

Elucidation of the substrates of mycosin 3, an essential protease of *Mycobacterium tuberculosis*

By

Zhuo Fang

*Thesis presented in partial fulfilment of the requirements for the degree
Master of Science in Medical Sciences (Medical Biochemistry) at the
University of Stellenbosch*



Supervisor: Prof. Nicolaas Claudius Gey van Pittius
Department of Biomedical Science
Co-supervisor: Prof. Robin Mark Warren
Department of Biomedical Science

March 2011

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 authorship owner thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2011

Copyright © 2011 Stellenbosch University

All rights reserved

Abstract

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), infects one third of the world's population and kills 1.7 million people per year. The increasing prevalence of multi- and extensively drug resistant *M. tuberculosis* strains means that there is an urgent need to develop new anti-TB drugs. The genome of *M. tuberculosis* has five copies of the ESAT-6 gene clusters (ESX-1, -2, -3, -4 and -5), which are essential for the survival (ESX-3) and pathogenicity (ESX-1 and ESX-5) of the bacterium. The ESX clusters encode for proteins which form a novel secretion system which has been shown to secrete small T-cell antigens of the *esx* gene family, as well as other proteins such as the PE and PPE's. The mycosins are a family of genes situated in the ESX clusters which encode for putative subtilisin-like serine proteases. These proteins are the most conserved proteins within the five clusters. Apart from their conserved protein sequence, mycosin-3 is also an essential protein specific to the mycobacteria, which makes it an attractive potential drug target. Identifying the substrate(s) of mycosin-3 could help to understand the function of this enzyme and discover novel inhibitors from which new drugs could be designed. We hypothesize that the secreted products of the ESX system could be potential substrates for the mycosins. Specifically, we hypothesize that PE5, PPE4, *esxG* and *esxH* (all found in ESX-3) might be the substrates for mycosin-3. Mycosin-3, PE5, PPE4, *esxG* and *esxH* were thus cloned, expressed and purified respectively. The four substrates were used for protease assays using mycosin-3 as the protease. The protease-substrate mixture were subsequently separated on 2-D SDS-PAGE gels to check whether there were any cleavage of the four substrates. Although all the target fusion proteins were cloned and expressed successfully, the protease assay results showed no cleavage for any of the four substrates. Possible explanations for the failure of cleavage were: (1) impure enzyme and substrate(s); (2) inappropriate buffer conditions; (3) the hypothesized substrates might not be the substrates of mycosin-3; and (4) incorrect folding or modification of the target fusion proteins might have taken place. Future research will aim to address these possible limitations in order to fully elucidate the function and substrate specificity of mycosin-3 and to use this information for the design of novel drugs against *M. tuberculosis*.

Opsomming

Mycobacterium tuberculosis, die organisme wat tuberkulose (TB) veroorsaak, infekteer 'n derde van die wêreld se bevolking en veroorsaak die dood van 1.7 miljoen mense per jaar. Die verhoogde voorkoms van multi- en ekstensiewe middelweerstandige stamme van *M. tuberculosis* beteken dat daar 'n ernstige nood is om nuwe anti-TB middels te ontwikkel. Die genoom van *M. tuberculosis* het vyf kopieë van die ESAT-6 geengroepe (ESX-1, -2, -3, -4 en -5), wat essensieel is vir die oorlewing (ESX-3) en patogenisiteit (ESX-1 and ESX-5) van die bakterium. Die ESX-groepe encodeer vir proteïene wat 'n nuwe uitskeidingsstelsel vorm wat bewys is om klein T-sel antigene van die esx-geenfamilie, sowel as ander proteïene soos die PE en PPE proteïene uit te skei. Die mycosins is 'n familie gene wat in die ESX-geengroepe voorkom en wat waarskynlik encodeer vir subtilisin-agtige serine proteases. Hierdie proteïene is die mees gekonserveerde proteïene in die vyf geengroepe. Mycosin-3 is 'n essensiële proteïen wat spesifiek in die mikobakteriële voorkom, sodat dit 'n aantreklike teken vir die ontwikkeling van middels is. Die identifisering van die substrate van mycosin-3 kan moontlik help om die funksie van die ensiem te verstaan en om nuwe inhibeerders vir die ensiem te ontdek, wat kan lei tot die ontwikkeling van nuwe middels. Ons hipotese is dat die uitgeskeide proteïene van die ESX-sisteem moontlik die substrate van die mycosin proteïene kan wees. Meer spesifiek, ons hipotiseer dat die proteïene PE5, PPE4, esxG en esxH (wat almal in ESX-3 voorkom) die substrate vir mycosin-3 kan wees. Mycosin-3, PE5, PPE4, esxG en esxH is afsonderlik gekloneer, uitgedruk en gesuiwer. Die vier substrate is gebruik vir protease proewe met mycosin-3 as die protease. Die protease-substraat mengsel is hierna deur middel van 2-D SDS-PAGE geanaliseer om te kyk of daar enige kliewing van die vier substrate voorgekom het. Alhoewel al die teken fusieproteïene suksesvol gekloneer, uitgedruk en gesuiwer is, het die protease proewe geen kliewing getoon vir enige van die vier potensiële substrate nie. Moontlike verklarings vir hierdie negatiewe resultaat is die volgende: (1) ensiem en substrate was moontlik onsuiver; (2) bufferkondisies was moontlik nie korrek nie; (3) gehipotiseerde substrate mag moontlik nie substrate van mycosin-3 wees nie; en (4) nie-korrekte vouing of modifisering van die teken proteïene kon moontlik plaasgevind het. Toekomstige navorsing sal daarop gemik wees om hierdie beperkinge aan te spreek om sodoende die funksie en substrate van mycosin-3 te kan ontdek en nuwe middels teen *M. tuberculosis* te ontwerp.

Acknowledgements

The author records his appreciation to:

Professor Nico Gey van Pittius, Division of Molecular Biology and Human Genetics, for invaluable supervision, guidance, patience and encouragement;

Professor Paul van Helden, Professor Robin Warren and Professor Ian Wiid, Division of Molecular Biology and Human Genetics, for invaluable advice and technical support;

Dr. Don Hayward and Dr. Monique Williams, Division of Molecular Biology and Human Genetics, Dr. Rabia Johnsons, Medical Research Council, for thoughtful advice and technical support;

Miss Mae Newton-foot, Mrs Magaretha de Vos, Mr Ruben van der Merwe, Miss Suereta Fortuin, Miss Michelle Smit, Miss Natalie Bruiner, Division of Molecular Biology and Human Genetics, for thoughtful advice and peer support;

Special thanks to Professor Paul van Helden, Division of Molecular Biology and Human Genetics, for generous research funding and bursary.

Table of Content

Declaration	i
Abstract	ii
Opsomming	iii
Acknowledgement	iv
List of Abbreviations	vii
List of Figures	ix
List of Tables	xiv
Chapter 1 Literature Review	1
1.1 Status of Tuberculosis in the world	1
1.2 Emergence of drug-resistant tuberculosis	1
1.3 Discovery of new anti-TB drugs	3
1.4 Genome of <i>Mycobacterium tuberculosis</i>	5
1.5 ESAT-6 gene cluster encoding for TypeVII secretion system	5
1.6 ESAT-6, CFP-10 and their homologs in ESAT-6 gene cluster region 3	10
1.7 PE and PPE genes and their gene products	11
1.8 Mycosin, a subtilisin-like serine protease	16
1.9 Potential drug target candidate	17
1.10 Catalytic characteristics and stability of novel subtilisin-like serine proteases	17
1.11 Methods to screen protease specificity	19
1.12 Problem Statement	22
1.13 Hypothesis	22
1.14 Aims of the project	22
Chapter 2 Methods and Materials	23
2.1 Strains and Plasmids	23
2.2 Primer Design	24
2.3 Polymerase Chain Reaction	27
2.4 Cloning of the Genes into the Expression Vector	28
2.4.1 Cloning of the genes into the pGEM-T Easy cloning vector	28
2.4.2 Cloning of inserts into exprssion vectors	29
2.5 Test Expression and Protein Purification	30
2.5.1 Test Expression of the pET-28a <i>E. coli</i> expression vector constructs in the <i>E. coli</i> BL21(DE3) strain	30
2.5.2 Test Expression of the p19Kpro and pDMN1 mycobacterial expression vector constructs in <i>M. smegmatis</i> mc ² 155	31
2.5.3 SDS-PAGE, Gel Staining and Western Blotting	32
2.5.3.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)	32
2.5.3.2 Polyacrylamide Gel Staining Technique	33
2.5.3.3 Western Blotting	34
2.5.4 Large-Scale Protein Expression and Purfication	35
2.5.5 Protein Assays	35
2.6 Protease Assay, 2-Dimensional gel electrophoresis and mass spectrometric analysis	36
2.6.1 Protease Assay	36

2.6.2 Two-Dimensional gel electrophoresis and Mass Spectrometric analysis	36
Chapter 3 Results	38
3.1 Cloning of mycP3 and potential substrates	38
3.2 Expression and Purification	39
3.3 Protein quantitation, protease assay, 2-D PAGE and Mass Spectrometric analysis	45
Chapter 4 Discussion	50
Chapter 5 Conclusion	59
Reference	60
Addendum	73
A.1 Plasmid map of expression vectors used in this project	73
A.2 Detailed results of colony PCR confirmation of the clones	74
A.2.1 Results for pET-28a expression vector	74
A.2.2 Results for p19Kpro expression vector	75
A.2.3 Results for pDMN1 expression vector	75
A.3 Detailed results of expression and purification of His-tagged fusion proteins	76
A.3.1 Results for pET-28a expression vector	76
A.3.2 Results for pDMN1 <i>M. smegmatis</i> expression vector	78
A.4 Detailed results of 2-D SDS-PAGE experiments and their Western blots	79
A.4.1 Protein quantification using Bradford assay	79
A.4.2 The 2-D SDS-PAGE gels in the control group and their Western blots using anti-mycosin 3 antibody	80

List of Abbreviations

ABC	ATP Binding Cassette
APS	Ammonium Persulphate
ATP	Adenosine Tri Phosphate
ATPase	Adenosine Tri-Phosphatase
BCG	Bacille Calmette et Guérin
bp	Base Pairs
BSA	Bovine Serum Albumin
CaCl ₂	Calcium Chloride
CD	Cluster of Differentiation
CFP-10	Culture Filtrate Protein 10
dNTP	Dephosphorylated nucleotide triphosphate
DNA	DioxyRibonucleic Acid
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMB	Ethambutol
ESAT-6	Early Secretory Antigenic Target of 10 kDa
ESX	ESAT-6 secretion system
GC	Guanine Cytosine
GO	Gene Ontology
GST	Glutathione S-Transferase
HIV	Human Immuno-deficiency Virus
HCl	Hydrogen Chloride
HRP	Horse Radish Peroxidase
IEF	Isoelectric Focusing
IM	Inner Membrane
INH	Isoniazid
IPG	Immobilized Ph Gradient
IPTG	Isopropyl-β-D-thiogalactopyranoside
IS	Insertion Sequence
kDa	Kilo Dalton
kV	Kilo Volt
LB	Luria Bertani

LC	Liquid Chromatography
MDR	Multi-Drug Resistant
MgCl ₂	Magnesium Chloride
MM	Mycomembrane
MPTR	Major Polymorphic Tandem Repeats
MS	Mass Spectrometry
MycP1	Mycosin-1
MycP3	Mycosin-3
NaCl	Sodium Chloride
Ni-NTA	Nickel-Nitrilotriacetic Acid
NTM	Non-Tuberculous Mycobacteria
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDA	Piperazine Diacrylamide
PE	Proline Glutamate
PGRS	Polymorphic G+C-Rich Sequence
PPE	Proline Proline Glutamate
PZA	Pyrazinamide
RD	Region of Difference
RIF	Rifampicin
SAP	Shrimp Alkaline Phosphatase
SDS	Sodium Dodecyl Sulphate
SOC	Super Optimal Catabolite
TAE	Tris-Acetic acid-EDTA
TB	Tuberculosis
TBS-T	Tris Buffered Saline – Tween 80
TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
UK	United Kingdom
USA	United States of America
UV	Ultraviolet
XDR	Extensive Drug Resistant
X-gal	5-bromo-4-chloro-3-indolyl-β-galactoside

List of Figures

Figure 1.1 Diagrammatic representation of the five ESAT-6 gene clusters and their evolutionary change

Figure 1.2 Working model for the ESX-1 secretion system

Figure 1.3 Solution structure of the CFP-10/ESAT-6 complex

Figure 1.4 Diagrammatic representation of the gene structures of the PE and PPE gene families

Figure 1.5 Crystal structure of the *M. tuberculosis* PE/PPE protein complex using the Rv2431c/Rv2430c pair as an example

Figure 1.6 Surface hydrophobicity of the PE/PPE protein complex

Figure 1.7 Schematic representation of substrate/inhibitor binding to a subtilisin-like serine protease

Figure 1.8 Diagrammatic representations of 2D DiGE, 2D SDS-PAGE and PROTO-MAP

Figure 3.1 Agarose gel electrophoresis image of colony PCR confirmation of recombinant expression vector pET-28a containing inserts of interest

Figure 3.2 Silver-stained SDS-PAGE images of purified expressed recombinant *M. tuberculosis* proteins

Figure 3.3 Western blots of SDS-PAGE gels in Figure 3.2

Figure 3.4 Western blots of Ni-NTA superflow column-purified His-tagged mycosin-3 with or without the transmembrane region fusion proteins using rabbit anti-mycosin-3 antibody

Figure 3.5 Comparison of the growth of *M. smegmatis* pDMN1-PPE4

Figure 3.6 Image of silver-stained SDS-PAGE gel where *M. smegmatis* expressed fusion proteins were loaded

Figure 3.7 2-D SDS-PAGE results of the experimental group using His-tagged PE5 fusion protein as the substrate

Figure 3.8 2-D SDS-PAGE results of the experimental group using His-tagged PPE4 fusion protein as the substrate

Figure 3.9 2-D SDS-PAGE results of the experimental group using His-tagged combined esxG and esxH fusion protein as the substrate

Figure A.1 Plasmid map of *E. coli* expression vectors pET-28a

Figure A.2 Plasmid map of *M. smegmatis* expression vector p19Kpro

Figure A.3 Plasmid map of *M. smegmatis* expression vector pDMN1

Figure A.4 Image of colony PCR on pET-28a-mycosin-3 without hydrophobic tail (lane 1 – 9, 1058 bp) or with hydrophobic tail (lane 11 – 15, 1241 bp) transformants

Figure A.5 Image of colony PCR on pET-28a-PE5 transformants (lane 1, 3, 4, 5 and 6, 311 bp)

Figure A.6 Image of colony PCR on pET-28a-PPE4 (1544 bp) transformant

Figure A.7 Image of colony PCR on pET-28a-esxH (lane 1, 293 bp) transformant, pET-28a-PE5&PPE4 (lane 4 – 6, 1855 bp) and pET-28a-esxG (lane 7 – 10, 296 bp) transformants

Figure A.8 Image of colony PCR on pET-28a-esxG&esxH (lane 2 – 5, 589 bp) transformants

6.2.2 Results for p19Kpro expression vector

Figure A.9 Image of colony PCR on p19Kpro-mycosin-3 (lane 1, 2 and 3) transformants

Figure A.10 Image of colony PCR on p19Kpro-esxG (lane 1, 2 and 5) transformants

Figure A.11 Image of colony PCR on p19Kpro-esxH (Lane 1 – 4) transformants

Figure A.12 Image of colony PCR on p19Kpro-PE5 (Lane 1, 2 and 3) transformants

Figure A.13 Image of colony PCR on p19Kpro-PPE4 (Lane 1, 2 and 3) transformants

Figure A.14 Image of colony PCR on p19Kpro-PE5/PPE4 (Lane 1, 2, 3, 5 and 6) transformants

Figure A.15 Image of colony PCR on pDMN1-mycP3 (Lane 1 – 5) transformants

Figure A.16 Image of colony PCR on pDMN1-esxG (Lane 1, 2, 4 and 5) transformants

Figure A.17 Image of colony PCR on pDMN1-esxH (Lane 1 – 7) and pDMN1-esxG/esxH (Lane 8 – 10) transformants

Figure A.18 Image of colony PCR on pDMN1-PE5 (Lane 1 and 3) transformants

Figure A.19 Image of colony PCR on pDMN1-PPE4 (Lane 1 – 4) and pDMN1-PE5/PPE4 (Lane 5 and 6) transformants

Figure A.20 Image of representative SDS-PAGE gel of test expression of pET-28a-esxG at 25°C with [IPTG] at 0.5 mM for 12 hours in *E. coli* (the area where the His-tagged esxG should be found is indicated in the graph)

Figure A.21 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxG under 9 different conditions where purified His-tagged esxG fusion proteins were loaded

Figure A.22 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxH under 9 different conditions where purified His-tagged esxH fusion proteins were loaded

Figure A.23 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxG/esxH under 9 different conditions where purified His-tagged esxG/esxH fusion proteins were loaded

Figure A.24 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PE5 under 9 different conditions where purified His-tagged PE5 fusion proteins were loaded

Figure A.25 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PPE4 under 9 different conditions where purified His-tagged PPE4 fusion proteins were loaded

Figure A.26 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PE5/PPE4 under 9 different conditions where purified His-tagged PE5/PPE4 fusion proteins were loaded

Figure A.27 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-mycP3-without hydrophobic tail under 9 different conditions where purified His-tagged mycP3-WT fusion proteins were loaded

Figure A.28 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-mycP3-with hydrophobic tail under 9 different conditions where purified His-tagged mycP3-T fusion proteins were loaded

Figure A.29 Image of silver-stained SDS-PAGE gel of expression of His-tagged fusion proteins in *M. smegmatis* at 37°C. The names of the proteins loaded in the lane are specified. Only combined esxG and esxH could be expressed and purified.

Figure A.30 Image of colourimetric western blot of the duplicate polyarylamide gel shown in Figure 6.25. It confirms that only combined esxG and esxH was successfully expressed and purified from *M. smegmatis*.

Figure A.31 The Standard curve of Bradford assay on a range of concentrations of Bovine Serum Albumin where the absorbance at 595 nm versus the concentration (mg/ml), the equation of this curve is $y = 0.8333x + 0.0518$, the correlation factor is 0.9703

Figure A.32 Silver-stain 2-D SDS-PAGE gel for the purified His-tagged mycosin-3 without transmembrane region fusion protein (a) and its western blot (b)

Figure A.33 Silver-stained 2-D SDS-PAGE gel for the purified His-tagged mycosin-3 with transmembrane region fusion protein (a) and its western blot (b)

Figure A.34 Silver-stained 2-D SDS-PAGE gels for two different batches of purified His-tagged PE5 fusion protein (the most identifiable spot is the one pointed by an arrow)

Figure A.35 Silver-stained 2-D SDS-PAGE gel for the purified His-tagged PPE4 fusion protein (the spot indicated by the square surrounded)

Figure A.36 Silver-stained 2-D SDS-PAGE gel for two different batches of purified His-tagged combined esxG and esxH fusion proteins (the spots indicated by the arrows are probably esxG and esxH)

List of Tables

Table 1.1 Targets of anti-mycobacterial agents and associated genetic loci

Table 2.1 Sequence of primers designed for cloning genes into the pET-28a expression vector

Table 2.2 Sequence of primers designed for cloning genes into the p19Kpro expression vector

Table 2.3 Sequence of primers designed for cloning genes into the pDMN1 expression vector

Table 2.4 Method of test expression of pET-28a in BL21(DE3) *E. coli* strain, same amount of inoculant was made for each flask

Table A.1 The amounts of freshly prepared His-tagged fusion proteins using Ni-NTA Superflow column

Chapter 1 Literature Review

1.1 Status of Tuberculosis in the world

Tuberculosis (TB) is a contagious pulmonary disease, the causative agent of which is *Mycobacterium tuberculosis*, a member of the high G+C gram-positive bacteria (Grange, 2009). The organism is air-borne, thus the infection results from inhaling air containing bacteria mostly originating from coughing TB patients. *M. tuberculosis* is a very successful human pathogen carrying out sophisticated defense mechanisms, such as inhibiting acidification and maturation of the phagosome, and persisting inside human macrophages (Grange, 2009). Additionally, *M. tuberculosis* is well adapted to harsh environments (including limited nutrient supplies). *M. tuberculosis* infects one third of the world's population (Global Tuberculosis Control, WHO, 2009). Infected persons may get diseased when their immune systems get compromised. Not surprisingly, there is a strong correlation between TB and HIV infection. TB causes more deaths than any other single infectious agent in human history (Grange, 2009). It has become a global health threat. According to World Health Organization (WHO) statistics, there were about 9.4 million incident cases of TB globally in 2008, and an estimated 1.3 million people died from TB in the same year. The highest number of deaths was in the South-East Asia Region, while the highest mortality per capita was in the Africa Region (Global Tuberculosis Control, WHO, 2009). Such a phenomenon indicates that TB has a high morbidity and mortality rate in populated and under-developed regions of the world.

1.2 Emergence of drug-resistant tuberculosis

M. tuberculosis evolved from non-tuberculous mycobacteria (NTM) which are mostly environmental mycobacteria (den Dooren de Jong, 1939). Dutch microbiologist den Dooren de Jong described tubercle bacillus as “the wayward son of honourable parents” in 1939. How *M. tuberculosis* evolved to become a human pathogen is still a mystery. Research on the interactions between environmental mycobacteria and protozoa provided some evidence. Primm and coworkers proposed that some environmental mycobacteria, including members of the *M. avium* complex, are able to replicate within various protozoans and also to survive within amoebic cysts under conditions of environmental stress (Primm *et al.*, 2004). Such an intracellular replication may play an essential role in the evolution of mycobacteria to become intracellular pathogens.

Humans have been combating TB for centuries, although the causative agent was only discovered in 1882. The standard therapy for TB involves treatment with antibiotics, namely isoniazid (INH), rifampicin (RIF), pyrazinamide (PZA) and ethambutol (EMB) for two months, then INH and RIF alone for another four months. The patient is considered as being cured after this regimen. These four drugs used for drug-sensitive *M. tuberculosis* strains are grouped as first line drugs because of their effectiveness and primary use. Unfortunately, drug-resistant TB emerged over the years due to poorly managed TB care, incorrect drug prescribing practices by providers, poor quality drugs, erratic supply of drugs, and also patient non-adherence (WHO, 2006). *M. tuberculosis* gradually adapted to the toxic effect of these drugs and become resistant through mutations of the genes encoding drug targets (shown in Table 1.1).

Table 1.1 Targets of anti-mycobacterial agents and associated genetic loci (Grange, 2009)

Agent	Target molecule or function	Genes and regions encoding for resistance
Isoniazid	Mycolic acid synthesis	<i>katG</i> , <i>inhA</i> and its promoter region. <i>oxyR-ahpC</i> intergenic region
Rifampicin	DNA-dependent RNA polymerase	<i>rpoB</i>
Pyrazinamide	Cell membrane energy function	<i>pncA</i>
Ethambutol	Arabinogalactan synthesis	<i>embA</i> , <i>embB</i> , <i>embC</i>
Streptomycin	Ribosomal protein S12	<i>rpsL</i>
Ethionamide and prothionamide	Mycolic acid synthesis	<i>ethA</i> , <i>inhA</i> and its promoter region
Capreomycin, Kanamycin, Amikacin and Viomycin	50S and 30S ribosomal subunit	<i>vicA</i> (50S), <i>vicB</i> (30S), <i>rrs</i> (16S)
Cycloserine	Peptidoglycan synthesis	<i>alrA</i>
Clofazimine	? RNA polymerase	Unknown
p-aminosalicylic acid	Folic acid synthesis	Unknown
Dapsone	Folic acid synthesis	Unknown

Multi-drug resistant (MDR) *M. tuberculosis* strains have recently emerged as a new threat which makes this global burden even worse. MDR-TB is caused by strains of *M. tuberculosis* that are resistant to at least the two main first-line TB drugs, namely INH and RIF. Epidemics of drug-resistant disease can be generated by three interrelated mechanisms: (1) conversion to drug-resistant strains from wildtype pan-susceptible strains during treatment (acquired resistance); (2) increasing development of resistance in drug-resistant strains due to inappropriate chemotherapy (amplified resistance); and (3) transmission of drug-resistance cases (transmitted resistance) (Blower *et al.*, 2004). This process is still ongoing and extensive drug resistant (XDR) TB [defined as MDR-TB that is also resistant to any fluoroquinolone, and to at least one of three injectable second line anti-TB drugs used in TB treatment (capreomycin, kanamycin and amikacin)] has now emerged.

1.3 Discovery of new anti-TB drugs

Scientists have realized the urgency of efficiently combating TB and have thus been either designing derivatives of old drugs (making them more effective against drug-resistant TB) or looking for new alternative drug targets. Many new drugs have been discovered in the last few years, such as SQ109, PA-824, OPC-67683 and TMC207.

SQ109 is a diamine analogue of ethambutol, which shows excellent *in vitro* activity against *M. tuberculosis*, including strains resistant to EMB, INH and RIF (Protopopova *et al.*, 2005). This drug has a long half-life suggesting that once-a-week dosing is achievable. SQ109 is believed to target cell wall synthesis but with a different mechanism from EMB, perhaps a different target. The combination of SQ109 and RIF against TB is so potent that 99% of the growth was inhibited at very low concentrations (Chen *et al.*, 2006).

PA-824 is a lead compound of bicyclic nitroimidazo[2,1-b]oxazines, which was found to be active against *M. tuberculosis* without any unfavourable mutagenic features. PA-824 is able to inhibit *M. tuberculosis* cell wall lipid and protein synthesis (Stover *et al.*, 2000). However, substitution of RIF or PZA with PA-824 appeared to deactivate the drug and a high relapse rate was observed after six months of a RIF, INH and PA-824 regimen. The sterilizing ability of the drug was limited (Nueremberger *et al.*, 2006).

The drug, OPC-67683, which belongs to the 6-nitro-2,3-dihydroimidazo[2,1-b]oxazole series, targets mycolic acid. It inhibits myoxy-mycolic and keto-mycolic acid synthesis, but not α -mycolic acid synthesis, at significantly lower concentration than INH. A combined treatment of OPC-67683 with RIF and PZA for a 2 months followed by a combination with RIF for another 2 months eliminates all lung bacteria within 3 months. It suggests that this drug has a powerful sterilizing ability and may be effective in shorter treatments (Matsumoto *et al.*, 2006).

TMC207 is the most potent molecule among the series of diarylquinolines (DARQs). It exhibits excellent activity against drug-susceptible, MDR and XDR *M. tuberculosis* strains, with no cross-resistance to first-line drugs (Andries *et al.*, 2005). TMC207 acts by inhibiting mycobacterium membrane-bound ATP synthase. This mechanism has little similarity with mycobacterial and human proteins encoded by the *atpE* gene that codes for the c subunit of ATP synthase, so it has great potential (Andries *et al.*, 2005). Combination of TMC207 with first-line drugs results in elimination of the bacteria in mouse models within two months.

More recently, a group of 36 3-methylquinoxaline-2-carboxamide 1,4-di-N-oxide derivatives have been developed for screening as anti-TB agents (Ancizu *et al.*, 2010). This stemmed from the efficacy of quinoxaline and quinoxaline 1,4-di-N-oxide derivatives displaying excellent anti-parasitic, anticancer, antiviral, and antibacterial activities in many different therapeutic areas.

Most anti-TB drugs target enzymes that are essential for biochemical reactions in *M. tuberculosis* and these reactions are vital for bacterial survival. Structural genomics encourages scientists to look for other essential enzymes as potential new anti-TB drug targets, including extracellular proteins involved in virulence and persistence determinants, secreted antigens and proteins involved in iron acquisition (Ahmed and Hasnain, 2004).

