glycolipid transport in bacterial cell wall biogenesis	(Drage et al. 2010)

Rv1483	*fab*G1	β-ketoacyl reductase	Long-chain fatty acid biosynthesis: NADPH-dep. reduction of long-chain β-ketoacyl derivatives.	(Cohen-Gonsaud et al. 2002)
Rv1484	*inh*A	NADH-dependent 2-trans-enoyl-ACP reductase	Mycolic acid elongation. Drug target of isoniazid (INH).	(He et al. 2006)
Rv2217	*lip*B	Octanoyl-[acyl carrier protein]-protein acyltransferase	Lipolyation: octamoic acid transfer to lipoyl domains for derivative lipolation by LipA.	(Ma et al. 2006)
Rv2243-Rv0649	*fab*D-*fab*D2	Malonyl coA-acyl carrier protein transacylase	Fatty acid biosynthesis: malonate transfer from malonyl-CoA to holo-ACP	(Li et al. 2007) (Ghadbane et al. 2007)
Rv2244	*acp*M	Acyl carrier protein M, AcpM	Type II fatty acid synthesis: Intermediate shuttling.	(Wong et al. 2002)
Rv2245	*kas*A	β-ketoacyl-ACP synthase (KAS)	Fatty acid biosynthesis: condense acyl-AcpM and malonyl-AcpM in each elongation cycle of the FAS-II pathway	(Luckner *et al.*, 2009)(Schiebel *et al.*, 2013)
Rv2266	*cyp*124	Cytochrome P450 124	Hydroxylates methyl-branched lipids at chemically disfavoured ω-position	(Johnston et al. 2009)
Rv2276	*cyp*121	Cytochrome P450 121	Unknown function: Possibly involved in lipid or polyketide metabolism	(McLean et al. 2008)
Rv2498c	*cit*E	Citrate lyase β-subunit	Fatty acid biosynthesis: Protein bound citryl-CoA cleavage to oxaloacetate and protein-CoA derivative.	(Goulding et al. 2007)
Rv2701c	*suh*B	Inositol monophosphatase (IMPase)	Myo-inositol-I-phosphate dephosphorylation to myo-inositol for cell-wall phosphatidyl inositol (PI) synthesis.	(Brown et al. 2007)
Rv2870c	*dxr*	1-Deoxy-D-xylulose-5-phosphate reductoisomerase	Isopentenyl diphosphate synthesis: 2nd step. Precursor of isoprenoids such as steroid hormones, carotenoids, cholesterol and ubiquinone	(Henriksson et al. 2007)
Rv2941	*fad*D28	Fatty acyl-AMP ligase D28	Fatty acyl activation and transfer to polyketide synthases for complex lipid production	(Goyal et al. 2012)

Rv3089	*fad*D13	Fatty acid CoA ligase 13	Binds CoA to fatty acids	(Goyal et al. 2012)
Rv3279c	*bir*A	Biotin acetyl-CoA carboxylase ligase	Lipid biosynthesis: 1st step. Biotinylates acetyl-CoA carboxylase	(Gupta et al. 2010)
Rv3280	*acc*D5	Acyl-CoA carboxylase	Mycolic acid synthesis: acyl-CoA carboxylation to produce extender subunits for fatty acids and polyketide natural product biosynthesis.	(Holton et al. 2006)
Rv3378c	Not assigned yet	Tuberculosinyl transferase	Biosynthesis of 1-tuberculosinyladenosine (1-TbAd)	(Layre *et al*., 2014)
Rv3518c	*cyp*142	Cytochrome P450 142	Sterol oxidation: Converts cholesterol to 5-cholestenoic acid or cholest-4-en-3-one to cholest-4-en-3-one-27-oic acid.	(Driscoll et al. 2010)
Rv3545c	*cyp*125A1	Cytochrome P450 125A1	Cholesterol side-chain transfer to cellular lipids.	(McLean et al. 2009)
Rv3566c	*nho*A	Arylamine N-acetyltransferase	Cholesterol catabolism: Links n-propionyl-CoA to cholesterol metabolism	(Abuhammad, et al., 2013)
Rv3569c	*hsa*D	2-Hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase	Cholesterol catabolism: Hydrolyses DSHA.	(Lack et al. 2008) (Lack et al. 2010)
Rv3582c	*isp*D	Cytidyl transferase	Isoprenoid biosynthesis (MEP pathway): 3rd step.	(Bjorkelid, et al., 2011)

Table S4.5. Proteins in nucleotide metabolism and modification

Accession Number	Gene Name	Protein Name	Function	Reference
Rv0233	*nrd*B	Mn/Fe R2-like ligand-binding oxidase protein	Deoxyribonucleotide biosynthesis	(Andersson & Hogbom 2009) (Andersson, Berthold & Hogbom 2012)
Rv0321	*dcd*	Deoxycytidine triphosphate deaminase	Deoxycytodine triphosphate (dCTP) deamination to deoxyuridine triphosphate (dUMP).	(Helt et al. 2008)
Rv0733	*adk*	Adenylate kinase	ATP terminal phosphate transfer to AMP producing 2 ADP.	(Bellinzoni et al. 2006)
Rv0805	Not assigned	Cyclic nucleotide phosphodiesterase	Cleaves cNMPs to NMPs	(Shenoy et al. 2006)
Rv0957	*pur*H	5-amino-imidazole-4-carboxamide ribonucleotide transformylate/inosine monophosphate cyclohydrolase (ATIC)	Purine biosynthesis: Last two steps	(Le Nours et al. 2011)
Rv1900c	lipJ (mis-annotated)	Adenylyl cyclase	cAMP and cGMP synthesis	(Sinha et al. 2005)
Rv2179c	Not Assigned yet	Exoribonuclease	Possess exoribonuclease activity, cleaving 3’ single strand overhangs of duplex RNA. It was annotated as hypothetical protein	(Abendroth *et al*., 2013)
Rv2465c	*rpi*B	Ribose-5-phosphate isomerase	Nucleotide biosynthesis: Ribose-5-phosphate/ribulose-5-phosphate	(Roos et al. 2004) (Roos et

