

**STRUCTURE, PROTEOLYSIS, AND EVOLUTION OF  
SECRETED TUBERCULOSIS VIRULENCE FACTORS**

by

Matthew Morris Solomonson

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF  
DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies  
(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

August 2015

© Matthew Morris Solomonson

# Abstract

*Mycobacterium tuberculosis* uses the ESX-1 type VII secretion system to export proteins to its cell surface, which permeabilize the host macrophage phagosomal membrane, allowing the bacterium to escape and spread to new cells. The structure of the type VII membrane complex and how it mediates this function is unknown, but it is hypothesized that some of the secreted proteins form an extracellular appendage that facilitates membrane lysis or direct secretion of virulence factors into the host cytoplasm. This thesis investigates the structural relationship between one of these secreted proteins, EspB, and a protease that processes it, MycP1. The x-ray crystallographic structures of both proteins are determined and described. EspB is shown to form a multimer with heptameric stoichiometry, and an EM reconstruction of this multimer is generated and used to create a model of the oligomer using symmetric Rosetta docking. The final model is supported by mass spectrometry-based detection of chemically cross-linked peptides from adjacent subunits. We use mass spectrometry to determine how EspB is proteolytically processed during secretion and discuss the effect of this processing event on the EspB ultrastructure. Finally, the structure of one of the membrane apparatus proteins, EccB1 is determined, revealing structural homology to a phage lysin. The combination of x-ray crystallography, EM, modeling, and mass-spectrometry provides an exciting first glimpse at the structure and function of the type VII secretion system - a critical factor in the TB pathogenesis cycle.

# Preface

The research described in Chapter 2 was published in “Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., and Strynadka, N.C. (2013). Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J Biol Chem* 288, 17782-17790.” I cloned and purified the MycP1 gene and generated point mutants and the N-terminal deletion series. For the mass spectrometry experiments, I extracted the EspB cleavage products and the mass spectrometry data collection and analysis was carried out by Dr. Pitter Huesgen. Greg Wasney performed the differential scanning fluorimetry experiments on the N-terminal deletion series. I wrote most of the paper, which was edited by my supervisor, Dr. Natalie Strynadka.

The research described in Chapter 3 was published in Solomonson, M., Setiapura, D., Makepeace, K.A., Lameignere, E., Petrotchenko, E.V., Conrady, D.G., Bergeron, J.R., Vuckovic, M., DiMaio, F., Borchers, C.H., et al. (2015). Structure of EspB from the ESX-1 Type VII Secretion System and Insights into its Export Mechanism. *Structure* 23, 571-583. Marija Vuckovic cloned the *M. tuberculosis* EspB gene, and I made subsequent mutants. I carried out the cloning and mDHFR experiments with materials kindly provided by Dr. Yossi Av-gay. Dr. Julien Bergeron and I carried out initial electron microscopy experiments, and Dheva Setiapura and Calvin Yip collected and processed the EM EspB presented in the paper. Emilie Lameignere helped with the static light scattering experiments and Deborah Conrady with the analytic ultracentrifugation experiments. Karl Makepeace and Jenya Petrochenko carried out the cross-linking/mass spectrometry analysis. Dr. Frank DiMaio performed the Rosetta modelling. I wrote most of the paper, which was edited by Dr. Natalie Strynadka.

The research described in Chapter 4 will constitute a future publication. I carried out the cloning, crystallography, structure solution, and analysis.

# Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
List of Abbreviations	ix
Nomenclature	xii
Acknowledgements	xiii
Dedication	xiv
Chapter 1. Introduction	1
1.1. <i>Mycobacterium tuberculosis</i> and <i>Homo sapiens</i>	1
1.2. The life cycle of <i>Mycobacterium tuberculosis</i>	3
1.2.1. Understanding the enemy	3
1.2.2. A newly infected macrophage	4
1.2.3. Life in the phagosome	6
1.2.4. Escaping the phagosome and triggering cell death	8
1.2.5. Granulomas and caseation	10
1.3. The role of mycobacterial secretion in virulence	11
1.3.1. The mycobacterial cell wall	11
1.3.2. Secretion systems of mycobacteria	13
1.3.3. Identification of ESAT-6 (EsxA)	14
1.3.4. Discovery of the ESX-1 secretion system	15
1.3.5. Possible intracellular functions of ESX-1	19
1.3.6. Structure, secretion, and proposed functions of EsxAB	23
1.3.7. The PE/PPE family proteins	27

1.3.8. Secreted Esp proteins of the ESX-1 locus .....	29
1.3.9. Type VII secretion chaperones .....	34
1.3.10. ESX-1 involvement in DNA conjugation .....	36
1.4. Research scope .....	38
<b>Chapter 2. Structure of a Mycosin Protease .....</b>	<b>40</b>
2.1. Introduction .....	40
2.2. Methods .....	43
2.2.1. Expression and purification .....	43
2.2.2. Crystallization and x-ray crystallographic analysis .....	43
2.2.3. Cleavage assay .....	44
2.2.4. Determination of EspB cleavage products .....	45
2.2.5. Edman sequencing .....	45
2.2.6. Peptide docking .....	45
2.2.7. Differential scanning fluorometry .....	46
2.2.8. Circular dichroism .....	46
2.3. Results .....	46
2.3.1. MycP1 cleaves EspB at two sites .....	46
2.3.2. Overall architecture of MycP1 .....	50
2.3.3. MycP1's catalytic triad .....	53
2.3.4. Active Site Cleft .....	54
2.3.5. N-terminal extension .....	57
2.4. Discussion .....	58
<b>Chapter 3. Structure of EspB and Insights Into its Export Mechanism .....</b>	<b>61</b>
3.1. Introduction .....	61
3.2. Methods .....	65
3.2.1. Purification, crystallization, and structure determination of EspB .....	65
3.2.2. Protein fragment complementation assay .....	66
3.2.3. Multi-angle light scattering .....	67
3.2.4. Analytical ultracentrifuge .....	67

3.2.5. Negative stain electron microscopy.....	67
3.2.6. Oligomeric modeling.....	68
3.2.7. Crosslinking and mass spectrometry.....	69
<b>3.3. Results.....</b>	<b>70</b>
3.3.1. Crystal structure of EspB.....	70
3.3.2. Structure of EspB's secretion signal.....	74
3.3.3. Biophysical analysis of EspB quaternary structure.....	78
3.3.4. Data-guided computational model of EspB oligomer.....	81
3.3.5. Proposed evolutionary scenario for ESX-1-secreted proteins.....	84
<b>3.4. Discussion.....</b>	<b>84</b>
<b>Chapter 4. Structure of EccB1.....</b>	<b>88</b>
<b>4.1. Introduction.....</b>	<b>88</b>
<b>4.2. Methods.....</b>	<b>89</b>
4.2.1. Purification of EccB1.....	89
4.2.2. Crystallization and structure determination of EccB1.....	90
<b>4.3. Results.....</b>	<b>91</b>
4.3.1. Overall structure of EccB.....	91
4.3.2. Pseudosymmetric modularity of EccB.....	94
4.3.3. Comparison with EspB <sub>tb</sub> structure.....	97
4.3.4. EccB surface conservation and electrostatics.....	98
<b>4.4. Conclusion.....</b>	<b>100</b>
<b>Chapter 5. Concluding Chapter.....</b>	<b>101</b>
<b>5.1. Summary.....</b>	<b>101</b>
<b>5.2. The two faces of MycP1.....</b>	<b>102</b>
<b>5.3. The PE/PPE fold.....</b>	<b>103</b>
<b>5.4. Visualization of the T7SS apparatus.....</b>	<b>104</b>
<b>Bibliography.....</b>	<b>107</b>

# List of Tables

<b>Table 2.1.</b> X-ray statistics for MycP1 structure.....	<b>52</b>
<b>Table 3.1.</b> X-ray statistics for EspB structures.....	<b>73</b>
<b>Table 3.2.</b> Crosslinks between adjacent EspB molecules.....	<b>82</b>
<b>Table 4.1.</b> X-ray statistics for EccB1 structure.....	<b>93</b>

# List of Figures

<b>Figure 1.1.</b> A diagram of the possible intracellular fates of <i>Mtb</i> .....	7
<b>Figure 1.2.</b> The mycobacterial cell wall.....	12
<b>Figure 1.3.</b> <i>Mycobacterium bovis</i> BCG region of difference 1.....	15
<b>Figure 1.4.</b> Schematic organization of the ESX-1 secretion system.....	18
<b>Figure 1.5.</b> Summary of hypothesized ESX-1 function.....	21
<b>Figure 1.6.</b> Structure of the EsxAB heterodimer.....	24
<b>Figure 1.7.</b> Structure of the PE/PPE heterodimer.....	28
<b>Figure 1.8.</b> Secreted proteins from the ESX-1 locus.....	31
<b>Figure 1.9.</b> PE25/PPE41 in complex with the EspG5 chaperone.....	35
<b>Figure 2.1.</b> Sequence alignment of MycP1 from <i>Mtb</i> and <i>M. smegmatis</i> .....	42
<b>Figure 2.2.</b> Identification of the MycP1 cleavage site of EspB.....	48
<b>Figure 2.3.</b> Overall structure of MycP1.....	51
<b>Figure 2.4.</b> MycP1's catalytic triad.....	54
<b>Figure 2.5.</b> Active site cleft of MycP1.....	56
<b>Figure 2.6.</b> The N-terminal extension.....	58
<b>Figure 3.1.</b> Overview of ESX-1 secretion.....	64
<b>Figure 3.2.</b> Crystal structure of <i>M. smegmatis</i> EspB.....	72
<b>Figure 3.3.</b> Structure of the bipartite composite secretion signal.....	75
<b>Figure 3.4.</b> Conserved secretion signals in other Esp proteins.....	77
<b>Figure 3.5.</b> Purification and size determination of EspB <sub>tb</sub> .....	79
<b>Figure 3.6.</b> Electron microscopy characterization of EspB.....	81
<b>Figure 3.7.</b> Heptameric EspB <sub>tb</sub> model fit to negative stain 3D reconstruction.....	83
<b>Figure 3.8.</b> Summary of ESX-secreted protein specialization in mycobacteria.....	84
<b>Figure 4.1.</b> Structure of <i>Mycobacterium smegmatis</i> EccB1.....	92
<b>Figure 4.2.</b> Pseudo-symmetric modularity of EccB.....	95
<b>Figure 4.3.</b> Domain differences between <i>Mtb</i> and <i>smegmatis</i> EccB.....	98
<b>Figure 4.4.</b> EccB conservation and surface electrostatics.....	100

# List of Abbreviations

AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AIM2	absent in melanoma 2
APBS	adaptive Poisson-Boltzmann solver
ATP	adenosine-5'-triphosphate
BCG	Bacillus Calmette–Guérin
BLAST	basic local alignment search tool
CBDPS	cyanur-biotin-dimercapto-propionyl-succinimide
CCD	charge-coupled device
CD	cluster of differentiation
CD	circular dichroism
CFP-10	culture filtrate protein 10
CFU	colony-forming unit
CI-2	chymotrypsin inhibitor-2
CTD	c-terminal domain
CV	column volume
CXMS	cross-linking coupled to mass spectrometry
DNA	deoxyribonucleic acid
DSF	differential scanning fluorimetry
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EEA1	early endosome antigen 1
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
EMDB	electron microscopy data bank
ERK	extracellular-signal-regulated kinases
ESAT-6	6 kDa early secreted antigen
Esp	ESX secretion-associated protein

ESX	Esat-6 secretion system
FA	formic acid
FPLC	fast protein liquid chromatography
FRET	Förster resonance energy transfer
FSC	Fourier shell correlation
GCN4	general control protein 4
GTP	guanosine-5'-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography
HR	homology region
HT	helical tip
IFN- $\gamma$	interferon gamma
IL	interleukin
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
IRF3	interferon regulatory factor 3
LAM	lipoarabinomannan
LAMP	lysosome-associated membrane glycoprotein
LC-ESI	liquid chromatography coupled to electrospray ionization
LLO	listeriolysin O
MALS	multi-angle static light scattering
MAPK	mitogen-activated protein kinases
mDHFR	murine dihydrofolate reductase
MES	2-(N-morpholino)ethanesulfonic acid
MHC	major histocompatibility complex
MPB	maltose-binding protein
MS	mass spectrometry
Mtb	<i>Mycobacterium tuberculosis</i>
MycP	mycosin protease