Analysis of the *M. tuberculosis* genome sequence provides important information for the identification of new potential drug targets.

1.4 Genome of *Mycobacterium tuberculosis*

The complete genome of the *M. tuberculosis* H37Rv reference strain was sequenced in 1998 (Cole *et al.*, 1998), and re-annotated in 2002 (Camus *et al.*, 2002). This showed that the genome of this strain contained 4,411,529 base pairs (bp) and around 4000 genes. It has a high guanine + cytosine (G+C) content (65.6%) uniformly throughout the genome. Several regions showing higher than average G+C content were detected, and it was found that these regions consist of genes which belong to a large gene family that includes the polymorphic G+C-rich sequences (PGRSs). One special feature of the genome is that around 250 genes are involved in synthesis and metabolism of lipids. They enable the bacteria to synthesize a very complex lipid-rich cell wall. Another very unusual feature is the fact that 10% of the coding capacity of the genome encodes for two gene families, PE and PPE. PE and PPE genes are mostly found in the mycobacteria. About 51% of the genome has arisen through gene duplication; and 3.4% of the genome is composed of insertion sequences (IS) and prophages (Cole *et al.*, 1998). Three of the most well-known *M. tuberculosis* gene families have arisen through gene duplication, namely the PE, PPE and ESAT-6 gene families.

1.5 ESAT-6 gene cluster encoding for Type VII secretion system

The genome of *Mycobacterium tuberculosis* contains five copies of the ESAT-6 gene cluster referred to as ESX-1, ESX-2, ESX-3, ESX-4 and ESX-5. Each cluster in turn contains members of the CFP-10 (Culture Filtrate Protein, 10 kDa) and ESAT-6 (Early Secreted Antigenic Target, 6 kDa) gene families, the PE (Proline-Glutamate) and PPE (Proline-Proline-Glutamate) gene families, secreted, cell-wall-associated subtilisin-like serine proteases (mycosins), putative ABC transporters, ATP-binding proteins and other membrane-associated proteins (Figure 1.1) (Gey van Pittius *et al.*, 2001).

Phylogenetics and comparative genomics analysis suggested that ESAT-6 gene cluster region 4 is ancestral and that all the other regions were duplicated from it. The five regions evolved through duplication which probably took place in the following order: region 3 (Rv0282-0292) => region 1 (Rv3866-3883c) => 2 (Rv3884c-3895c) => and 5 (Rv1782-1798) (demonstrated in Figure 1.1) (N.C. Gey van Pittius, personal communication).

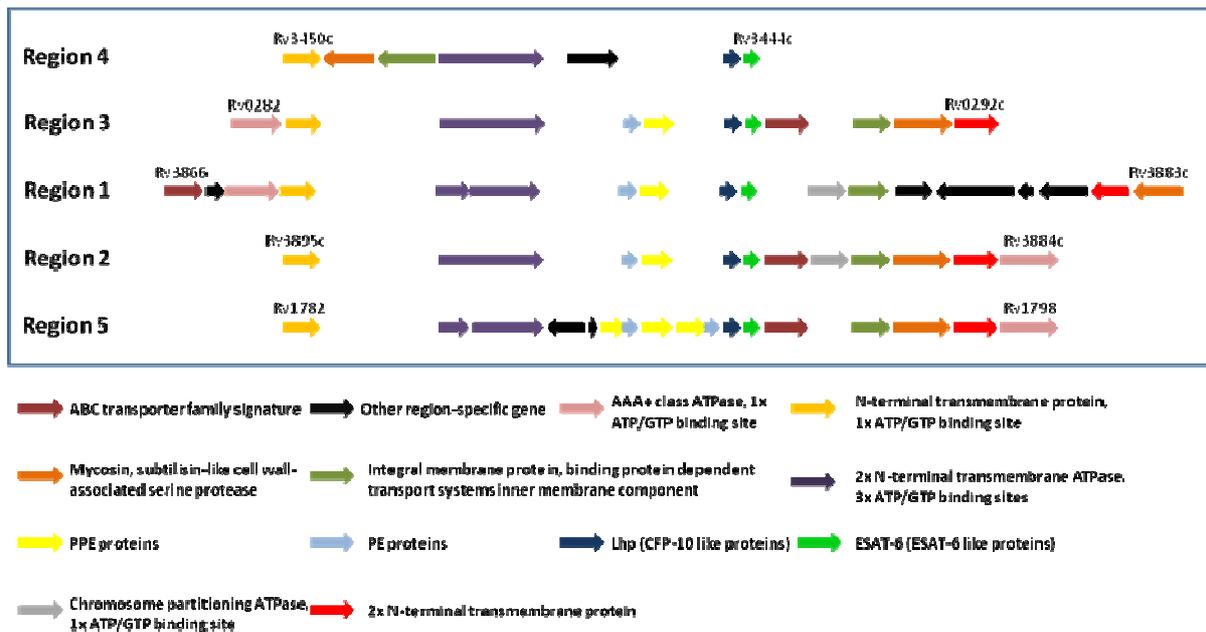


Figure 1.1 Diagrammatical representations of components of five ESAT-6 gene clusters and their evolutionary change (the figure is adapted from Gey van Pittius *et al.*, 2001)

The five ESAT-6 gene clusters encode for a special type of bacterial secretion system on the membrane, named the Type VII secretion system, which functions to secrete members of the ESAT-6 and CFP-10 protein families, PE and PPE family proteins, EspA, EspB and perhaps other unknown substrates (Abdallah, *et al.*, 2007). Unlike the other six types of bacterial secretion systems discovered previously in Gram-negative bacteria, Type VII secretion system is primarily found within the mycobacteria (Gram-positive bacteria) due to the special characteristics of their cell envelopes, although there are homologues systems in closely related genera (Abdallah, *et al.*, 2007). The mycobacterial cell envelope is very complex compared to the cell walls of other gram-positive bacteria, and are characterized by the presence of mycolic acids which are large hydroxylated branched-chain fatty acids (Grange, 2009). Another special feature of this secretion system is the target secreted proteins (CFP-10 and ESAT-6 in Esx-1 for instance) do not contain a Sec-signal sequence (Abdallah, *et al.*, 2007).

The structure and the secretion mechanism of the Type VII secretion system have been described using ESX-1 as a model (Teutschbein *et al.*, 2006; Abdallah *et al.*, 2007). ESAT-6 and CFP-10 form a 1:1 complex in solution after being co-expressed (Renshaw *et al.*, 2002). Although the complex lacks a secretion signal sequence, the C-terminal 7 amino acids of CFP-10, which is not associated with ESAT-6, is crucial for binding to Rv3871 (which

encodes for a transmembrane ATPase) (Renshaw *et al.*, 2005). Once the ESAT-6/CFP-10 complex is recognized by Rv3871, Rv3871 in turn delivers the complex to Rv3870, which belongs to the same protein family as Rv3871 (Figure 1.2), whereafter it is translocated to the secretion machinery on the cell membrane, probably Rv3877 (which is a multi-transmembrane protein that might constitute the inner-membrane secretion channel). Similarly, EspA (Rv3616c), a gene which is an additional non-ESX-1 gene necessary for secretion (acting like ESAT-6), binds to Rv3615c (EspC, acting like CFP-10) prior to secretion. Rv3615c was found to be recognized by another cytosolic AAA ATPase (Rv3868) (other than Rv3871 and Rv3870) through its C-terminus (the same recognition mechanism as CFP-10 with Rv3871) (Champion *et al.*, 2009). EspA was found to bind to CFP-10/ESAT-6 complex as well. They can be co-secreted via ESX-1 and it is believed that EspA interacts with both CFP-10 and ESAT-6 (Callahan *et al.*, 2009). Very surprisingly, without EspA or EspC, the ESAT-6/CFP-10 dimer can be produced but not secreted (MacGurn *et al.*, 2005; Fortune *et al.*, 2005). Without the expression of ESAT-6/CFP-10, none of the other known substrates of ESX-1 is secreted. Similarly, strains lacking EspB, EspC or EspR fail to secrete the ESAT-6/CFP-10 dimer (McLaughlin *et al.*, 2007; Xu *et al.*, 2007; Raghavan *et al.*, 2008). This implies that the ESX-1 substrates are dependent on one another for secretion. It is possible that Type VII substrates are only secreted as multimeric complexes (Abdallah *et al.*, 2007); alternatively these four substrates might be components of the secretion system and form some sort of pilus or extracellular structure (Ize and Palmer, 2006). The Rv3781-Rv3780 complex could form a hexameric ring structure with a central cavity that propels ESX-1 substrates through the secretion channel. The functions of other components of ESX-1 are more difficult to predict.

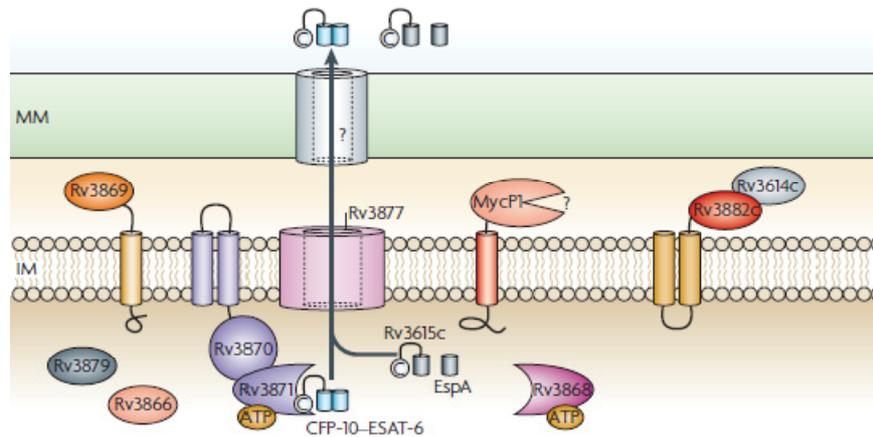


Figure 1.2 Working model for the ESX-1 secretion system. The secretion of Rv3616c (EspA) is interdependent on the presence of ESAT-6/CFP-10 complex. The ESAT-6/CFP-10 complex is recognized by Rv3871 which binds to the C-terminal of CFP-10. Rv3871 is situated with the inner membrane (IM) by interacting with Rv3870. The translocation channel in the IM is probably formed by Rv3877. It is unknown which protein forms the channel in the mycomembrane (MM). The AAA+ chaperone-like protein Rv3868 could be involved in the biogenesis of the secretion machinery. The function of MycP1 is essential in the secretion process and one substrate, EspB, has been discovered by Ohol and coworkers (taken from Abdallah *et al.*, 2007).

Even though there are five copies of the ESAT-6 gene cluster within the genome of *M. tuberculosis*, these five ESX systems do not complement one another fully, although a small degree of partial complementation has been observed. The reason for this might be that each one of them has a different signal for their secretion and they also differ in their regulation patterns (Champion *et al.*, 2006). They might also have evolved different functions. ESX-1 genes are downregulated when the culture is starved whereas ESX-2 genes are upregulated under these conditions (Betts *et al.*, 2002). ESX-3 is regulated by the availability of zinc and iron ions, as part of the *ideR* and *Fur* regulon (Rodriguez and Smith, 2003; Maciag *et al.*, 2007). ESX-4 genes are regulated by the alternative sigma factor SigM (Agarwal *et al.*, 2007). Besides regulation differences among these five gene clusters, it was also found that ESX-1, ESX-2 and ESX-4 can be disrupted by knockouts, but ESX-3 and ESX-5 cannot, which suggests the essentiality of those two clusters for the growth of the culture (Sasseti *et al.*, 2003). However, it was possible to knock out ESX-3 in *M. smegmatis* (Siegrist. *et al.*, 2009).

Among the five ESX systems, ESX-1 and ESX-5 are responsible for virulence. The ESX-1 secretion system can be found in many mycobacteria but ESX-5 is only found in slow-growing mycobacteria which are mostly pathogenic (Abdallah *et al.*, 2006; Gey van Pittius *et al.*, 2006). ESX-1 transports and secretes ESAT-6 and CFP-10 which are potent T-cell

antigens from *M. tuberculosis* (Sorensen *et al.*, 1995). Experimental evidence showed that a knockout of ESX-1 in *M. tuberculosis* attenuates the pathogen and that complementation of ESX-1 in the attenuated *M. bovis* BCG strain makes the bacteria partially regain its virulence (Pym *et al.*, 2002). In fact, deletion of any of the genes situated in ESX-1 attenuates *M. tuberculosis* (Guinn *et al.*, 2004). Knock-out mutations of the genes Rv3868-Rv3872 and Rv3877 in ESX-1 also abolishes secretion of ESAT-6 and CFP-10 (Lewis *et al.*, 2003; Pym *et al.*, 2003; Stanley *et al.*, 2003; Gao *et al.*, 2004; Guinn *et al.*, 2004). Inactivation of *esxA* and *esxB*, the genes of ESAT-6 and CFP-10 also attenuates *M. tuberculosis*. However, inactivating genes Rv3876 and Rv3873 did not appear to prevent the secretion of ESAT-6 or CFP-10 (Brodin *et al.*, 2004; Demangel *et al.*, 2004). The same characteristics were revealed in ESX-1 in *Mycobacterium smegmatis*, where disruption of any of the genes Sm3866, Sm3883c, Sm3882c and Sm3869 (except Sm3868) abolished the secretion of SmESAT-6 and SmCFP-10 (Converse and Cox, 2005). In the same study, it was found that SmESAT-6 and SmCFP-10 were secreted when *M. smegmatis* was grown in Sauton's medium but not in normal 7H9 medium. The reason for this is unknown (Converse and Cox, 2005). Moreover, in spite of the evolutionary distance between *M. tuberculosis* and *M. smegmatis*, the *M. smegmatis* secretion system can secrete the *M. tuberculosis* ESAT-6 and CFP-10 proteins, suggesting that substrate recognition is also conserved between the two species. Interestingly ESX-1 in non-pathogenic *M. smegmatis* was found to carry out DNA transfer, i.e. conjugation (Coros *et al.*, 2008). ESX-5 is also responsible for virulence like ESX-1 but it mostly carries out the transport and secretion of the PE_PGRS and PPE-MPTR proteins (Abdallah *et al.*, 2009). ESX-3 is required for growth both in medium under iron-limited conditions and macrophages within which the iron is also scarce (Siegrist *et al.*, 2009). This study showed a close relationship between ESX-3 and the mycobactin pathway. The authors concluded that ESX-3 is essential for mycobactin-mediated iron acquisition. In a study done by Serafini and coworkers, it was suggested that ESX-3 encodes a novel iron/zinc uptake system or it has a strong effect on mycobacterial cell surface permeability to iron and zinc (Serafini *et al.*, 2009). They also proposed that ESX-3 must be secreting some unrecognized factors required for the optimal uptake of iron and zinc. It suggests that ESX-3 might be involved in zinc and iron homeostasis. The functions of ESX-2 and ESX-4 have not been studied in detail because their relationship with pathogenicity is not as strong as that of the other three clusters.

1.6 ESAT-6, CFP-10 and their homologs in ESAT-6 gene cluster region 3

Many studies have been done on the ESAT-6/CFP-10 complex of ESX-1 since ESX-1 has been shown to be important for virulence. ESAT-6 and CFP-10 are co-transcribed through a single promoter and interact with each other forming a complex (Berthet *et al.*, 1998). The direct interaction between ESAT-6 and CFP-10 is very strong ($K_d < 1.1 \times 10^{-8}$ M) (Renshaw *et al.*, 2002; Meher *et al.*, 2006). The elucidation of the molecular structure of the ESAT-6/CFP-10 1:1 complex revealed that the core of the complex consists of two helix-turn-helix hairpin structures formed from two individual proteins, which have an extensive hydrophobic contact surface and lie anti-parallel to each other to form a four-helix bundle (Figure 1.3). Both proteins have disordered N-termini as well as C-termini, which form long flexible arms at both ends of the four-helix bundle core. The surface of this complex has a very uniform distribution of positive and negative charge, with no hint of a significant hydrophobic patch (Renshaw *et al.*, 2005). There is no significant cleft in the surface of the complex which indicates that there is no active site for an enzyme. This suggests that there is no catalytic role for this complex. The surface feature of the complex, in fact, suggests a role based on specific binding to one or more target proteins, probably in pathogen-host cell signaling. This prediction was proven by imaging the interaction of fluor-labeled CFP-10/ESAT-6 complex with U937 monocytes. It was confirmed that the CFP-10/ESAT-6 complex mediated the binding of fluorescently labeled CFP-10/ESAT-6 complex to the surface of U937 cells and that the flexible C-terminal arm of CFP-10 formed an essential part of the cell surface receptor binding site. These findings imply that the CFP-10/ESAT-6 complex plays a possible signaling role in which binding to cell surface receptors leads to modulation of host cell behaviour (Renshaw *et al.*, 2005). CFP-10 dissociates from ESAT-6 when the pH of the solution is lowered to acidic levels (pH 4 or 5). These levels are often encountered in the phagosome, allowing the binding of ESAT-6 to the membrane and lysing it (CFP-10 does not have the same membrane lysis ability) (de Jonge *et al.*, 2007). It is proposed that CFP-10 might function as a chaperone for ESAT-6. CFP-10 is involved in the transport and protection of ESAT-6 until it reaches the phagosomal compartment (de Jonge *et al.*, 2007). Derrick and coworkers found that the ESAT-6 protein also induces apoptosis of macrophages by activating caspase expression (Derrick and Morris, 2007).

The homologues of CFP-10 and ESAT-6 in ESX-3 are named *esxG* (Rv0287) and *esxH* (Rv0288). ESAT-6 gene cluster region 3 was proven to be essential for the growth of *M. tuberculosis*, but not for pathogenicity *in vivo* (Siegrist, *et al.*, 2009). *esxH* expression was

found to be markedly down-regulated in an attenuated strain of *M. tuberculosis* H37Ra (Rindi *et al.*, 1999). Protein *esxH* is also a potent T-cell antigen strongly recognized in *M. tuberculosis*-infected humans (Skjot *et al.*, 2002). However, studies on *esxG* have been very limited. It was found in short-term culture filtrates of *M. tuberculosis* in proteomic studies (Rosenkrands *et al.*, 2000). Like CFP-10 and ESAT-6, these two genes are also co-operonic (Okkels and Andersen, 2004). Moreover, they interact with each other just like the interaction between ESAT-6 and CFP-10; however, cross-interaction was not observed (Okkels and Andersen, 2004). The same study also found that the ESAT-6 proteins (ESAT-6, CFP-10 and *esxH*) interact directly with PPE68 (Rv3873), and the binding was specific. Apart from this, not much research has been done on *esxG* and *esxH*. Their structure and function remain unknown.

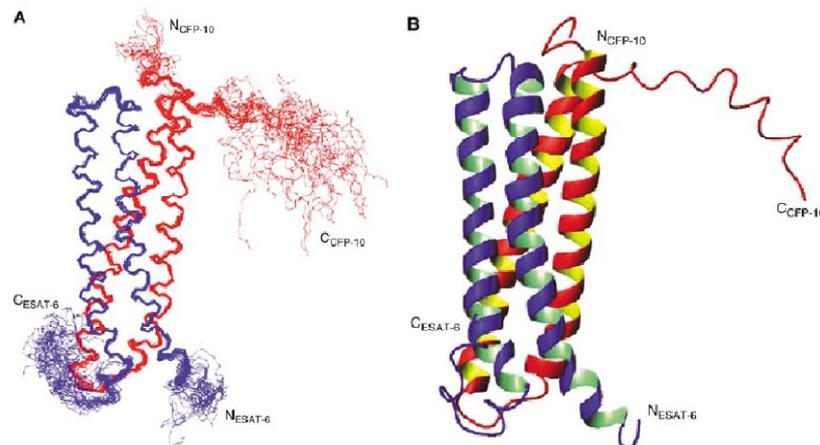


Figure 1.3 Solution structure of the CFP-10/ESAT-6 complex. (A) A best-fit superposition of the protein backbone for the family of 28 converged structures obtained, with CFP-10 shown in red and ESAT-6 in blue. The long flexible C-terminal arms of both proteins are identifiable, as is the propensity to helical structure in this region of CFP-10. (B) A ribbon representation of the backbone topology of the CFP-10/ESAT-6 complex based on the converged structure closest to the mean, which illustrates the two helix-turn-helix hairpin structures formed by the individual proteins. The orientation of the complex is identical to that shown in panel A, with CFP-10 in red and ESAT-6 in blue. The helical propensity of the section in the flexible C-terminus of CFP-10 can be clearly seen in the top right of the figure (Take from Renshaw *et al.*, 2005).

1.7 PE and PPE genes and their gene products

The PE and PPE family proteins are essential components of the ESAT-6 gene clusters except for the ancestral region 4. They can also be found outside the ESAT-6 gene clusters. The genes encoding PE and PPE proteins are frequently clustered (Gey van Pittius, *et al.*,

2006). They are often based on multiple copies of the polymorphic repetitive sequences (PGRSs) and major polymorphic tandem repeats (MPTRs), respectively (Figure 1.4). The names PE and PPE are derived from the N terminal motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) at positions 8-9, or 8-10, in the amino acid sequences respectively (Gordon *et al.*, 1999). The PE protein family has 99 members, all of which have a highly conserved N-terminal domain of about 110 amino acids, followed by a C-terminal segment that varies in size, sequence and repeat copy number (Cole, *et al.*, 1998). The sizes of the PE proteins vary from 110 (contain N-terminal motif only) to 1500 residues (Cole, *et al.*, 1998). Therefore, based on these variations, the PE protein family is divided into three subfamilies. The first family containing 29 members only has the PE domain; the second one of 8 members contains the PE domain followed by a unique sequence; the third one of 67 members contains the PE domain followed by multiple repetitive tandem repeats of Gly-Gly-Ala or Gly-Gly-Asn, the so-called PGRS (Gordon *et al.*, 1999). The PPE protein family has 68 members and also has a conserved N-terminal domain that comprises about 180 amino acids, followed by C-terminal segments that vary markedly in sequence and length. There are four subfamilies of the PPE protein family (Adindla and Guruprasad, 2003). The largest family has 24 members, which are characterized by the motif Gly-X-X-Ser-Val-Pro-X-X-Trp between position 300 and 350 in the amino acid sequence. The second largest subfamily (23 members), also termed as major polymorphic tandem repeat (MPTR) PPE subfamily, contains multiple C-terminal repeats of the motif Asn-X-Gly-X-Gly-Asn-X-Gly, encoded by a consensus repeat sequence GCCGGTGTTG. The third subfamily (10 members) is characterized by a conserved 44 amino acid residue region in the C-terminus comprising of highly conserved Gly-Phe-X-Gly-Thr and Pro-X-X-Pro-X-X-Trp sequence motifs (Adindla and Guruprasad, 2003). The fourth PPE subfamily (12 members) consists of proteins with a low percentage of homology at the C-terminus (Gordon *et al.*, 1999).

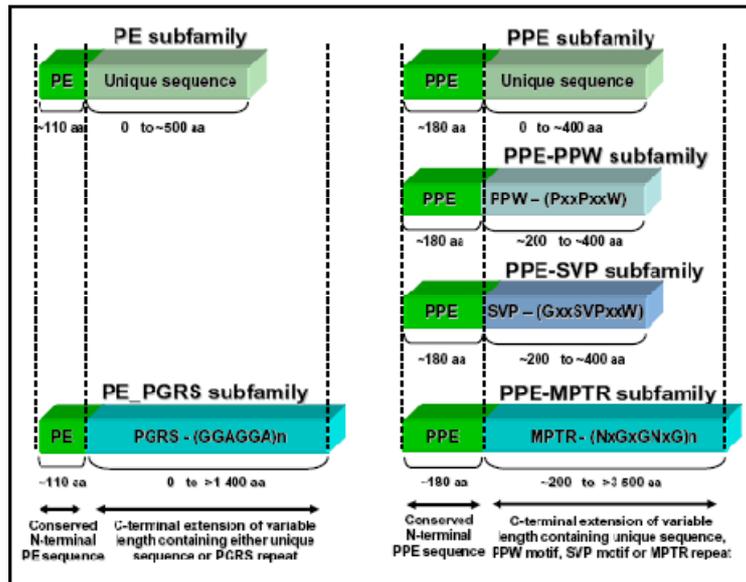


Figure 1.4 Diagrammatic representation of the gene structures of the PE and PPE gene families, showing conserved N-terminal domains, motif positions and differences among different subfamilies found in these two families (Gey van Pittius *et al.*, 2006).

PE and PPE genes are organized in operons where PE genes are usually (28 out of 41 operons) upstream to PPE genes in the genome. Within these operons, the PE and PPE genes are separated by less than 90 bp (Tundup *et al.*, 2006 and Strong *et al.*, 2006). Their expression is co-operonic where one promoter controls the expression of both PE and PPE genes in the pair. Such an arrangement happens frequently in the *M. tuberculosis* genome. For instance, the PPE gene Rv0915c is downstream from PE gene Rv0916c, and they are separated by a 14 bp intergenic region (Skeiky *et al.*, 2000); PPE gene Rv1787 is separated from a PE gene Rv1788 by 78 bp (Li *et al.*, 2005) and PE gene Rv2431c precedes its PPE gene partner Rv2430c by 46 bp (Tundup *et al.*, 2006).

Interestingly, recombinant PE/PPE proteins rRv2431c and rRv2430c form inclusion bodies when over-expressed on its own in *Escherichia coli*, but they appeared in soluble fraction when they were co-expressed. There is evidence that they interact with each other similar to ESAT-6 and CFP-10 (Tundup *et al.*, 2006 and Strong *et al.*, 2006). They form oligomers when alone, but exist as a heteromer when present together (Tundup *et al.*, 2006).

The 3-D structures of individual PE and PPE proteins are difficult to elucidate because of the solubility problems described above (Strong *et al.*, 2006; Tundup *et al.*, 2006). The structure of this 1:1 complex could only be defined when they are co-expressed and co-purified

(Strong *et al.*, 2006). The complex is highly α -helical and it is heterodimeric, containing one PE and one PPE protein. The PE protein is a two-helix bundle; together with two of the five helices of the PPE protein, they form a four-helix bundle (Figure 1.5). The PE protein is composed of two α -helices (residues 8-37 and 45-84) which run anti-parallel to each other, connected by a loop (residues 38-44), with both the N and C termini at the top of the complex. This PE loop is stabilized by the interactions with helices 2 and 5 of the PPE protein. The conserved Pro-Glu (PE) sequence motif is located at the N-terminus of the PE protein (residues 8-9). The PPE protein is entirely helical. The conserved Pro-Pro-Glu (PPE) sequence motif is located near the N-terminus of the PPE protein (residue 7-9). Helices $\alpha 2$ (residues 21-53) and $\alpha 3$ (residues 58-103) of the PPE protein run anti-parallel and form the interaction interface with the PE protein (Strong *et al.*, 2006).