			interconversion.	al. 2008)
Rv2613c	Not assigned	Diadenosine 5',5'''-P1,P4-tetraphosphate phosphorylase	Diadenosine polyphosphate (ApnA) conversion to ATP and ADP.	(Mori et al. 2011)
Rv2754c	*thy*X	Flavin-dependent thymidylate synthase	Thymidylate biosynthesis: Interconversion of dUMP/dTMP.	(Sampathkumar et al. 2005) (Sampathkumar et al. 2006)
Rv2763c	*dfr*A	Dihydrofolate reductase	Purines, thymidylic acid and amino acid biosynthesis: Converts dihydrofolate to tetrahydrofolate. Target for isoniazid.	(Argyrou, et al., 2006)
Rv2793c	*tru*B	ψ55 tRNA pseudouridine synthase	Modifies uridine in tRNA, rRNA or snRNA to pseudouridine	(Chaudhuri et al. 2004)
Rv2883c	*pyr*H	Uridine monophosphate kinase	Converts UMP plus ATP to UDP plus ADP.	(Labesse et al. 2011)
Rv2966c	*rsm*D	RsmD-like methyltransferase	Nucleotide methylation	(Kumar et al. 2011)
Rv3628	*ppa*	Inorganic pyrophosphatase	pyrophosphate to orthophosphate conversion	(Tammenkoski, et al., 2005) (Benini & Wilson, 2011)

Table S4.6. Proteins in glycan biosynthesis and metabolism

Accession Number	Gene Name	Protein Name	Function	Reference
Rv1018c	*glm*U	Probable UDP-N-acetylglucosa-mine pyrophosphorylase	Peptidoglycan and lipid A synthesis: Last 2 steps in UDP-N-acetyl-D-glucosamine synthesis	(Zhang et al. 2009)
Rv1449c	*tkt*	Transketolase	Cell wall synthesis: Links non-oxidative pentose phosphate pathway to pentose sugar (arabinan) biosynthesis	(Fullam et al. 2011)
Rv1477	*rip*A	Peptidoglycan endopeptidase	Cleaves peptidoglycan peptide crosslinks during cell division.	(Ruggiero et al. 2010)
Rv1566c	*rip*D	Peptidoglycan hydrolase	Unlike other peptidoglycan hydrolases, ripD has evolved to a non-catalytic PG-binding PH.	(Böth *et al*., 2014)
Rv2158c	*mur*E	UDP-N-acetylmuramoylalnyl-D-glytamate-2,6-diaminopimelate ligase	Cell wall synthesis: Incorporates amino acids such as m-DAP and L-lysine into peptidoglycan.	(Basavannacharya et al. 2010) (Basavannacharya et al. 2010)
Rv3717	Not assigned yet	*N*-Acetylmuramyl-*L*-alanine amidase	Peptidoglycan fragment recycling: hydrolyzes muramyl dipeptide (MDP) releasing *N*-acetyl muramide and *L*-Ala-iso-*D*-Gln dipeptide products.	(Prigozhin *et al*., 2013)
Rv3808c	*glf*T2	Galactofuranosyltransferase	Cell wall synthesis: Generates galactan for mycolyl-arabinogalactan complex.	(Wheatley et al. 2012)

Rv3910	*mvi*N	Peptidoglycan biosynthetic pseudokinase	Peptidoglycan biosynthesis: Regulates FhaA by phosphorylation-dependent complex formation.	(Gee, et al., 2012)

Table S4.7. Proteins in metabolism of cofactors and vitamins

Accession Number	Gene Name	Protein Name	Function	Reference
Rv0548c	*men*B	1,4-dihydroxy-2-naphthoyl-CoA synthase	Menaquinone (vitamin K2) biosynthesis: 4th step	(Truglio et al. 2003)
Rv1092c	*pan*K	Pantothenate kinase	Coenzyme A (CoA) biosynthesis: 1st step. ATP-dependent phosphorylation of pantothenate.	(Das et al. 2005) (Chetnani et al. 2010) (Bjorkelid, et al., 2013)
Rv1207	*fol*P2	Putative dihydropteroate synthase	Folate biosynthesis	(Gengenbacher et al. 2008)
Rv1568	*bio*A	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase	Biotin biosynthesis: 2nd step	(Shi et al. 2011)
Rv1596	*nad*C	Quinolinic acid phosphoribosyltransferase	NAD biosynthesis.	(Sharma, Grubmeyer & Sacchettini 1998)
Rv2447c	*fol*C	Folylpolyglutamate synthase	Poly-glutamation of folate for improved cell retention	(Young et al. 2008)

			and affinity of folate-dependent enzymes.	
Rv3248c	*sah*H	S-adenosylhomocysteine hydrolase	Reversibly hydrolyses S-adenosylhomocysteine (SAH) to adenosine and L-homocysteine, to limit feedback inhibition of SAM-dependent methyltransferases	(Reddy et al. 2008)
Rv3607c	*fol*B	7,8-dihydroneopterin aldolase	Folate biosynthesis: 2nd step	(Goulding et al. 2005)
Rv3608c	*fol*P1	6-Hydroxymethyl-7,8-dihydro-pteroate synthase	Produces the folate precursor 6-hydroxymethyl-7,8-di-hydropteroate (H2PtPP).	(Baca et al. 2000)