NADP	nicotinamide adenine dinucleotide phosphate
NFκB	nuclear factor kappa-light-chain-enhancer of activated B cells
NLRP3	NOD-like receptor family, pyrin domain containing 3
NMR	nuclear magnetic resonance
OD	optical density
ORF	open reading frame
PAMP	pathogen-associated molecular pattern
PDB	protein data bank
PDIM	phthiocerol dimycocerosate
PEG	polyethylene glycol
PI3P	phosphatidylinositol 3-phosphate
PRM	proline recognition motif
PVDF	polyvinylidene fluoride
RMSD	root-mean-square deviation
SDS-PAGE	sodium dodecyl sulfate gel electrophoresis
Sec	general secretory pathway
SEC	size exclusion chromatography
Sm	<i>Mycobacterium smegmatis</i>
STING	stimulator of interferon genes
T3SS	Type III secretion system
T4SS	type IV secretion system
T7SS	type VII secretion system
TB	Tuberculosis
TBK	TANK-binding kinase 1
TCEP	tris(2-carboxyethyl)phosphine
TDIM	trehalose dimycolate
TLR	toll-like receptor
TM	transmembrane
TNF-α	tumor necrosis factor alpha
TRIS	2-amino-2-hydroxymethyl-propane-1,3-diol

# Nomenclature

Proteins are capitalized and genes are italicized.  $\Delta$  indicates a genetic mutant strain, or in the context of a particular protein, a stretch of deleted residues. Numbers enclosed by square brackets in subscript refer to boundaries of a referenced protein or structure, e.g. MycP1 [24-407]. Proteins belonging to a particular organism are also denoted by subscript, for example MycP1<sub>tb</sub>[24-407] refers to the *Mycobacterium tuberculosis* variant

# Acknowledgements

I thank my supervisor, Dr. Natalie Strynadka for her guidance, great patience, and the privilege of working in an exciting field surrounded by incredible researchers. I acknowledge my supportive committee members, Drs Yossi Av-Gay and Filip Van Petegem for their valuable feedback, expertise, equipment, and materials.

Thanks to collaborators and postdoctoral mentors. Drs Pitter Huesgen and Chris Overall taught me about proteomics. Thanks to Dr. Calvin Yip and Dheva Setiাপutra for the EM data and expertise. Thanks to Dr. Julien Bergeron, a close friend and mentor. Drs Rob Gruninger and Gerd Prehna guided me through the early stages of crystallography, and to Drs Emilie Lameignere and Deb Conrady helped with techniques that measure the size of things. Thank you Marija Vuckovic for that tough clone, Liza Decastro and John Mendes for all the things you did that made my day easier (Liza: I'm sorry about the huge tryptone mess I left on the weighing bench that one time). Thanks to Greg Wasney for inspiring me to develop better organizational skills — my lab book started looking much when you were hired. I am also grateful to staff at the Canadian Macromolecular Crystallography Facility for assistance with x-ray data collection

I acknowledge support from the University of British Columbia Four Year Fellowship, the Richard A. Robertson Memorial Award in Medicine, the Faculty of Graduate Studies Travel award, and the Biochemistry and Molecular Biology Travel Award.

Finally I thank my family, friends and all colleagues for their constant support.

# Dedication

**To Mom and Dad:**

For being so supportive throughout my life  
I could not imagine better parents

**To my brother, Luke:**

For 'goats in boats'

**To Trevor:**

For philosophical discussions, sharpie moustaches, bubble hockey

**To my roommates, Trev, Jessie, Rachelle, G:**

For some of the best years. I'm glad we moved to Vancouver together!

**To my sun and stars, Alina:**

For taking me to Venice, Ajisai, Panorama Ridge, Cannon Beach, (and Boston?)

Love you all

# Chapter 1: Introduction

## 1.1 *Mycobacterium tuberculosis* and *Homo sapiens*

Unlike pathogens that inadvertently infect and kill humans through chance encounters, *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis (TB), has purposefully waged battle with *Homo sapiens* for thousands of years (Brosch et al., 2002). These two species are now so intertwined that it is thought *Mtb* accompanied humans during the 'Out-of-Africa' migrations around 70,000 years ago (Comas et al., 2013). *Mtb*, also known as tubercle bacillus, appears to adapt to the changing parameters of human society: it minimizes transmission in smaller communities so as to not completely wipe out its host, while at times maximizing infectivity in crowded cities (Clark et al., 2002). In this way, *Mtb* is frequently credited as the world's most successful pathogen for having infected over two billion people today and reaching all corners of the globe.

The fascinating story of *Mtb*'s global colonization is matched by its place in the histories of art, medicine, and science. All-time great physicians such as Hippocrates and Galen sought to understand the cause of TB, but it was Robert Koch who undoubtedly made the biggest mark. His 1882 paper, '*The Etiology of Tuberculosis*', which described the criteria by which bacteria cause disease, almost singlehandedly birthed the field of microbiology. Koch invented the core techniques the discipline still uses today, such as liquid and solid bacterial culturing, photomicrography, steam sterilization, and animal infection models. He also created methods to diagnose TB including the tuberculin skin test and visualization by sputum-smear microscopy. Following the discoveries of Koch, TB became the target of the largest vaccination effort ever carried out, with an estimated 3 billion doses of the Bacillus Calmette–Guérin (BCG) vaccine administered by 1989. Finally, as a result of intense funding and research, new drugs were discovered in the 1940s and 50s that led many to predict the end of TB. In the second half of the 20th century, the disease was in steady decline. It seemed that TB would soon be defeated

through the ingenuity of individuals such as Robert Koch and publicly-funded TB research essentially halted (Kaufmann and Parida, 2007).

However, in the 1980s *Mtb* re-emerged as a result of the human immunodeficiency virus (HIV) pandemic, catching the world off-guard. The lack of ongoing basic TB research created a huge gap in knowledge related to the fundamental biology of this organism, and the tools to control the outbreak were limited to those from the Koch era (Comas and Gagneux, 2009). Methods of diagnosis were impractical and technically inadequate (Young et al., 2008). No new drugs had been discovered in decades, treatment regimes were long and difficult to adhere to (Young et al., 2008), and the BCG vaccine proved unable to protect adults from pulmonary disease. The World Health Organization declared TB an emergency in 1993 and today there are approximately 10 million new TB infections per year, with 2 billion people latently infected (Young et al., 2008). Nearly 2 million people die of TB each year, a third of which are co-infected with HIV (Kaufmann and Parida, 2007). Emerging multi- and extensively-drug resistant strains are essentially untreatable (Kaufmann and Parida, 2007).

Our lack of understanding of the fundamental disease mechanisms of *Mtb* and our premature declaration of its defeat played a significant part in our inability to curb the re-emergence of TB. Developing new methods to diagnose, treat, and prevent TB requires deep knowledge of how this pathogen interacts with its host (Brites and Gagneux, 2015). The aim of this thesis is to contribute toward understanding a key molecular system used by *Mtb* during parasitization of its human host. This system, called ESX-1, influences most stages of the *Mtb* infection cycle including its ability to spread infection, manipulate the immune system, and cause lung pathology. Determining how ESX-1 functions at a molecular level will improve our basic knowledge of this organism and may help us develop novel therapeutics and epidemiological tools.

## 1.2 The life cycle of *Mycobacterium tuberculosis*

### 1.2.1 Understanding the enemy

Basic knowledge related to the *Mtb* life cycle and the molecular mechanisms that allow it to manipulate its human host may lead to better drugs, diagnostics, and vaccines. Of paramount importance is our ability to locate the bacillus over the course of the infection cycle, as this might reveal pressure points relevant to vaccine design or drug screening. For example, as the microenvironment of the pathogen changes, so will the availability of carbon sources, metals, vitamins, cofactors, and oxygen (Young et al., 2008). The bacillus will modify gene expression accordingly, and drug screens must be designed to mimic these conditions such that essential proteins are targeted at critical time points.

The microenvironment of the pathogen also influences the immune response; different eukaryotic compartments are monitored by distinct pathogen sensors which are tightly coupled to antigen processing and presentation, cytokine secretion, and the shaping of long term responses. Surface structures or secreted proteins discovered to be highly antigenic and accessible to the immune system could be included in novel subunit vaccines. Alternatively, novel vaccine strains can be engineered to perturb or over-present surface structures in order to elicit an optimal adaptive immune response. We must also study countermeasures employed by the bacteria to evade or subvert the host immune response, as there may be a way to augment the natural immune response or inhibit detrimental signalling processes via therapeutics that target human proteins.

Unfortunately, we do not fully understand the infection process, let alone many of the intricate details of how *Mtb* interacts with its host. However significant progress has been made in recent decades. In the following section, a brief overview of the *Mtb* infection cycle is presented in relation to its primary ecological niche, the macrophage. In section 1.3.4, the influence of the ESX-1 system on these processes is discussed.

### 1.2.2 A newly infected macrophage

An *Mtb* infection begins when an individual with active disease coughs, expelling the bacillus from the airway to the surrounding environment where it can remain suspended in air for several hours. *Mtb* poses a high threat to individuals in crowded, poorly ventilated environments, which is why the disease became more problematic when humans started living in cities after the Neolithic expansion (Comas et al., 2013). Interestingly, the aerosol droplet size is an important determinant of infectivity: a smaller droplet with about 6 bacteria travels more deeply through the airway (Wells, 1948), bypassing crowded parts of the mucosal airway that are teeming with other bacteria to reach the lower lung (Cambier et al., 2014). Here, the bacillus is engulfed by patrolling alveolar macrophages that destroy invaders and digest debris. While *Mtb* also infects other cell types, the macrophage is discussed here.

The macrophage is a type of white blood cell that plays a critical role in non-specific innate immunity. It also helps initiate adaptive immunity by activating other immune cells such as T lymphocytes so that the body creates long-term memory and protects itself from future infections. Macrophages wrap their cell membrane around their prey to internalize them in a vesicle, a process called phagocytosis (Desjardins et al., 2005). This digestive process degrades the engulfed bacterium and liberates complex molecules through conserved processes. Macrophages react to what they are internalizing and communicate potential danger to neighbouring cells. This is mediated by various surface receptors that trigger specific intracellular events. Macrophage receptors known to bind *Mtb* during uptake recognize either opsonic or non-opsonic ligands. *Mtb* is covalently opsonized by C3b from the complement pathway. In contrast, non-opsonic ligands are structures found naturally on the bacterial surface, called pathogen-associated molecular patterns (PAMPs) (Vieira et al., 2002). PAMP receptors include Toll-like receptors, C-type lectin receptors, scavenger receptors, and complement receptors (Philips and Ernst, 2012). Stimulation of particular surface receptors will have specific effects; for example, one receptor may reduce phagosome maturation, while another may up-regulate antimicrobial responses

aimed at destroying the endocytosed cargo and preferentially send the vesicle to the lysosome (Philips and Ernst, 2012).