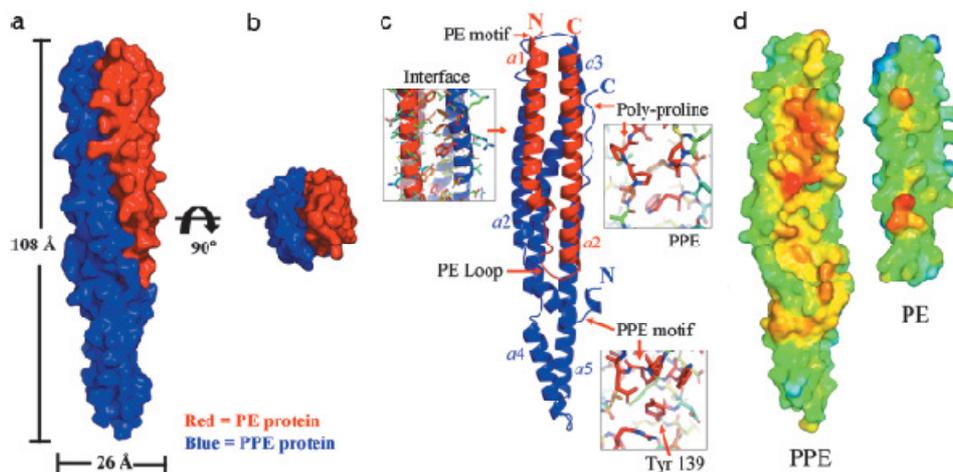


Figure 1.5 Crystal structure of the *M. tuberculosis* PE/PPE protein complex using the Rv2431c/Rv2430c pair as an example. (a) Surface representation of the PE/PPE protein complex. The PE protein Rv2431c is shown in red and the PPE protein Rv2430c is in blue. (b) The longitudinal view of PE/PPE protein complex. (c) Ribbon diagram of the PE/PPE protein complex. (d) Interface hydrophobicity of the PPE and PE proteins. The strength of hydrophobicity increases from the colour blue to red (taken from Strong *et al.*, 2006).

The PE and PPE proteins are predicted to carry out interactions among cells and have immunological importance. The PE-PGRS protein encoded by Rv1818c is able to mediate cell-cell adhesion because the disruption of this gene causes a great reduction of bacterial clumping (Brennan *et al.*, 2001). Moreover, the phagocytosis of such mutant cells by macrophages was also reduced (Brennan *et al.*, 2001). Another PE-PGRS protein, Rv1759c is able to bind fibronectin and could thus mediate bacterial attachment to host cells (Espitia *et al.*, 1999). Immunization with the PE domain of PE-PGRS protein, Rv1818c, was proven to induce Th1-type responses that were not found when the complete PE-PGRS protein was

used. Instead, the PGRS part of the protein elicited antibodies and suppressed the Th1 response induced by the PE domain (Delogu and Brennan, 2001). It was found that there is some similarity between structural proteins of insects, such as silk, and the PGRS domain. This suggests that the role of the PE-PGRS proteins may be purely structural (Banu *et al.*, 2002). Some PE proteins may be essential for virulence. Mutations of two of the PE_PGRS genes of *Mycobacterium marinum*, which are the homologues of *M. tuberculosis* Rv3812 and Rv1651c, rendered *M. marinum* strains incapable of replication in macrophages and also resulted in decreased persistence in granulomas (Ramakrishnan *et al.*, 2000).

As mentioned previously, the PPE proteins of the MPTR subfamily also shows variability, however, not much evidence has been found concerning their possible functions. One member of the PPE protein family, Rv1917c, was shown to be cell-wall-associated and surface-exposed (Sampson *et al.*, 2001). PPE68 (Rv3873) was also found to be associated with the cell envelope (Okkels *et al.*, 2003). The PPE family protein Rv2608, which is a member of the major polymorphic tandem repeat (MPTR) subfamily, was found to elicit high humoral responses and low T-cell responses in TB patients (Chakhaiyar *et al.*, 2004). The PPE gene Rv0951c was found to develop both CD4 and CD8 specific T cell responses and could provide protection against *M. tuberculosis* comparable to *M. bovis* BCG vaccination when immunized in C57BL/6 mice (Skeiky *et al.*, 2000). Overall, it can be concluded that both the PE_PGRS and PPE-MPTR proteins may act as variable surface antigens (Banu *et al.*, 2002).

Besides the functions of individual PE and PPE family proteins that have been discovered so far, the PE/PPE protein complex was also found to elicit humoral and cell mediated immune responses. The PE25/PPE41 (Rv2431c/Rv2430c) complex induces significant B cell responses in sera derived from TB patients (Tundup *et al.*, 2008). The complex may play a different role other than eliciting immune responses. When the protein structure of the PE and PPE complex was elucidated, it was found that an apolar stripe appears along one side of the complex, suggesting a docking site for another protein (Figure 1.6) (Strong *et al.*, 2006). Strong's group used the metaserver ProKnow to predict the function of the PE/PPE complex. ProKnow infers functions for proteins based on sequence homology and structural similarity to other proteins of known function in the Protein Data Bank. The possible functions were expressed as Gene Ontology (GO) terms, each given with a Bayesian weight. The highest scoring GO term for biological process of this PE/PPE protein complex is "signal

transduction” with a probability of 75%. A similar result was concluded by using a combinatorial extension programme which identified protein structures with similar 3-D structures (Strong *et al.*, 2006). This predicted role of “signal transduction” still needs to be proved experimentally.

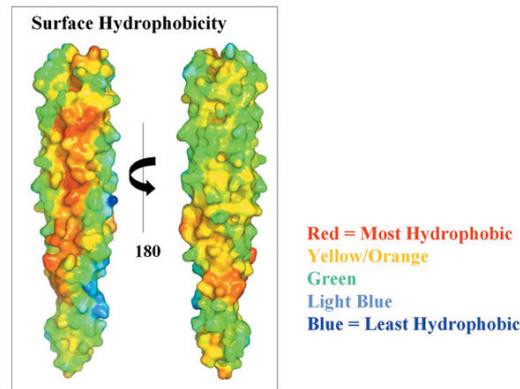


Figure 1.6 Surface hydrophobicity of the PE/PPE protein complex. One face of the PE/PPE complex surface has a stretch of apolar amino acids that may suggest a putative binding surface for other protein–protein interactions. Interestingly, this hydrophobic stretch overlaps with the conserved polyproline stretch and conserved PPE region of the complex. (Strong *et al.*, 2006)

Importantly PE-PGRS and PPE proteins, like ESAT-6 and CFP-10 acting as the substrates of ESX-1, were also found to be the substrates of ESX-5 secretion system (Abdallah *et al.*, 2009). It is not sure whether the PE and PPE proteins in other regions (ESX-2 and ESX-3) happen to be the substrates of their regions as well.

1.8 Mycosin, a subtilisin-like serine protease

Another protein family associated with the ESX gene cluster is the mycosins. The five mycosins belong to a family of transmembrane serine proteases. They contain a conserved catalytic triad (Asp, His, Ser), which is typical for the proteases of the subtilisin family (Brown *et al.*, 2000). One member, mycosin 1, whose gene is situated 3700 bp from the RD1 deletion region in the genome of the attenuated vaccine strain *M. bovis* BCG, was later confirmed to be a cell wall-associated extracellular protein expressed during infection of macrophages (Dave *et al.*, 2002). Mycosins are believed to modify the substrates upon their secretion (Gey van Pittius, *et al.*, 2001) or regulate the secretion and virulence (Ohol *et al.*, 2010). The substrates for the mycosins, however, have never been identified until a recent

study in which Ohol and coworkers found that EspB (which is an ESX-1 substrate) is a target of MycP1 *in vitro* and *in vivo*. Questions were soon raised about this discovery because EspB is only found in ESAT-6 gene cluster region 1 and thus not present in any of the other 4 ESX regions. Since the five mycosins evolved from duplication and are very conserved, their substrates are presumably from ESAT-6 gene clusters regions, such as ESAT-6, CFP-10 or even PE and PPE proteins. Unfortunately, studies on the mycosins are very limited. So far their structures, substrate specificity, and functions have not been revealed.

1.9 Potential drug target candidate

The mycosins are the most conserved enzymes encoded within the five virulence-associated ESAT-6 gene clusters and as such they are potential drug target candidates (Gey van Pittius, *et al.*, 2006). Their functions are unknown, but it is known that they are vital for the pathogenicity of *M. tuberculosis*. Their actual substrates have also not been identified or fully understood. Among the five of them, mycosin-3 (MycP3) was the only one found to be essential for *M. tuberculosis* growth and the gene cluster containing it could not be knocked out in this organism. Mycosin-3 thus provides a very important new potential drug target for treating TB.

1.10 Catalytic characteristics and stability of novel subtilisin-like serine proteases

Since there are not much known about mycosins, it is useful to look at their homologues in other organisms. It is known that MycP3, together with the other four mycosins, belong to the subtilase family (the superfamily of subtilisin-like serine proteases). It is hypothesized that the mycosins have similar characteristics to other subtilases even though previous protease substrate identification experiments were not successful (Dave *et al.*, 2002; N.C. Gey van Pittius, personal communication).

Subtilases can be found in numerous prokaryotes and eukaryotes such as gram-positive bacteria, slime molds, plants, insects, nematodes, mollusks, amphibians, fish, mammals, and even viruses. According to the sequence homology of their catalytic domains, this superfamily can be divided into 6 families namely subtilisin (A), thermitase (B), Proteinase K (C), lantibiotic peptidase (D), Kexin (E) and pyrolysin (F). Among all of them, only four residues in the catalytic region are conserved, including the catalytic triad residues Asp (D)

32, His (H) 64, Ser (S) 221, and a single glycine residue (G) 219. The substrate binding region of the subtilase proteases was presented diagrammatically by Siezen and coworkers (Figure 1.7).

The substrate/inhibitor binding pocket shown in Figure 1.7 accommodates six amino acid residues (P4, P3, P2, P1, P1' and P2'). The specificity largely depends on the interactions between P4-P1 residue side chains and S4-S1 clefts respectively. S1 and S4 binding sites in subtilisin (A) and thermitase (B) are large and hydrophobic, which explains the broad specificity of both enzymes with a preference for aromatic or large nonpolar P1 and P4 substrate residues (Gron *et al.*, 1992). Most of the subtilases have binding regions similar to those in subtilisin and thermitase so they should also have a broad specificity and can be considered as general-purpose proteases. Modeling and engineering studies have shown that a high density of negative charge at the substrate binding site, and in particular at S1, S2 and S4 sites, is responsible for high substrate selectivity (Lipkind *et al.*, 1995; Perona *et al.*, 1995; Siezen *et al.*, 1994). This suggests that if there are more electrostatic interactions between the binding pocket and the substrate residues instead of hydrophobic interactions, the subtilase is going to have a narrower selectivity for the substrate (Siezen and Leunissen, 1997).

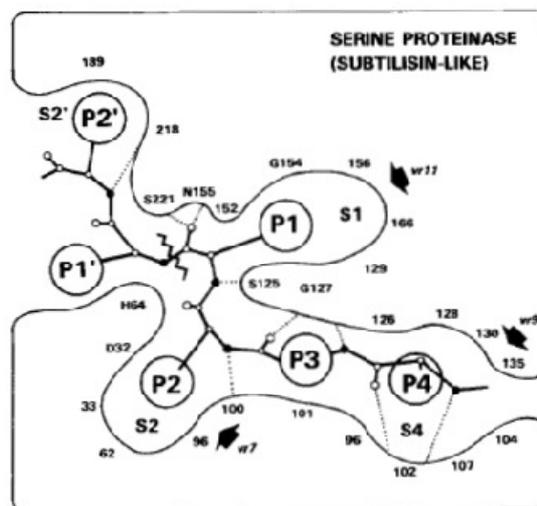


Figure 1.7 Schematic representation of substrate/inhibitor (bold lines) binding to a subtilisin-like serine protease. Side chains of the P4-P2' residues are shown as large spheres; position of the enzyme residues that may interact with these P4-P2' side chains are shown surrounding the binding sites (S1, S2, S2' S4). Hydrogen bonds between enzyme and substrate/inhibitor are shown as dotted lines, and the scissile bond is shown by a jagged line. Catalytic residues D32, H64, and S221, and oxyanion-hole residue N155 are indicated (taken from Siezen *et al.*, 1997).

Calcium ions are essential for the stability and activity of subtilases. There are usually four calcium binding sites in a typical subtilase indicating the importance of divalent ion for structural stability (Siezen and Leunissen, 1997). Disulfide bonds are much less important because subtilases do not rely on highly conserved disulfides for stabilization, and in fact, most subtilases do not have any disulfides (Siezen and Leunissen, 1997).

1.11 Methods to screen protease specificity

As described above, the mycosins are grouped under the domain of subtilases based on sequence similarity. However, their protease activities seem to be different from other members of these families and they seem to have very narrow substrate selectivity. In order to elucidate their substrates, protease specificity screening experiments need to be conducted. Several methods for identifying protease substrates were summarized by Agard and coworkers in 2009.

The first method which can be employed is two-dimensional differential gel electrophoresis (2D-DiGE), which separates proteins from a cell lysate into resolvable spots. The proteins in the first dimension resolve based on their pI value; the second dimension is run under the principle of SDS-PAGE thus the proteins separate based on size. Comparative staining reveals differences between proteolyzed and control samples which can be analyzed by mass spectrometry (MS) (Tonge *et al.*, 2001) (Figure 1.8A). A modified 2D-PAGE method employs SDS-PAGE on both dimensions with an intermediate in-gel proteolysis treatment step to identify protease targets (Shao *et al.*, 2007). The proteins which migrate off the diagonal in the second dimension SDS-PAGE gel are probably the substrates (Figure 1.8B). A one-dimensional SDS-PAGE-based method called PROTO-MAP was developed during Jurkat cell apoptosis (Dix *et al.*, 2008). In this method, apoptotic and control cell lysates were run parallel on one SDS-PAGE gel. The gel lanes were sliced into bands (Figure 1.8C). After in-gel trypsin digestion, the proteins in each band were identified by liquid chromatography-mass spectrometry (LC-MS), and quantified by spectral counting. Proteins from apoptotic cells that decreased in intensity or shifted from higher to lower apparent molecular weight were presumed to be caspase substrates.

Another approach for screening protease substrate specificity is to use a short peptide library where short peptides with 3 or 4 residues are constructed in a completely random manner

with all 20 amino acids (Thomas *et al.*, 2006). Usually there are several glycine residues on each side of the tri- or tetra-peptides, and a donor group and a quencher group on each end respectively are paired and this pair is in a quenched state prior to cleavage. When the short peptide is cleaved by the protease, the donor group will leave the quencher group resulting in the emission of fluorescence. By monitoring the increase of fluorescence over time, the protease specificity over certain peptides will be elucidated and the kinetic data can also be generated. Using this method, cleavage sites of the protease can be elucidated. The substrate specificity of MycP1 was recently identified by such an approach using a library of fluorogenic tetrapeptide substrates (Ohol *et al.*, 2010).

Knowledge about the substrate specificity of mycosin-3 and the sequences of the protease recognition and cleavage sites could be used to compare this with the known substrates of subtilases in order to find potential protease inhibitors.

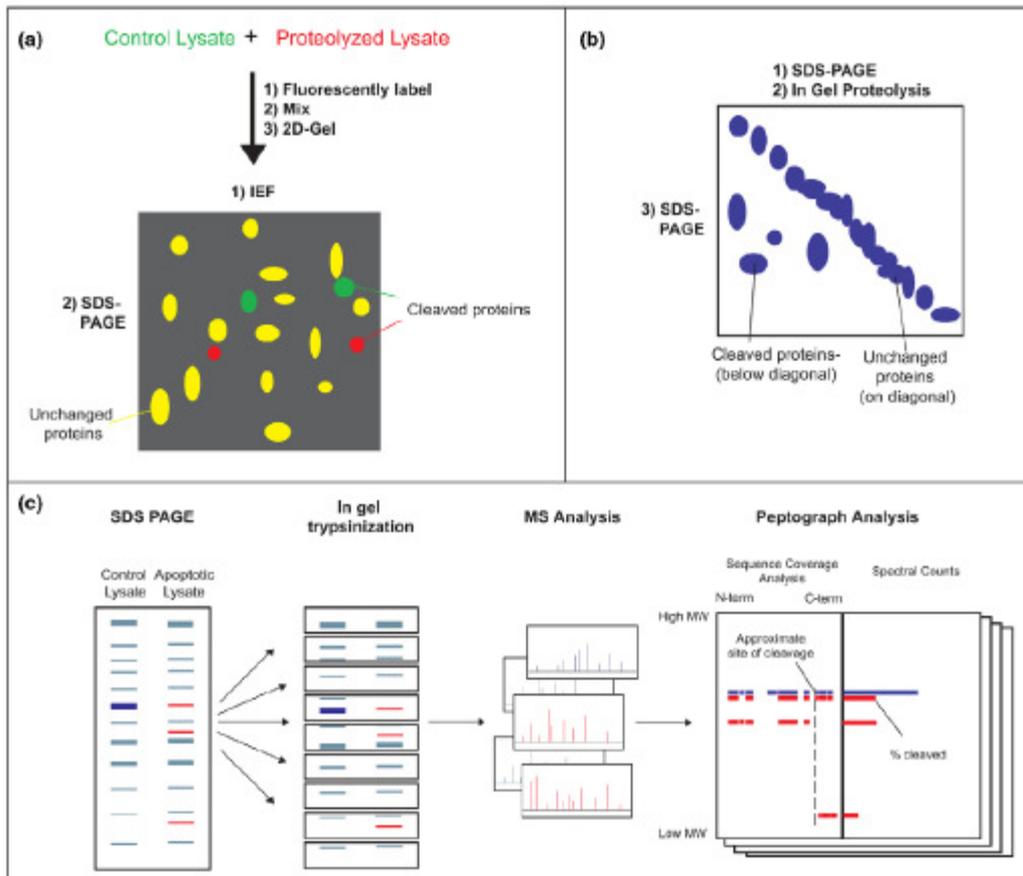


Figure 1.8 (A) 2D DiGE: A control and a proteolyzed lysate are fluorescently labeled, mixed and analyzed by a 2D-gel. Spots with unequal fluorescence ratios are picked up as potential substrates for subsequent MS analysis; (B) 2D SDS-PAGE: a lysate is resolved on one SDS-PAGE first, then in-gel treated with protease of interest. After the 2nd SDS-PAGE analysis, any spots below the diagonal are identified as substrate; (C) PROTO-MAP: control and proteolyzed cell lysates are analyzed side-by-side via SDS-PAGE. The gel is cut into bands, trypsinized, and peptides are identified by LC-MS. For each protein, peptides are analyzed by peptographs, which are analysis tools that display the sequence coverage and intensity in each band, revealing the approximate site(s) and extent of cleavage. (Taken from Agard and Wells, 2009)

1.12 Problem Statement

The ESAT-6 gene cluster has duplicated five times during evolution. The conservation of the gene clusters over time shows the importance and essentiality of the five secretion systems towards the survival of mycobacteria. The mycosins are the enzymes in the ESAT-6 gene cluster with the most conserved sequences. Among the five of them, mycP3 is essential for the growth of the organism, and an ESAT-6 gene cluster region 3 knock-out strain can not be generated. Studying the functions and elucidating the substrates of mycP3 will aid in the understanding of the mechanisms of the ESX-3 secretion system. This could lead to the design of new drugs to target the system in *M. tuberculosis*.

1.13 Hypothesis

We hypothesize that the substrates of mycosin-3 in *Mycobacterium tuberculosis* are the secreted proteins encoded by ESAT-6 gene cluster 3 (ESX-3), namely PE5 and PPE4, or esxG and esxH.

1.14 Aims of the project

1. To clone the mycosin-3 (with/without hydrophobic tail), PE5, PPE4, combined PE5 and PPE4, esxG , esxH, and combined esxG and esxH genes in an *E. coli* expression vector
2. To express the proteins in *E. coli* as His-tagged fusion proteins
3. To clone the genes of mycosin-3 without hydrophobic tail, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, combined esxG and esxH in a mycobacterial expression vector
4. To express the proteins in *M. smegmatis* as His-tagged fusion proteins
5. To purify all His-tagged fusion proteins using nickel columns
6. To conduct protease assays using purified mycosin-3 and selected potential substrate proteins PE5, PPE4, esxG and esxH in an appropriate buffer condition
7. To identify the substrate of mycosin-3 by conducting 2D-PAGE and Mass Spectrometry

Chapter 2 Materials and Methods

2.1 Strains and Plasmids

The *Escherichia coli* expression vector pET-28a (Novagen, Germany), as well as the *Mycobacterium smegmatis* expression vectors p19Kpro (Imperial College Medical School, UK) and pDMN1 (a kind gift from Dr. Ros Chapman, University of Cape Town, South Africa) were selected as the vectors for expressing target fusion proteins. The plasmid DNA maps of these expression vectors are given in Addendum 6.1. The reason for using two different mycobacterial expression vectors was because the strengths of the promoters of these two expression vectors are different. Vector pDMN1 contains the P_{smyc} promoter which is a much stronger promoter than the 19 kDa promoter (P_{19kDa}) in vector p19Kpro. Vector pDMN1 was used in the case where p19Kpro could not express the protein of interest in sufficient amounts. *E. coli* XL-1 blue strain was used for gene manipulation for all three expression vectors. *E. coli* BL21 (DE3) pLysS strain was used for recombinant fusion protein expression for the pET-28a expression vector. BL21 (DE3) pLysS is lysogenic for λ-DE3, which contains the T7 bacteriophage gene I, encoding T7 RNA polymerase under the control of the *lac* UV5 promoter (Studier and Moffatt, 1986). BL21 (DE3) pLysS also contains a plasmid, pLysS, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG (Studier and Moffatt, 1986). Therefore, BL21 (DE3) pLysS is an ideal strain of *E. coli* for protein expression compared to XL-1 blue strain. *M. smegmatis* mc²155 strain was used for recombinant fusion protein expression for the p19Kpro and pDMN1 expression vectors.

Electro-competent XL1 cells were used for recombinant plasmid DNA transformation and they were prepared as follows: Fresh XL1 single colonies were inoculated into 50 ml Luria Bertani (LB)-tetracycline (50 µg/ml) broth and incubated overnight at 37°C. The next morning, 20 ml of the overnight culture was re-inoculated into 2 L LB-tetracycline broth and grown at 37°C with constant shaking until the OD_{600nm} reached 0.6. The cells were harvested by centrifugation at 4000 xg for 10 min at 4°C, and the pellet was resuspended in a volume of ice cold sterile 10% glycerol equal to the original culture volume. The resuspension was centrifuged again at the same conditions as previously and the washing step was repeated. Thereafter, the supernatant was carefully poured off and the cells were resuspended in the

volume of glycerol remaining in the centrifuge bottle. The cells in all centrifuge bottles were pooled in a 50 ml sterile centrifuge tube and spun at 4000 xg at 4°C for 10 min. The supernatant was poured off and the pellet was resuspended in 2–3 ml of 10% glycerol. The competent cells were aliquoted into eppendorf tubes (100 µl per tube), snap frozen in liquid nitrogen and stored at -80°C.

Electro-competent *M. smegmatis* was used for recombinant *M. smegmatis* expression vector transformation. They were prepared as follows: One liter 7H9 broth was prepared by mixing 4.7 g 7H9 powder (Becton Dickinson, USA) in 970 ml H₂O, supplemented with 10 ml filter-sterilized 20% Tween 80 (Sigma, Germany), 10 ml filter-sterilized 50% glucose (Merck, Germany) and 10 ml filter-sterilized 50% glycerol (Merck, Germany). One ml of *M. smegmatis* stock culture was inoculated into 100 ml 7H9 broth and incubated overnight at 37°C. As soon as the OD_{600nm} reached 0.5, the overnight culture was incubated on ice for 1 hour. Cells were harvested by centrifugation at 4000 xg at 4°C for 10 min. The supernatant was discarded and the pellet was washed with 50 ml 10% glycerol after which cells were collected again by centrifugation as described previously. The washing step was repeated once and the pellet was re-suspended in 4 ml of 10% glycerol. Finally, the fresh electro-competent *M. smegmatis* cells were snap-frozen at 200 µl per aliquot.

2.2 Primer Design

Genes encoding mycosin-3 (mycP3, Rv0291), PE5 (Rv0285), PPE4 (Rv0286), esxG (Rv0287), and esxH (Rv0288) were selected for cloning. The primers designed for cloning purposes are given below. As PE5 and PPE4 are co-operonic in the genome, it was decided to clone both genes together into the three expression vectors. esxG and esxH are also co-operonic and the same strategy was followed for these genes. Pairs of primers were designed for PCR amplification of the genes into the three different expression vectors pET-28a, p19Kpro and pDMN1 (See Table 2.1, Table 2.2 and Table 2.3).

Table 2.1 Sequences of primers designed for cloning genes into pET-28a expression vector.

Genes	Sense primer	Anti-sense primer
Mycosin-3 without hydrophobic tail	5'-CCATGGATGCGTGCAGCGTCT-3'	5'CTCGAGTTGCCAGGTCAGGGCC-3'
Mycosin-3 with hydrophobic tail	5'-CCATGGATGCGTGCAGCGTCT-3'	5'-CTCGAGTTCGGTGGGCTCCCTTCG-3'
PE5	5'-CCATGGTTACGTTGCGAGTGGTT-3'	5'-CTCGAGCACGACCCCGTACGT-3'
PPE4	5'-CCATGGTTCCCATCTGGATGGC-3'	5'-CTCGAGCTTGCTGTCGTGCGGTAAG-3'
Combined PE5 and PPE4	Same as PE5 sense primer	Same as PPE4 anti-sense primer
esxG	5'-CCATGGTTAGCCTTTTGGATGC-3'	5'-CTCGAGGGTATAGGTGACGCG-3'
esxH	5'-TCATGACTTCGCAAATCATGTAC-3'	5'-CTCGAGGCCCCATTTGG-3'
Combined esxG and esxH	5'-TCATGATTAGCCTTTTGGATGCTC-3'	5'-CTCGAGGCCCCATTTGGC-3'

All sense primers contained *NcoI* restriction sites (5'-CCATGG-3') except the ones for esxH and combined esxG and esxH, where the *BspHI* restriction site (5'-TCATGA-3') was used instead. The reason for this was because the gene sequence of esxH contains an *NcoI* site which could thus not be used in the primer. *BspHI* was used so that the cloned genes could be restricted intact and it generates identical overhang termini to *NcoI*. Since the *NcoI* restriction site contains a start codon ATG, two nucleotide residues were introduced between the 3' end of the restriction site and the 5' end of the gene so that the cloned genes would be in frame inside the expression vector. All anti-sense primers contained an *XhoI* restriction site (5'-CTC GAG-3').

Table 2.2 Sequences of primers designed for cloning genes into the p19Kpro expression vector.

Genes	Sense primer	Anti-sense primer
Mycosin-3 without hydrophobic tail	5'-GGATCCAGGCGGAGAACAAA TGGGTGCGTGCAGCGTCTC-3'	5'-ATCGATTTAGTGGTGGTGGTGG TGGTGTGCCAGGTCAGGGCCG-3'
PE5	5'-GGATCCAGGCGGATCAGTCA TGACGTTGCGAGTGGTTCC-3'	5'- ATCGATTTAGTGGTGGTGGTGGTGG TGGCCACGACCCCGTACGTAGC-3'
PPE4	5'- GATATCAGGCGGACTGAGCATG GCCGCGCCCAT-3'	5'- ATCGATTTAGTGGTGGTGGTGGTGG TGT GCTTGCTGTCGTGCGGTAAG-3'
Combined PE5 and PPE4	Same as PE5 sense primer	Same as PPE4 anti-sense primer
esxG	5'- GGATCCAGGCGGAGATGTTATG A GCCTTTTGGA-3'	5'- ATCGATTTAGTGGTGGTGGTGGTGG TGT GGAACCCGGTATAGGTCGACG-3'
esxH	5'- GGATCCAGGCGGACTTGTGATG TC GCAAATCATGTACAACACTACCCC G-3'	5'- ATCGATTTAGTGGTGGTGGTGGTGG TGT GGCCCCATTTGGCGGCTTCGG-3'
Combined esxG and esxH	Same as esxG sense primer	Same as esxH anti-sense primer

The mycobacterial expression vector p19Kpro does not contain a ribosomal binding site (rbs), a start codon (ATG), a 6 x Histidine tag or a stop codon. Thus, in all the sense primers except the ones for PE5, PPE4 and the combination of the two, a *Bam*HI restriction site (5'-GGATCC-3) was designed at the 5' end, followed by a rbs "AGGCGGA" (Table 2.2). Six nucleotides were incorporated between the rbs and the start codon. For the sense primers of PE5, PPE4 and the combined PE5 and PPE4, an *Eco*RV restriction site was used since the genes for PE5 and PPE4 contain *Bam*HI restriction sites within their sequences. In all anti-sense primers, a 6x GTG sequence of which the compliment sequence encodes a 6x His tag was designed after the gene sequence, followed by a stop codon and a *Cl*aI restriction site (5'-ATCGAT-3').