Table S4.8. Protein in bacterial self-defense and detoxification

Accession Number	Gene	Protein	Function	Reference
Rv0253	*nir*D	Nitrite reductase	Nitrate assimilation/anaerobic respiration: Subunit of nitrite reductase to reduce nitrite to ammonia	(Izumi, Schnell & Schneider 2012)
Rv0262c	*aac*	Aminoglycoside 2'-N-acetyltransferase	Exact function unknown. CoA-dependently acetylates 2-hydroxy or amino group of aminoglycolsides and possibly mycothiol to regulate cellular redox potential	(Hedge, Javid-Majd & Blanchard 2001) (Vetting et al. 2002)
Rv0407	*fgd*1	F420-dependent glucose-6-phosphate dehydrogenase	Exact function unknown. Reduces F420 for low-redox potential reactions during anaerobic survival and persistence. Rv3547 uses reduced F420 to activate prodrug PA-824.	(Bashiri et al. 2008)

Rv0467	*icl*	Isocitrate lyase	Diverts acetyl-coA from fatty acid β-oxidation to glyoxylate shunt pathway for net carbon gain.	(Sharma, et al., 2000)
Rv0819	*msh*D	Mycothiol synthase	Mycothiol biosynthesis: Last step. Transfers acetyl to the cysteinyl amine of cysteine-glucosamine-inositol.	(Vetting et al. 2003) (Vetting et al. 2006)
Rv1876	*bfr*A	Bacterioferritin	Iron storage and detoxification: Critical for adaptation and survival of tubercule bacilli in the host.	(Gupta et al. 2009)
Rv1908c	*kat*G	Catalase-peroxidase	Convert pro-drug isoniazid to its active form so it will interfere with the enzyme inhA which is involved in Fatty Acid Synthase II.	(Zhao *et al*., 2013)
Rv1938	*eph*B	Epoxide hydrolase B	Hydrolyses epoxides to trans-dihydrodiols during drug metabo-lism and signalling molecule processing.	(Biswal et al. 2006) (Biswal et al. 2008)
Rv2043c	*pnc*A	Pyrazinamidase	Activates first-line drug pyrazinamide converting it to pyrazinoic acid to inhibit fatty acid synthase I	(Petrella et al. 2011)
Rv2068c	*bla*C	β-Lactamase BlaC	Hydrolyzes β-lactam related antibiotics	(Tremblay, Fan & Blanchard 2010)
Rv2275	*alb*C	Cyclodityrosine synthetase	Converts 2 Tyr-tRNA to cyclo(L-Tyr-T-Tyr) of potential anti-bacterial, antifungal and antitumor activity.	(Vetting, Hedge & Blanchard 2010)
Rv2416c	*Eis*	Enhanced intracellular survival protein	It is a structurally and functionally unique acetyltranferase which unusually acetylates aminoglycosides at multiple positions, conferring drug resistance against kanamycin for *M. tuberculosis*.	(Chen *et al*., 2011)(Houghton *et al*., 2013)

Rv2428	*ahp*C	Alkyl hydroperoxidase, AhpC	Part of NADH-dependent peroxidase and peroxynitrite reductase.	(Guimaraes et al. 2005)
Rv2429	*ahp*D	Alkyl hydroperoxidase AhpD	Same function as AhpC despite unrelated amino acid sequence	(Nunn et al. 2002) (Koshkin et al. 2003)
Rv2461c	*clp*P1	Caseinolytic protease	Degrades misfolded, short-lived, dysfunctional, denatured and aggregated proteins.	(Ingvarsson, et al., 2007)
Rv2579	*dha*A	Haloalkane dehalogenase	Cleaves carbon-halogen bonds of halogenated alphatic hydrocarbons yielding a halide anion and an aliphatic alcohol.	(Mazumdar et al. 2008)
Rv2740	*eph*G	Epoxide hydrolase B EphG	Detoxifies and processes signalling molecules by converting reactive epoxides to diols.	(Johansson et al. 2005)
Rv3682	*Pon*A2	Penicillin binding protein	Its C-terminal PASTA domain has untypical binding ability compared to other PASTA domains	(Calvanese *et al*., 2013)

Table S4.9. Proteins involved in transcriptional regulation

Accession number	Gene	Protein	Function	Reference
Rv0300-Rv0301	*vap*B3-*vap*C3	Toxin-antitoxin system VapBC-3	VapC: endoribonuclease cleaving mRNA to stop cell growth. VapB inhibits VapC.	(Min et al. 2012)
Rv0626-Rv0627	*vap*B5-*vap*C5	Toxin-antitoxin system VapBC-5	As VapBC-3	(Miallau et al. 2009)