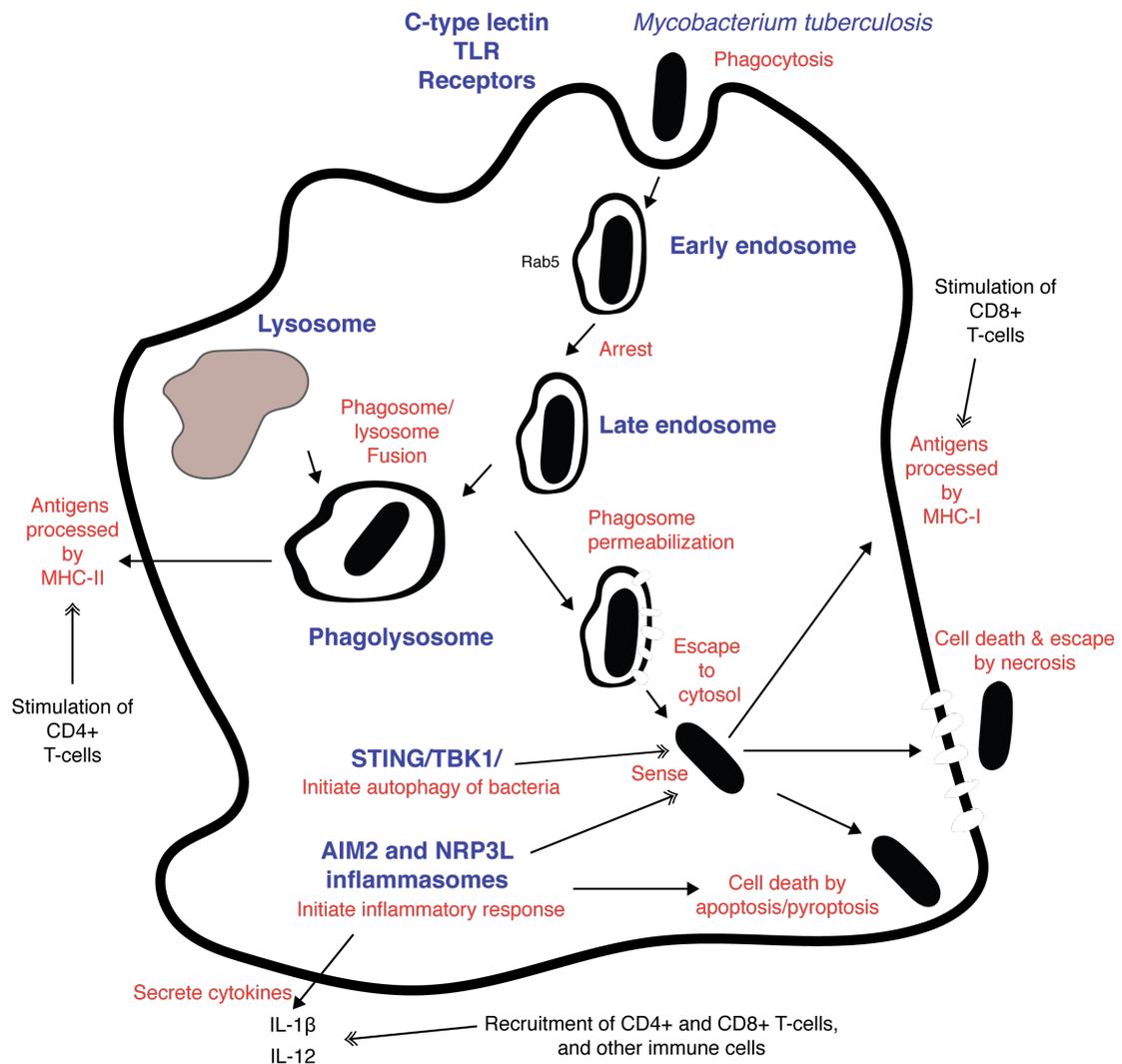
Surface receptors also set off molecular events that trigger secretion of pro-inflammatory agents known as cytokines which alert neighbouring cells. Extensive cross-talk occurs at the site of infection. The goal is to eliminate the bacterium and create lasting memory of the pathogen throughout the body. Macrophages release cytokines that recruit and activate surrounding immune cells, most importantly CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Macrophages also process and present antigens to these recruited T cells. Once bacteria are enclosed by vesicles of the endocytic pathway, their antigens are processed primarily through the major histocompatibility complex class II pathway (MHC-II), which interacts with CD4<sup>+</sup> T cells. However, if bacteria escape to the cytoplasm, their antigens are processed via the MHC class I pathway (MHC-I), which interacts primarily with CD8<sup>+</sup> T cells. When screening for antigens to include in novel TB vaccines, it is important to evaluate if a given antigen triggers strong CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses and produces an optimal cytokine profile that helps the immune system control *Mtb*. Some of these cytokines include interleukins IL-1 $\beta$ , IL-18, IL-12, tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ), and interferon- $\gamma$  (IFN- $\gamma$ ). Therefore, tracking the bacterium through the various host cell compartments is critical to understanding aspects such as antigen-presentation and receptor stimulation. Immediately after uptake, for example, the relevant compartment is the phagosome.

### **1.2.3 Life in the phagosome**

Once internalized, ingested particles are processed by the endocytic pathway in an endosome subtype called the phagosome. If left unimpeded, the phagosome undergoes a maturation process that fills the vesicle with damaging oxidative species and acidifies the vesicle contents. The last step of maturation is lysosome fusion. The lysosome carries a collection of hydrolytic enzymes that degrade ingested bacteria, freeing up organic molecules for nutrients. Phagosome maturation takes about 90 minutes to complete.

During the maturation process, the phagosome undergoes a complex set of interactions with organelles and vesicles from the endocytic pathway, including recycling endosomes, the Golgi/Sec pathway, multi-vesicular bodies, and the lysosome. Vesicles fuse and bud, and cargo is sent and delivered. The mechanistic details of these events are performed by fusion trafficking proteins such as SNAREs, NSF, and SNAPs, but the actual directors of endocytic traffic are GTPases known as Rabs. The Rabs confer identity to the various types of vesicles, allowing the cell to sort vesicles to their target destination. Different Rabs are characteristic of different endosomal stages; an early endosome will have Rab4, Rab5, and Rab11, while a late endosome is rich in Rab7 and Rab9. For example, as the phagosome matures, Rab5 is exchanged for Rab7 through a process known as Rab conversion. When Rab5 is converted from a GDP- to GTP-bound state, it recruits a protein called EEA1, which in turn recruits the vesicle fusion machinery. Another group of key regulators of endosomal traffic consist of phosphoinositide 3-kinases (PI3-kinases) such as hVPS34. Their enzymatic product, the regulatory lipid phosphatidylinositol-3-phosphate (PI3P), accumulates on the vesicle, earmarking it for lysosomal fusion by recruiting trafficking proteins that have PI3P recognition domains (Vergne et al., 2005).

Intracellular pathogens facing the threat of phagosome maturation and lysosome fusion have several evasion strategies: (1) lyse the phagosomal membrane and escape to the cytoplasm, (2) arrest phagosome maturation, or (3) withstand the acidic/oxidative/lytic environment (Finlay and Falkow, 1997). The classic theory is that *Mtb* utilizes the second option, however in recent years it has been suggested that *Mtb* employs a combination of all three strategies. The degree of phagosomal maturation can be determined by tracking fluorescently-labelled markers that are acquired through interaction with other vesicles. These markers include Rab5 (an early endosomal marker), Rab7 (a late endosomal marker), and LAMP-2 or Cathepsin D (lysosomal markers) (Seto et al., 2009). Fluorescent dyes such as LysoTracker are used to selectively stain acidic compartments (Via et al., 1998). Vesicle markers are critical for locating *Mtb* in infected cells, however the degree to which these markers represent the true state of the endosome has been debated (Russell et al., 2009).



**Figure 1.1: A diagram of the possible intracellular fates of *Mtb*.**

How does the bacillus deal with the threats of the endocytic pathway? The first challenge is the superoxide burst during which the NADPH complex assembles on the vesicle membrane to generate superoxide within the compartment for 15-20 minutes (Russell, 2011). Superoxide breaks down into toxic hydrogen peroxide and hypervalent iron species. To combat this, pathogenic mycobacteria are equipped with enzymes such as superoxide dismutase and scavenging lipoglycans that neutralize superoxide radicals. Next, the phagosome is usually acidified through the accumulation of vacuolar ATPases, which pump hydrogen ions into the vesicle, rapidly lowering the pH to ~4.8 over the

course of 15 minutes. The pH drop is required to activate hydrolytic enzymes deposited once the phagosome fuses with the lysosome. However, *Mtb* inhibits vacuolar ATPases to arrest the pH at ~6 (Wong et al., 2011). There is still debate as to what proportion of *Mtb*-containing phagosomes ultimately fuse with the lysosomes (van der Wel et al., 2007).

A number of bacterial lipid and protein molecules contribute to phagosome maturation arrest. The cell wall lipid lipoarabinomannan (LAM) incorporates into the phagosomal membrane to inhibit hVPS34, preventing PI3P accumulation. Other lipids involved in arresting phagosome maturation include phenolic glycolipid phenolphthiocerol (PGL-1), isoprenoid edaxadiene, phthiocerol dimycocerosate (PDIM), and trehalose dimycolate (TDIM). *Mtb* also secretes eukaryote-like serine/threonine protein kinases such as PknG that may inhibit phagosome maturation (Chao et al., 2010), and phosphatases PtpA, PtpB, and SapM, which inhibit vacuolar ATPase trafficking (Bach et al., 2008; Wong et al., 2011), subvert immune responses (Wong et al., 2013), and inhibit PI3P production, respectively (Vergne et al., 2005).

#### **1.2.4 Escaping the phagosome and triggering cell death**

To continue its life cycle, *Mtb* must escape the phagosome, exit the macrophage, and infect new host cells. The classic view is that *Mtb* exists in the phagosome for most of its life cycle and does not spend a significant amount of time in the cytoplasm. However, recent evidence suggests that *Mtb* permeabilizes the phagosome and escapes to the cytosol prior to exiting the cell. A controversial paper was published by van der Wel et al., who, through careful quantitation, found that a sizeable number of bacteria translocate to the cytosol a few days after infection (van der Wel et al., 2007). The significance of cytosolic mycobacteria is now a topic of intense debate (Russell et al., 2010).

Inside the cytoplasm, pathogens face a distinct set of challenges and opportunities. Eukaryotic cells invest significant effort in order to maintain strict compartmentalization; if compartments such as the phagosome are compromised, cytoplasmic-facing receptors detect unsequestered bacterial DNA, RNA, and protein (O'Riordan et al., 2002). Stimulation of these cytosolic surveillance pathways initiates a robust inflammatory

response that activates processes such as autophagy and apoptosis in order to dispose of cytoplasmic bacteria (Fredlund and Enninga, 2014). However, *Mtb* appears to tolerate and even encourage an inflammatory response in order to control immune cell recruitment to the infection site and cause host cell death.

Cytosolic sensing of *Mtb* proceeds through intracellular pathways such as the NLRP3 or AIM2 inflammasome platforms (Dorhoi et al., 2011). Inflammasomes are oligomeric complexes that activate an inflammatory response by processing caspase-1, which induces maturation of pro-inflammatory cytokines such as IL-1 $\beta$  and IL-18 and triggers cell death. Recently, *Mtb* was also found to be detected by the STING/TBK1/IRF3 DNA sensing axis, which targets *Mtb* for elimination via autophagy (Watson et al., 2012). The STING pathway appears to sense extracellular mycobacterial double-stranded DNA (dsDNA), which is becoming increasingly appreciated as a key mycobacterial stimulant of the immune system (Yamashiro et al., 2014).

Cytosolic surveillance pathways can also influence the mechanism by which the infected host cell dies. Infected cells have three possible modes of death: apoptosis, pyroptosis, and necrosis. Apoptosis and pyroptosis are forms of programmed cell death activated by the host in a regulated manner. Pyroptosis is associated with antimicrobial infection, triggered by the inflammasome-mediated release of caspase-1. During apoptosis and pyroptosis, the host cell attempts to contain the infection by initiating a controlled suicide that destroys invading bacteria with reactive oxygen species produced by mitochondria. This keeps the plasma membrane intact, trapping the pathogen inside apoptotic bodies that are subsequently engulfed by other macrophages (Cambier et al., 2014). In contrast to apoptosis and pyroptosis, necrosis is an uncontrolled form of cell death resulting from compromised plasma membrane integrity. *Mtb* actively facilitates necrosis, permeabilizing the plasma membrane while simultaneously preventing the host from repairing the holes (Divangahi et al., 2009). *Mtb* also appears to promote the necrotic pathway by up-regulating TNF- $\alpha$ .

To summarize, *Mtb*'s escape from the phagosome exposes the bacterium to cytosolic surveillance systems that trigger proinflammatory responses leading to cytokine secretion, autophagy, and programmed cell death. However, *Mtb* actively manipulates these

pathways in order to drive formation of the hallmark tissue pathology of TB: the granuloma.