The mycobacterial expression vector pDMN1 contains a start codon "ATG" 19 base pairs upstream from a *Hind*III restriction site. All sense primers thus contained a 5' end *Hind*III restriction site, as well as 2 nucleotides in between the start codon of the gene and the 3' end

of the *Hind*III restriction site in order for the gene to be cloned in frame (Table 2.3). All anti-sense primers contained a complimentary 6 x Histidine sequence and a stop codon “TGA”.

Table 2.3 Sequences of primers designed for cloning genes into pDMN1 expression vector.

Genes	Sense primer	Anti-sense primer
Mycosin-3 without hydrophobic tail	5'- AAGCTTTTCGCGGTGCGTGCA GCGTCTCCGGTGTTATCCCGGG C-3'	5'-GTAACTCAGTGGTGGTGGTG GTGGTGTGTCAGGTCAGGGCCG C-3'
PE5	5'- AAGCTTTTATGACGTTGCGAGT G GTTCCGGAGGGGCTGGCCGAG -3'	5'-GTAACTCAGTGGTGGTGGTG GTGGTGGCCGCCACGACCCCGT AC-3'
PPE4	5'- AAGCTTTTATGGCCGCGCCCAT CT GGATGGCTTCGCCCGGAGG- 3'	5'-GTAACTCAGTGGTGGTGGTG GTGGTGCTTGCTGTCGTGCGGTAA G-3'
Combined PE5 and PPE4	Same as PE5 sense primer	Same as PPE4 anti-sense primer
esxG	5'- AAGCTTTTATGAGCCTTTTGGAT G CTCATATCCCACAGTTGGTGG-3'	5'- GTAACTCAGTGGTGGTGGTGGT G GTGGAACCCGGTATAGGTCTGA-3'
esxH	5'- AAGCTTTTATGTCCGAAATCAT GT ACAAC TACCCCGCGATGTTG-3'	5'- GTAACTCAGTGGTGGTGGTGGT G GTGGCCGCCCATTTGGC-3'
Combined esxG and esxH	Same as esxG sense primer	Same as esxH anti-sense primer

2.3 Polymerase Chain Reaction

All oligonucleotide primers were purchased from Whitehead Scientific (Pty) Ltd, South Africa. Lyophilized primers were dissolved in TE (Tris-EDTA) buffer (10 mM Tris, 1 mM EDTA, pH 8.0) so that the final concentration of the oligonucleotide stock was 100 µM. The working concentration of the oligonucleotide was 10 µM. For each PCR experiment (one tube, 25 µl in volume), 14.8 µl DNase-free water (Promega, USA), 2.5 µl 10x buffer containing MgCl₂(Roche, Switzerland), 1 µl dNTP (Promega, USA), 0.2 µl Taq polymerase (Roche, Switzerland), 5 µl 5x GC solution (Roche, Switzerland), 0.5 µl of forward and

reverse primer respectively and 0.5 µl *M. tuberculosis* H37Rv reference strain genomic DNA were added. The PCR programme was set up as follows, stage 1: 95°C for 15 minutes; stage 2 (which was repeated 40 times): 94° for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute; stage 3: 72°C for 10 minutes; and stage 4: 4°C for 10 minutes. After the PCR amplifications were finished, the PCR products were subjected to agarose gel electrophoresis for separation and ethidium bromide staining for visualization under the ultraviolet (UV) light (Aaij and Borst, 1972).

2.4 Cloning of the Genes into the Expression Vectors

2.4.1 Cloning of the genes into the pGEM-T Easy cloning vector

The PCR products were separated on a 1% [w/v in TAE (Tris-Acetate-EDTA) buffer] agarose gel by electrophoresis at 100 volts for 70 min. For a liter of 50x TAE buffer, 242g Tris, 57.1 ml glacial acetic acid, and 100 ml 0.5 M EDTA were mixed; the pH was adjusted to 8.0 and the volume was made up to 1 liter. The DNA molecules were stained inside the gel with 0.001 mg/ml (w/v) ethidium bromide (Sigma, Germany) [stock solution 10 mg/ml (w/v)]. DNA bands of appropriate sizes were cut out of the agarose gel. The DNA was purified from the agarose gel using the Wizard[®] SV Gel and PCR Clean-Up System (Promega, USA).

The concentrations of the PCR products after the purification were quantified by NanoDrop[®] (WhiteSci, South Africa). Ligations of the PCR products into the pGEM-T Easy vectors were carried out in a molar ratio of 3:1. The formula used for calculating the volume of DNA added in one ligation reaction is as follows: (ng of vector × kb size of insert) × (insert : vector molar ratio) = ng of insert × kb size of vector. Using the PCR product of Rv0285 (PE5) as an example, the amount of insert added in one ligation reaction was (50 ng of pGEM-T Easy vector × 0.311 kb) × 3 / 3 kb = 15.55 ng of Rv0285 PCR product. The same method was applied to the other types of PCR products. The ligations were carried out at 4°C overnight.

An electroporator (Biorad, USA) was used for transforming the recombinant pGEM-T Easy constructs into electrocompetent *E. coli* XL1 cells. One microliter of ligation mix containing

the recombinant DNA, together with 15 µl sterile 10% glycerol was mixed with 50 µl of XL1 competent cells and transferred into a pre-chilled electroporation cuvette (Bio-Rad, USA). Electroporation was carried out at 2.5 kV with 200 ohms resistance. The time constant was 3.9 s.

After electroporation, 600 µl of Super Optimal Broth with glucose (SOC) medium [2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄·7H₂O and 20 mM glucose (Merck, Germany)] was added to the cuvette. The cells and the medium were mixed and transferred to a sterile eppendorf tube, after which it was incubated in a shaking 37°C incubator for 60 min. An aliquot of 75 µl of the cells were plated out on LB-ampicilin (100 ng/µl, Roche, Germany) plates which were previously treated with 100 µl of 0.1 mM IPTG (Invitrogen, USA) and 100 µl of 20 mg/ml X-gal (Roche, Germany) to enable blue/white colony selection. The plates were incubated overnight at 37°C. White colonies were picked and used in colony PCR to confirm the presence of the inserts and the success of the pGEM-T Easy cloning. During colony PCR, the picked colonies acted as the template DNA and were mixed with the PCR reaction mix for the amplification reaction.

2.4.2. Cloning of inserts into expression vectors

Colonies confirmed to contain the appropriate inserts in the recombinant pGEM-T Easy constructs were inoculated into 5 ml LB+Ampicilin broth and incubated overnight at 37°C. Recombinant plasmid DNA was isolated from the mini-cultures using the Wizard[®] SV Plus Miniprep DNA Purification System (Promega, USA). The isolated recombinant vector was double digested with *NcoI/BspHI* and *XhoI* (Fermentas, USA) for cloning of the insert into the pET-28a expression vector, *BamHI/EcoRV* and *ClaI* (Roche, Germany) for cloning of the insert into the p19Kpro expression vector, and *HindIII* (Roche Germany) and *HpaI* (Fermentas, USA) for cloning of the insert into the pDMN1 expression vector.

The digested reaction mix was run on a 1% agarose gel and the inserts were cut out of the agarose gel and purified as previously described. The inserts were dephosphorylated using shrimp alkaline phosphatase (SAP) (Roche, Germany) by incubation at 37°C for 45 min. SAP was deactivated by incubation at 65°C for 15 min, and the treated inserts were ligated with

double-digested expression vectors using T4 DNA ligase (Promega, USA). Ligation reactions were carried out at 4°C overnight.

An aliquot of 1.5 µl of the ligation mix was used for electroporation. All the cells used for electroporation were plated onto LB agar plates with an appropriate antibiotic added [Kanamycin (Sigma, Germany) for pET-28a and pDMN1, 30 ng/µl; Hygromycin (Invitrogen, USA) for p19Kpro, 50 ng/µl]. The plates were incubated at 37°C overnight. Colony PCR was applied to check whether the colonies grown on the plates contained the insert DNA. Clone verification was also done by plasmid DNA sequencing with the concentration of plasmid DNA sample at 100 ng/µl (DNA Sequencing Facility at Stellenbosch University).

2.5 Test Expression and Protein Purification

2.5.1 Test Expression of the pET-28a E. coli expression vector constructs in the E. coli BL21 (DE3) strain

Recombinant expression vector pET-28a containing insert DNA of mycosin-3 with hydrophobic tail, mycosin-3 without hydrophobic tail, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, and combined esxG and esxH were all electroporated into BL21 (DE3) *E. coli* (Promega, USA) electro-competent cells. Preparation of BL21 (DE3) electro-competent cells and electroporation conditions were the same as that of XL1 cells described previously. The BL21 (DE3) *E. coli* transformants containing specific recombinant DNA were inoculated into 5 ml Rich Broth (1% tryptone, 0.5% yeast extract, 1% sodium chloride, 0.1% glucose) and incubated at 37°C overnight. The following day, a 500 µl aliquot from the 5 ml culture was re-inoculated into 9 conical flasks individually, each of which contained 50 ml Rich Broth, and incubated at 37°C. Upon reaching OD_{600nm} ~ 0.6, IPTG was added to 3 sets of 3 conical flasks to a concentration of 0.25 mM, 0.5 mM, and 1 mM respectively. After the addition of IPTG, expression was started by incubating the 9 flasks at three different temperatures (24°C, 30°C and 37°C) (See Table 1.4). The expression was allowed to carry on for 4 hours.

Table 2.4 Method of Test Expression of pET-28a in BL21 (DE3) *E. coli* strain, same amount of inoculant was made for each flask

	24°C	30°C	37°C
[IPTG] = 0.25 mM	Flask1	Flask2	Flask3
[IPTG] = 0.5 mM	Flask4	Flask5	Flask6
[IPTG] = 1 mM	Flask7	Flask8	Flask9

The protein-expressing culture was harvested by centrifugation at 4000 xg at 4°C for 15 min. The supernatant was discarded and the pellet was stored at 80°C. The pellets containing His-tagged fusion proteins were treated subsequently using the MagneHis™ Protein Purification System (Promega, USA) for small scale His-tagged fusion protein purification according to the manufacturer's instructions. The samples of purified His-tagged proteins were analyzed by SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE, Laemmli, 1970) (see Section 2.3.3.1 for SDS-PAGE preparation). The polyacrylamide gel was stained by either Coomassie blue stain or Silver stain (see section 2.3.3.2 for gel staining methods). Optimal expression conditions were scored by the intensity of the protein band on the stained SDS-PAGE gel. These optimal conditions were subsequently applied for the protein expression of large scale His-tagged fusion protein purification using the Ni-NTA Superflow column (Qiagen, Germany) (see Section 2.5.4).

2.5.2. Test Expression of the p19Kpro and pDMN1 mycobacterial expression vector constructs in *M. smegmatis* mc²155

Recombinant expression vectors p19Kpro and pDMN1 containing insert DNA of mycosin-3 without hydrophobic tail, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, and combined esxG and esxH were electroporated into electro-competent *Mycobacterium smegmatis* mc²155 cells.

An aliquot of 200 µl electro-competent *M. smegmatis* cells was used for one electroporation reaction. Electroporation was carried out at 2.5 kV with 1000 ohms resistance. The time constant was usually between 15 s and 18 s. After electroporation, 1 ml 7H9 broth was added and the cells were incubated at 37°C for 3 hours before a 100 µl aliquot was plated onto 7H11 agar plates. The 7H11 agar medium was prepared by mixing 19g of 7H11 powder (Becton Dickinson, USA) with 970 ml H₂O; glycerol, glucose and Tween 80 were added in the same way as 7H9 broth. Kanamycin (30 ng/µl) or hygromycin (50 ng/µl) were also added for positive transformants selection. The plates were incubated at 37°C for at least 60 hours.

Since both p19Kpro and pDMN1 mycobacterial expression vectors were non-inducible vectors, the His-tagged fusion proteins were extracted directly from a volume of 100 ml *M. smegmatis* transformants culture which was grown and collected by centrifugation at 4000 xg at 4°C. After centrifugation, the supernatant was discarded and the fusion proteins were extracted using the MagneHis™ Protein Purification System according to the manufacturer's instructions. However, instead of chemical lysis of the culture as instructed on the manual of the MagneHis™ Protein Purification System, ribolysis was applied because it is more efficient for breaking open the cell walls of *M. smegmatis*. The protocol for ribolysis was as follows. The pellet (collected from 100 ml *M. smegmatis* culture) was resuspended in 1 ml ice cold PBS (phosphate-buffered saline) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.76 mM KH₂PO₄, pH 7.4). The resuspension was divided and transferred into two 2 ml screw cap microtubes (Quality Scientific Plastics, USA) and an equal volume of 0.1 mm glass beads (Roth, Germany) was added to the screw cap microtubes. Another 500 µl PBS was added and the mixture was ribolyzed using the ribolyser (Hybaid, UK) with the programme set for 6 repeats of 20 sec (ribolysation step) and a subsequent 60 sec at 4°C (incubation step). The microtubes were centrifuged at 14000 rpm at 4°C for 10 min and the supernatants were pooled and subjected to the His-tagged fusion protein purification procedure using the MagneHis™ Protein Purification System. The extracted protein samples were analyzed by SDS-PAGE and Western blotting.

2.5.3. SDS-PAGE, Gel Staining and Western blotting

2.5.3.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels were composed of two parts: a 15% resolving gel and a 3% stacking gel. The 15% resolving gel was prepared first, from a master mix containing 2.35 ml H₂O, 5 ml acrylamide/bisacrylamide 29:1 solution (Bio-rad, USA), 2.5 ml 1.5 M Tris-HCl (Merck, Germany) pH 8.8, 0.1 ml 10% SDS (Merck, Germany), 10 µl TEMED (Sigma, Germany) and 50 µl 10% ammonium persulphate (Sigma, Germany). After the resolving gel was set, a 3% stacking gel was loaded onto the resolving gel. It was prepared from a master mix consisting of 7.4 ml H₂O, 650 µl acrylamide/bisacrylamide 29:1 solution, 1.25 ml 1.5 M Tris-HCl pH 6.8, 0.1 ml 10% SDS, 10 µl TEMED and 50 µl 10% ammonium persulphate. After the stacking gel was set, the tank was filled up with SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS).

Prior to sample loading, 10 μ l protein sample was mixed with 10 μ l sample buffer which was prepared as a master mix containing 3.4 ml 1M Tris, pH 6.8, 2 ml glycerol, 3 ml 20% SDS, 0.5 ml 0.75% bromophenol blue and 1 ml β -mercaptoethanol. The 20 μ l protein sample mix was heated at 100°C for 5 min. The protein samples and protein pre-stained marker (Fermentas, USA) were loaded parallel in the same gel. The gel was run at 150 Volts until the dye front reached the bottom of the gel.

2.5.3.2 Polyacrylamide Gel Staining Technique

Two staining techniques were used to visualize the proteins in the polyacrylamide gel, namely Coomassie blue staining and silver staining.

Two versions of the Coomassie blue staining methods were used in this project. The standard version was applied to stain mini-SDS-PAGE gels; while the modified version (which is suitable for mass spectrometric analysis of the protein spots) was applied to 2D-SDS-PAGE gels. With the standard Coomassie blue staining technique, the post-run polyacrylamide gel was immersed in Coomassie staining solution [0.25% brilliant blue R250 powder (Sigma, Germany), 45% methanol, 10% glacial acetic acid] and shaken on a rotator for 2 hours. The gel was destained with destaining solution (45% methanol, 10% glacial acetic acid) for three hours and then stored in water. With the modified Coomassie blue staining technique, the 2-D SDS-PAGE gel was firstly immersed in 250 ml Coomassie 1 solution [0.025% (w/v) Brilliant blue R-250, 25% (v/v) isopropanol, and 10% (v/v) acetic acid] and heated in a microwave for 2 min. It was placed on a shaker for fixation overnight. The solution was decanted and the gel was immersed in 250 ml Coomassie II solution [0.003% (w/v) Brilliant Blue R-250, 10% (v/v) isopropanol, 10% (v/v) acetic acid] and heated in a microwave for 2 min. It was placed on a shaker for 30 min. The solution was decanted and the gel was immersed in 250 ml Coomassie III solution [0.003% (w/v) Brilliant blue R-250, 10% (v/v) acetic acid] and heated in a microwave for 2 min. Thereafter, it was placed on a shaker for 30 min. The solution was decanted again and the gel was immersed finally in 250 ml destaining solution [10% (v/v) acetic acid, 1% (v/v) glycerol] overnight. The gel was stored in distilled water after destaining.

With the silver staining technique, the post-run polyacrylamide gel was soaked in fixation solution [50% (v/v) ethanol, 12% (v/v) glacial acetic acid, 0.05% (v/v) formalin] at room

temperature for at least 2 hours. The fixation solution was discarded and the gel was immersed in washing solution [20% (v/v) ethanol] for 20 min. The washing solution was changed twice during the 20 min. The washing solution was discarded and sensitizing solution [0.02% (w/v) sodium thiosulfate] was added. The gel was soaked in sensitizing solution for 2 min and washed twice with cold deionized water twice, each for 1 min. The gel was subsequently soaked in fresh cold silver staining solution [0.2% (w/v) silver nitrate, 0.076% (v/v) formalin] at 4°C for 1 hour with constant shaking. After the staining, the gel was rinsed twice with H₂O, 1 min each. The development was done by immersing the gel in developing solution [6% (w/v) Na₂CO₃, 0.0004% (w/v) Na₂S₂O₃, 0.05% (v/v) formalin] until the light brown coloured bands showed up on the gel. The developing reaction was stopped by adding 50 ml termination solution [12% (v/v) glacial acetic acid]. The gel was stored in 1% (v/v) acetic acid solution.

2.5.3.3 Western blotting

Post-run polyacrylamide gels were also used for Western blotting for detecting His-tagged fusion proteins. As soon as the electrophoresis was completed, the Western blotting apparatus was set up in the following order: 1 pad, 1 Whatman sheet, polyacrylamide gel, HybondTM-ECL nitrocellulose membrane (GE Healthcare, USA), 1 Whatman sheet, and 1 pad. The sandwich was closed and inserted into a Western blotting apparatus (Bio-rad, USA), together with an ice pack. The buffer chamber was filled with Transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3), a magnetic bar was placed inside the chamber and the whole apparatus was placed onto a magnetic stirrer. The transfer was done at 100 volts for 60 min at 4°C. The apparatus was dismantled and the membrane was taken out and reversibly stained with Ponceau red dye for verification of protein transfer. The membrane was blocked overnight with blocking buffer (20 mM Tris, 137 mM NaCl, 0.1% Tween 80, 10% milk powder, 1% BSA) at room temperature. The blocked membrane was rinsed three times with TBS-T buffer (20 mM Tris, 137 mM NaCl, 0.1% Tween 80), and incubated with mouse anti-6x histidine primary antibody (Invitrogen, USA), previously diluted in TBS-T buffer at 1:5000, for 60 min at room temperature. The membrane was rinsed twice with TBS-T buffer and then washed 3 times, each for 10 min, in TBS-T buffer with shaking. Horse-radish peroxidase (HRP)-conjugated goat anti-mouse secondary antibody (Santa-cruz, USA), was applied to the membrane with in a 1:1000 dilution in TBS-T buffer. The membrane was incubated for 60 min at room temperature. The washing step was repeated as described

previously, and the signal was developed using ECL detection fluid (GE healthcare, USA). The membrane and an autoradiographic film were placed into an autoradiography cassette (GE healthcare, USA) for between 10 sec and 1 hour. The film was developed using a Hyper Processor developer (Amersham, USA).

An alternative Western blotting technique was also performed using the HisDetectorTM Western Blot Kit (KPL, USA) in cases where antibody detection through ECL were problematic. Additionally, anti-mycosin-3 antibody (a kind gift from N.C. Gey van Pittius) was used for detecting the His-tagged fusion mycosin-3.

2.5.4. Large-scale protein expression and purification

The optimal expression conditions obtained in the test expression were applied when large-scale protein purification was required. For large-scale His-tagged fusion protein purification, large cultures (1-4 liter) of transformed *E. coli* or *M. smegmatis* was prepared for expression according to the specific expression vector used. After the expression was completed, the cells were collected by centrifugation at 4000 xg at 4°C. The supernatant was discarded and the pellets were stored at -80°C. The frozen pellets were prepared and run over the Ni-NTA Superflow Columns (Qiagen, Germany) according to the manufacturer's instructions, for purification of the His-tagged proteins. The end products, purified His-tagged fusion proteins, were stored at -20°C if it was not for immediate use.

2.5.5 Protein Assays

Protein was quantified using the Quick StartTM Bradford Protein Assay Kit (Bio-rad, USA). An aliquot of 10 µl protein sample (or the protein standards) was mixed with 200 µl of the Quick StartTM Bradford dye reagent in one well on a 96-well microtiter plate. The plate was incubated at room temperature for 5 min. The absorbance was measured at a wavelength of 595 nm using 96-well plate reader (BioTek, USA). Protein standards were prepared by dissolving bovine serum albumin (BSA) powder in distilled water making up the solutions with concentrations of 2 mg/ml, 1.5 mg/ml, 1 mg/ml, 0.75 mg/ml, 0.5 mg/ml and 0.25 mg/ml.

2.6. Protease assay, 2-Dimensional gel electrophoresis and mass spectrometric analysis

2.6.1 Protease Assay

The protease assay method used was published by Yamini, *et al.* (2010). The buffer conditions of the His-tagged fusion proteins in the eluant of the large-scale protein purification were changed using a PD-10 desalting column (GE health science, USA). In order for the proteins to be optimally visualized on 2-D gel with Coomassie blue staining, the proteins were concentrated using centrifugal filter devices (Millipore, USA). MycP3 activity assays were performed in a mycosin activity buffer (20 mM Tris pH 8, 100 mM NaCl, 2 mM CaCl₂, and 0.01% Brij-35) (Ohol *et al.*, 2010), together with its potential substrates (molar ratio was 1:1). The reaction was incubated at 37°C for 60 min to allow protease cleavage to take place.

2.6.2 Two-Dimensional Gel Electrophoresis and Mass Spectrometric Analysis

Prior to isoelectric focusing (IEF) (first dimension gel electrophoresis), protein samples had to undergo a clean-up procedure using the ReadyPrep™ 2-D Cleanup Kit (Bio-Rad, USA) in order to remove any salts in the protein sample which might interfere with the conductivity of the IEF run. For mass spectrometric analysis, the amount of protein loaded onto the IEF strip was around 200 µg to be visualized by Coomassie blue stain. For silver staining, 60 µg of protein was sufficient.

One hundred microliters of cleaned-up protein sample was mixed with 100 µl rehydration buffer (Bio-Rad, USA). The 200 µl mixture was transferred onto one lane of the tray. The mixture was covered by one ReadyStrip IPG strip, pH 4-7 (Bio-Rad, USA); and incubated at room temperature for one hour. Four milliliters of mineral oil was added on top of the strip allowing dehydration overnight. The strip was transferred into one lane of the IEF chamber, and the lane was filled with 3 ml mineral oil. The run was set at 8000 volts for 40000 voltage hours (Protean IEF cell, Bio-Rad, USA). After the run, the strip was stored at -80°C until being used for the second dimension polyacrylamide gel electrophoresis.

Prior to the second dimension SDS-PAGE, the strip was treated by immersing into two equilibration buffers and shaken for 10 min at room temperature, 10 ml for each, sequentially.

The first buffer contained 0.375 M Tris-HCl, pH 8.8, 6 M urea, 20% glycerol, 2% SDS, and 2% DTT (Bio-Rad, USA). In the second buffer, instead of 2% DTT, 2% iodoacetamide was used.

The 15% acrylamide gel was prepared from a master mix made from 60.16 ml 1.5 M Tris-HCl, pH 8.8, 120 ml acrylamide/PDA [30% acrylamide and 0.8% PDA, deionised with mixed resin TMD-8 (Sigma, Germany) and then filtered out], 56.9 ml H₂O, 0.92 ml 20% sodium thiosulphate, 123 µl TEMED and 1.85 ml 10% APS. The gel was overlaid with water saturated butanol. After the gel was set, the wells were washed with water and dried by blotting with Whatman paper.

The strip, together with a small piece of whatman paper (1cm length and the same width as the strip) which was wet with 8 µl pre-stained protein ladder solution (Fermentas, USA), was loaded onto the gel. One milliliter molten 0.5 % agarose (Bio-Rad, USA) was overlaid onto the top of the strip. After the agarose was set, the apparatus was set up. The top section was filled with cathode buffer [192 M glycine pH 8.3 (adjusted with Tris), 0.1% SDS]; and the tank was filled with a mixture of 1.5 L anode buffer [0.375 M Tris, pH 8.8 (adjusted with glacial acetic acid)] and 1.5 L distilled water. The electrophoresis was carried out at 180 volts at 12°C for 6 hours or until the bromophenol blue dye front reached the bottom.

The 2-D polyacrylamide gel was stained with Coomassie blue or silver nitrate (See Section 2.5.3.2 for details). Gels containing only mycosin-3 or the individual potential substrates acted as the control gels, while gels with protease activity assay products acted as the result group. These two groups were compared in order to identify spots of difference. Identified spots would be cut out and sent for mass spectrometric analysis for protein sequencing and identification (Mass Spectrometry Service, Department of Biotechnology, University of Western Cape).

Chapter 3 Results

3.1 Cloning of mycP3 and potential substrates

Three groups of primers were successfully designed in this study. In Group 1, 8 pairs of primers were designed for amplifying genes of mycosin-3 with transmembrane region, mycosin-3 without transmembrane region, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, and combined esxG and esxH, for cloning into *Escherichia coli* expression vector pET-28a. In Group 2, 7 pairs of primers were designed for amplifying genes of mycosin-3 without transmembrane region, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, and combined esxG and esxH, for cloning into *Mycobacterium smegmatis* expression vector p19Kpro. In Group 3, the same strategy was applied as in Group 2; but they were designed for cloning into the other *Mycobacterium smegmatis* expression vector pDMN1.

All primers mentioned above were successfully used to generate the correct amplification products through PCR. PCR products were all successfully cloned into the pGEM-T Easy cloning vector and subsequently subcloned into the respective expression vectors. Colony PCR and sequencing of the constructs confirmed all cloning steps and verified that all 22 clones were generated successfully. Owing to the large amount of experimental results for this cloning section, only the confirmatory colony PCR results for the recombinant pET-28a expression vectors are given in this section (see Figure 3.1). The sizes of the PCR fragments of each gene for cloning into the expression vectors p19Kpro and pDMN1 were basically the same as those for cloning into pET-28a, so colony PCR results for these two mycobacterial expression vectors are not given in this chapter. Detailed results are shown in Addendum 6.2.

Verification of these clones was also done by plasmid DNA sequencing (results not shown). All clones contained the insert of interest with no mutations detected by sequencing.