Rv0678	*mar*R	Transcriptional regulator MarR	Transcriptional regulation of MmpS5-MmpL5 transporter	(Radhakrishnan *et al*., 2014)
Rv0735-Rv0736	*sig*L-*rsl*A	Sigma factor L and anti-sigma factor RslA	Transcriptional regulation of cell envelope and secreted protein modification genes.	(Thakur, Praveena & Gopal 2010)
Rv1219c	Not assigned yet	TetR-like transcriptional regulator	Transcriptional regulation of multidrug efflux system Rv1217c-Rv1218c	(Kumar *et al*., 2014)
Rv1267c	*emb*R		Putative arabinosyltransferase transcriptional regulator.	(Alderwick et al. 2006)
Rv1657	*arg*R	Arginine repressor	Arginine biosynthetic transcription regulator.	(Lu et al. 2007) (Cherney et al. 2010) (Cherney et al. 2008) (Cherney et al. 2009) (Cherney et al. 2008)
Rv1846c	blaI	β-Lactamase operon transcriptional repressor	Transcriptional repression of β-lactamase genes	(Sala et al. 2009)
Rv1985c	*arg*P	Arginine permease	Transcription regulator for arginine export and transport/metabolism of other amino acid	(Zhou et al. 2010)
Rv1994c	*cmt*R	Cd^{2+}/Pb^{2+}-sensing ArsR-S*M. tuberculosis* repressor	Regulator of P1-type ATPase for export of toxic metal ions.	(Cavet et al. 2003)
Rv2069	*sig*C	Sigma factor C	Sigma C-dependent DNA transcription regulator virulence associated genes such as *hsp*X, *sen*X3 and *mtr*A	(Thakur, Joshi & Gopal 2007)

Rv3291c	*lrp*A	Leucine-responsive global regulator	Global transcriptional regulator during starvation and stationary phase transition	(Shrivastava & Ramachandran 2007) (Reddy et al. 2008) (Shrivastava, Dey & Ramachandran 2009)
Rv3849	*esp*R	Virulence regulator	ESX-1 secreted activator of espACD operon of virulence associated ESX-1 type VII secretion system;	(Blasco et al. 2011)
Rv3854c	*eth*R	TetR/CamR family repressor	*eth*A repressor, preventing ethionamide (ETH) activation by EthA, conferring drug resistance.	(Dover et al. 2004) (Carette et al. 2011)

Table S4.10. Proteins in protein synthesis

Accession Number	Gene	Protein	Function	Reference
Rv0009	*ppi*A	Peptidyl-prolyl cis-trans isomerase A	Peptide cis-trans isomerase for protein folding, signalling, cell surface recognition, chaperoning and heat-shock response	(Henriksson et al. 2004)
Rv1014c	*pth*	Peptidyl-tRNA hydrolase	Cleaves peptides from tRNA to avoid peptidyl-tRNA toxicity	(Selvaraj et al. 2007) (Pulavarti et al. 2008)
Rv2969c	*Dsb*A	Disulfide bond forming protein	It facilitates proper folding and disulfide bond formation of periplasmic and secreted proteins.	(Chim *et al*., 2013)
Rv3417c	*gro*EL1	Chaperonin	Facilitates protein folding.	(Sielaff, Lee & Tsai 2010) (Sielaff, Lee & Tsai 2011)

Rv3462c	*inf*A	Translation initiation factor I	IF1/2/3 constitute the translation initiation complex	(Hatzopoulos & Mueller-Dieckmann 2007) (Hatzopoulos & Mueller-Dieckmann 2010)

Table S4.11. Proteins in DNA replication and repair

Accession Number	Gene	Protein	Function	Reference
Rv0002	*dna*N	DNA polymerase IIIβ sliding clamp	Component of DNA polymerase III for genome replication	(Gui et al. 2011) (Kukshal et al. 2012)
Rv0005-Rv0006	*gyr*B-*gyr*A	DNA gyrase	DNA topology regulation. Drug target for quinolone.	(Piton et al. 2010) (Tretter & Berger 2012)
Rv0058	*dna*B	DNA helicase	ATP-dependent helicase unwinding parental duplex DNA during replication	(Biswas & Tsodikov 2008)
Rv0938	*lig*D	Non-homologous end-joining (NHEJ) DNA repair polymerase	Catalyses NHEJ-mediated DNA repair	(Gong et al. 2004) (Pitcher et al. 2005) (Akey et al. 2006) (Pitcher et al. 2007) (Brissett et al. 2007)
Rv1316c	*ogt*	O^6-methylguanine-DNA methyltransferase	Repairs O^6-alkylguanine in double-stranded DNA preventing G:C to A:T transition mutations.	(Miggiano et al. 2013)
Rv2593c	*ruv*A	Holliday junction DNA helicase	Processes/resolves Holliday junctions with ruvB and ruvC for homologous recombination orientated DNA repair.	(Prabu et al. 2006)
Rv2976c	*ung*	Uracil-DNA glycosylase	First step in uracil excision-repair pathway.	(Singh et al. 2006) (Kaushal et al. 2010)
Rv3053c	*nrdH*	NrdH. Probable glutaredoxin electron transport component	Provides electrons for ribonucleotide reduction (RR) for DNA replication and repair	(Phulera & Mande, 2013)