### **1.2.5 Granulomas and caseation**

The granuloma, or “tubercle”, is the structure from which tuberculosis gets its name. It is a tightly interdigitated wall of immune cells surrounding the caseum: a population of bacteria residing in a necrotic core of dead macrophages. The granuloma was once thought to be tissue intentionally remodelled by the host in order to contain and sterilize the site of infection. However this is now believed to be an anthropocentric view; the pathogen actively directs granuloma formation (Russell, 2007). The macrophage cytokines involved in this process include IL-1 $\beta$ , IL-2, TNF- $\alpha$ , CCL2, and CXCL10. These cytokines recruit neutrophils, which in turn secrete cytokines that recruit CD4+ and CD8+ T cells as well as B-lymphocytes. The recruited lymphocytes release IFN- $\gamma$  which inhibits the pro-inflammatory cascade. The final result is a series of stratified layers of infected-macrophages and lymphocytes with caseum and bacteria in the centre (Russell, 2007). It has been suggested that *Mtb* exists extracellularly in the caseum’s rich deposits of cholesterol and triacylglycerol (Kim et al., 2010), which may be ideal grounds for replication (Cambier et al., 2014).

Why does the granuloma suddenly reactivate? Pathogens must facilitate transmission to new hosts, and for this, tissue destruction is required (Brites and Gagneux, 2015). Granulomatous rupture, called caseation, releases bacteria into the airway, which are exhaled into the atmosphere to continue the infection. Caseation occurs most frequently in immunocompromised patients and appears to be regulated locally at the level of individual granulomas. The bacterial factors involved are not entirely clear. Host genes involved in lipid sequestration and metabolism may be involved (Kim et al., 2010), and the human matrix metalloprotease-9 (MMP9) is found at high levels in the serum of transmitting individuals, suggesting that tissue is actively destroyed by this protease (Russell, 2007). During caseation, the granuloma becomes more necrotic, liquefies, and bursts into the airway. The infection begins again when a new host inhales the bacillus.

## 1.3 The role of mycobacterial secretion in virulence

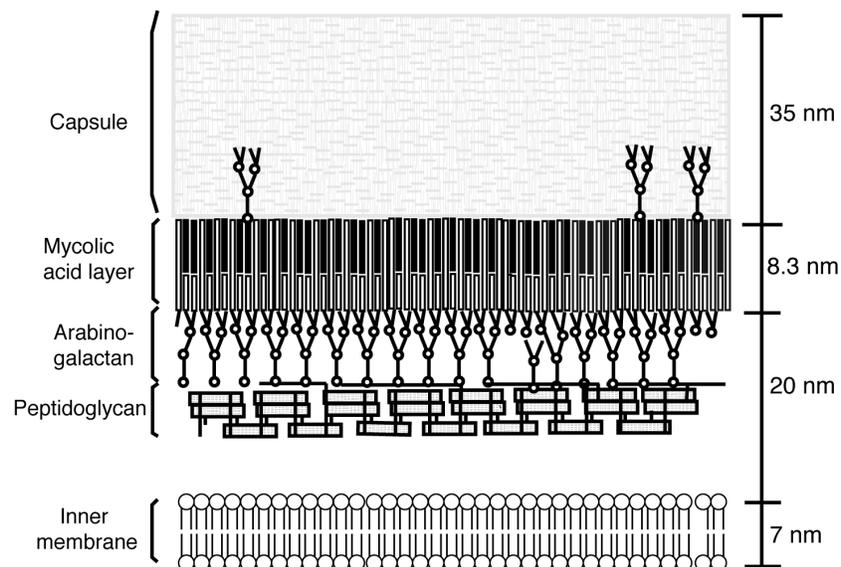
### 1.3.1 The mycobacterial cell wall

The mycobacterial cell wall is a highly impermeable barrier to the outside world and a major virulence factor. *Mtb* mutant strains lacking cell wall biosynthetic enzymes exhibit drastically reduced virulence. These enzymes are also targeted by TB drugs such as ethambutol and isoniazid (Hoffmann et al., 2008). The first layer in the mycobacterial cell wall is the plasma membrane (**Figure 1.2**). Next, is a layer of peptidoglycan, composed of alternating N-acetylglucosamine and modified muramic acid residues, cross-linked by L-alanyl-D-isoglutaminyl-meso-diaminopimelyl-D-alanine (L-Ala, D-Glu, DAP, D-Ala) tetrapeptide sidechains. The majority of linkages occur between the carboxyl group of the terminal D-Ala and the amino group of the D-center of the DAP residue of an adjacent unit (Brennan and Crick, 2007). The 6-carbon of some muramic acid residues link to the next layer: a heterogeneous network of arabinogalactan, which is an elemental sugar unique to mycobacteria. Finally, the arabinogalactan is linked to mycolic acids, an  $\alpha$ -branched  $\beta$ -hydroxylated fatty acid (Brennan and Crick, 2007)

Mycolic acid is unique to mycobacteria and is the primary constitute of the mycolic acid outer membrane, sometimes dubbed the “mycomembrane” (Brennan and Crick, 2007). Despite early controversies, cryo-EM studies have established the existence of the mycolic acid layer (Hoffmann et al., 2008; Sani et al., 2010). Like the Gram-negative outer membrane, the mycobacterial outer membrane creates a hydrophobic barrier to the outside world (Hoffmann et al., 2008). The width of the mycolic acid layer is between 6-8 nm, slightly wider than the Gram-negative outer membrane. Some mycobacteria have unique oligomeric goblet-shaped membrane-embedded porins located in the mycolic acid layer (Faller et al., 2004). Together, the peptidoglycan, arabinogalactan, and mycolic acid layers are essential for cell viability; enzymes involved in synthesizing these layers are important drug targets (Brennan and Crick, 2007).

The outermost layer is called the capsule, also known as the “electron transparent zone” in classic EM studies (Armstrong and Hart, 1975). It is a poorly defined mixture of extractable non-covalently linked glycans, lipids, and proteins that presumably intercalate

with mycolic acids. Lipids and proteins of the capsule appear to interact directly with the host and are essential to virulence; the presence of the capsule dampens pro-inflammatory cytokine responses (Sani et al., 2010). The capsule's existence was initially questioned as it was not observed in some EM studies. However, its absence in these studies can be explained by their use of detergents such as tween-80, which prevents undesirable bacterial clumping during growth but inadvertently causes the capsule to shed (Sani et al., 2010). The fragile nature of the capsule makes it difficult to ascertain whether exported proteins identified in the growth media are naturally separated from the bacterial surface or rather these proteins were originally located in the capsule but shed due to chemical disruption or physical agitation. The capsule's existence was definitively shown using detergent-free culture media followed by cryo-EM analysis of the bacterial cell wall and immunolabelling of its primary sugar moiety,  $\alpha$ -glucan (Sani et al., 2010). These authors also established that the capsular layer contains ESX-1-secreted proteins (Sani et al., 2010). As the capsule is an important virulence determinant, elucidating the mechanisms through which the capsule's chemical constituents interact with infected host may provide important insight into *Mtb* pathogenesis.



**Figure 1.2: The mycobacterial cell wall.**

### 1.3.2 Secretion systems of mycobacteria

The impermeable cell wall that protects mycobacteria from environmental stress also poses a challenge to the bacteria: how does it transport its own proteins across? These proteins are needed to manipulate the outside world and gather nutrients. Mycobacteria have a general secretory translocase (SecYEG) that translocates proteins possessing an N-terminal Sec signal recognized by the SecA1 ATPase. In mycobacteria, Sec signals are longer than those found in Gram-negative bacteria; upwards of 60 residues instead of 20 (Champion and Cox, 2007). A few immunologically relevant *Mtb* T cell antigens are secreted through the Sec pathway including the antigen 85 complex. Mycobacteria are unusual in having a second Sec ATPase called SecA2, which is required for intracellular growth of *Mtb*. It is suggested that SecA2 exists to accommodate post-translationally modified substrates (Champion and Cox, 2007). Some virulence effector kinases (van der Woude et al., 2014) and phosphatases (Wong et al., 2013) may be secreted through SecA2.

Mycobacteria also have a twin arginine translocase (Tat) pathway, which translocates cofactor-containing proteins that are required to first fold in the cytoplasm. Proteins exported via Tat have a hallmark twin-arginine signal sequence. The Tat pathway is essential for *Mtb* growth in culture, and between 11 and 31 potential Tat substrates have been identified (Champion and Cox, 2007). One important group of proteins secreted through Tat are phospholipases C enzymes, which are thought to act on host signalling pathways and membranes (Champion and Cox, 2007).

Finally, mycobacterial genomes encode Esat-6 secretion systems, or ESX/type VII secretion systems (T7SS) (Abdallah et al., 2007). Mycobacteria have up to five ESX gene clusters that have arisen through gene duplication, denoted ESX-1 through ESX-5 (Houben et al., 2013). ESX-1 is an *Mtb* virulence determinant and is the subject of this thesis. ESX-3 is involved in iron acquisition (Serafini et al., 2009; Siegrist et al., 2009). ESX-5 secretes the PE and PPE protein families which modulate the immune response (Abdallah et al., 2008). The functions of ESX-2 and ESX-4 are unknown (Gey van Pittius et al., 2006).

### 1.3.3 Identification of ESAT-6 (EsxA)

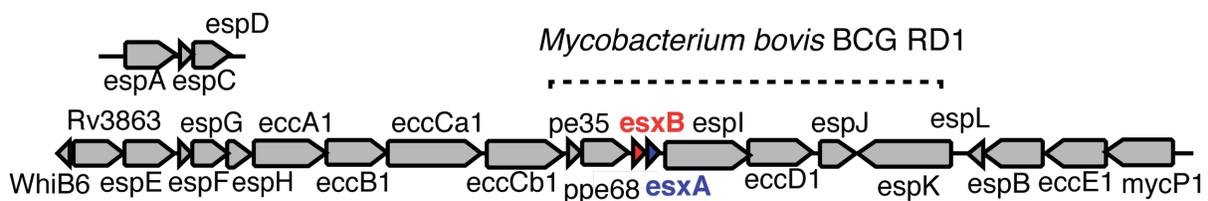
In 1975, a large vaccine trial in India proved the BCG vaccine to be relatively ineffective at protecting adults from pulmonary TB (Rook, 1987). TB incidence in the developing world was also on the rise, indicating drug treatments were failing to completely eliminate the disease (Rook, 1987). This led to renewed efforts to identify proteins secreted by *Mtb* that might be useful for improving the BCG vaccine, and new molecular biology tools were emerging that could help. It was of special interest to identify *Mtb* antigens recognized by circulating human T cells (Rook, 1987), as these were thought to be most relevant to vaccine design and development of novel diagnostics (Rook, 1987). One approach toward achieving this goal involved assaying the reactivity of T cell clones raised against dead or sonicated bacilli followed by phage display to identify the reactive antigens (Young et al., 1985). However it soon became apparent that the antigens recognized from dead/sonicated samples correlated poorly with T cell clones found in human blood (Rook et al., 1986), indicating that the antigens released by disrupted bacilli are not the same as those presented in a live context. These antigens are unlikely to be of use in engineered TB vaccines as they would not be accessible to the immune system during infection. To identify surface-displayed antigens secreted by live bacteria, Abouhzeid et al. developed an approach to growing *Mtb* for a short time (4-7 days) followed by analysis of [<sup>35</sup>S]-labeled secreted proteins; the labeling strategy allowed bona fide secreted proteins to be distinguished from those released by cell lysis (Abouzeid et al., 1986). Soon, many secreted antigens of varying size were identified (Abouzeid et al., 1988; Andersen et al., 1991a; Andersen et al., 1991b; Baird et al., 1989; Barnes et al., 1989; Collins et al., 1988) and a list was compiled (Young et al., 1992).