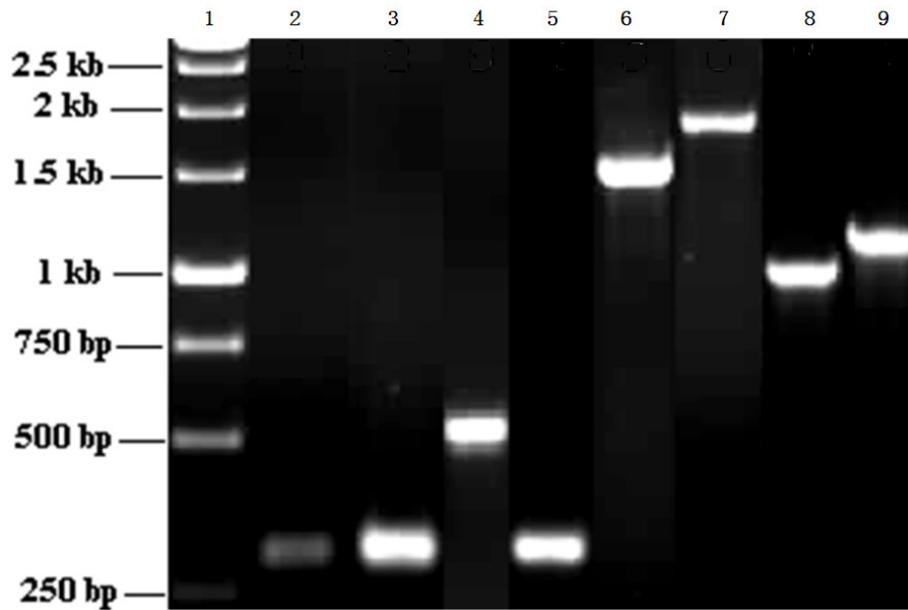


Figure 3.1 Agarose gel electrophoresis image of colony PCR confirmation of recombinant expression vector pET-28a containing inserts of interest, Lane 1: 1kb GeneRuler[®] DNA markers (Fermentas, USA); Lane 2: esxG (296 bp); Lane 3: esxH (293 bp); Lane 4: combined esxG and esxH (619 bp); Lane 5: PE5 (311 bp); Lane 6: PPE4 (1544 bp); Lane 7: combined PE5 and PPE4 (1855 bp); Lane 8: mycosin-3 without hydrophobic tail (1058 bp) and Lane 9: mycosin-3 with hydrophobic tail (1241 bp)

3.2 Expression and Purification

Test expression of the pET-28a expression vector recombinant clones were initially conducted by running crude cell lysate of a 1 ml sample, which was taken from the expression culture in each hour interval after induction using IPTG, on SDS-PAGE. The over-expressed protein was supposed to show as a band with increasing intensity over time compared to the background. However, such a pattern was not identifiable on the polyacrylamide gels (an example is shown in Figure 6.20 in the Addendum).

In order to increase the visibility of the expressed proteins, Western blotting was conducted using the anti-6x Histidine primary antibodies. For unclear reasons, the Western blotting failed to visualize the bands of interest after autoradiography. A test experiment was done to investigate this problem. Cell lysate and primary antibody were applied separately as two spots on the same piece of nitrocellulose membrane. The Western blot and detection were conducted on the two spots in parallel and the results showed that the spot of primary antibody expressed a very strong signal while the spot of cell lysate gave no signal (results

not shown). It was thus unclear why expression was undetectable, but it is assumed that expression levels were too low to detect in a crude sample.

To address this problem, the strategy was changed and we decided to purify His-tagged fusion protein in order to concentrate it for SDS-PAGE, so that there were sufficient amounts of expressed fusion proteins to be seen on the gel and the background could be largely reduced. Interestingly, the His-tagged esxG, and His-tagged esxH could not be stained visibly by Coomassie Blue R 250 (results not shown). However, they could be visualized by silver staining. The expression levels of the mycosins (with or without transmembrane region), PPE4, and combined PE5 and PPE4 were so low that they could not even be visualized by silver staining. However, using previously-generated anti-mycosin-3 antibodies the presence of the expressed mycosins was confirmed by Western blotting. Expressed and purified fusion proteins, PE5 and combined esxG and esxH, could be stained by both staining methods.

Test expressions were done under 9 different expression conditions (as indicated in Chapter 2, Materials and Methods), and the optimal condition (0.5 mM IPTG, 30°C) results are shown in Figure 3.2. Only the silver stained polyacrylamide gels are shown. In these gels, the proteins PE5, esxG, esxH and the combined esxG and esxH can be clearly seen as thick bands. However, the bands for mycosin-3 without transmembrane region, mycosin-3 with transmembrane region, PPE4, and the combined PE5 and PPE4 cannot be identified because of the background and their low expression levels.

The His-tag protein purification kit was not able to separate the His-tagged protein from all the background proteins, resulting in multiple bands appearing on the gel. In order to identify the correct bands, Western blotting using the HisDetector™ Western Blot Kit was applied. The results are shown in Figure 3.3.

The Western blot kit also detected some background proteins. However, the proteins of interest could be identified by their sizes and relative abundance.

Anti-mycosin-3 antibodies were used for the detection of the purified His-tagged mycosin-3 fusion proteins when neither the Coomassie blue or silver staining methods nor the Colourimetric Western Blot kit were able to detect the proteins due to the extremely low expression levels. The anti-mycosin-3 antibody Western blotting results are shown in Figure

3.4. Both purified His-tagged mycosin-3 with transmembrane region and His-tagged mycosin-3 without transmembrane region were able to be detected by Western blotting using anti-mycosin-3 antibodies in the pellet, flow through (when column binding capacity is reached) and the eluant.

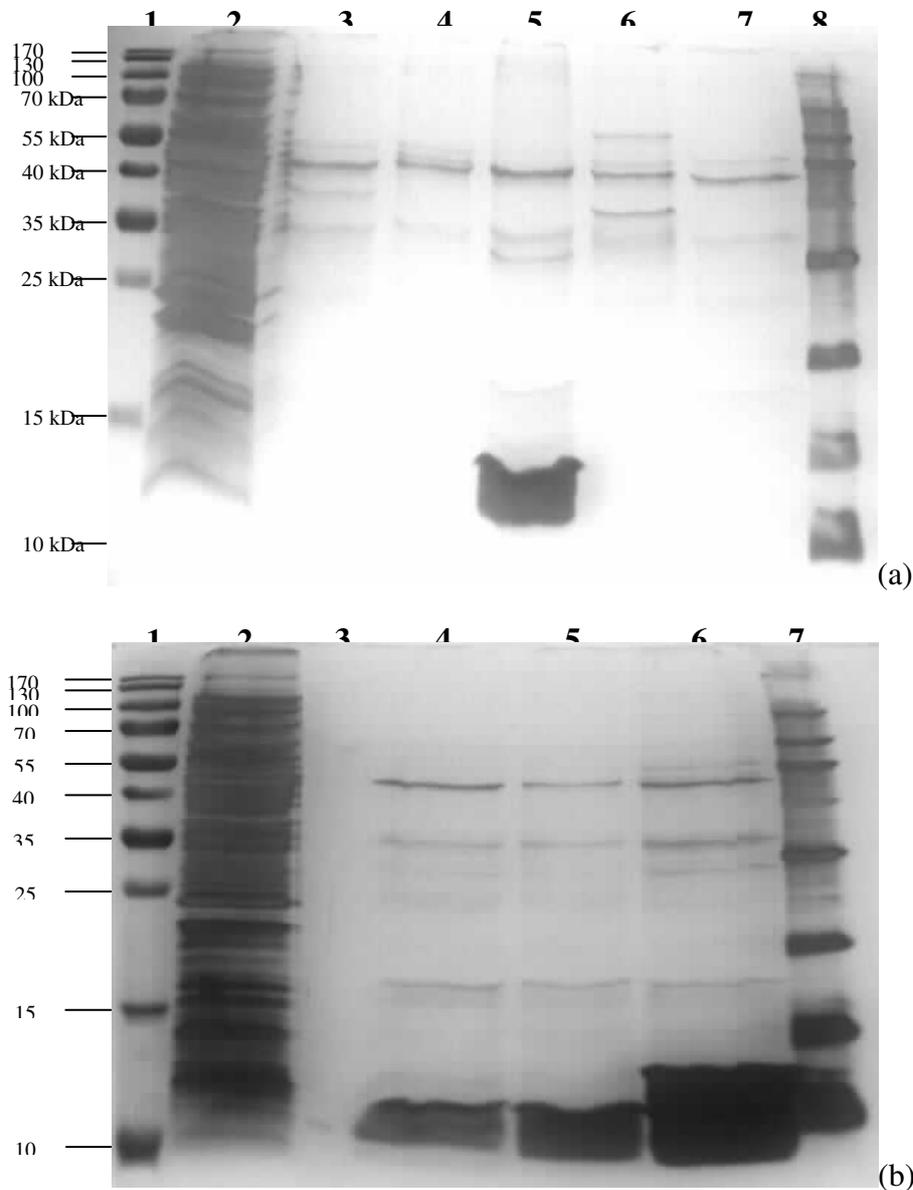


Figure 3.2 Silver-stained SDS-PAGE image of purified expressed recombinant *M. tuberculosis* proteins, (a) Lane 1: PageRuler[®] prestained protein ladder (Fermentas, USA); Lane 2: cell lysate; Lane 3: purified His-tagged mycosin-3 without the transmembrane region (~36.2 kDa); Lane 4: purified His-tagged mycosin-3 with the transmembrane region (~42.2 kDa); Lane 5: purified His-tagged PE5 (~10.6 kDa); Lane 6: purified His-tagged PPE4 (~52.8 kDa); Lane 7: purified His-tagged combined PE5 and PPE4; Lane 8: His-tagged protein marker (Invitrogen, USA); (b) Lane 1: PageRuler[®] prestained protein ladder (Fermentas, USA); Lane 2: cell lysate; Lane 3: empty lane; Lane 4: purified His-tagged esxG (~10.7 kDa); Lane 5: purified His-tagged esxH (~11.3 kDa); Lane 6: purified His-tagged esxG and esxH; Lane 7: His-tagged protein marker (Invitrogen, USA).

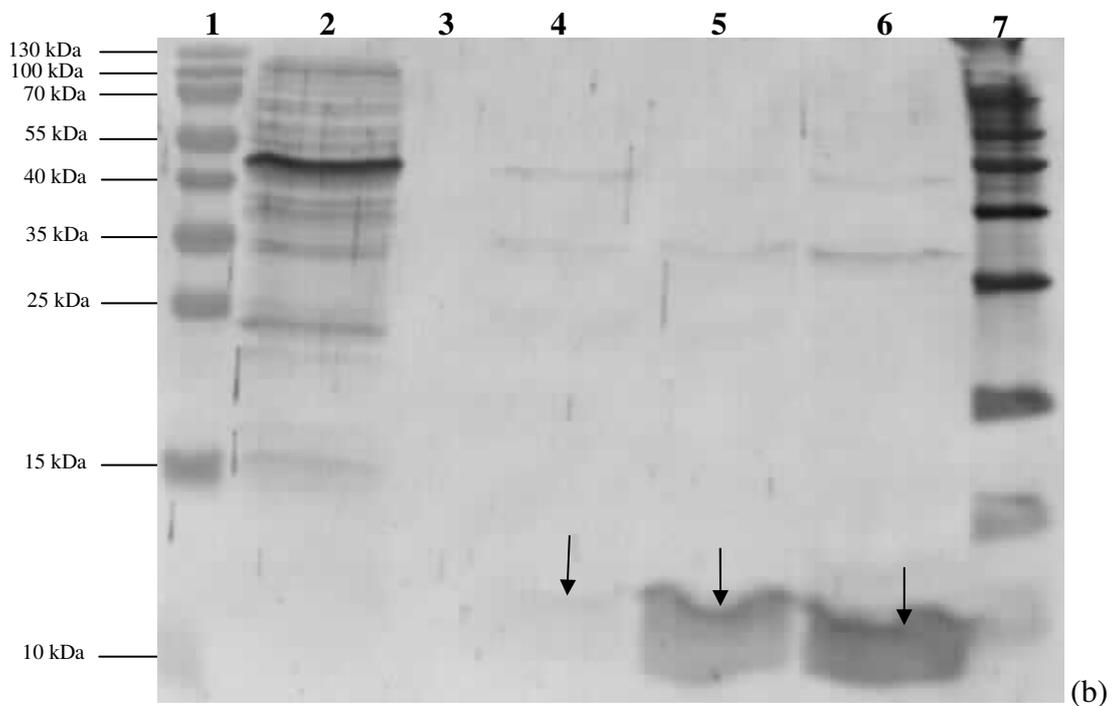
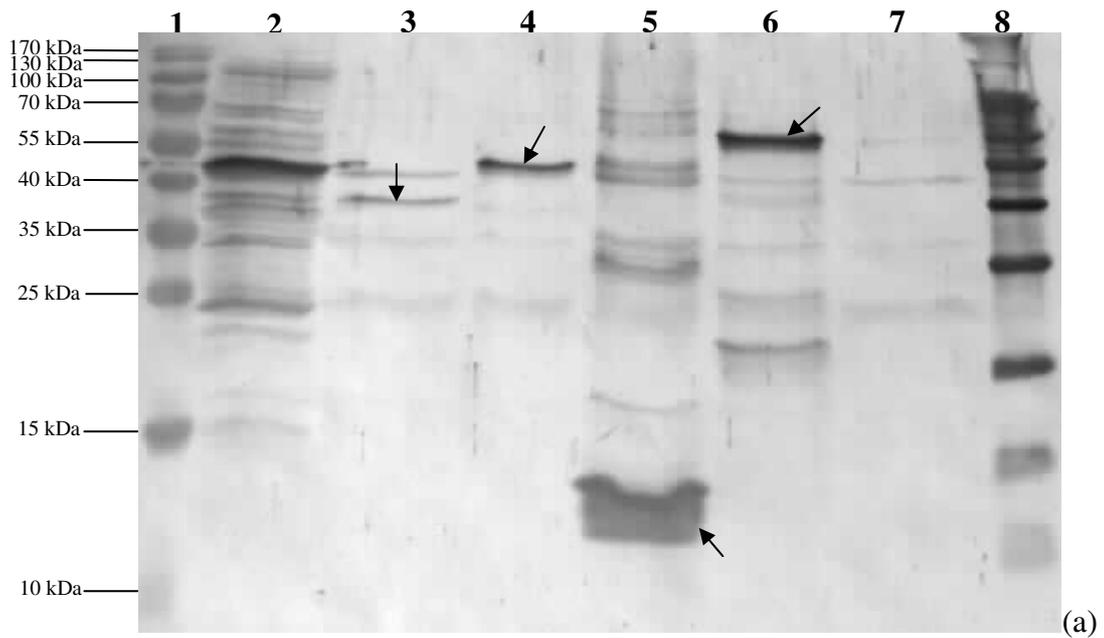


Figure 3.3 Western blots of SDS-PAGE gels in Figure 3.2. Bands indicated with arrows show the target His-tagged fusion proteins. (a) Lane 1: PageRuler[®] prestained protein ladder (Fermentas, USA); Lane 2: cell lysate; Lane 3: purified His-tagged mycosin-3 without the transmembrane region (~36.2 kDa); Lane 4: purified His-tagged mycosin-3 with the transmembrane region (~42.2 kDa); Lane 5: purified His-tagged PE5 (~10.6 kDa); Lane 6: purified His-tagged PPE4 (~52.8 kDa); Lane 7: purified His-tagged combined PE5 and PPE4; Lane 8: His-tagged protein marker; (b) Lane 1: PageRuler[®] prestained protein ladder (Fermentas, USA); Lane 2: cell lysate; Lane 3: empty lane; Lane 4: purified His-tagged esxG (~10.7 kDa); Lane 5: purified His-tagged esxH (~11.3 kDa); Lane 6: purified His-tagged esxG and esxH; Lane 7: His-tagged protein marker.

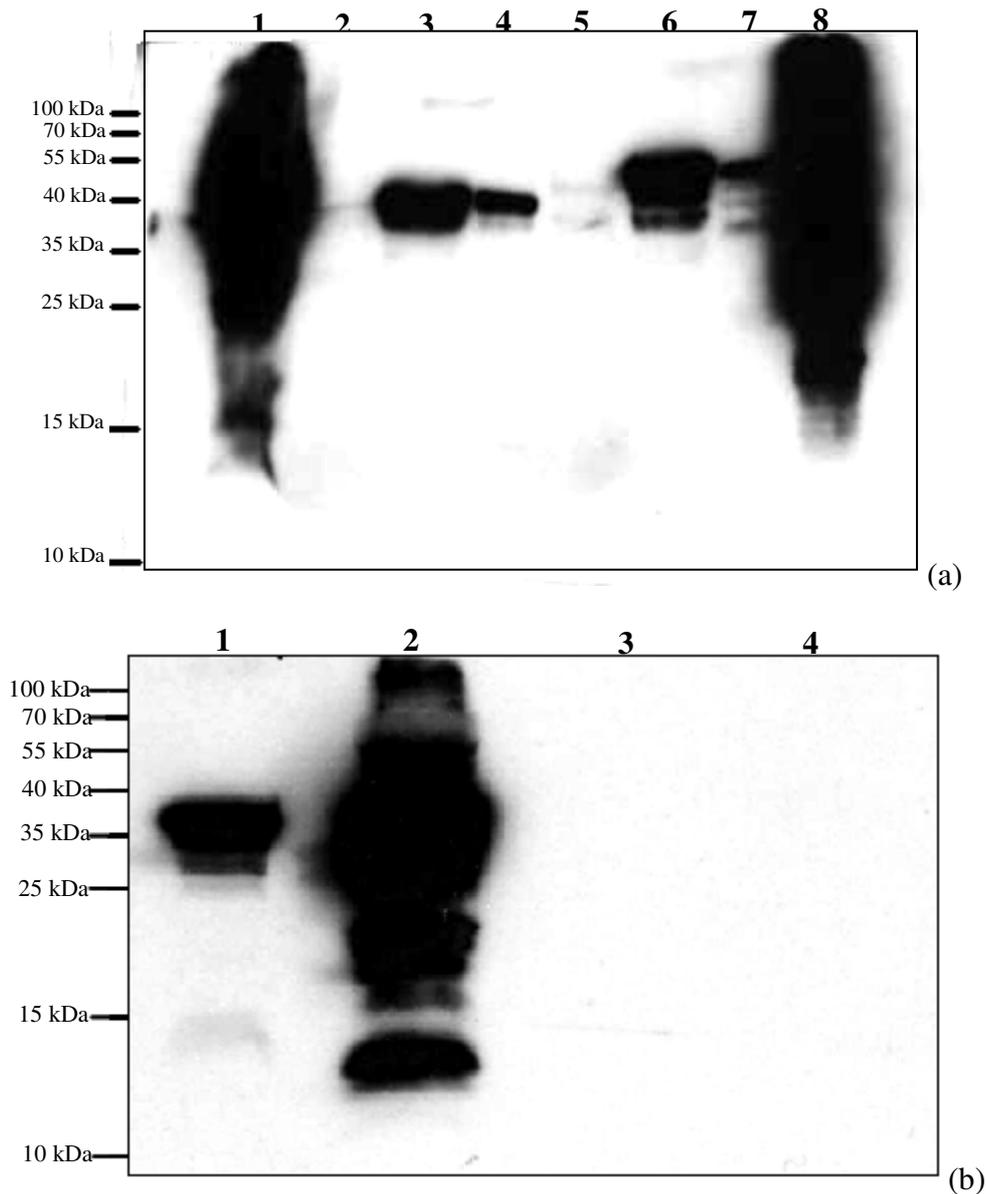


Figure 3.4. Western blots of Ni-NTA superflow column-purified His-tagged mycosin-3 with or without the transmembrane region fusion proteins using rabbit anti-mycosin-3 antibody. (a) Lane 1: Pellet from cell lysate of pET-28a-mycosin-3 recombinant plasmid DNA transformants; Lane 2: Flow through of cell lysate of transformants expressing mycosin-3 without transmembrane region; Lane 3: The eluant of mycosin-3 without the transmembrane region using MagneHis Purification System; Lane 4: The eluant of mycosin-3 without transmembrane region using Ni-NTA superflow column; Lane 5: Flow through of cell lysate of transformants expressing mycosin-3 with the transmembrane region; Lane 6: The eluant of mycosin-3 with the transmembrane region using MagneHis Purification System; Lane 7: The eluant of mycosin-3 with transmembrane region using Ni-NTA superflow column; Lane 8: GST-tagged mycosin-3 fusion protein as a positive control. (b) Lane 1: flow through of cell lysate of 2 liters of transformants expressing mycosin-3 without the transmembrane region; Lane 2: highly concentrated eluant (from 6 liters of *E. coli* culture) of mycosin-3 without the transmembrane region; Lane 3: flow through of cell lysate of transformants expressing the combined esxG and esxH; Lane 4: eluant of the His-tagged combined esxG and esxH fusion proteins.

The test expression of the p19Kpro mycobacterial expression vector was done by purifying His-tagged fusion protein from a small volume of mycobacterial culture and running it on SDS-PAGE with silver staining. For unknown reasons, none of the 7 different His-tagged fusion proteins could be isolated in high enough amounts to be visualized by silver staining (results not shown).

The test expression of the pDMN1 mycobacterial expression vector was done in the same manner as p19Kpro. Unlike the normal consistent growth of *M. smegmatis* for all 7 of the clones containing the p19Kpro recombinant expression vectors, the growth of *M. smegmatis* for 7 of the clones containing the pDMN1 recombinant expression vectors varied on both the plate and the broth. On 7H11 agar plates, *M. smegmatis* with pDMN1-esxG, pDMN1-esxH, pDMN1-esxG/esxH, pDMN1-PE5, and pDMN1-PE5/PPE4 showed equally good growth; but *M. smegmatis* with pDMN1-mycP3 grew much slower. Very interestingly, the survival of *M. smegmatis* with pDMN1-PPE4 was very low on solid media compared to the other transformants (Figure 3.5). The bacteria with pDMN1-PPE4 that did survive grew very slowly on both 7H11 agar and in 7H9 broth. In 7H9 broth, the growth of *M. smegmatis* with different pDMN1 transformants also varied in each culturing attempt. The *M. smegmatis* culture with the pDMN1-PPE4 construct also tended to form aggregates at the start of the culture. It took around 6 days for these cultures to grow to an OD_{600nm} of 0.6. The others had differential growth, growing well at one time, but very slowly at another. Sometimes, the culture even died (such as one culture of the pDMN1-esxH transformant). Apart from the differences in growth rate, no His-tagged fusion proteins could also be purified from most of these recombinants (results are shown in Addendum, Figure 6.29 and 6.30). However, the co-expression and co-purification of His-tagged esxG and esxH was successful (Figure 3.6).

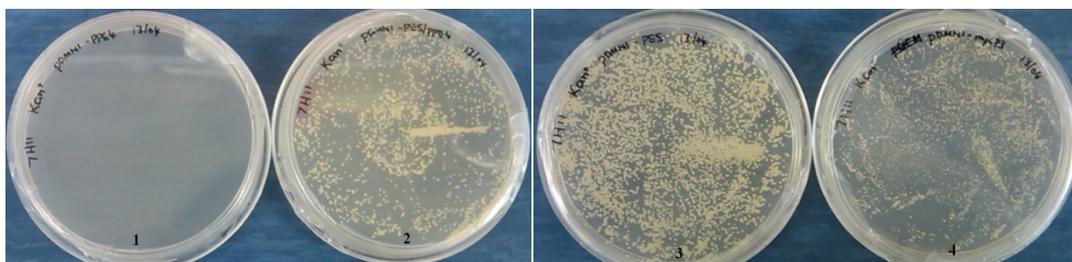


Figure 3.5 Comparison of the growth of *M. smegmatis* pDMN1-PPE4 (1), pDMN1-PE5/PPE4 (2), pDMN1-PE5 (3) and pDMN1-mycP3 (4) transformants on 7H11 agar plates after 72 hours of incubation at 37°C

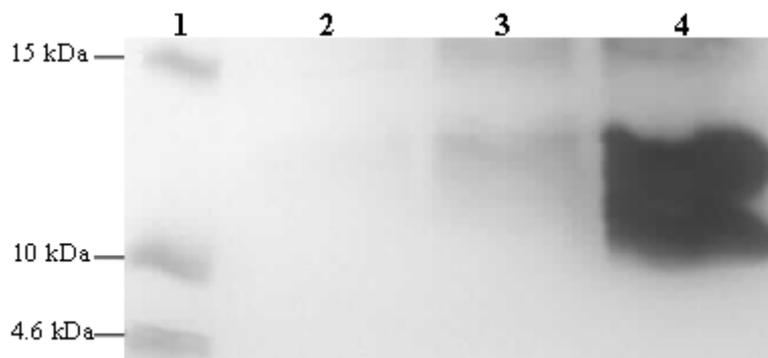


Figure 3.6 Image of silver-stained SDS-PAGE gel. Lane 1: PageRuler[®] prestained protein ladder (Fermentas, USA); Lane 2: His-tagged esxG isolated from *M. smegmatis* pDMN1 recombinant clone; Lane 3: His-tagged esxH isolated from pDMN1 recombinant clone; Lane 4: the combined esxG and esxH isolated from pDMN1 recombinant clone (esxH was His-tagged)

3.3 Protein quantification, Protease Assay, 2-D PAGE and Mass Spectrometric Analysis

Since mycobacterial expression vectors p19Kpro and pDMN1 were not able to express any of the target fusion proteins except for the combined esxG and esxH, only *E. coli* expressed fusion proteins were used for this part of the project. However, according to the test expression results, esxG and esxH individually could not be expressed in significant amount using the MagneHis[™] Protein Purification System, and they showed little yield using Ni-NTA Superflow columns; while the expression of combined PE5 and PPE4 was not detectable at all. Therefore, single esxG, single esxH, and the combined PE5 and PPE4 could not be used in this part of the study either.

The concentrations of the freshly prepared His-tagged mycosin-3 without transmembrane region, His-tagged mycosin-3 with transmembrane region, His-tagged PE5, His-tagged PPE4, and His-tagged combined esxG and esxH are given in the Addendum (Table 6.1). The five different fusion proteins were run individually on 2-D SDS-PAGE to act as the control group. All 5 gels in the control group were silver-stained. The gels loaded with mycosin-3 were used for Western blotting as well. These Western blots also belong to the control group. The control group is shown in Addendum 6.4. The experimental group contained the silver-stained 2-D SDS-PAGE gels of protein products from the protease assays (mycosin-3 with/without transmembrane region + PE5, mycosin-3 with/without transmembrane region + PPE4, mycosin-3 with/without transmembrane region + the combined esxG and esxH). The gels in the experimental group are shown in Figure 3.7, Figure 3.8 and Figure 3.9.