Table S4.12. Proteins in cellular information processing and substrate transport

Accession Number	Gene	Protein	Function	Reference
Rv0014c	*pkn*B	Serine/threonine kinase B	Unknown substrates. Regulates mycobacterial life cycle.	(Ortiz-Lombardia et al. 2003) (Barthe et al. 2010)
Rv0491	*reg*X3	Transcriptional regulatory protein RegX3	Part of two component system RegX3/SenX3	(King-Scott et al. 2007)
Rv0757	*pho*P	Transcriptional regulatory protein PhoP	Part of two component system PhoP/PhoQ. A probable positive regulator for phosphate regulon for intracellular growth	(Wang, Engohang-Ndong & Smith 2007)
Rv0899	*omp*A	Outer membrane protein A	Pore-forming protein for uptake of water-soluble compounds.	(Li et al. 2012)
Rv0934	*pst*S1	Phosphate-specific transport substrate binding protein-1	Primary extracellular receptor for the ABC phosphate transporter regulating phosphate uptake.	(Vyas, Vyas & Quiocho 2003)
Rv1266c	*pkn*H	Serine/threonine kinase H	Phosphorylation by PknH enhances Rv1267 binding to the promoter of the embCAB arabinosyltranferase genes that promote cell wall synthesis. PknH also phosphorylate kasA, kasB and MtFabH promoting mycolic acid synthesis	(Cavazos, Prigozhin & Alber 2012)
Rv1626	Not assigned	Putative transcriptional	It is the sensor part of a two component system. It provides the linkage between the phosphorylation cascade of the two component	(Morth et al. 2004)

	yet	antiterminator	systems with the antitermination event in the transcriptional machinery	
Rv1743	*pkn*E	Serine/threonine kinase E	The function of PknE is unknown. But *in vitro* the activated kinase domain phosphorylate Rv1747, an ABC transporter of unknown specificity, and GarA, a conserved hypothetical protein	(Gay, Ng & Alber 2006)
Rv2027c	*dos*T	Histidine kinase DosT	The N-terminal sensory domain of DosT directly binds to oxygen gas, NO, and CO. NO or CO binding to or O_2 dissociate from the N-terminal GAF domain initiates autophosphorylation by the C-terminal histidine kinase domain, with subsequent phosphoryl transfer to the Asp residue of the transcriptional regulator DosR. DosT functions as a hypoxia sensor, mediating an early *M. tuberculosis* response on the host immune factor (NO).	(Sardiwal et al. 2005) (Podust, Ioanoviciu & Ortiz de Montellano 2008)
Rv2386c	*mbt*I	Salicylate synthase	The 1st enzyme in the siderophore mycobactin synthetic pathway, catalysing the 1st step from chorismate to salicylate	(Harrison et al. 2006)
Rv2919c	*gln*B	Nitrogen regulatory PII protein	It acts as nitrogen sensor. The precise function is unknown. It might play a key role in the nitrogen regulatory pathway	(Shetty et al. 2010)
Rv3132c	*dos*S	Histidine kinase DosS	DosS is sensor kinase, responsible for recognition of hypoxia in *M. tuberculosis*. It is structurally similar to DosT but DosS is a redox sensor while DosT is direc oxygen sensor	(Cho et al. 2009) (Cho et al. 2011) (Cho et al. 2013)
Rv3220	*pdt*aS	Histidine kinase PdtaS	PdtaS is cytosolic sensor protein. The two component system PdtaR/PdtaS structurally and functionally is equivalent to the EutW/EutV TCS involved in regulatory ethanolamine catabolism in many gram positive bacteria	(Preu et al. 2012)

Rv3676c	*crp*	Cyclic AMP receptor protein	It acts as a sensor o cAMP level in cells. cAMP mediates a large variety of cellular signalling processes	(Akif et al. 2006) (Gallagher et al. 2009) (Kumar et al. 2010)

Table S4.13. Proteins that are recognized as virulence factors

Accession Number	Gene Name	Protein Name	Function	Reference
Rv0153c	*ptp*B	Tyrosine phosphate	It is secreted into the host cells where it disrupts unidentified signalling pathways	(Grundner, Ng & Alber 2005) (Grundner et al. 2007)
Rv0198c	*zmp*1	Zinc-dependent metalloprotease-1	It plays a key role in the process of phagosome maturation inhibition and emerged as an important player in pathogenesis	(Ferraris et al. 2011)
Rv2111c	*mpa*	Mycobacterium proteasome	Upon binding to ubiquitin-like protein, Mpa delivers proteins into the mycobacterial proteasome for destruction	(Wang, Darwin & Li 2010)
Rv2430c-Rv2431c	*pe*25-*ppe*41	PE25 and PPE41	The function is unknown. They are cell wall associated, immunoantigenic, secreted virulence factors	(Strong et al. 2006)
Rv2945c	*lpp*X	Probable conserved lipoprotein	It is required for the translocation of phthiocerol dimycocerosates (DIM) from cell wall layer to outside OM. The exact function is still unknown. Attenuation is observed when knocked out in mouse model	(Sulzenbacher et al. 2006)

Rv3361c	Not assigned yet	Fluoroquinolone resistance protein	It binds to DNA gyrase and inhibits its activity. Such inhibition causes the resisitance to ciprofloxacin and sparfloxacin (fluoroquinolone)	(Hedge et al. 2005)
Rv3671c	Not assigned yet	Putative serine protease	It maintains the bacteria near neutral intrabacterial pH. It confers resistance to acidification and protection from oxidative stress	(Biswas et al. 2010)
Rv3874-Rv3875	*cfp*10-*esat*6	Culture filtrate protein 10-Early secretory antigenic target 6	They are secreted antigenic virulence factors. They are believed to form larger complex with Esp protein upon secretion via ESX-1	(Renshaw et al. 2005)
Rv3914	*trx*	Thioredoxin C	It regulates cellular redox homeostasis, enabling resistance to host cell phagocytosis	(Hall et al. 2011)

Table S4.14. Hypothetical protein and proteins with unrecognized possible functions