Next, Andersen et al. sought to determine which secreted antigens are recognized specifically by memory T cells, as these may help the body establish long-term immunity against TB (Andersen and Heron, 1993). They hypothesized that the T cell responses provoked by actively growing bacteria at the height of infection might be different from those initiated by long term memory T cells; this subset of T cells may better protect against a second infection. Their efforts led to the identification of a low molecular weight

protein fraction from the short term culture filtrate (Andersen and Heron, 1993) that could provide mice with a level of protective immunity that was comparable to BCG (Andersen, 1994). While isolating the antigens responsible for this protective effect, Andersen purified a 6 kDa protein from the short-term culture filtrate (Sorensen et al., 1995). Using the phage display approach (Young et al., 1985), he identified the antigen as a 95-residue protein and named it the 6 kDa early secretory antigenic target, or ESAT-6, which was later renamed EsxA. EsxA was present as multiple apparent bands on a gel, localized to the culture filtrate and cell wall fractions, and could elicit release of IFN- $\gamma$  from memory T cells (Sorensen et al., 1995). EsxA would be the subject of intense research for the next 20 years and was soon discovered to be secreted through the ESX-1 system.

### 1.3.4 Discovery of the ESX-1 secretion system

The discovery of the ESX-1 system began with development of the infamous BCG vaccine between 1908 and 1921 by Albert Calmette and Camille Guérin. The protective agent comprising the BCG vaccine is an avirulent strain of *Mycobacterium bovis*, a close relative of *Mtb* that causes TB in cattle. The attenuated *M. bovis* strain, later named BCG, emerged after many rounds of cultured growth in ox bile media. BCG's inability to revert back to virulence hinted that an irrecoverable genetic deletion was responsible for the attenuation; genome hybridization (Mahairas et al., 1996) and microarray experiments (Behr et al., 1999) later confirmed this to be the case. Of all the chromosomal deletions that had occurred in the numerous BCG substrains, only one deletion was common to all of them: a 9.5 kilobase stretch of DNA spanning several open reading frames (ORFs) termed region of difference 1 (RD1) (**Fig. 1.3**).



**Figure 1.3.** *Mycobacterium bovis* BCG region of difference 1 (RD1)

Strikingly, RD1 was found to encode EsxA: the secreted T cell antigen identified earlier by Andersen (Sorensen et al., 1995). When the complete *Mtb* genome sequence became available (Cole et al., 1998), it was revealed that the *esxA* gene is flanked by several ORFs predicted to encode a number of membrane proteins and ATPases; these were immediately hypothesized to form the secretion system responsible for EsxA's Sec-independent export (Tekaiia et al., 1999). The genome sequence also helped identify a second antigen secreted to the short term culture filtrate named, 'culture filtrate protein 10' (CFP-10, later renamed EsxB), and the *esxB* gene is co-transcribed with *esxA* (Berthet et al., 1998). This newly-identified gene cluster, hypothesized to encode a novel type of secretion system that exports immunologically-relevant substrates, received a considerable amount of attention from microbiologists, geneticists, bioinformaticians, and immunologists.

The floodgates opened for RD1-related experiments. In the following years, many major discoveries related to *Mtb* pathogenesis involved some form of RD1 mutant. Of all the BCG regions of difference identified, only reintroduction of RD1 was found to restore *M. bovis* BCG virulence (Pym et al., 2002). Multiple groups used targeted gene deletions and signature-tagged mutagenesis to confirm that *Mtb* strains with a variety of RD1 disruptions (collectively designated here as *Mtb*- $\Delta$ RD1) are attenuated to the same extent as *M. bovis* BCG. Compared to wild type bacteria, *Mtb*- $\Delta$ RD1 strains exhibit markedly reduced virulence in multiple types of assays designed to probe various aspects of the *Mtb* infection cycle. *Mtb*- $\Delta$ RD1 strains have a reduced ability to kill mice (Hsu et al., 2003; Lewis et al., 2003; Stanley et al., 2003) and display less severe lung pathology in mouse model infections (Hsu et al., 2003; Lewis et al., 2003). *Mtb*- $\Delta$ RD1 are less cytotoxic, have reduced ability to lyse infected cells, and are unable to spread to new cells as measured in human macrophage and epithelial infection models (Gao et al., 2004; Guinn et al., 2004; Hsu et al., 2003; Lewis et al., 2003). *Mtb*- $\Delta$ RD1 strains also exhibit reduced capacity to replicate in cells, as measured by CFU studies in both cellular and mouse models (Lewis et al., 2003; Stanley et al., 2003). Furthermore, *Mtb*- $\Delta$ RD1 strains are less able to arrest phagosome maturation (MacGurn and Cox, 2007) and their ability to modulate immune

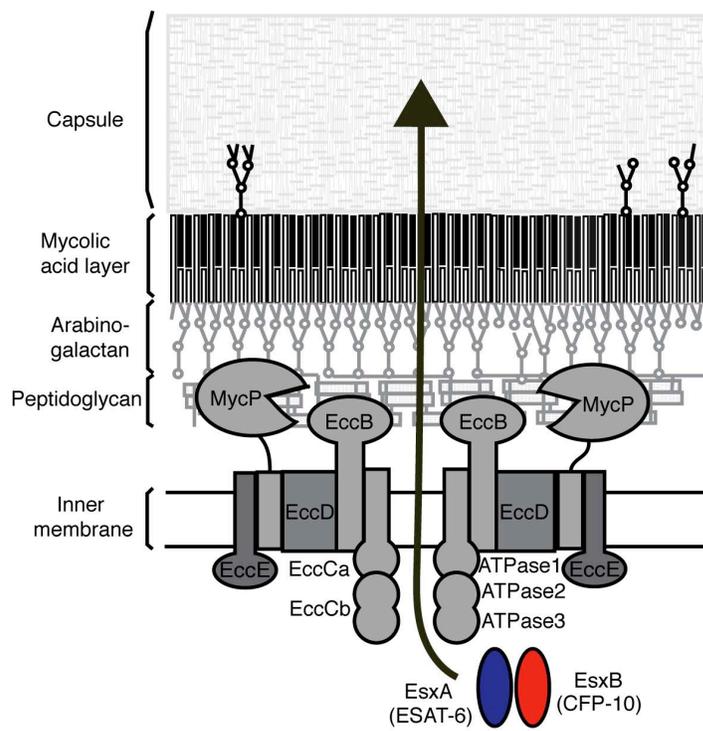
responses such as cytokine release and nitric oxide production is perturbed (Stanley et al., 2003). Clearly, RD1 had an important role in pathogenesis.

Many groups also sought to test the prediction that RD1 encoded a novel secretion system by carrying out experiments related to RD1's predicted transcripts and ORFs. The RD1 region can be broken down into 4 transcriptional units (Pym et al., 2003). To determine which RD1 genes are most essential to virulence, different combinations of these transcriptional units were reintroduced into *M. bovis* BCG, however only a plasmid carrying the entire deleted region resulted in recovery of virulence (Pym et al., 2003). Several ORFs surrounding the *esxA* and *esxB* genes were deemed essential for secretion of EsxA and EsxB, adding support for the system's proposed role in protein export (Guinn et al., 2004; Hsu et al., 2003; Pym et al., 2003; Stanley et al., 2003). The fish pathogen *Mycobacterium marinum* proved especially useful for establishing the contributions of individual ORFs to the system (Gao et al., 2004). Many of the single-gene disruptions had similar phenotypes as deletions of the complete RD1 region (Brodin et al., 2006; Guinn et al., 2004). These studies also served to expand the accepted boundaries of the functional gene cluster to include additional genes adjacent to the RD1 region (Gao et al., 2004). Finally, yeast two-hybrid experiments identified interactions between multiple proteins encoded in the locus (Stanley et al., 2003). The cumulative result was that, rather than individual genes acting independently, the gene cluster disrupted by the RD1 deletion likely encodes a multi-component secretion apparatus, perhaps similar to those commonly involved in the virulence mechanisms of Gram-negative bacteria.

The secretion system received a new designation: ESX-1. The field also introduced a systematic genetic nomenclature along with a more general designation "type VII secretion system" (T7SS) (Bitter et al., 2009a; Bitter et al., 2009b). The christening of T7SS was not without controversy; experts in Gram-negative bacterial secretion systems claimed that the "type *n* secretion" designation should be reserved strictly for bacteria with definitive outer-membranes (Desvaux et al., 2009a, b). In this nomenclature, the core components of the apparatus are labelled with the prefix Ecc, for **Esx core component**, namely, EccA, EccB, EccC, EccD, and EccE. Other ESX-associated proteins were assigned the prefix Esp (ESX secretion-associated protein). These include EspA through EspL, and

the transcriptional regulator EspR. To distinguish between proteins from the 5 paralogous ESX systems, the protein names are followed by numerals 1 through 5 (e.g. EccB1, EccB2, EccB3, etc.). This standardized naming scheme appears to have mitigated problems seen in other bacterial secretion systems where gene/protein names were not standardized from the outset. These other systems often have a confusing set of names that vary by species.

Bioinformatic analyses of the RD1 genes provided a few hints regarding the function of some of the ESX membrane proteins (Tekai et al., 1999). EccC, which in some ESX systems is broken down into two genes, EccCa and EccCb, is a FtsK-SpoIIIE-like ATPase responsible for supplying energy for protein translocation. EccD is a multi-spanning integral transmembrane protein thought to be a channel through which secreted substrates pass (Abdallah et al., 2007). Each of the five ESX gene clusters also encode subtilisin-like serine proteases, termed mycosins, a linguistic blend of the genus **mycobacteria** and the subtilisin protease family (Brown et al., 2000). Mycosins are also sometimes called mycosin proteases and are abbreviated MycP followed by the numerals 1 through 5 depending on their associated ESX system (Bitter et al., 2009a).



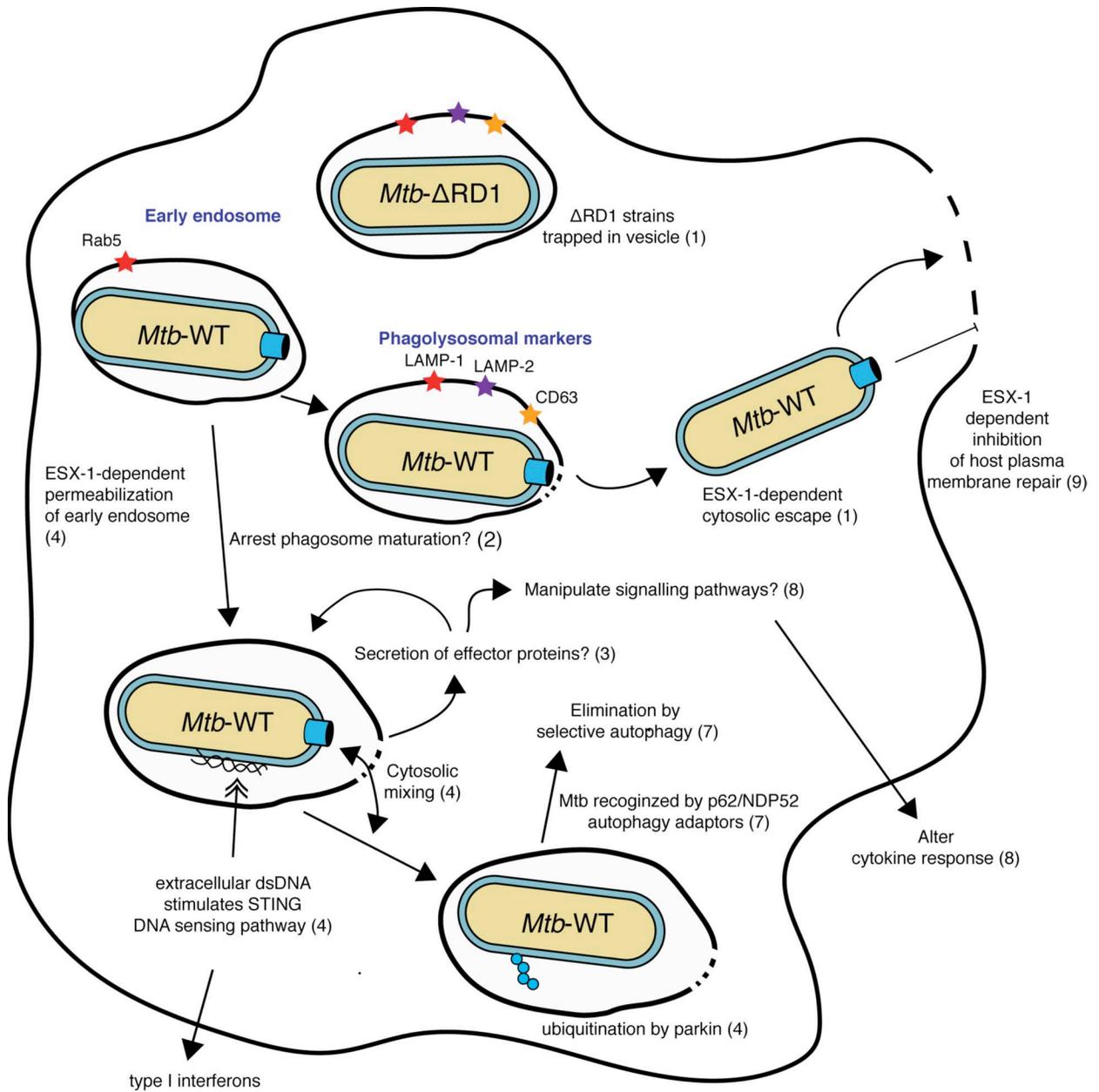
**Figure 1.4. Schematic organization of the ESX-1 secretion system.**