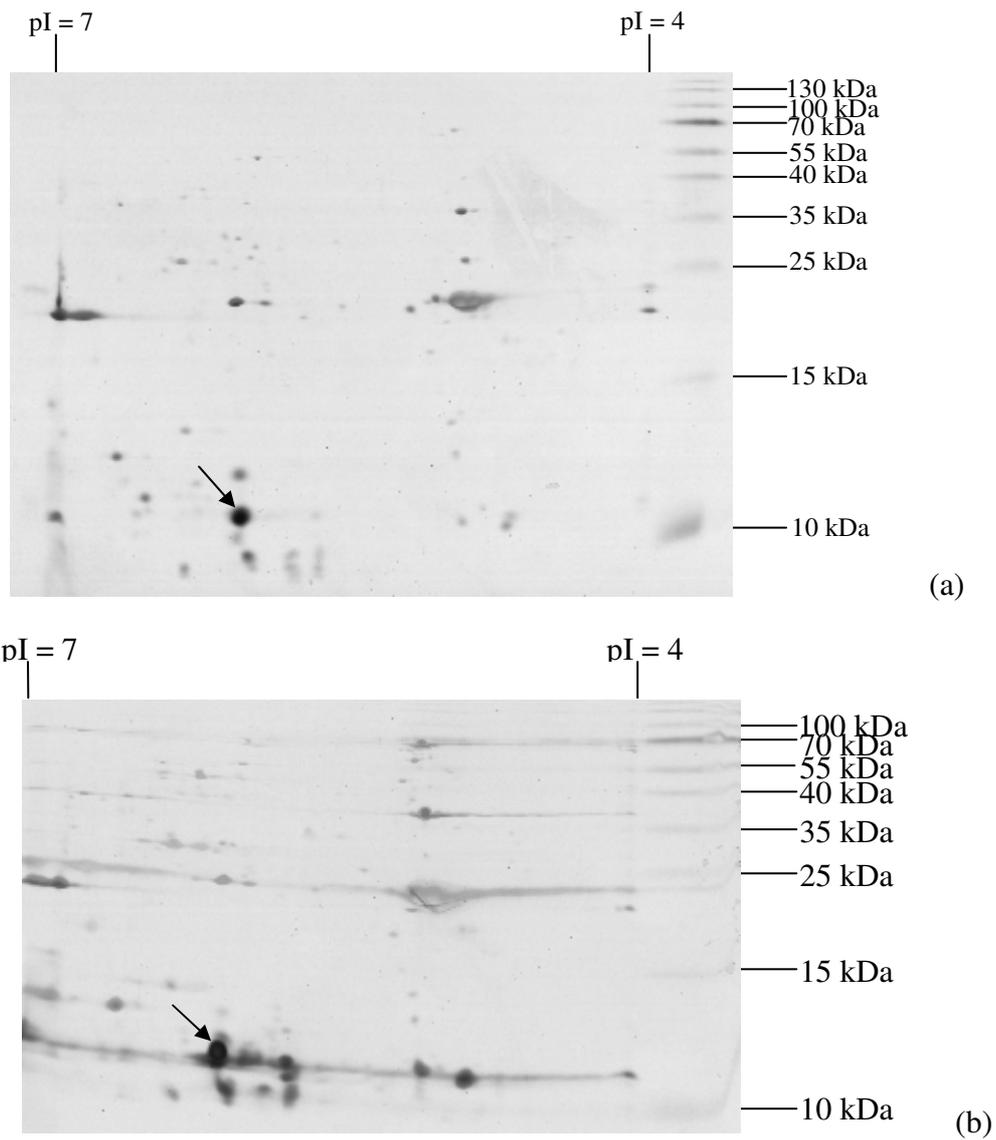
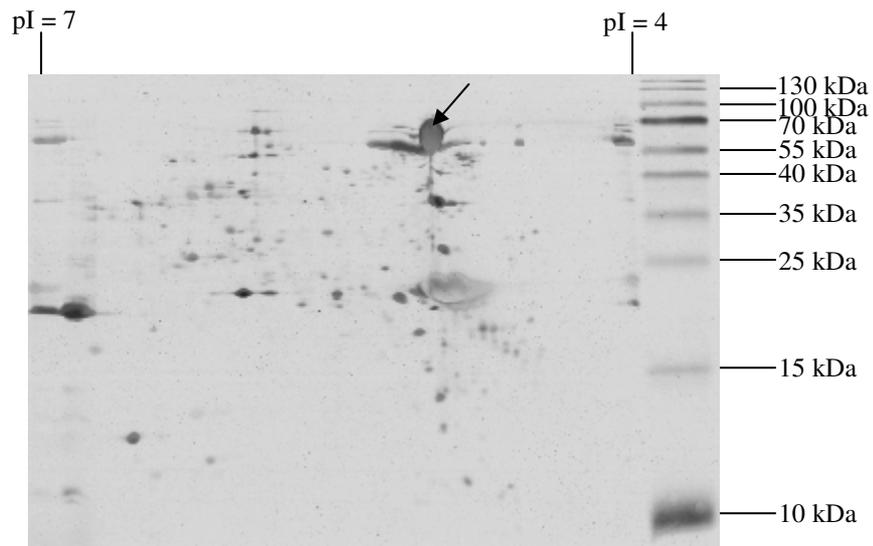
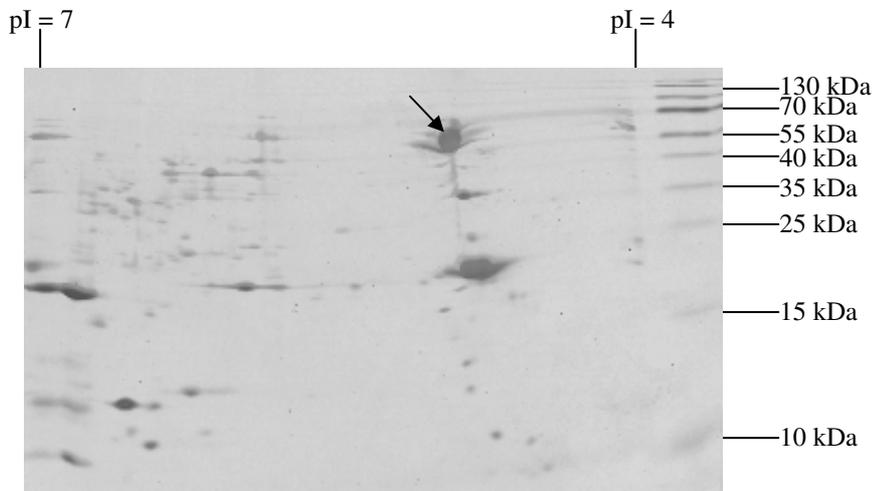


Figure 3.7 2-D SDS-PAGE results of the experimental group using His-tagged PE5 fusion protein as the substrate (the arrows indicate the spots of fusion target proteins). (a) Mycosin-3 without the transmembrane region + PE5; (b) Mycosin-3 with the transmembrane region + PE5.

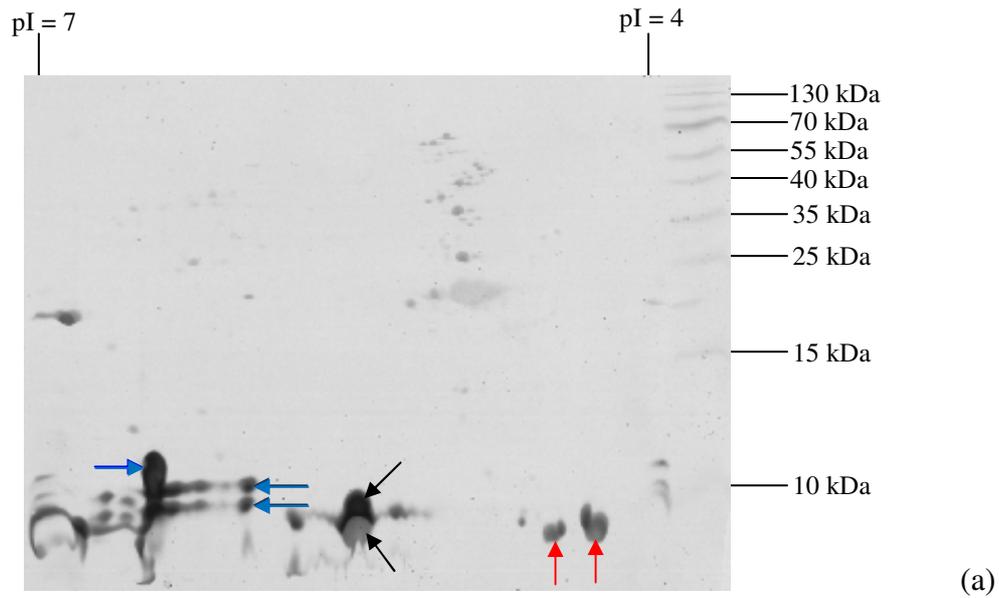


(a)

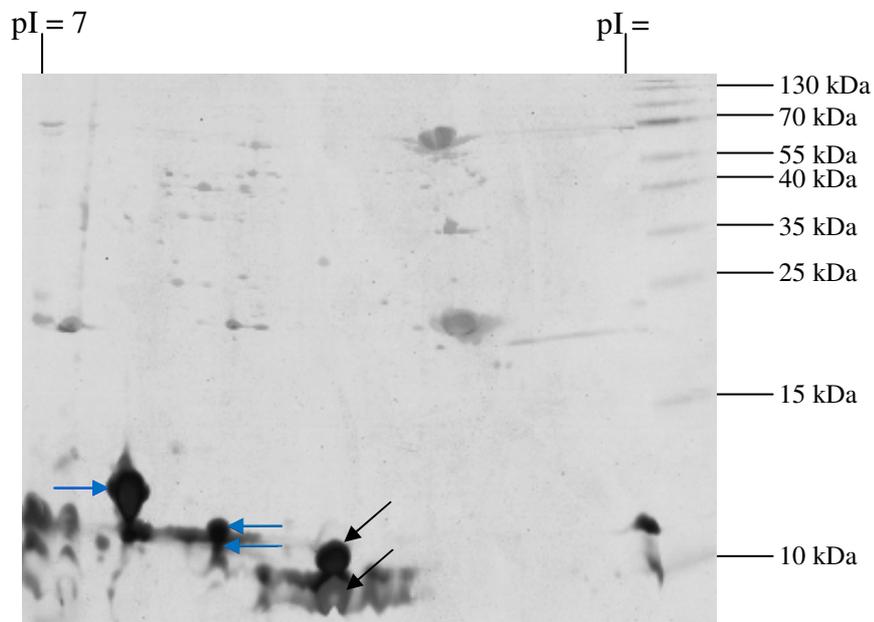


(b)

Figure 3.8 2-D SDS-PAGE results of the experimental group using His-tagged PPE4 fusion protein as the substrate (the arrows indicate the spots of fusion target proteins). (a) Mycosin-3 without the transmembrane region + PPE4; (b) Mycosin-3 with the transmembrane region + PPE4;



(a)



(b)

Figure 3.9 2-D SDS-PAGE results of the experimental group using His-tagged combined esxG and esxH fusion protein as the substrate (the arrows indicate the spots of fusion target proteins). (a) Mycosin-3 without the transmembrane region + combined esxG and esxH; (b) Mycosin-3 with the transmembrane region + combined esxG and esxH. Blue arrows indicate common spots among the two gels while red arrows indicate the two spots only appearing in one 2-D experiment but not the other.

Because of the large amount of background on the gel, spots representing the target fusion proteins were not easily identifiable. By comparing the spot patterns between the control gels and the experimental gels, as well as considering the size of the fusion proteins, the spots for the target fusion proteins were identified (Figure 3.7 – 3.9). However, the positions of those spots on the experimental gels did not deviate from the corresponding spots on the control gels (i.e. the positions of the spots did not deviate after the treatment of the putative purified substrates PE5, PPE4, combined esxG and esxH with mycosin-3).

Due to the absence of any identifiable variation between the protease assay gels and the control gels, we were unable to perform gel spot protein sequencing using Mass Spectrometry.

Chapter 4 Discussion

The genes encoding mycosin-3 with/without transmembrane region, PE5, PPE4, combined PE5 and PPE4, esxG, esxH, and combined esxG and esxH were cloned from the *Mycobacterium tuberculosis* H37Rv reference strain into both *E. coli* and mycobacterial expression vectors. The cloning was successful but the expression and purification were troublesome. The His-tagged target fusion proteins were not detectable in a 1 ml aliquot of the *E. coli* culture collected at each hour interval after the initiation of the expression (with addition of IPTG), using either Coomassie blue stain or silver stain on SDS-PAGE gels (Figure 6.20). Western blotting was supposed to amplify the 6x His tag signal by at least a thousand times. However, the Western detection using anti-His-tag (C-terminal) primary antibody failed. A test experiment was conducted as described in Section 3.2. The positive result from the spot of primary antibody implies that the binding between the primary antibody and the secondary HRP-conjugated anti-mouse antibody was successful and that the detection kit worked as well. The spot of test expression culture lysate showed a negative result. Apparently in this case, the binding between the target His-tagged fusion protein and the commercial anti-His-tag (C-terminal) primary antibody did not occur. It might be due to the following reasons: 1. the C-terminal 6x His-tag was prevented from binding to the anti-His antibody even though the protein sample was pre-treated with SDS-PAGE sample buffer before running SDS-PAGE and Western blotting (implying that the folding of the target protein by *E. coli* did not favour such a detection method); 2. the expression level in the 1 ml aliquot of the expression culture was so low that even Western blotting was not able to detect the target protein; or 3. the quality of the purchased primary anti-His antibody was not good.

In order to overcome the difficulties in the detection of the target His-tagged fusion protein expression, the expression and detection protocols were altered bearing in mind the three possible problems mentioned above. Purification of the His-tagged fusion proteins under denaturing conditions might have sorted out the problem regarding the hindrance of the binding site and thereby increase the yield of soluble proteins. However, this method does not favour the downstream experiments where the target proteins were required to be in their native form and active so that functional studies could be performed. Re-folding of denatured proteins is frequently unsuccessful and troublesome. Each protein may have its own unique folding mechanism. Therefore it was considered not worthwhile to try to generate denatured target proteins and refold them. We decided that an attempt to get more concentrated target

fusion protein for the detection was a better option. The alternative strategy was thus to perform small-scale His-tagged fusion protein purification before the detection using the staining methods and Western blotting.

This alternative strategy seemed to work well (Figure 3.2 and 3.3; Figure 6.21 – 6.28). However, the expression levels remained very low, even though the His-tagged fusion protein was purified and the background was largely reduced (compare to Figure 6.20). The expression levels for His-tagged mycosin-3 with/without transmembrane region, His-tagged PPE4 and the combined PE5 and PPE4 were exceedingly low (not identifiable with either Coomassie blue stain or Silver stain). The small secreted proteins, His-tagged PE5, His-tagged esxG, His-tagged esxH, and His-tagged combined esxG and esxH, could be detected by silver stain, and also by Coomassie blue stain if a larger culture was used for protein extraction or the purified protein was concentrated. It implies that *E. coli* is not capable of expressing large *M. tuberculosis* proteins such as mycosin-3, PPE4 and combined PE5 and PPE4 in significant amounts (Daugelat *et al.*, 2003). The possible reasons for this will be discussed later in this chapter. However, small secreted proteins (~10 kDa) PE5, esxG, esxH and even the combined esxG and esxH (20 kDa) were much easier for *E. coli* to express. Our study confirmed that silver stain is a more sensitive stain than Coomassie blue stain. Coomassie brilliant blue G-250 dye is able to bind to Phe and Tyr due to hydrophobic interaction, and Lys, His and Arg in a lesser extent due to electrostatic interaction (Georgiou *et al.*, 2008). Upon protein binding, the protein-dye complex shifts its absorption peak to 615 nm to 625 nm compared to the original absorption peak of the dye (450 nm). In the other hand, silver nitrate molecules bind and form complexes with the proteins by interacting with sulfhydryl (Met and Cys) and carboxyl groups (Asp, Glu and C-terminus of the protein) (Rabilloud, 1990). The silver ions are reduced to silver metal in slight alkaline conditions (formaldehyde in sodium or potassium carbonate solution). The reduction rate will increase when there is some silver metal available giving a better stained image. Silver stain increases the sensitivity from μg to ng compared to Coomassie blue stain (Rabilloud, 1990). It explains the reason why esxG and esxH individually could be stained by silver nitrate but not by Coomassie blue when a small culture was used (~10 ml) (results not shown).

For the small-scale His-tagged fusion protein purification, the MagneHisTM Protein Purification System was used. This kit uses a batch procedure for the purification where the affinity matrix resin is incubated in the cell lysate solution before the wash and the elution.

For some reason, this batch purification method was able to generate a better purification profile than the column purification method (Ni-NTA Superflow columns) (See Figure 3.4 for the comparison). Moreover, it was found that individually expressed esxG and esxH could not be purified by the column method (only when they were co-expressed) but the batch method could purify all three (results not shown). This might be due to the binding mechanism of the fusion proteins and the immobilized nickel ions on the matrix resins. It suggests that the free beads onto which the nickel ions were immobilized have a better chance for the His-tagged fusion proteins to access than the stationary matrix in the column. In other words, in this case, the bead purification method was more efficient than the column purification method. It was also found that the expression of esxG and esxH were dependent on each other. Although they could be expressed alone, this was at a very low level. However, the expression levels were much higher when they were co-expressed. It is interesting to note that the genes encoding esxG and esxH belong to a single operon and that they are co-operonically regulated. The proteins also interact with each other after being produced (Figure 6.36). The same phenomenon was observed with the expression of PE25 protein (Rv2431c) and PPE41 protein (Rv2430c) where the genes encoding for these two proteins are also organized in an operon and they also form a complex (Tundup *et al.*, 2006). Since the MagneHisTM protein purification system is only suitable for small-scale protein purification, it is very labour-intensive if it is used for large-scale His-tagged fusion protein purification. Therefore Ni-NTA Superflow columns were used for the large-scale fusion protein purification in spite of the lower purification efficiency compared to the MagneHisTM protein purification system.

Apart from the problem of low expression discussed above, there were two other major problems associated with the expression and purification of the His-tagged fusion proteins. These were associated with the rare codons in the *M. tuberculosis* genome and the poor purity of the purified His-tagged fusion proteins. These two problems resulted in unavoidable large amounts of background proteins, insoluble proteins forming inclusion bodies, and truncated proteins after purification. They made the identification of the target fusion proteins on 2-D gels very difficult. The target fusion proteins were also potentially vulnerable due to the presence of *E. coli* proteases. No protease inhibitors could be used in the purification as it could potentially inhibit the activity of mycosin-3.

M. tuberculosis, like all other species in the genus *Mycobacterium*, has a high GC content (65.6%), unique codon preferences and proteins which tend to be rich in glycine, alanine, proline and arginine. All these factors can cause problems for overexpression of mycobacterial proteins in *E. coli*. Numerous studies have shown that *E. coli* struggles to express mycobacterial proteins (e. g. Daugelat *et al.*, 2003). In this project, *E. coli* was able to express the small secreted proteins (soluble in their native form), PE5, esxG and esxH. It struggled to express larger proteins, especially the ones with hydrophobic regions such as PPE4 and mycosin-3 with/without the transmembrane region. When the target gene was too long, such as the gene for the combined PE5 and PPE4, the expression was almost abolished (no detection whatsoever). Mycobacteria have specific codon preferences which are different from *E. coli*. When mycobacterial genes are being expressed in *E. coli*, there are low concentrations of respective tRNAs in the cell (Lkemura, 1981). Depletion of those specific tRNAs during protein synthesis results in a slow translation rate (Varenne *et al.*, 1984). The translation might stop if there is no tRNA with the rare codon coming into the protein synthesis machinery. This leads to the formation of incomplete target fusion proteins. The full length target fusion proteins might also get cleaved at their N- or C-termini by proteases in the cell. These factors result in the production of truncated products of the original proteins. The truncated product is described as a protein that does not achieve its full length or obtain its proper conformation, and thus lack some of the amino acid residues that are present in a normal protein. A truncated protein generally cannot perform its normal function because its structure is incapable of doing so. Evidence for the production of truncated products of mycosin-3 is shown in Figure 6.32 and Figure 6.33 where a number of spots representing not only the mycosin-3 but also truncated version of it were picked up by the anti-mycosin-3 antibodies. However, some of these spots might also be degradation products of mycosin-3.

There are several ways of avoiding truncated products and reducing the production of inclusion bodies while doing expressing foreign genes in *E.coli*. One way is to use *E. coli* strains supplemented with genes encoding for tRNA molecules which are capable of recognizing rare codons. These *E. coli* strains are usually commercially available. However, these strains are not made solely for *M. tuberculosis* gene expression. A better option is to synthesize a gene which encodes for the same protein but with all the rare codons used in *M. tuberculosis* changed to the ones which are preferred in *E. coli*. This codon-optimized gene can then be ligated into an expression vector and expressed. However, this is a very costly method, especially if large numbers of proteins are needed to be expressed or if protein sizes

are large. The third method is to clone and express the *M. tuberculosis* gene in *M. tuberculosis* itself. Owing to the extremely slow growing behaviour of *M. tuberculosis*, *M. smegmatis* is usually used as the expression host for this purpose as it is a fast-growing mycobacterium. A number of mycobacterial expression vectors have been generated, which is able to generate correctly folded and post-translationally modified proteins using *M. smegmatis* as the host (Daugelat *et al.*, 2003). In this project, the third method was applied. The results will be discussed later in this chapter.

The second major problem was the poor purity of the purified target His-tagged fusion proteins. It is well known that His-tagged fusion protein purification usually generate impure co-eluates. Nonspecific binding of untagged proteins is usually the problem. Although histidine occurs relatively infrequently, some cellular proteins contain two or more adjacent histidine residues. These proteins have an affinity for the IMAC (immobilized metal-affinity chromatography) matrix [such as nickel-nitriloacetic acid (Ni^{+2} -NTA)-immobilized resins used in this project] and may co-elute with the protein of interest, resulting in significant contamination of the final product. In *E. coli*, proteins observed to co-purify with His-tagged target proteins can be divided into four groups: (i) proteins with natural metal-binding motifs, (ii) proteins with histidine clusters on their surfaces, (iii) proteins that bind to heterologously expressed His-tagged proteins, for example by a chaperone mechanism, and (iv) proteins with affinity to agarose-based supports (Bolanos-Garcia and Davies, 2006). Many of these impurities are stress-responsive proteins, suggesting that the cultivation conditions and the bacterial strain have an influence on their abundance and consequently their appearance as a contaminating species in the target protein preparation; it is therefore recommended to induce as little stress as possible during cultivation of *E. coli* cells (e.g. adding 0.1% glucose in the culture medium or using shake flasks without baffles) (Block *et al.*, 2009).

Disulfide bond formation between the protein of interest and other proteins can also lead to contamination. Ten mM 2-mercaptoethanol can be used in the loading, wash, and elution buffers for eliminating this potential problem (Bornhorst and Falke, 2000). Nonspecific hydrophobic interactions can also cause some copurification with the desired protein. Including low levels (up to 1%) of the non-ionic detergent Tween 20 or Triton X-100 in the protein buffers can reduce these interactions without substantially affecting the binding of the tagged protein to the Ni^{2+} -NTA matrices. The addition of salt (up to 500 mM NaCl), glycerol (up to 20%), or low levels of ethanol (up to 20%) can also reduce non-specific hydrophobic

protein interactions with these matrices (Bornhorst and Falke, 2000). The washing and elution buffer conditions need to be optimized experimentally for individual proteins.

There are several other options to improve the purification of His-tagged fusion proteins, besides the optimization methods mentioned above. These include additional purification steps using ion-exchange chromatography or size exclusion chromatography; the use of an engineered host strain that does not express certain proteins (however, this is not ideal since some of these proteins which may be co-purified with the target protein may be vital for the well-being of the host strain and the knock-out mutant could result in slow growth and low expression); and tag cleavage followed by reverse chromatography where the un-tagged target protein will be in the flow-through and the tag together with other co-purified contaminants are trapped in the column (Block *et al.*, 2009).

We attempted to express the target His-tagged fusion proteins in a mycobacterial host in order to produce these proteins using their native transcriptional and translational machinery. However, the expression vector p19Kpro could not express these proteins in significant amounts, probably because the promoter on this expression vector is not strong enough. Although this might have been solved by culturing a large volume of *M. smegmatis* for a large scale target fusion protein expression, we decided to rather use another mycobacterial expression vector (pDMN1) which has a very strong promoter P_{smyc} isolated from *M. smegmatis*. The results showed that even though this vector has a very strong promoter, only the combined esxG and esxH could be expressed in significant amounts using this vector. The other constructs appeared to be not functional, even though all constructs were verified to be correct and in frame using plasmid DNA sequencing.

We hypothesize that overexpression of the proteins might be toxic since some of them are secreted T-cell antigens (e.g. esxG and esxH) which might exert cell lysis effects on host cells. The host *M. smegmatis* was possibly stressed and thus slowed down its growth or even shut down the target protein expression. We observed that, in an extreme case, the culture even died out (Figure 3.5, transformants of recombinant pDMN1-mycP3). The target protein could not be extracted even though the culture was initially able to grow. It is possible that the promoter sequence on the expression vector was mutated by the host so that the host could not be harmed by the incoming overexpressed toxic proteins (personal communication with M. Williams, unpublished results). Similar to what we observed in the expression of

esxG and esxH in *E. coli*, individually expressed esxG and esxH in *M. smegmatis* seemed to also be so toxic to the host, that the host shut down the expression; However when they were co-expressed, the toxicity seemed to be reduced and the host allowed them to be produced together in large amounts.

The toxic effect of the overexpressed proteins to the host cells during expression could be reduced by making use of inducible expression. Since the mycobacterial expression vector pDMN1 is not an inducible expression vector, the toxic target fusion proteins were probably produced and began to harm the host from the moment when the transformant was cultured. An inducible expression vector is able to overcome such a problem, since the induction only begins after the culture is in its mid-log phase where the *M. smegmatis* host cells are relatively strong. This increases the production of target fusion proteins. There are several inducible expression systems that have been described so far (Sasseti *et al.*, 2009). Future studies will be done with parallel expression in inducible expression vectors to ensure that toxic effects are identified and minimized.

In spite of the low expression and poor purity of the mycosin-3 with/without transmembrane region, the His-tagged mycosin-3 was concentrated and used for protease assays immediately after the purification. It was assumed that no matter how low the quantity of mycosin-3 was, it should be able to perform its function when in solution with its potential substrate, provided that the enzyme was produced and folded correctly and was at its correct buffer conditions. Enzymes are made to catalyze biochemical reactions, so a small amount of mycosin-3 should still be able to cleave a significant amount of the substrate molecules if functional and under the correct conditions.

After allowing for cleavage under the proposed optimal conditions, the samples were loaded onto 2-D PAGE gels. The results showed that the spots of interest (PE5, PPE4, combined esxG and esxH) in the experimental group did not deviate from the corresponding spots in the control group, irrespective of whether mycosin-3 with or without transmembrane region was used (Figure 3.7 – 3.9). There are a few possible reasons why the results showed no significant findings: (1) the buffer conditions that were used might not have favoured the activity of mycosin-3 with/without transmembrane region; (2) mycosin-3 and the potential substrates were not folded in its correct conformation or they were modified incorrectly and were thus no longer functional (*E. coli* is not an ideal expressing host for mycobacterial

proteins); (3) PE5, PPE4, the combined esxG and esxH are not the substrates of mycosin-3; (4) the 6x histidine tag on the target fusion proteins might interfere with the binding affinity between the mycosin-3 and the substrates; (5) mycosin-3 is not a real protease; (6) the contaminating proteins may exert an inhibitory effect on mycosin-3 (e.g. binding to mycosin-3) so substrates could not access the binding pocket due to steric hindrance; or (7) the cleavage products were too small to be detected.

Looking at all the 2-D gel results, a very large number of spots representing contaminating proteins could be observed, as well as breakdown products of large proteins and truncated products of the target proteins (section 6.4.2). In all the different gels, most spots appeared at the same location (i.e. same pI and molecular weight) and they are thus assumed to be common contaminants. By comparing the spots among different gels, the unique spots were assumed to be the target His-tagged fusion proteins, as well as their truncated products and modified products. In the gel where purified His-tagged PE5 fusion protein was loaded (Figure 6.34), there were a number of spots surrounding the predicted PE5 protein spot showing similar pI and molecular weight. These could not be found on other gels and they might be the truncated and modified products of PE5. ESAT-6 has been shown to have different protein species produced by *M. tuberculosis* including a form with an acetylated N-terminus or a truncated C-terminus (Jungblut *et al.*, 2008). PE5 might have similar different protein species as it is also a small secreted protein of the ESX gene cluster. Mycosin-3 with/without transmembrane region is a much larger protein than PE5. Because of its low expression, it could not be identified as a clear spot on the 2-D gel stained with silver nitrate. Western blotting revealed a large amount of breakdown products (Figure 6.32b and Figure 6.33b), as well as truncated and modified products with different pI from the original mycosin-3. Mycosin-3 without transmembrane region (Figure 6.32b) had more breakdown products than mycosin-3 with transmembrane region. The reason for this might be because the transmembrane region and the extracellular linker region have a protective effect on the functional domain of the protein.