Accession Number	Gene Name	Protein Name	Possible function	Reference
Rv0020c	*fha*A	Conserved protein with FHA domain	It might be linked to cell wall biosynthesis regarding the functions of its neighbouring genes in the genome	(Roumestand et al. 2011)
Rv0130	*htd*Z	Probable 3-hydroxyl-thioester dehydratase	It has R-specific hydratase motif, hydrating a trans-2-enoyl-coA moiety	(Johansson et al. 2006)
Rv0216	Not assigned yet	Double hotdog hydratase	It is able to catalyse the R-specific hydration of 2-enoyl coenzyme A	(Castell et al. 2005)

Rv0760c	Not assigned yet	Resemble bacterial ketosteroid isomerases	It catalyzes the allylic rearrangement of Δ5-3-ketosteroids to Δ4-3-ketosteroids. It might be a steroid binding protein	(Cherney, Garen & James 2008)
Rv0802c	Not assigned yet	Possible succinyltransferase	It belongs to GNC5-related N-acetyltransferase family. Substrate is unknown	(Vetting, Errey & Blanchard 2008)
Rv0813c	Not assigned yet	Resemble eukaryotic fatty acid binding protein	It plays a role in the recognition, transport, and/or storage of small molecules	(Shepard et al. 2007)
Rv1155	Not assigned yet	Resemble flavin mononucleotide (FMN)-binding protein	It has no binding activity towards FMN. It might have evolved a new unknown function	(Canaan et al. 2005)
Rv1873	Not assigned yet	Unknown	Unknown	(Garen et al. 2006)
Rv1980c	*mpt*64	Immunogenic protein Mpt64	It contains a protein-protein interaction motif. Function is unknown. It might be a virulence factor	(Wang et al. 2007)
Rv2140c	Not assigned yet	Unknown	It shares strong structural homology with the phos- phatidylethanolamine-binding family of proteins. It might be involved in lipid metabolism	(Eulenburg, et al., 2013)
Rv2175c	Not assigned yet	Unknown	It is a DNA binding protein. It can be phosphorylated by Ser/Thr kinase. It might act as a transcriptional regulator	(Cohen-Gonsaud et al. 2009)
Rv2543	*lpp*A	Probable conserved lipoprotein	It might play a role in host-pathogen interaction	(Grana et al. 2010)
Rv2704	Not assigned	Putative σ^A regulator	It might be involved in metabolic interactions	(Thakur, Praveena

	yet			& Gopal 2010)
Rv2714	Not assigned yet	Unknown	A representative of a new nucleoside phosphorylase-like family of duplicated actinobacterial proteins	(Grana et al. 2009)
Rv2827c	Not assigned yet	Unknown	It has a unique fold. It might be a DNA-binding protein	(Janowski et al. 2009)
Rv3214	*gpm*2 (mis-annotated)	A broad specificity phosphatase	It is important for mycobacterial phosphate metabolism. Exact function is unknown	(Watkins & Baker 2006)
Rv3547	ddn	Deazaflavin-dependent nitroreductase	It catalyzes the reduction of nitroimidazoles such as PA-824, resulting in intracellular release of lethal reactive nitrogen species. Exact function is unknown	(Cellitti et al. 2012)

Table S4.15. Structural research groups involved in structural work on *M. tuberculosis* H37Rv strain proteins

Name of P. I.	Institution	Proteins investigated
Tom Alber	Department of Molecular and Cell Biology, University of California, Berkeley, USA	Ser/Thr Kinases PknH and PknE, Tyrosine phosphatase PtpB, Protein lysine acetyltransferase, pseudokinase (peptidoglycan biosynthesis) , N-acetylmuramyl L-analine amidase
Pedro, M. Alzari	Unite´ de Biochimie Structurale, Institut Pasteur, Paris, France	Adenylate kinase, lipoprotein LppA, peroxidase AhpC, fumarase, ser/thr kinase PknB, lactamase transcriptional regulator BlaI, fatty acid binding protein Rv0813, glucosyl-3-phosphoglycerate synthase

Kristina Bäckbro	Department of Cell and Molecular Biology, Uppsala University, Sweden	Hypothetical proteins Rv0130 and Rv0213
Edward, N. Baker	Maurice Wilkins Centre for Molecular Biodiscovery and School of Biological Sciences, University of Auckland, New Zealand	F420-dependent glucose-6-phosphate dehydrogenase FGD1, mycobactin biosynthetic enzyme MbtI, α-isopropylmalate synthase, purine biosynthetic enzyme PurE, TrpD (tryptophan biosynthesis), Phosphatase Rv3214, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, folypolyglutamate synthase, GlmU (UDP-GlcNAc synthesis), Phosphoserine aminotransferase (SerC, serine biosynthesis), NAD^+-dependent -1-pyrroline-5-carboxylic dehydrogenase (PruA)
Marco Bellinzoni	Unité de Microbiologie Structurale, Institut Pasteur, Paris, France	Adenylate kinase, lipoprotein LppA, Rv2714, Class II fumarase Rv1098c
James, M. Berger	Division of Biochemistry, Biophysics, and Structural Biology, Department of Molecular and Cell Biology, University of California, Berkeley, USA	DNA ligase D, DNA gyrase
Sanjib Bhakta	Institute of Structural and Molecular Biology, Department of Biological Sciences, University of London, UK	MurE (Peptidoglycan biosynthesis)
John S. Blanchard	Department of Biochemistry, Albert Einstein College of Medicine, New York, USA	aminoglycoside 2'-N-acetyltransferase (AAC(2')-Ic), Fluoroquinolone resistance protein, β-lactamase, succinyltransferase, cyclodityrosine synthetase, Mycothiol synthase, Dihydrofolate reductase (isoniazid target)
Yves Bourne	Architecture et Fonction des Macromole´cules Biologiques, CNRS, Marseille, France	Hypothetical protein Rv1155, lipoprotein LppX