### 1.3.5 Possible intracellular functions of ESX-1

With ESX-1's importance to promoting *Mtb* virulence established, researchers began to look more closely at the impact of the system in the context of the infected cell. Various intracellular effects have been observed that can perhaps be grouped into three broad infection models: (1) ESX-1 may allow for *Mtb* to full-on escape from the phagosome, (2) ESX-1 may permeabilize the phagosomal membrane, allowing for some mixing with the cytoplasm, and (3) ESX-1 may secrete proteins into the cytoplasm that modulate host cytosolic factors. The common theme of the models that ESX-1 permits the bacterium some degree of access to the host cytoplasm. However, the three models fall on a spectrum with respect to the extent in which they disrupt the previously-accepted scenario in which *Mtb* resides exclusively in phagosomes. The specific intracellular role of ESX-1 is hotly debated in the field, and there is by no means a consensus.

In 2007, van der Wel et al. published a paper that challenged the notion that *Mtb* spends most of its time in the phagosome; rather, they suggested that *Mtb* escapes to the cytosol two days after the start of infection (van der Wel et al., 2007). They conducted a cryo-electron microscopy analysis to survey the subcellular locations of bacteria. Is *Mtb* found in the phagosome or in the cytosol, and at which time point during infection? The group evaluated whether the bacterium was encapsulated in phagosome by tracking markers of early endosomes (EEA1 and transferrin receptor) or markers associated with lysosomes (LAMP-1, LAMP-2, CD63, and cathepsin D). They also observed the presence or absence of phagosomal membrane using EM. The results indicated that a portion bacteria had escaped from phagosomes. Importantly, this was not observed with BCG or heat-killed *Mtb*, and EsxA/EsxB secretion was required for escape. The same group later published a second report using a wider range of mycobacterial species, and demonstrated only the pathogenic species (*Mtb*, *M. marinum*, and *M. leprae*) can escape phagosomes (Houben et al., 2012a). Intriguingly, although some non-pathogenic bacteria like *M. smegmatis* have intact ESX-1 systems that actively secrete EsxA, they are incapable of phagosomal escape. A separate group added quantitative support for the notion that *Mtb* gains cytoplasmic access in an ESX-1-dependent manner using a single-cell

fluorescence resonance energy transfer (FRET)-based assay (Simeone et al., 2012). Further evidence for phagosomal escape was presented by researchers who discovered that *M. marinum* infecting the amoeba *Dictyostelium* were able to escape from vacuoles and spread to new cells in a non-lytic manner using a polymerized actin-based structure called the “ejectosome”; this was entirely dependent on having an intact ESX-1 (Hagedorn et al., 2009; Stamm et al., 2003). Macrophage infection experiments combined with contact-dependent red blood cell haemolysis assays indicated *M. marinum* also used ESX-1 to form pores and escape to the cytosol (Smith et al., 2008). It is unclear to what extent the *M. marinum* results can be extrapolated to *Mtb*. Finally, ESX-1-dependent *Mtb* escape was also determined to occur *in vivo* in a mouse-tail infection model (Carlsson et al., 2010). However there is still much controversy associated with *Mtb* phagosomal escape; it is unknown if this occurs in human lungs.



**Figure 1.5. Summary of various ESX-1-dependent effects.** 1:(van der Wel et al., 2007) 2: (Bach et al., 2008), 3: (Stanley et al., 2007), 4: (Manzanillo et al., 2013), 5: (Manzanillo et al., 2012), 6: (Kurenuma et al., 2009). 7: (Watson et al., 2012). 8: (Koo et al., 2008), 9: (Divangahi et al., 2009).

A second school of thought is more conservative, not necessarily going so far as to say that ESX-1 helps *Mtb* completely escape phagosomes, but rather that ESX-1 generates pores that permeabilize phagosomal membranes, allowing the phagosome contents to mix with the cytosol. This hypothesis partly stems from observations that ESX-1 is required for

*Mtb* to trigger host cytosolic surveillance pathways like the STING DNA-sensing pathway (Manzanillo et al., 2012). This pathway is thought to sense extracellular DNA secreted by *Mtb* during biofilm formation. The STING pathway results in *Mtb* being marked for ubiquitination by the parkin ubiquitin ligase, targeting the bacterium for selective autophagy and destruction by the autophagolysosome (Watson et al., 2012). Interaction with STING and autophagic targeting would require direct cytosolic contact (Manzanillo et al., 2013). The work also provided an interesting link between nucleic acid sensing and autophagy as a mechanism of control of intracellular pathogens. It is worth noting that stimulation STING also impacts cytokine release, which would influence the immune response and granuloma formation. Interestingly, the ESX-1-dependent transcriptional profile of *Mtb*-infected bone marrow-derived macrophages was found to be nearly identical to that of macrophages infected with *Listeria monocytogenes*, which enters the cytosol using a pore-forming toxin called listeriolysin O (LLO) (Manzanillo et al., 2012). Both of these pathogens appear to activate a transcription factor called interferon regulatory factor 3 (IRF3) which initiates the production a class of cytokines called type I interferons including IFN-1 $\beta$  typically associated with viral infections.

Finally, it has been proposed that rather than disputing phagosomal membranes, ESX-1 secretes protein effectors directly into the macrophage host cytoplasm to modulate host responses (Stanley et al., 2007). For example, such secreted proteins may interact with the NLRP3/ASC inflammasome, leading to cytokine secretion required for bacterial spread. The inflammasome processes caspase-1, which is responsible for cleavage and secretion of cytokines IL-1 $\beta$  and IL-18. ESX-1 stimulates the release of these factors from infected macrophages (Koo et al., 2008), possibly by causing a potassium ion influx that results in cathepsin-B release from lysosomes (Kurenuma et al., 2009). IL-1 $\beta$  and IL-18 secretion are hypothesized to promote recruitment of other immune cells for continued bacterial spread, and a number of groups have observed the inability of  $\Delta$ ESX-1 mutant strains to form structured granulomas that caseate normally (Carlsson et al., 2010; Volkman et al., 2004).

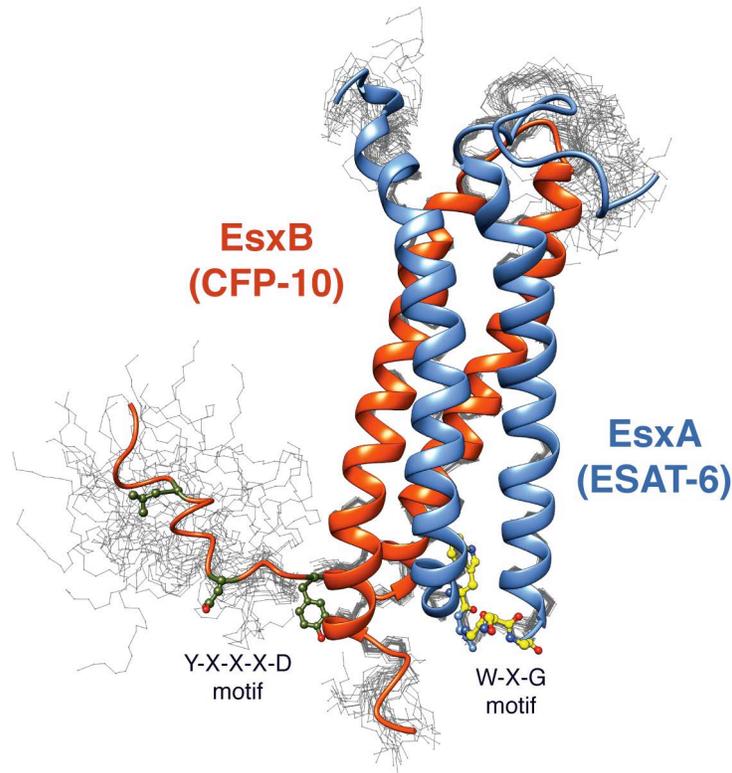
The issue of ESX-1-dependent access to the cytoplasm has practical implications related to vaccine design. The system appears to influence the route through which *Mtb*

antigens are presented on the surface of the infected macrophage. If *Mtb* is sequestered to the phagosome, its antigens will be processed through MHC-II pathway to prime CD4+ T cells. However escape to the cytosol would mean some of its antigens are processed via MHC-I and presented to CD8+ T cells. For example, EsxB peptides are processed through the MHC-I pathway (Lewinsohn et al., 2006). Taken together, *Mtb*'s ESX-1-dependent access to the cytosol has important effects on the life cycle of the pathogen and the host immune response, which may be important to consider in the context of vaccine engineering.

### **1.3.6 Structure, secretion, and proposed functions of EsxA and EsxB**

As ESX-1's role in virulence and intracellular pathogenesis was elucidated, other important steps were made toward understanding how the system operates at a molecular level. These steps began with the biophysical characterization of EsxA and EsxB combined with the use of purified protein in assays developed with goal of discovering the direct functions of these proteins. There was also considerable interest as to how substrates are targeted for secretion through ESX; locating a secretion signal could help identify other secreted proteins.

On its own, EsxA is ~75% helical but with minimal tertiary structure as measured by circular dichroism, tryptophan fluorescence, and NMR experiments (Renshaw et al., 2002). In contrast, EsxB adopts a random coil (Renshaw et al., 2002). While EsxA and EsxB are largely unstructured when separated, they may never be found this way in nature because their genes are co-transcribed (Okkels and Andersen, 2004) and their protein products form a tight, fully folded heterodimeric complex (Renshaw et al., 2002) that exists both inside and outside the cell (Brodin et al., 2005). Herein, this complex will be referred to as EsxAB. The solution NMR structure of EsxAB revealed an antiparallel four-helix bundle (**Fig 1.6**) (Renshaw et al., 2005). Of particular interest are the N and C termini of both proteins, which form long, flexible arms at either end of the four-helix bundle (Renshaw et al., 2005). Mutations to residues in these regions disrupt secretion and virulence (Brodin et al., 2005).



**Figure 1.6. Structure of the EsxAB heterodimer.** One representative structure from the ensemble depicted as ribbon, the rest as backbone stick trace. W-X-G and Y-X-X-X-D motifs indicated (PDB: 1WA8, (Renshaw et al., 2005)).