The gels on which purified combined esxG and esxH were loaded showed some spots (indicated with arrows in Figure 3.9) which could not be found in all other 2-D gels. EsxG and esxH were co-expressed and co-purified, and it is known from the literature that they tend to form a 1:1 complex. This complex was retained on the IEF gel because the conditions on the IEF strip is not denaturing. The complex migrated on the IEF gel according to the

combined pI of esxG and esxH. Theoretically, the esxG and esxH should have only dissociated from each other on the 2nd dimension SDS-PAGE. Therefore the black arrows indicating the two connecting spots are very likely to be the complex of esxG and esxH (Figure 3.9). The spots indicated with blue arrows probably represent the modified esxG or esxH since the molecular mass of this spot is very similar to esxG and esxH. It is interesting to note that different preparations of the purified combined esxG and esxH gave different 2-D gel profiles as indicated by the two spots with red arrows (Figure 3.9). These two spots are present in one 2-D gel [Figure 3.9 (a)] but absent in the other 2-D gel [Figure 3.9 (b)], indicating that they are not the target fusion proteins. They are likely the proteins whose production is cell culture stress dependent, i.e. stress-responsive proteins. Future studies will focus on confirming the identities of the spots on the 2-D PAGE gels by sending it for mass spectrometric sequencing.

Chapter 5 Conclusion

We have successfully generated clones of mycosin-3 with/without the transmembrane region, PE5, PPE4, the combined PE5 and PPE4, esxG, esxH and the combined esxG and esxH in pET-28a *E. coli* expression vector as well as p19Kpro and pDMN1 mycobacterial expression vectors. All 22 clones were verified by colony PCR and plasmid DNA sequencing. We managed to express and purify these 9 different His-tagged fusion target proteins in *E. coli* and the combined esxG and esxH in *M. smegmatis* using two different purification methods; and we successfully conducted protease assays, IEF, 2-D SDS-PAGE and Western blotting. However, no substrate cleavage was observed in this project. The *in vitro* identification of potential substrates for mycosin-3 has not been successful thus far. This could be due to the following possible limitations: (1) impure enzyme and substrate(s); (2) inappropriate buffer conditions; (3) hypothesized substrates might not be the substrates of mycosin-3; and (4) incorrect folding or modification of the target fusion proteins.

For future studies, we recommend that either the genes be cloned and expressed in mycobacteria using an appropriate inducible mycobacterial expression vector or the genes be codon-optimized for *E. coli* expression. Different affinity tags could be used instead of 6x Histidine since 6x His protein purification does not generate pure target fusion proteins. Alternatively, one or more chromatography steps can be performed after His-tag affinity chromatography. The substrates of mycosin-3 can be identified *in vivo* as well. Furthermore, the structure of mycosin-3, as well as the other four homologues will also be elucidated and analyzed; hence the enzyme binding pockets of mycosins will be modeled so that enzyme inhibitors can be designed for drug development.

Reference

Aaij, C. and Borst, P., 1972, The gel electrophoresis of DNA, *Biochimica et Biophysica Acta*, 269: 192

Abdallah, A. M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R. J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., Luirink, J. and Bitter, W., 2006, A specific secretion system mediate PPE41 transport in pathogenic mycobacteria, *Molecular Microbiology*, 62: pp667-679

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. and Bitter, W., 2007, Type VII secretion – mycobacteria show the way, *Nature Reviews, Microbiology*, 5: pp883-891

Abdallah, A. M., Verboom, T., Weerdenburg, E. M., Gey van Pittius, N. C., Mahasha, P. W., Jimenez, C., Parra, M., Cadieux, N., Brennan, M. J., Appelmelk, B. J. and Bitter, W., 2009, PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5, *Molecular Microbiology*, 73 (3): pp329-340

Adindla, S. and Guruprasad, L., 2003, Sequence analysis corresponding to the PPE and PE proteins in *Mycobacterium tuberculosis* and other genomes, *Journal of Bioscience*, 28: pp169-179

Agard, N. J. and Wells, J. A., 2009, Methods for the proteomic identification of protease substrates, *Current Opinion in Chemical Biology*, 13: pp503-509

Agarwal, N., Woolwine, S. C., Tyagi, S. and Bishai, W. R., 2007, Characterization of the *Mycobacterium tuberculosis* sigma factor SigM by assessment of virulence and identification of SigM-dependent genes, *Infection and Immunity*, 75: pp452-461

Ahmed, N. and Hasnain, S. E., 2004, Genomics of *Mycobacterium tuberculosis*: Old threats and new trends, *Indian Journal of Medical Research*, 120: pp207-212

Ancizu, S., Moreno, E., Solano, B., Villar, R., Burguete, A., Torres, E., Pérez-Silanes, S., Aldana, I., Monge, A., 2010, New 3-methylquinoxaline-2-carboxamide 1,4-di-N-oxide derivatives as anti-*Mycobacterium tuberculosis* agents, *Bioorganic and Medicinal Chemistry*, 18(7): pp2713-2719

Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H. W. H., Neefs, J-M., Winkler, H., Gestel, J. V., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N. and Jarlier, V., 2005, A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*, *Science*, 307 (5707): pp223-227

Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M. and Cole, S. T., 2002, Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Molecular Microbiology*, 44: pp9-19

Berthet, F. X., Rasmussen, P. B., Rsenkrands, I., Andersen, P. and Gicquel, B., 1998, A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10), *Microbiology*, 144: pp3195-3203

Betts, J. C., Lukey, P. T., Robb, L. C., Mcadam, R. A. and Duncan, K., 2002, Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology*, 43: pp717-731

Block, E., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. and Schafer, F., 2009, Immobilized-metal affinity chromatography (IMAC): A Review, *Methods in Enzymology*, 463: 439-473

Blower, S. M. and Chou, T., 2004, Modeling the emergence of the “hot zones”: tuberculosis and the amplification dynamics of drug resistance, *Nature Medicine*, 10: pp1111-1116

Bolanos-Garcia, V. M. and Davies, O. R., 2006, Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilized metal affinity chromatography, *Biochimica et Biophysica Acta*, 1760: 1304-1313

Bornhorst J. A. and Falke J. J., 2000, Purification of Proteins Using Polyhistidine Affinity Tags, *Methods in Enzymology*, 326: 245-254

Brennan, M. J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M. and Jacobs, W. R., Jr, 2001, Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells, *Infection and Immunology*, 69: pp7326-7333

Brodin, P., Rosendrans, I., Andersen, P., Cole, S. T., Brosch, R., 2004, ESAT-6 proteins: protective antigens and virulence factors? *Trends of Microbiology*, 12: pp500-508

Brown, G. D., Dave, J. A., Gey van Pittius, N. C., Stevens, L., Ehlers, M. R. W. and Beyers, A. D., 2000, The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases, *Gene*, 254: pp147-155

Callahan, B., Nguyen, K., Cllins, A., Valdes, K., Caplow, M., Crossman, D. K., Steyn, A. J. C., Eisele, L. and Derbyshire, K. M., 2010, Conservation of structure and protein-protein interactions mediated by the secreted mycobacterial proteins EsxA, EsxB, and EspA, *Journal of Bacteriology*, 192: pp326-335

Camus, J-C., Pryor, M. J., Medigue, C. and Cole, S. T., 2002, Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv, *Microbiology*, 148: pp2967-2973

Chakhaiyar, P., Nagalakshmi, Y., Aruna, B., Murthy, K. J., Katoch, V. M. and Hasnain, S. E., 2004, Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. *Journal of Infectious Disease*, 190: pp1237-1244

Champion, P. A., Stanley, S. A., Champion, M. M., Brown, E. J. and Cox, J. S., 2006, C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*, *Science*, 313: pp1632-1636

Champion, P. A. D., Champion, M. M., Manzanillo, P. and Cox, J. S., 2009, ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria, *Molecular Microbiology*, doi:10.1111/j.1365-2958.2009.06821.x

Chen, R., Gearhart, J., Protopopova, M., Einck, L. and Nacy, C. A., 2006, Synergistic interaction of SQ109, a new ethylene diamine, with front-line antitubercular drugs *in vitro*, *Journal of Antimicrobial Chemotherapy*, 58 (2): pp332-337

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.*, 1998, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature*, 393: pp537-544

Converse, S. E. and Cox, J. S., 2005, A Protein Secretion Pathway Critical for *Mycobacterium tuberculosis* Virulence is Conserved and Functional in *Mycobacterium smegmatis*, *Journal of Bacteriology*, 187(4): pp1238-1245

Coros, A., Callahan, B., Battaglioli, E. and Derbyshire, K. M., 2008, The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*, *Molecular Microbiology*, 69 (4): pp794-808

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

Dave, J. A., Gey van Pittius, N. C., Beyers, A. D., Ehlers, M. R. W. and Brown, G. D., 2002, Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages, *BioMed Central Microbiology*, 2 : pp30-38

De Jonge, M. E., Pehau-Arnaudet, G., Fretz, M. M., Romain, F., Bottai, D., Brodin, P., Honore, N., Marchal, G., Jiskoot, W., England, P., Cole, S. T. and Brosch, R., 2007, ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity, *Journal of Bacteriology*, 189: pp6028-6034

Delogu, G. and Brennan, M. J., 2001, Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*, *Infectious Immunology*, 69: pp5606-5611

Demangel, C. P., Brodin, P., Cockle, P. J., Brosch, R., Majlessi, L., Leclerc, C., *et al.*, 2004, Cell envelope protein PPE68 contributes to *Mycobacterium tuberculosis* RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6, *Infection and Immunity*, 72: pp2170-2176

Den Dooren de Jong, L. E., 1939, De tuberculose van bacteriologisch standpunt bezien, *Nederl Tijdschr Geneeskinde*, 83: 5896-5906

Derrick, S. C. and Morris, S. L., 2007, The ESAT-6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression, *Cellular Microbiology*, 9: pp1547-1555

Dix, M. M., Simon, G. M. and Cravatt, B. F., 2008, Global mapping of the topography and magnitude of proteolytic events in apoptosis, *Cell*, 134: 679-691

Espitia, C., Lacleste, J. P., Mondragon-Palomino, M. and 7 other authors, 1999, The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology*, 145: 3487-3495

Fortune, S. M., Jaeger, A., Sarracino, D. A., Chase, M. R., Sasseti, C. M., Sherman, D. R., Bloom, B. R. and Rubin, E. J., 2005, Mutually dependent secretion of proteins required for mycobacterial virulence, *Proceedings of the National Academy of Sciences, USA*, 102: pp10676-10681

Gao, L. Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J. N. and Brown, E. J., 2004, A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion, *Molecular Microbiology*, 53: pp1677-1693

Georgiou, C. D., Grintzalis, K., Zervoudakis, G. and Papapostolou, I., 2008, Mechanism of Coomassie brilliant blue G-250 binding to proteins: a hydrophobic assay for nanogram quantities of proteins, *Analytical Bioanalytical Chemistry*, 391: 391-403

Gey van Pittius, N. C., Gamielien, J., Hide, W., Brown, G. D., Siezen, R. J. and Beyers, A. D., 2001, The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria, *genome biology*, 2 (10): research0044.1-0044.18

Gey van Pittius, N. C., Sampson, S. L., Lee, H., Kim, Y., van Helden, P. D. and 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 Evolutionary Biology*, 6: pp95-126

Gordon, S. V., Eiglmeier, K., Brosch, R., Garnier, T., Honore, N., Barrell, B. and Cole, S. T., 1999, Genomics of *Mycobacterium tuberculosis* and *Mycobacterium leprae*, In *Mycobacteria: molecular biology and virulence* Edited by: Ratledge, C. and Dale, J., Oxford, Blackwell Science Ltd; pp93-109

Grange, J. M., 2009, The genus *Mycobacterium* and the *Mycobacterium tuberculosis* complex, *Tuberculosis: A Comprehensive Clinical Reference*, Saunders, W. B., pp44-59

Gron, H., Meldal, M., Breddam, K., 1992, Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching, *Biochemistry*, 31: pp6011-6018

Guinn, K. M., Hickey, M. J., Mathur, S. K., Zakel, K. L., Grotzke, J. E., Lewinsohn, D. M., Smith, S. and Sherman, D. R., 2004, Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium Tuberculosis*, *Molecular Microbiology*, 51(2): pp359-370

Ize, B. and Palmer, T., 2006, *Mycobacteria's* export strategy, *Science, Microbiology*, 313: pp1583-1584

Jungblut, P. R., Holzhutter, H. G., Apweiler, R. and Schluter, H., 2008, The speciation of the proteome, *Chemistry Central Journal*, 2: 16-26

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

Lewis, K. N., Liao, R., Guinn, K. M., Hickey, M. J., Smith, S., Behr, M. A., *et al.*, 2003, Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacilli Calmette-Guerin attenuation, *Journal of Infectious Disease*, 187: pp117-123

Li, Y., Miltner, E., Wu, M., Petrofsky, M. and Bermudez, L. E., 2005, *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice, *Cellular Microbiology*, 7: pp539-548

Lipkind, G., Gong, Q. and Steiner, D. F., 1995, Molecular modeling of the substrate specificity of prohormone convertases SPC2 and SPC3, *Journal of Biological Chemistry*, 270: pp13277-13284

Lkemura, T., 1981, Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system, *Journal of Molecular Biology*, 151(3): 389-409

MacGurn, J. A., Raghavan, S., Stanley, S. A. and Cox, J. S., 2005, A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*, *Molecular Microbiology*, 57: pp1653-1663

Maciag, A. *et al.*, 2007, Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon, *Journal of Bacteriology*, 189: pp730-740

Matsumoto, M., Hashizume, H., Tomishige, T., Kawasaki, M., Tsubouchi, H., Sasaki, H., Shimokawa, Y. and Komatsu, M., 2006, OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice, *PLoS Medicine*, 3 (11): e466

McLaughlin, B., Chon, J. S., MacGun, J. A., Carlsson, F., Cheng, T. L., Cox, J. S. and Brown, E. J., 2007, A mycobacterium ESX-1-secreted virulence factor with unique requirements for export, *PLoS Pathogenicity*, 3: e105

Meher, A. K., Bal, N. C., Chary, K. V. R. and Arora, A., 2006, *Mycobacterium tuberculosis* H37Rv ESAT-6-CFP-10 complex formation confers thermodynamic and biochemical stability, *FEBS Journal*, 273: pp 1445-1462

Nuermberger, E., Rosenthal, I., Tyagi, S., Williams, K. N., Almeida, D., Peloquin, C. A., Bishai, W. R. and Grosset, J. H., 2006, Combination chemotherapy with the Nitroimidazopyran PA-824 and first-line drugs in a murine model of Tuberculosis, *Antimicrobial Agents and Chemotherapy*, 50 (8): pp2621-2625

Ohol, Y. M., Goetz, D. H., Chan, K., Shiloh, M. U., Craik, C. S. and Cox, J. S., 2010, *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence, *Cell Host and Microbe*, 7: pp210-220

Okkels, L. M., Brock, I., Follmann, F., Agger, E. M., Arend, S. M., Ottenhoff, T. H. M., Oftung, F., Rosenkrands, I. and Andersen, P., 2003, PPE protein (Rv3873) from DNA segment RD1 of *Mycobacterium tuberculosis*: strong recognition of both specific T-cell epitopes and epitopes conserved within the PPE family, *Infectious Immunity*, 71: pp6116-6123

Okkel, L. M. and Andersen, P., 2004, Protein-Protein Interaction of Proteins from the ESAT-6 Family of *Mycobacterium tuberculosis*, *Journal of Bacteriology*, 186(8): pp2487-2491

Perona, J. J. and Craik, C. S., 1995, Structural basis of substrate specificity in the serine proteases, *Protein Science*, 4; pp337-360

Primm, T. P., Lucero, C. A. and Falkinham, J. O., 2004, Health impacts of environmental mycobacteria, *Clinical Microbiology Reviews*, 17: pp98-106

Protopopova, M., Hanrahan, C., Nikonenko, B., Samala, R., Chen, P., Gearhart, J., Einck, L. and Nacy, C. A., 2005, Identification of a new antitubercular drug candidate, SQ109, from a combinatorial library of 1,2-ethylenediamines, *Journal of Antimicrobial Chemotherapy*, 56 (5): pp968-974

Pym, A. S., Brodin, P., Brosch, R., Huerre, M. and Cole, S. T., 2002, Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*, *Molecular Microbiology*, 46 (3): 709-717

Pym, A. S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., *et al.*, 2003, Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis, *Natural Medicine*, 9: pp533-539

Rabilloud, T., 1990, Mechanisms of protein silver staining in polyacrylamide gels: A 10-year synthesis, *Electrophoresis*, 11(10): 785-794

Raghavan, S., Manzanillo, P., Chan, K., Dovey, C. and Cox, J. S., 2008, Secreted transcription factor controls *Mycobacterium tuberculosis* virulence, *Nature*, 454: pp717-721

Ramakrishnan, L., Feederspiel, N. A. and Falkow, S., 2000, Granuloma specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE_PGRS family, *Science*, 288: pp1436-1439

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. and Carr M. D., 2005, Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6, *The EMBO Journal*, 24(14): pp2491-2498

Renshaw, P. S., Panagiotidou, P., Whelan, A., Gordon, S. V., Hewinson, R. G., Williamson, R. A. and 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, *Journal of Biological Chemistry*, 277: pp21598-21603

Rindi, L., Lari, N. and Garzelli, C., 1999, Search for the genes potentially involved in *Mycobacterium tuberculosis* virulence by mRNA differential display, *Biochem. Biophys. Res. Commun.* 258: pp94-101

Rodriguez, G. M. and Smith, I., 2003, Mechanisms of iron regulation in mycobacteria: role in physiology and virulence, *Molecular Microbiology*, 47: pp1485-1494

Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. and Smith, I., 2002, *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response, *Infection and Immunity*, 70 (7): pp3371-3381

Rosenkrands, I., King, A., Weldingh, K., Moniatte, M., Moertz, E. and Andersen, P., 2000, Towards the proteome of *Mycobacterium tuberculosis*, *Electrophoresis*, 21: pp3740-3756

Sampson, S. L., Lukey, P., Warren, R. M., van Helden, P. D., Richardson, M. and Everett, M. J., 2001, Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c, *Tuberculosis (Edinburg)*, 81: pp305-317

Sasseti, C. M., Boyd, D. H. and Rubin, E. J., 2003, Genes required for mycobacterial growth defined by high density mutagenesis, *Molecular Microbiology*, 48: pp77-84

Sasseti, C. M., 2009, Inducible expression systems for mycobacteria, *Methods in Molecular Biology*, 465: 255-264

Serafini, A., Boldrin, F., Palu, G. and Manganeli, R., 2009, Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: Essentiality and Rescue by iron and zinc, *Journal of Bacteriology*, 191 (20): pp6340-6344

Shao, W., Yeretssian, G., Goiron, K., Hussain, S. N., Saleh, M., 2007, The caspase-1 digestome identifies the glycolysis pathway as a target during infection and septic shock, *Journal of Biological Chemistry*, 282: pp36321-36329

Siegrist, M. S., Unnikrishnan, M., McConnell, M. J., Borowsky, M., Cheng, T. Y., Siddiqi, N., Fortune, S. M., Moody, D. B. and Rubin, E. J., 2009, Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition, *Proceedings of the National Academy of Sciences, USA*, 106 (44): 18792-18797

Siezen, R. J., Creemers, J. W. M. and van de Ven, W. J. M., 1994, homology modeling of the catalytic domain of human furin, A model for the eukaryotic subtilisin-like proprotein convertases, *European Journal of Biochemistry*, 222: pp255-266

Siezen, R. J. and Leunissen, J. A. M., 1997, Subtilases: The superfamily of subtilisin-like serine proteases, *Protein Science*, 6: pp501-523

Skeiky, Y. A., Owendale, P J., Jen, S., Alderson, Mr R., Dillon, D. C., Smith, S., Wilson, C. B., Orme, I. M., Reed, S. G. and Campos-Neto, A., 2000, T cell expression cloning of a *Mycobacterium tuberculosis* gene encoding a protective antigen associated with the early control of infection, *Journal of Immunology*, 165: pp7140-7149

Skjot, R. L. V., Brock, I., Arend, S. M., Munk, M. E., Theisen, M., Ottenhoff, T. H. M. and Andersen, P., 2002, Epitope mapping of the immunodominant antigen TB10.4 and the two homologous proteins TB10.3 and TB12.9, which constitute a subfamily of the esat-6 gene family, *Infectious Immunity*, 70: pp5446-5453

Sorensen, A. L., Nagai, S., Houen, G., Andersen, P. and Andersen, A., B., 1995, Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*, *Infectious Immunology*, 63: pp1710-1717

Stanley, S. A., Raghavan, S., Hwang W. W. and Cox J. S., 2003, Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialised secretion system, *Proc. Natl. Acad. Sci. USA*, 100: pp13001-13006

Stover, C. K., Warrenner, P., van Devanter, D. R., Sherman, D. R., Arain, T. M., Langhorne, M. H., Anderson, S. W., Towell, J. A., Yuan, Y., McMurray, D. N., Kreiswirth, B. N., Barry, C. E. and Baker, W. R., 2000, *Nature*, 405: pp962-966

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D and Eisenberg, D., 2006, Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. USA*, 103(21): pp8060-8065

Studier, F. W. and Moffatt, B. A., 1986, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of Molecular Biology*, 189 (1): 113-130

Teutschbein, J., Schumann, G., Mollmann, U., Grabley, S., Cole, S. T. and Munder, T., 2009, A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*, *Microbiological Research*, 164: pp253-259

Thomas, D. A., Francis, P., Smith, C., Ratcliffe, S., Ede, N. J., Kay, C., Wayne, G., Martin, S. L., Moore, K., Amour, A. and Hooper, N. M., 2006, A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates, *Proteomics*, 6: 0000-0000, DOI 10.1002/pmic.20050153

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I. and Davison, M., 2001, Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology, *Proteomics*, 1: pp377-396

Tundup, S., Akhter, Y., Thiagarajan, D. and 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 Letters*, 580: pp1285-1293

Tundup, S., Pathak, N., Ramanadham, M., Mukhopadhyay, S., Murthy, K. J. R., Ehtesham, N. Z. and Hasnain, S. E., 2008, The Co-operonic PE25/PPE41 Protein complex of *Mycobacterium tuberculosis* elicits increased Humoral and cell mediated immune response, *PLoS ONE* 3(10): e3586. doi:10.1371/journal.pone.0003586

Varenne S., Buc, J., Lloubes, R. and Lazdunski, C., 1984, Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains, *Journal of Molecular Biology*, 180(3): 549-576

Xu, J. O., Laine, M., Masciocchi, J., Manoranjan, J., Smith, J., Du, S. J., *et al.*, 2007, A unique mycobacterium ESX-1 protein co-secreted with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation, *Molecular Microbiology*, 66: pp787-800

Addendum

A.1 Plasmid maps of expression vectors used in this project

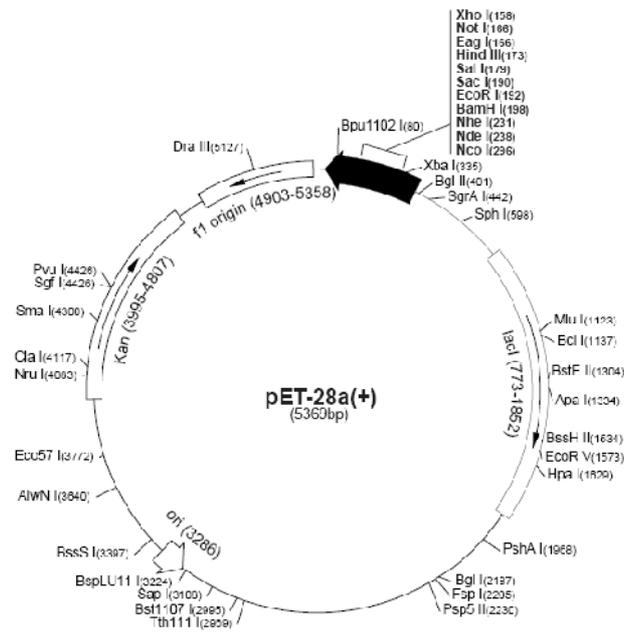


Figure A.1 Plasmid map of *E. coli* expression vector pET-28a

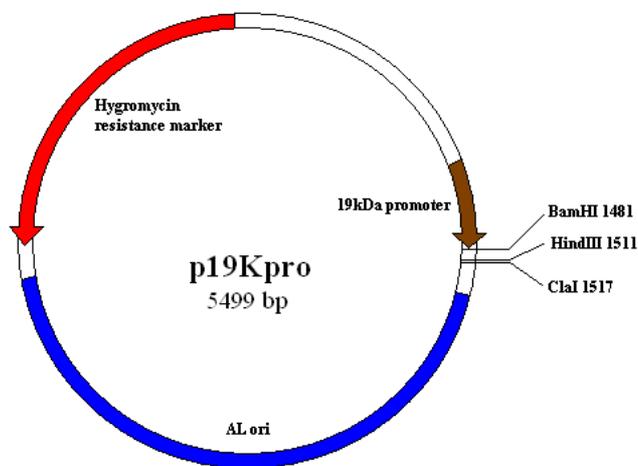


Figure A.2 Plasmid map of *M. smegmatis* expression vector p19Kpro

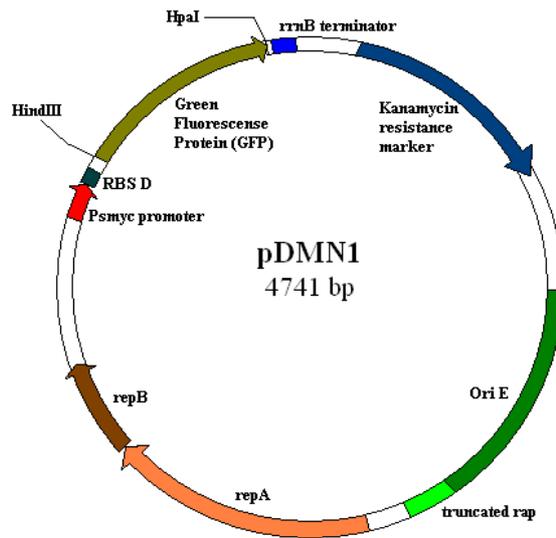


Figure A.3 Plasmid map of mycobacterial expression vector pDMN1

A.2 Detailed results of colony PCR confirmation of the clones

A.2.1 Results for pET-28a expression vector

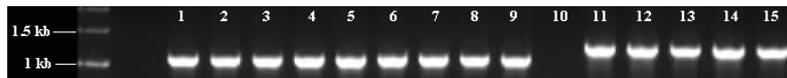


Figure A.4 Image of colony PCR on pET-28a-mycosin-3 without hydrophobic tail (lane 1 – 9, 1058 bp) or with hydrophobic tail (lane 11 – 15, 1241 bp) transformants



Figure A.5 Image of colony PCR on pET-28a-PE5 transformants (lane 1, 3, 4, 5 and 6, 311 bp)



Figure A.6 Image of colony PCR on pET-28a-PPE4 (1544 bp) transformant

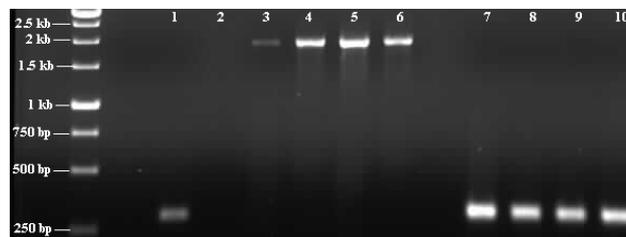


Figure A.7 Image of colony PCR on pET-28a-esxH (lane 1, 293 bp) transformant, pET-28a-PE5&PPE4 (lane 4 – 6, 1855 bp) and pET-28a-esxG (lane 7 – 10, 296 bp) transformants