Martin, Cohen-Gonsaud	Centre National de la Recherche Scientifique UMR 5048, Centre de Biochimie Structurale, Montpellier, France	Ser/Thr Kinase PknB, and its substrates Rv2175c and FhaA, MabA (long-chanin fatty acid biosynthesis)
Stewart T. Cole	Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	Virulence regulator EspR, transketolase, alkyl hydroperoxidase AhpC, CYP121, BlaI, fatty acid binding protein Rv0813c
Amit Kumar Das	Department of Biotechnology, Indian Institute of Technology, Kharagpur, West Bengal, India	Pantothenate kinase, β-ketoacyl coA reductase FabG4
Thomas Dick	Novartis Institute for Tropical Disease, Singapore	Ddn (deazaflavin-dependent nitroreductase), putative dihydropteroate synthase ortholog Rv1207
Snezana Djordjevic	Department of Biochemistry and Molecular Biology, University College London, UK	alkylhydroperoxidase AhpD, redox sensor DosS
Aidan, J. Doherty	genome damage and stability centre, University of Sussex, Brighton, UK	NHEJ polymerase, NHEJ DNA repair ligase
David Eisenberg	Howard Hughes Medical Institute, UCLA-Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, USA	Mycobacterial folate biosynthetic enzyme, Citrate lyase β-subunit CitE, toxin-antitoxin system VapBC5 and VapBC3, PE/PPE protein complex
Lindsay D. Eltis	Department of Microbiology and Immunology, Life Sciences Institute, University of British Columbia,	Steroid-degrading hydrolase HsaD, carbon-carbon hydrolase (cholesterol metabolism)

	Vancouver, Canada	
Klaus, Fűtterer	School of Biosciences, The University of Birmingham, UK	Ser/Thr kinase, Inositol monophosphatase, TetR/CamR family repressor EthR, mtFabD, Trehalose synthase (TreS) and Maltokinase (Pep2)
Lalit, C. Garg	Gene Regulation Laboratory, National Institute of Immunology, New Delhi, India	Phosphoglucose isomerase,
Balasubramanian Gopal	Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India	Sigma factor C, Rv2704, antisigma factor RslA
Marcelo E. Guerin	Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, USA	Fructose-1,6-bisphosphate aldolase, GpgS (methylglucose lipopolysaccharide biosynthesis), glucosyl-3-phosphoglycerate synthase
Kalervo J. Hiltunen	Biocenter oulu and Department of Biochemistry, University of Oulu, Finland	α-methyl-acyl-coA raemase,
Martin Hőgbom	Department of Cell and Molecular Biology, Uppsala University, Sweden	Mn/Fe R2-like ligand-binding oxidase protein (Mn/Fe R2lox protein), β-Carbonic Anhydrase
Wim, G. J. Hol	Howard Hughes Medical Institute, University of Washington, Seattle, USA	Thymidylate synthase ThyX, 6-Hydroxymethyl-7,8-dihydropteroate synthase (Folate biosynthesis)
Mary Jackson	Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, USA	Fructose-1,6-bisphosphate aldolase, GpgS (methylglucose lipopolysaccharide biosynthesis), lipoprotein LppX, glucosyl-3-phosphoglycerate synthase
Michael N. G.	Group in Protein Structure and Function,	Epoxide hydrolases A and B, Arginine repressor protein, N-acetyl-γ-

James	Department of Biochemistry, University of Alberta, Edmonton, Canada	glutamyl-phosphate reductase, Δ5-3-ketosteroid isomerase, hypothetical protein Rv1873, haloalkane dehalogenase, ArgC and ArgF (Arginine biosynthesis), ornithine carbamoyltranferase
Alwyn, T. Jones	Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Sweden	Hypothetical proteins Rv0130 and Rv0213, deazaflavin-dependent nitroreductase Ddn, β-carbonic anhydrase, transketolase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase, epoxide hydrolase glutamine synthetase, ribose-5-phosphate isomerase, cytidyl transferase IspD, pantothenate kinase
Beom Slk Kang	School of Life Science and Biotechnology, Kyungpook National University, Daegu, South Korea	Histidine kinase DosS
Jane E. Ladner	Biochemical Science Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, USA	Intracellular chorismate mutase, Secreted chorismate mutase
Shaun J. Lott	Centre for Molecular Biodiscovery and School of Biological Sciences, University of Auckland, New Zealand	Salicylate synthase (mycobactin biosynthesis), TrpD (lung colonization)
Shekhar, C. Mande	Centre for DNA Fingerprinting and Diagnostics, Nacharam, Hyderabad, India	cAMP receptor protein, chorismate mutase, NrdH, Probable glutaredoxin electron transport component
Claudine Mayer	Unite´ de Dynamique Structurale des Macromole´cules, De´partement de Biologie Structurale et Chimie, Institut	Glycine cleavage system protein H, DNA gyrase