The secretion signal for the T7SS system was soon located in the flexible C-terminus of EsxB (Champion et al., 2006). Yeast two-hybrid experiments demonstrated that EsxB is recruited to the membrane by the EccCb1 ATPase (Stanley et al., 2003). When the last seven residues of EsxB are deleted or mutated, interaction with EccCb1 is abolished and EsxB is not secreted (Champion et al., 2006). The sequence of the secretion signal was later refined to Tyr-x-x-x-Asp/Glu motif (YXXXD) and a few additional hydrophobic residue downstream from this (Daleke et al., 2012a) (**Fig. 1.6**, discussed in greater detail in **section 3.1**). The signal is required for export of a number of other ESX substrates including PE25, LipY, and EspB, but does not appear to target substrates to particular ESX systems; switching the signal does not change the ESX system through which they are secreted, indicating that this information is encoded elsewhere in the protein (Daleke et

al., 2012a). A separate group demonstrated that the signal is conserved in many Gram-positive bacteria such as *Bacillus subtilis* (Sysoeva et al., 2014). They suggested that the WXG motif – another characteristic motif of the EsxAB superfamily -- forms a bipartite signal in conjunction with the YXXXD motif (**Fig. 1.6**). The crystal structure of an EsxA homologue from *Streptococcus agalactiae* hinted that residues of the YXXXD secretion signal align on the same face of a stabilized helix when bound to the EccCb1 ATPase (Poulsen et al., 2014).

Alongside these structural studies, a number of groups attempted to determine the function of EsxAB using purified protein in various types of biochemical and cell-based assays. Does the EsxAB complex rupture membranes independent from the rest of the ESX-1 system? An early study showed that purified EsxA forms membrane pores in a planar lipid bilayer experiment (Hsu et al., 2003), however these experiments were not independently validated. A separate group used liposome flotation assays, surface plasmon resonance experiments, and cryo-EM to show that at low pH (such as the one encountered by *Mtb* in the phagolysosome), EsxA dissociates from EsxB and destroys liposomes (de Jonge et al., 2007). Furthermore, treatment of whole cells with purified EsxA creates pores that allows dye to permeate the cell (Derrick and Morris, 2007). Purified EsxA also causes various cells to lyse *in vitro*, including pneumocytes (Kinhikar et al., 2010) and red blood cells (Smith et al., 2008). Moreover, heterologous expression of EsxA in amoeba helps  $\Delta$ RD1 *Mycobacterium marinum* strains escape from phagosomes and proceed with non-lytic ejection from cells, again suggesting EsxA alone is sufficient for membrane rupture (Hagedorn et al., 2009). Finally, Smith et al. measured the size of the pores created by EsxA using a polyethyleneglycol osmoprotection assay (Smith et al., 2008) and calculated a pore size of ~4.5 nm. Together, these studies suggest that EsxA indeed forms pores on its own.

However, others have argued against EsxA pore-forming activity. Membrane-spanning proteins typically have surface-exposed hydrophobic residues with electropositive patches adjacent to an aromatic girdle; Renshaw et al. pointed out that EsxAB lacks such features (Renshaw et al., 2005). Moreover, EsxAB does not aggregate at concentrations over 2 mM in solution, an unusual observation for a pore-forming complex (Renshaw et al., 2005).

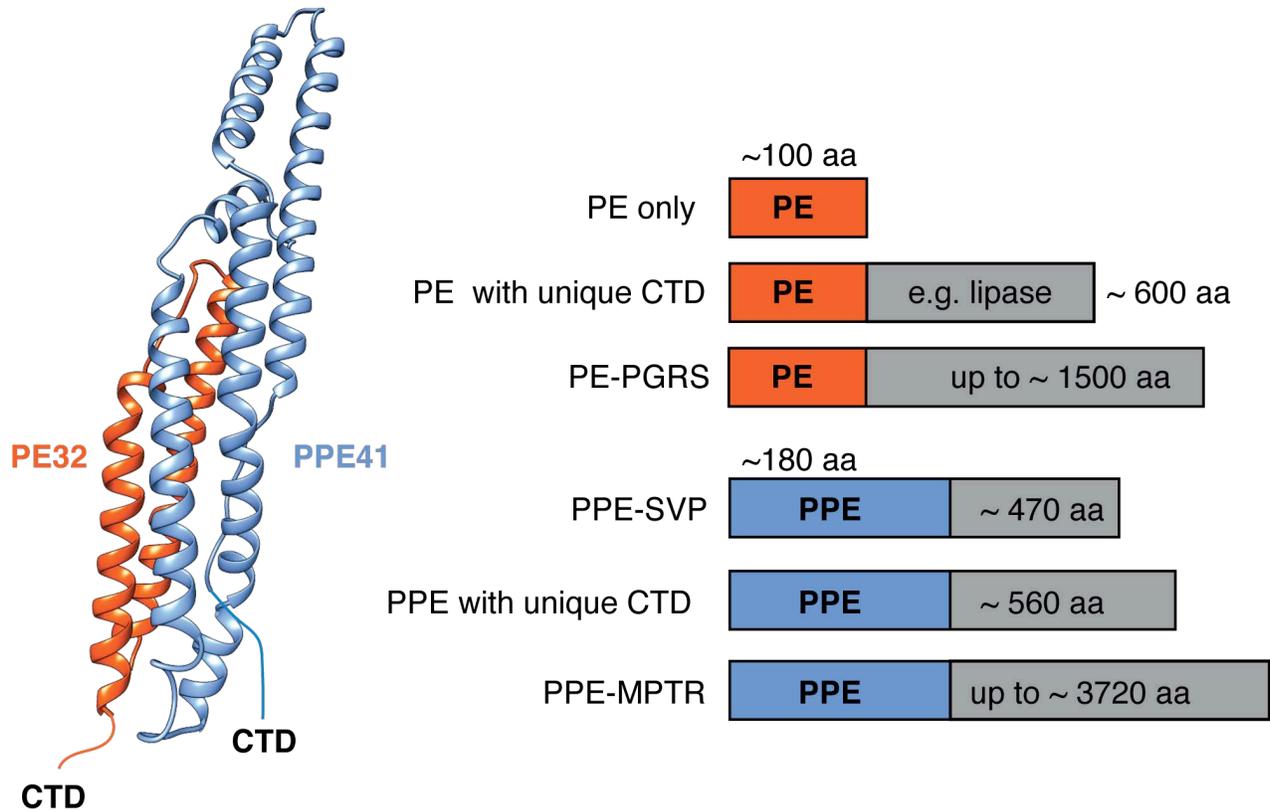
The group also showed that purified EsxAB does not disrupt macrophages membranes (Renshaw et al., 2005) and that EsxA does not dissociate from EsxB even at extremely low pH (Lightbody et al., 2008), refuting the earlier finding (de Jonge et al., 2007). Therefore, although there is evidence that EsxA forms pores, the issue is controversial and the molecular mechanism for how would occur has not been demonstrated.

Along with its proposed involvement in membrane permeabilization, EsxA is also suspected to function as an effector that manipulates host intracellular signalling pathways. Many distinct EsxA effector functions have been reported: it was found to interact with a regulator of intracellular traffic called syntenin-1 (Schumann et al., 2006), modulate phagosome maturation (Tan et al., 2006), and induce apoptosis through activation of caspase activity (Derrick and Morris, 2007). It was also observed to dampen lipopolysaccharide-induced TLR signalling pathways through modulation of ERK1/2 MAP kinases (Ganguly et al., 2007), decrease production of reactive oxygen species, and inhibit NFkB DNA-binding (Ganguly et al., 2008). EsxA was further observed to interact with TLR2 to inhibit the MyD88 signalling scaffold assembly. The C-terminal region of EsxA was deemed necessary for some these functions and was shown to bind host proteins directly (Pathak et al., 2007). Cumulatively, these effects are believed to help *Mtb* manipulate the infected macrophage (Samten et al., 2011).

How can a small 6-kDa protein interact with multiple host targets in addition to its role in permeabilizing membranes? One caveat of many of these studies is that they were carried out with purified EsxAB, separate from the rest of the secretion system complex. This may not accurately represent EsxAB's true physiological environment. An alternative possibility is that EsxA and EsxB are instead extracellular components of the secretion system that facilitate export of other effectors that are responsible for the ESX-dependent phenotypes that have been observed (Pallen, 2002). This hypothesis is more consistent with the observation that ESX systems are encoded in a diverse range of non-parasitic Gram-positive bacteria that presumably have no need to permeabilize phagosomal membranes or modulate the eukaryotic immune system. In this scenario, non-pathogens would use ESX to secrete proteins that are unrelated to virulence.

### 1.3.7 The PE and PPE protein families

There are two other ESX-secreted protein families that have received considerable attention. These have been termed the PE and PPE families for their characteristic Pro-Glu (PE) and Pro-Pro-Glu (PPE) motifs in their N-terminal regions, respectively (Sampson, 2011). PE and PPE paralogues together account for nearly 200 open reading frames and 10% of the coding capacity of *Mtb* - a finding that was one of the major surprises following the sequencing of the *Mtb* genome (Cole et al., 1998). Some members of this family appear to interact with host cellular targets to modulate the immune response (Choudhary et al., 2003; Demangel et al., 2004; Li et al., 2005; Mishra et al., 2008; Nair et al., 2009; Ramakrishnan et al., 2000). Many PE and PPE proteins are predicted to have large, disordered C-terminal domains (CTDs) that lack structural complexity. These CTDs are hypothesized to promote antigenic variation at the cell surface and may function similarly to proteins used by the Epstein Barr virus to confuse the immune system (Cole et al., 1998). There are also PE and PPE proteins with CTDs predicted to adopt enzymatic domains such as lipases (Daleke et al., 2011) and hydrolases (Sultana et al., 2011). The different types of PE and PPE family members are illustrated in **Figure 1.7**.



**Figure 1.7. Structure of the PE/PPE heterodimer.** PDB: 2G38 (Strong et al., 2006). Schematic adapted from (Sampson, 2011). **PGRS**: polymorphic GC-rich sequence. **SVP**: refers to a characteristic Ser-Val-Pro motif. **MPTR**: major polymorphic tandem repeat.

Aside from the lone PE/PPE pair encoded within the ESX-1 gene cluster which is secreted through ESX-1, most of the PE and PPE proteins appear to be secreted through a dedicated ESX-5 secretion system. ESX-5 is necessary for growth of pathogenic mycobacteria, and is hypothesized to form a transport system that helps mycobacteria acquire nutrients (Ates et al., 2015). The PE and PPE domains are required for their CTD cargo to reach the cell wall (Cascioferro et al., 2007; Sampson et al., 2001; Sani et al., 2010). Some CTDs are proteolytically cleaved following export, possibly by the ESX-5-associated MycP5 protease (Daleke et al., 2011). Although most members of the PE and PPE families are notoriously difficult to express and purify (Riley et al., 2008), PE25 and PPE41 were successfully purified and shown to adopt a heterodimeric complex (Strong et

al., 2006). Crystallographic analysis of the PE25/PPE41 complex revealed a helical bundle reminiscent of the EsxAB heterodimer (**Fig. 1.7**) (Strong et al., 2006).

### **1.3.8 Secreted Esp proteins of the ESX-1 locus**

Along with the EsxAB and PE/PPE families, there are a number of other proteins secreted through ESX-1 that are also required for virulence. While the literature is often conflicted regarding which Esp proteins are secreted, a complete list of proteins with strong evidence of secretion through ESX-1 include EspA (Fortune et al., 2005), EspB (Gao et al., 2004; McLaughlin et al., 2007; Xu et al., 2007), EspC (Millington et al., 2011), EspE (Sani et al., 2010), EspF (Sani et al., 2010), EspJ (Champion et al., 2014), EspK (Champion et al., 2014; Sani et al., 2010), PE32, and PPE68 (Chen et al., 2013b). Establishing that a protein is secreted is complicated by the high propensity for mycobacteria to lyse in culture, requiring strict experimental controls. The western blot is the most common method used to detect ESX-secreted proteins. However, more powerful mass spectrometry-based methods have been developed and employed to detect secreted proteins located specifically in the cell wall (Sani et al., 2010) or present in whole colonies (Champion et al., 2012). A promising new approach uses mass spectrometry-based multiple reaction monitoring (MRM) as a replacement for western blots, allowing for sensitive simultaneous detection of all known secreted substrates and identification of new ones (Champion et al., 2014).

Many ESX-1-secreted proteins require each other for secretion, making it difficult to determine their isolated functions (Fortune et al., 2005). However, it is an important topic, as these proteins are located outside the cell and are positioned to interact with the bacteria's environment. Particular secreted proteins may endow the five ESX systems with unique functionality. It is unclear if ESX-1-secreted proteins have separate, distinct effector functions, or if they operate as a complex. Experiments aimed at determining the co-secretion requirements, effects on virulence, and binding are complicated and sometimes conflicting. A schematic attempts to summarize the results (**Fig. 1.8**).