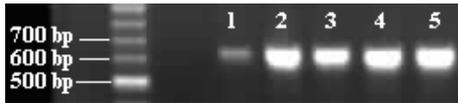


Figure A.8 Image of colony PCR on pET-28a-esxG&esxH (lane 2 – 5, 589 bp) transformants

A.2.2 Results for p19Kpro expression vector

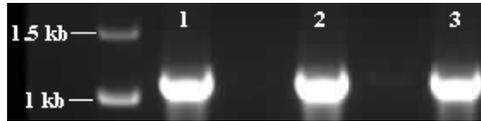


Figure A.9 Image of colony PCR on p19Kpro-mycosin-3 (lane 1, 2 and 3) transformants

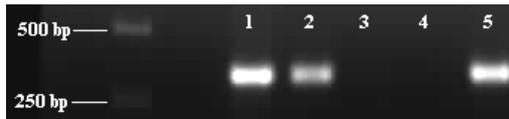


Figure A.10 Image of colony PCR on p19Kpro-esxG (lane 1, 2 and 5) transformants



Figure A.11 Image of colony PCR on p19Kpro-esxH (Lane 1 – 4) transformants



Figure A.12. Image of colony PCR on p19Kpro-PE5 (Lane 1, 2 and 3) transformants

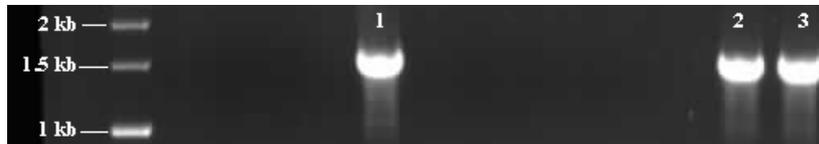


Figure A.13 Image of colony PCR on p19Kpro-PPE4 (Lane 1, 2 and 3) transformants

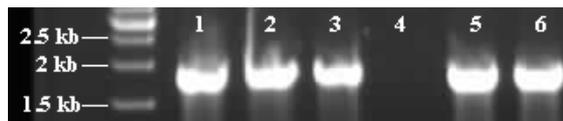


Figure A.14 Image of colony PCR on p19Kpro-PE5/PPE4 (Lane 1, 2, 3, 5 and 6) transformants

A.2.3 Results for pDMN1 expression vector

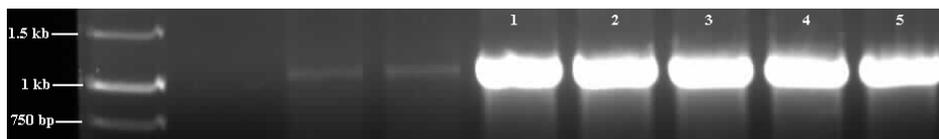


Figure A.15 Image of colony PCR on pDMN1-mycP3 (Lane 1 – 5) transformants

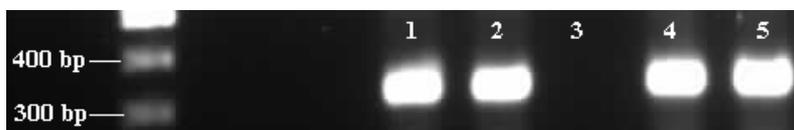


Figure A.16 Image of colony PCR on pDMN1-esxG (Lane 1, 2, 4 and 5) transformants

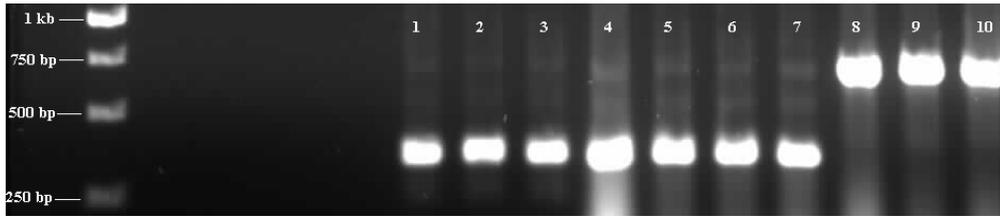


Figure A.17 Image of colony PCR on pDMN1-esxH (Lane 1 – 7) and pDMN1-esxG/esxH (Lane 8 – 10) transformants

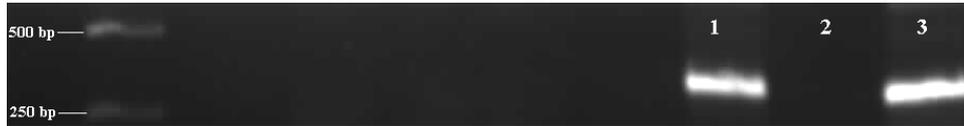


Figure A.18 Image of colony PCR on pDMN1-PE5 (Lane 1 and 3) transformants

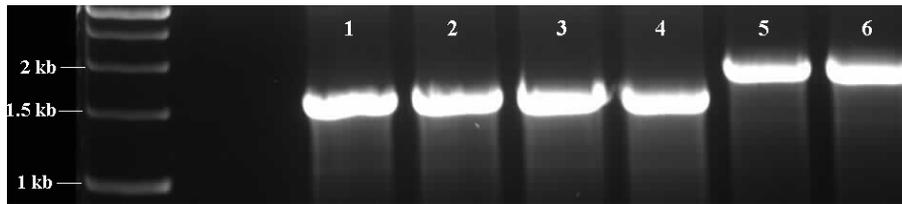


Figure A.19 Image of colony PCR on pDMN1-PPE4 (Lane 1 – 4) and pDMN1-PE5/PPE4 (Lane 5 and 6) transformants

A.3 Detailed results of expression and purification of His-tagged fusion proteins

A.3.1 Results for pET-28a expression vector (expression conditions are indicated)

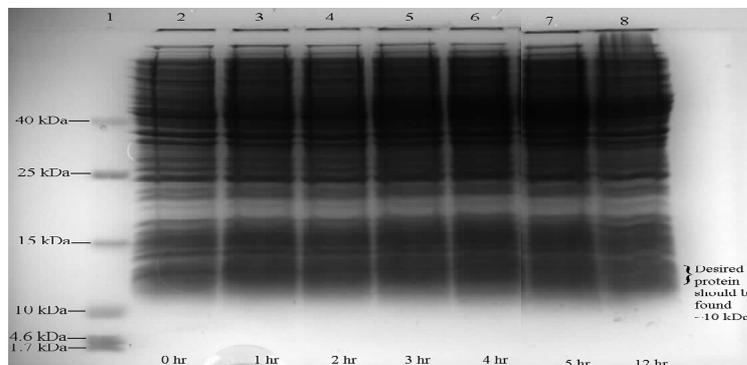


Figure A.20 Image of representative SDS-PAGE gel of test expression of pET-28a-esxG at 25°C with [IPTG] at 0.5 mM for 12 hours in *E. coli* (the area where the His-tagged esxG should be found is indicated in the graph)

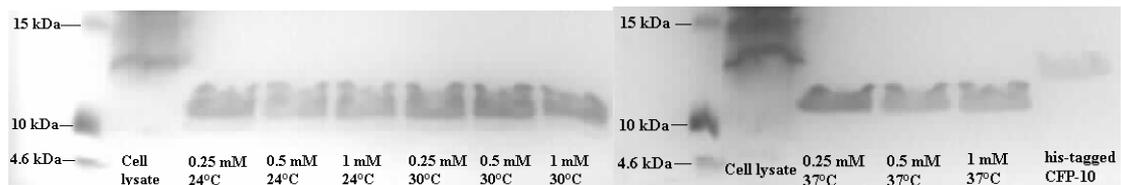


Figure A.21 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxG under 9 different conditions where purified His-tagged esxG fusion proteins were loaded

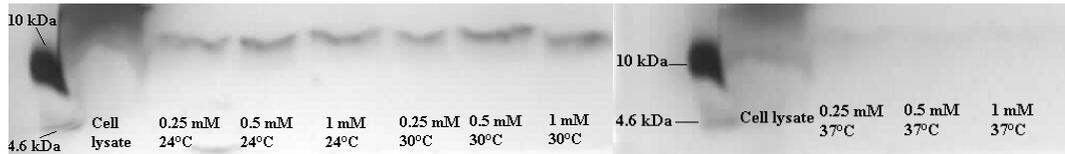


Figure A.22 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxH under 9 different conditions where purified His-tagged esxH fusion proteins were loaded

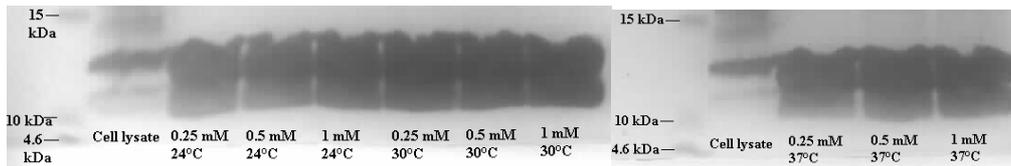


Figure A.23 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-esxG/esxH under 9 different conditions where purified His-tagged esxG/esxH fusion proteins were loaded

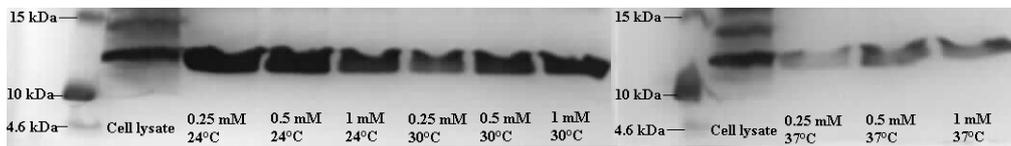


Figure A.24 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PE5 under 9 different conditions where purified His-tagged PE5 fusion proteins were loaded

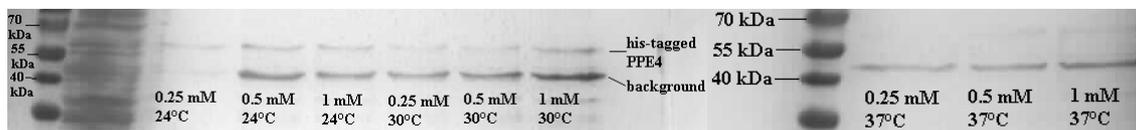


Figure A.25 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PPE4 under 9 different conditions where purified His-tagged PPE4 fusion proteins were loaded

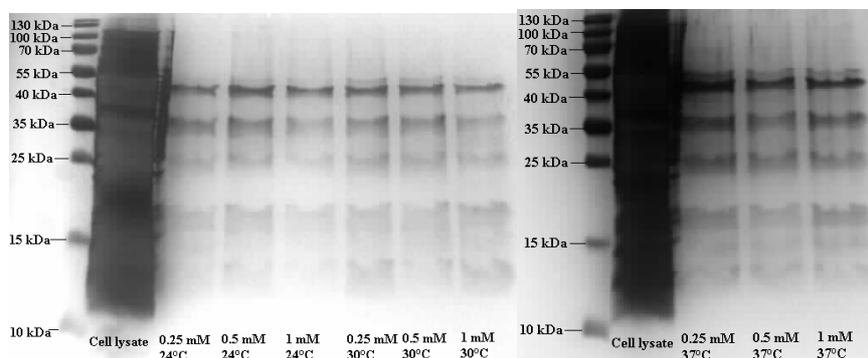


Figure A.26 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-PE5/PPE4 under 9 different conditions where purified His-tagged PE5/PPE4 fusion proteins were loaded

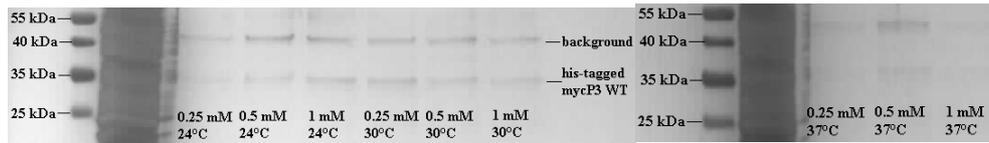


Figure A.27 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-mycP3-without hydrophobic tail under 9 different conditions where purified His-tagged mycP3-WT fusion proteins were loaded

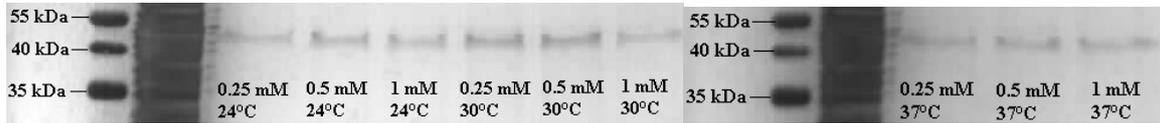


Figure A.28 Image of silver-stained SDS-PAGE gel of test expression of pET-28a-mycP3-with hydrophobic tail under 9 different conditions where purified His-tagged mycP3-T fusion proteins were loaded

A.3.2 Results for pDMN1 *M. smegmatis* expression vector

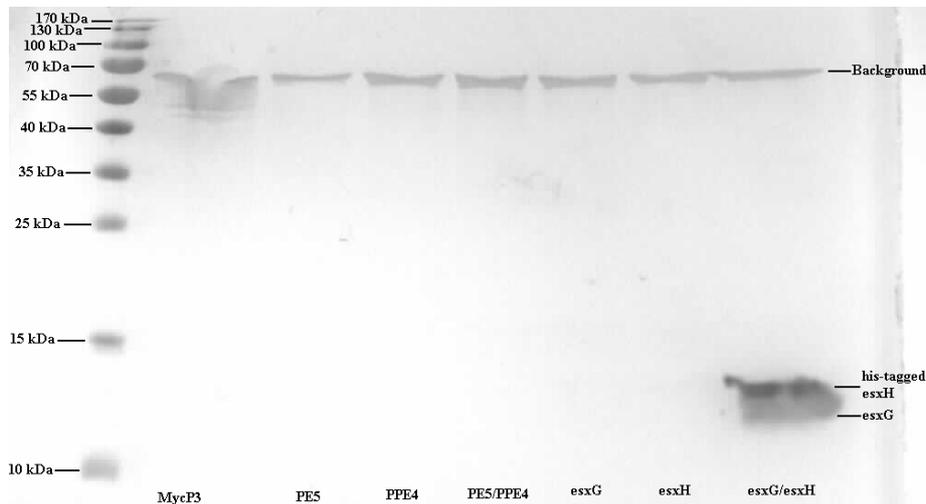


Figure A.29 Image of silver-stained SDS-PAGE gel of expression of His-tagged fusion proteins in *M. smegmatis* at 37°C. The names of the proteins loaded in the lane are specified. Only combined esxG and esxH could be expressed and purified.

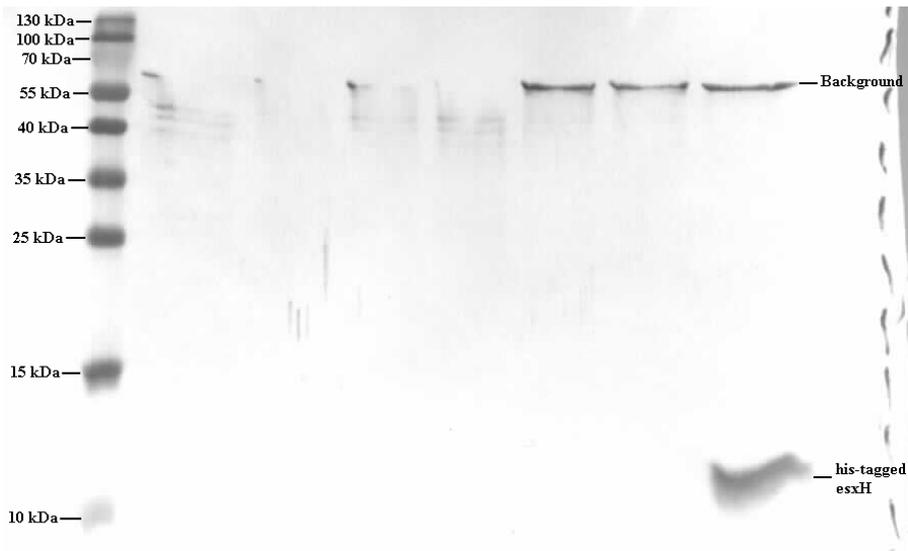


Figure A.30 Image of colourimetric western blot of the duplicate polyarylamide gel shown in Figure 6.25. It confirms that only combined esxG and esxH was successfully expressed and purified from *M. smegmatis*.

A.4 Detailed results of 2-D SDS-PAGE experiments and their western blots

A.4.1 Protein quantification using Bradford assay

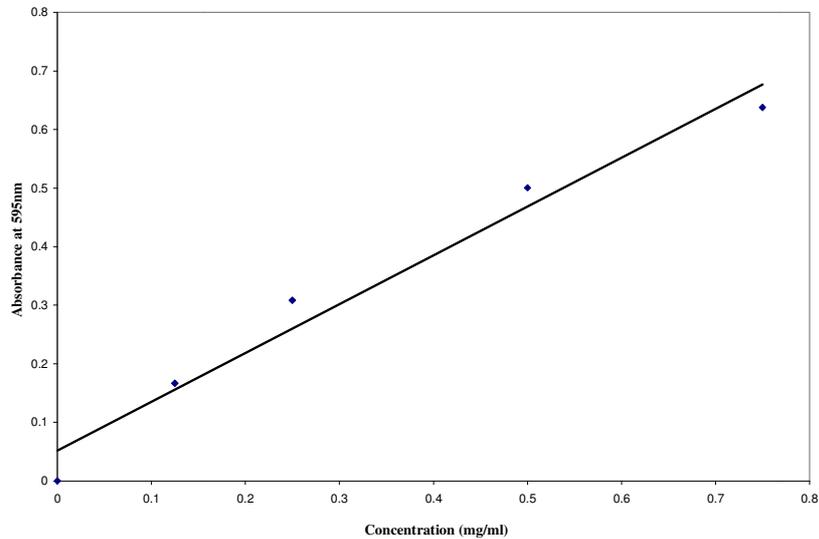


Figure A.31 The Standard curve of Bradford assay on a range of concentrations of Bovine Serum Albumin where the absorbance at 595 nm versus the concentration (mg/ml), the equation of this curve is $y = 0.8333x + 0.0518$, the correlation factor is 0.9703

Table A.1 The amounts of freshly prepared His-tagged fusion proteins using Ni-NTA Superflow column

Name	Volume of original culture (Liter)	Final Concentration (mg/ml)	Final Volume (concentrated) (ml)	Amount (mg)
Mycosin-3 with transmembrane region	3	13	0.6	7.8
Mycosin-3 without transmembrane region	3	15	0.6	9
PE5	2	11	0.7	7.7
PPE4	2	10	0.7	7
Combined esxG and esxH	1	15	0.6	9

A.4.2 The 2-D SDS-PAGE gels in the control group and their western blots using anti-mycosin-3 antibodies

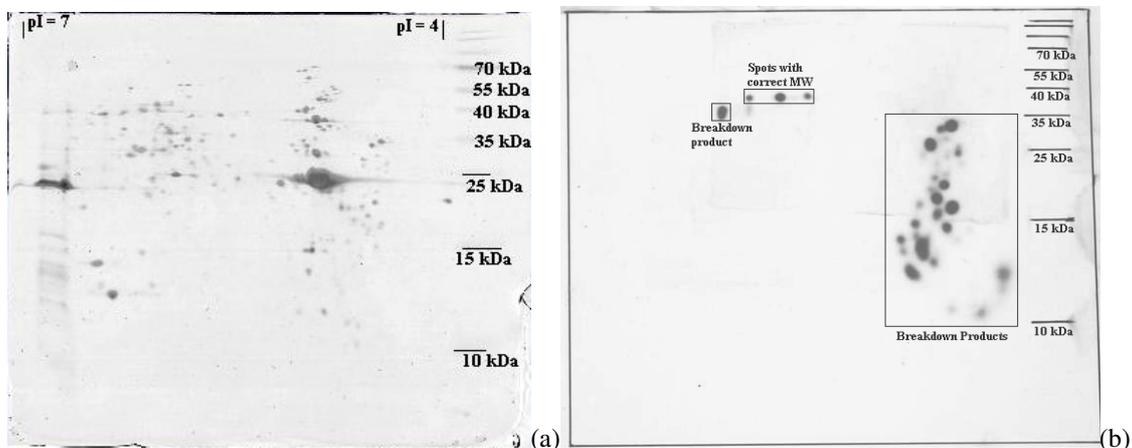


Figure A.32 Silver-stain 2-D SDS-PAGE gel for the purified His-tagged mycosin-3 without transmembrane region fusion protein (a) and its western blot (b)

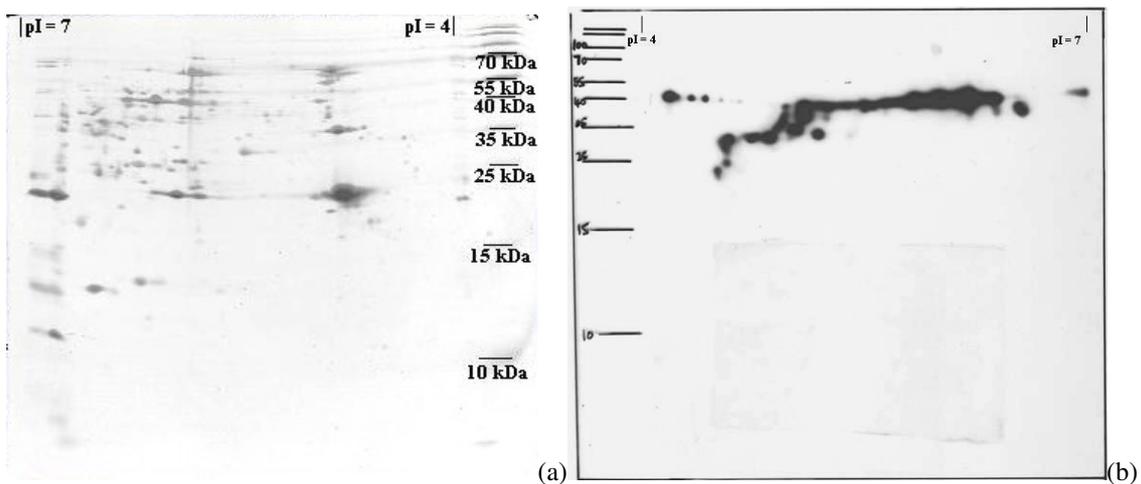


Figure A.33 Silver-stained 2-D SDS-PAGE gel for the purified His-tagged mycosin-3 with transmembrane region fusion protein (a) and its western blot (b)

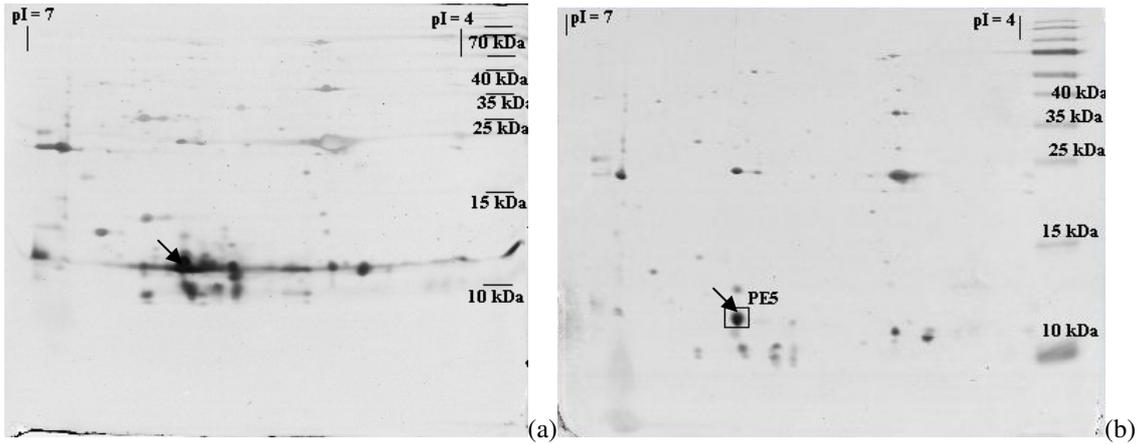


Figure A.34 Silver-stained 2-D SDS-PAGE gels for two different batches of purified His-tagged PE5 fusion protein (the most identifiable spot is the one pointed by an arrow)

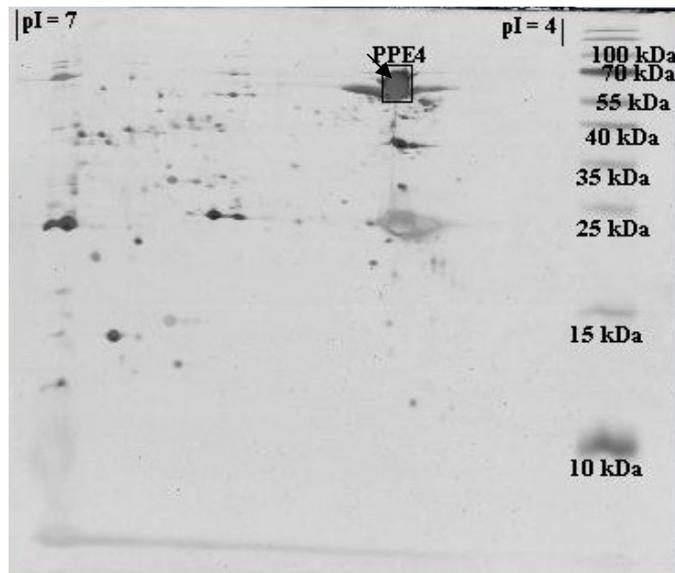


Figure A.35 Silver-stained 2-D SDS-PAGE gel for the purified His-tagged PPE4 fusion protein (the spot indicated by the square surrounded)

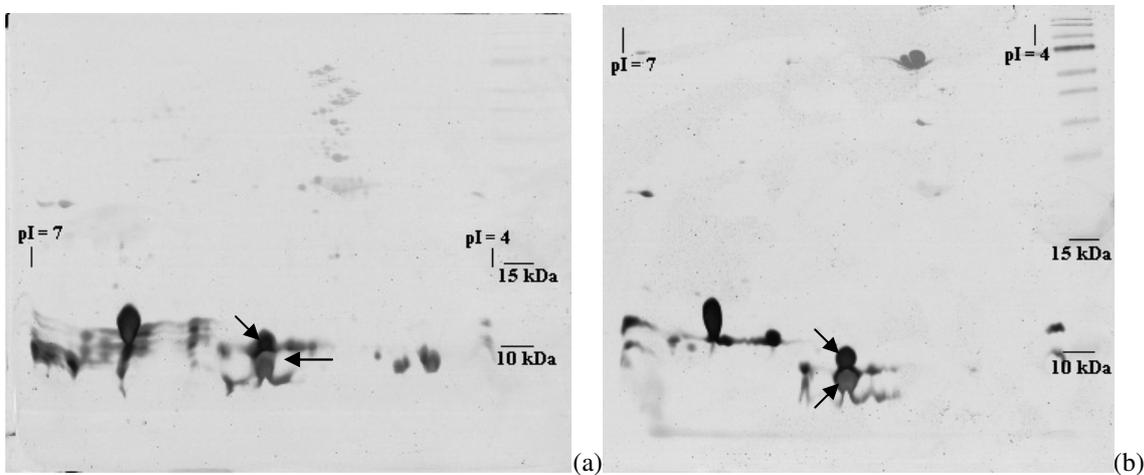


Figure A.36 Silver-stained 2-D SDS-PAGE gel for two different batches of purified His-tagged combined esxG and esxH fusion proteins (the spots indicated by the arrows are probably esxG and esxH)