	Pasteur, Paris, France	
Virginie Molle	Institut de Biologie et Chimie des Prote´ines, Universite´ Lyon, France	Response element for Ser/Thr kinase signalling EmbR, Ser/Thr kinase substrate Rv2175c, PknB interaction partner Rv0020c
Sherry L. Mowbray	Department of Molecular Biology, Biomedical Center, Swedish University of Agricultural Sciences, Uppsala, Sweden	carbonic anhydrase, peptidyl-prolyl cis-trans isomerase A, 1-deoxy-D-xylulose-5-phosphate reductoisomerase, epoxide hydrolase, glutamine synthetase, D-ribose-5-phosphate isomerase
Jochen Mueller-Dieckmann	EMBL Hamburg Outstation, Germany	Translation initiation factor I, DapB (Llysine biosynthesis), LeuB (leucine biosynthesis)
Hélène Munier-Lehmann	Institut Pasteur, Unite´ de Chimie et Biocatalyse, De´ partement de Biologie Structurale et Chimie, Paris, France	adenylate kinase, uridine monophosphate kinase
Andrew W. Munro	Manchester Interdisciplinary biocentre, faculty of Life Sciences, University of Manchester, UK	Cycochrome P450 CYP142, CYP121, CYP125
Paul R. Ortiz de Montellano	Department of Pharmaceutical Chemistry, University of California, San Francisco, USA	Enoyl acyl carrier protein reductase, Cytochrome P450 CYP124, CYP130, alkylhydroperoxidase AhpD, sensory histidine kinase DosT
Emily J. Parker	Centre of Structural Biology, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand	α-isopropylmalate synthase, 3-deoxy-D-arabinoheptulosonate 7-phosphate synthase (aromatic amino acid biosynthesis)

Florence Pojer	Global Health Institute, Ecole Polytechnique Federale de Lausanne, Switzerland	EspR, transketolase
Annaïk Quemard	Institut de Pharmacologie et de Biologie Structurale, Centre National de la Recherche Scientifique, Toulouse, France	S-adenosylmethionine-dependent methyltransferase, MabA (long chain fatty acid biosynthesis)
Florante A. Quiocho	Verna and Marrs McLean Department of Biochemistry and Molecular Biology and Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas, USA	ABC phosphate transport receptor
Ravishankar Ramachandran	Molecular and Structural Biology Division, Central Drug Research Institute (Council of Scientific and Industrial Research), Lucknow, Uttar Pradesh, India	DNA polymerase β-clamp, feast/famine regulatory protein Rv3291c, ε-aminotransferase, Alanine dehydrogenase
Zihe Rao	National Laboratory of macromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing, China	Arginine permease, MCAT (fatty acid biosynthesis)
Prasad T. Reddy	Biochemical Science Division, Chemical Science and Technology Laboratory,	cAMP receptor protein, intracellular chorismate mutase, secreted chorismate mutase

	National Institute of Standards and Technology, Gaithersburg, Maryland, USA	
Menico Rizzi	DISCAFF Department of Chemical, Food, Pharmaceutical and Pharmacological Sciences, University of Piemonte Orientale A. Avogadro, Novara, Italy	Zinc-dependent metalloprotease-1, O^6-methylguanine methyltransferase
Steven L. Roderick	Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York, USA	Fluoroquinolone resistance protein Rv3361, Aminoglycoside 2'-N-acetyltransferase, Mycothiol synthase
James C. Sacchettini	Department of Biochemistry and Biophysics, Texas A & M University, College Station, Texas, USA	ATP phosphoribosyltransferase, lipoprotein LprG, diaminopimelate decarboxylase (lysine biosynthesis), phosphoribosyl-ATP pyrophosphohydrolase, FadD10 (lipid metabolism), S-adenosyl-L-homocysteine hydrolase, leucine-responsive global regulator LrpA, quinolinic acid phosphoribosyltransferase, Nitrogen regulatory PII protein, Isocitrate lyase
Rajan Sankaranarayanan	Centre for Cellular and Molecular Biology, Council of Scientific and Industrial Research, Hyderabad, India	Fatty acyl-AMP ligase, ornithine carbamyoyltransferase
Gunter Schneider	Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden	Nitrite reductase, Sulfite reductase
Christopher J. Squire	School of Biological Sciences and Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland,	F420-dependent glucose-6-phosphate dehydrogenase FGD1, α-isopropylmalate synthase, GlmU (peptidoglycan and lipid A biosynthesis)

	New Zealand	
Francis T. F. Tsai	Departments of Biochemistry and Molecular Biology, and Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, USA	GroEL1 (chaperone protein)
Oleg, V. Tsodikov	Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, USA	Serine protease Rv3671c, DnaB helicase
Paul A. Tucker	EMBL Hamburg Outstation, Germany	Two-component system component RegX3, putative transcriptional antiterminator, cytosolic sensor kinase PdtaS
Anil K. Tyagi	Department of Biochemistry, University of Delhi South Campus, New Delhi, India	Biotin protein ligase, Bacterioferritin A
Mamannamana Vijayan	Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India	Pantothenate kinase, uracil-DNA glycosylase, Holliday junction DNA helicase RuvA, peptidyl-tRNA hydrolase
Rik, K. Wierenga	Biocenter Oulu and Department of Biochemistry, University of Oulu, Linnanmaa, Finland	α-methyl-acyl-coA racemase, β-oxidation trifuctional enzyme
Matthias Wilmanns	EMBL-Hamburg Unit, European Molecular Biology Laboratory, Germany	Histidine/tryptophan biosynthesis isomerase, Acyl-CoA carboxyltransferase, cysteine/lysine dyad acyltransferase LipB