The first protein identified as uniquely secreted through ESX-1 was EspA (Fortune et al., 2005). EspA not part of the ESX-1 gene cluster; rather, it is located in an upstream operon with EspC and EspD (Fortune et al., 2005; MacGurn et al., 2005). However, EspAC share sequence similarity with EspEF in the ESX-1 locus (MacGurn et al., 2005). The EspACD operon appears to be specific to pathogenic mycobacterial species and is required for EsxAB secretion, and EsxAB is required for EspAC secretion. *Mtb-ΔespA* mutants have similar virulence phenotypes as *Mtb-ΔRD1* mutants (Fortune et al., 2005; MacGurn et al., 2005). EspA has been shown to bind directly to an EsxAB fusion in a yeast two-hybrid assay (Callahan et al., 2009). The core ESX-1 EccC ATPase is required for EspA secretion, but interestingly, so is EspI, a homologue to a chromosomal segregation ATPase (Fortune et al., 2005). One intriguing aspect of EspA is that it appears to adopt a homodimer following secretion, mediated by a disulfide bond; the homodimer is required for *Mtb* virulence (Garces et al., 2010). EspA's mutually-dependent secretion requirements led to speculation that it is a component of the secretion apparatus, since deletion of either EsxAB or EspA reduces apparatus integrity. It is speculated that EspA may carry out a role similar to LcrV from the *Yersinia* type III secretion system, which simultaneously acts as an apparatus component and an effector (Fortune et al., 2005).

Does EspA have a direct impact on virulence, or is the loss of virulence a result of disrupted EsxAB secretion? To distinguish between these possibilities, two groups made attempts to decouple co-dependent secretion requirements by making specific point mutations in EspA. The first group mutated EspA's lone cysteine that is involved in dimerization; the mutant still secretes EsxAB but nonetheless reduces *Mtb* ability to cause disease in mouse models (Garces et al., 2010). The mutation also appeared to affect cell wall integrity in an SDS-susceptibility test, although this finding was later disputed (Chen et al., 2013a). Mutating residues surrounding EspA's WXX motif results in loss of EspA and EsxAB secretion, but for some reason does not attenuate *Mtb* virulence, providing the first ever report where EsxB secretion was unlinked to virulence (Chen et al., 2013a). This suggests that other virulence factors may be capable of compensating for EsxAB and EspA (Chen et al., 2013a). However, the exact function of EspA remains undetermined.

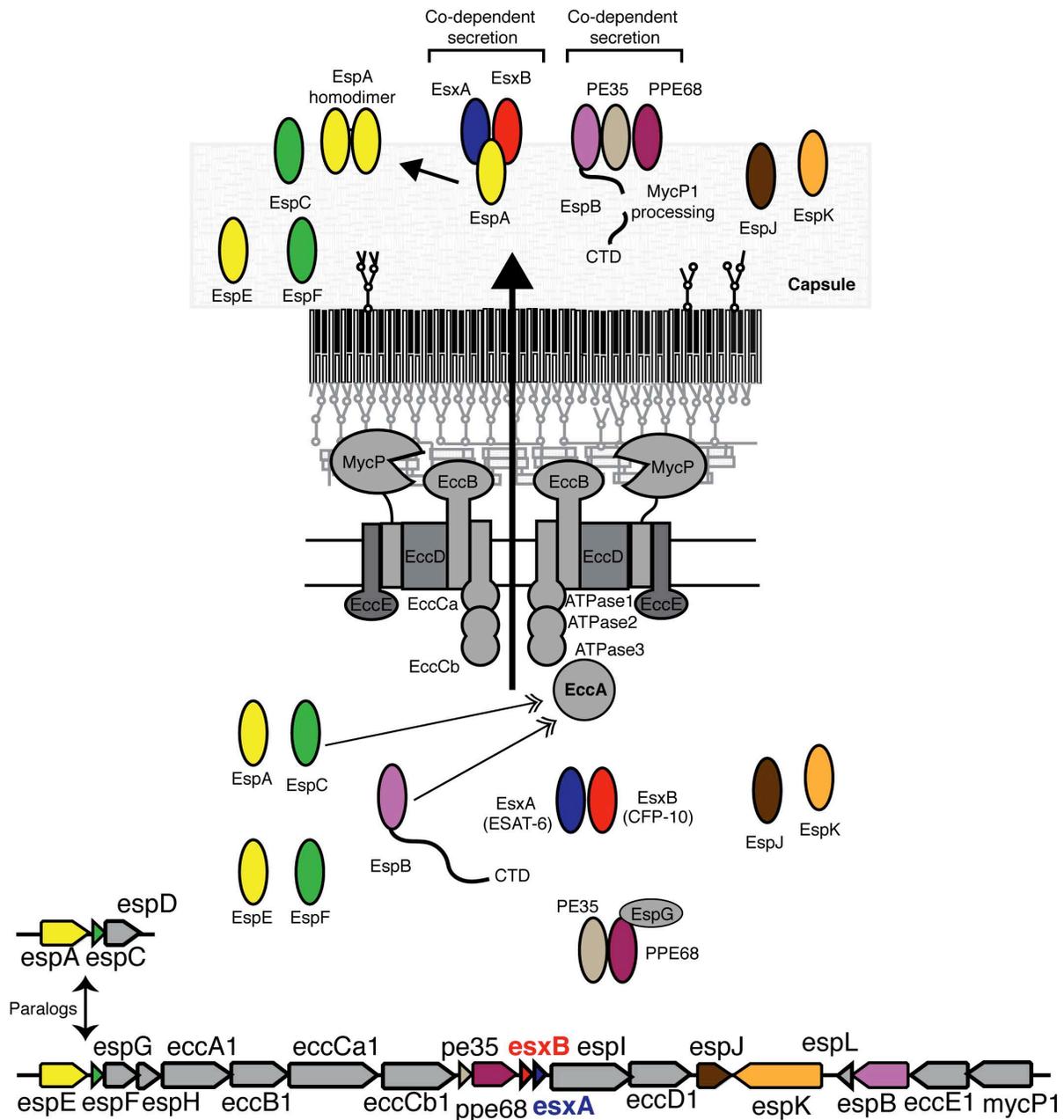


Figure 1.8. Secreted proteins from the ESX-1 locus

*EspA*'s operon partner *EspC* is also secreted and is required for *Mtb* virulence. In addition, studies related to *EspC* serve to reiterate the importance of studying ESX-1 secretion from an immunological perspective. Millington et al. assessed the immune response to *EspC* in patients with active TB using ELISA and flow cytometry (Millington et

al., 2011). They found that EspC is broadly recognized by CD4<sup>+</sup> and CD8<sup>+</sup> T cells that secreted IFN- $\gamma$  and IL-2, which are promising cytokine markers of vaccine potential. EspC is at least as immunodominant as EsxAB in patients with both latent and active TB, and is not recognized by BCG-vaccinated subjects, allowing infected individuals to be differentiated from vaccinated ones. For these reasons, EspC is a potential vaccine candidate or immunodiagnostic (Millington et al., 2011). The exact mechanism by which EspA and EspC are secreted is not known, although it has been demonstrated EspC has a secretion signal similar to EsxB but is targeted by the EccA ATPase (Champion et al., 2009).

The fourth ESX-1-secreted protein to be identified was EspB (McLaughlin et al., 2007). A *M. marinum* EspB transposon mutant is one of the most attenuated RD1-related mutants in zebrafish infection models, is unable to cause contact-dependent haemolysis of red blood cells (Gao et al., 2004), and is not cytotoxic in macrophage infection assays (Gao et al., 2004). EspB's importance to virulence was recapitulated in *Mtb* macrophage infections (McLaughlin et al., 2007). As was the case with  $\Delta$ espA mutants,  $\Delta$ espB mutants failed to secrete EsxAB, and EsxAB mutants do not secrete EspB, suggesting these substrates require each other for secretion (Chen et al., 2013b; Smith et al., 2008; Xu et al., 2007). However, this finding was subsequently disputed (Chen et al., 2013b). Also like EspA, EspB appears to require the EccA ATPase for export (McLaughlin et al., 2007; Xu et al., 2007). Additionally, EspB has its own unique requirements for export, which appear to include EspK (McLaughlin et al., 2007). Putative chaperones EspG and EspH may also be required for EspB secretion (Xu et al., 2007), although these proteins were not required in a separate study (McLaughlin et al., 2007). Recently, a dependence on EspL has been reported for EspB secretion (Champion et al., 2014), but a separate group reported EspL is not required (Stoop et al., 2011). Notably EspACD secretion is not required for EspB secretion (Chen et al., 2013b). Overall, the documented requirements for EspB secretion have been somewhat conflicted, but it is agreed this protein plays a critical role in *Mtb* virulence.

EspB is a unique ESX-1 substrate because it undergoes C-terminal processing during secretion (McLaughlin et al., 2007; Xu et al., 2007). Only its N-terminal cleavage product

can be located in the supernatant; the C-terminal fragment is not detectable in the culture filtrate and may be degraded (McLaughlin et al., 2007). One group found that the C-terminus of EspB is required for co-dependent secretion with EsxAB and that the EspB C-terminus binds EsxAB directly, however others have not observed this (Chen et al., 2013b; McLaughlin et al., 2007). It was subsequently found that MycP1 is responsible for the cleavage event, though the exact site of cleavage was not well defined (Ohol et al., 2010). The purpose of this cleavage event is unknown.

As is the case with EspA, the specific molecular function of EspB has been difficult to deconvolute from the rest of the ESX-1 system. Several possibilities have been proposed. EspB was initially suggested to be an apparatus component due to its dependence on other secreted proteins for its own secretion. In this hypothesis, EspB may be a part of a larger multi-component extracellular structure whose assembly is disrupted when any one of its constituent building-blocks is absent (Gao et al., 2004; Xu et al., 2007). Another hypothesis was put forth by Ohol et al., who suggested EspB cleavage serves as a post-translational control point for fine-tuning the quantity of ESX-secreted proteins (Ohol et al., 2010). This may be important for preventing over-stimulation of cytosolic surveillance as a result of too much highly antigenic secreted protein. A third hypothesis for EspB function is based on data provided by Chen et al. that indicates EspB can operate as an effector separately from the rest of the apparatus, perhaps by modulating the immune system directly (Chen et al., 2013b). These authors noted some residual cytotoxic activity in  $\Delta$ espA *Mtb* strains and attributed this to EspB, because the residual cytotoxicity could be reduced by adding EspB-specific antibodies that presumably interfered with functional molecular surfaces. Moreover, EsxAB secretion-deficient strains can be made more cytotoxic through exogenous addition of purified EspB (Chen et al., 2013b). EspB was also shown to bind the eukaryotic phospholipids phosphatidylserine and phosphatidic acid, but only in its processed form, indicating that the presence of the C-terminal domain may determine EspB's ability to bind to host membranes (Chen et al., 2013b).

The remaining ESX-1-secreted proteins to discuss are EspJ, EspK, and the lone PE/PPE pair encoded by the ESX-1 locus.  $\Delta$ espJ or  $\Delta$ espK mutants result in intermediate levels of EsxAB secretion, indicating that they are not strictly necessary for virulence and not all

ESX-1 disruptions have identical effects (Champion et al., 2014). Likewise, disruption of EspB or possible chaperones EspI and EspG result in intermediate levels of secreted EspJ and EspK (Champion et al., 2014). PE35 and PPE68 are also secreted, and a  $\Delta pe35$  mutant grows poorly in macrophages and is less cytotoxic (Chen et al., 2013b). PE35 secretion is required for secretion of EsxAB, EspA, and EspB (Chen et al., 2013b). Thus the lone PE/PPE pair of the ESX-1 systems appear to be quite important. A big question is: what is the relationship between this PE/PPE pair and the other 180 PE/PPE proteins in the *Mtb* genome?

To summarize, there has been significant work from many groups aimed at characterizing the proteins secreted through ESX-1. These studies have tried to address several non-trivial, tightly coupled questions: (1) what proteins are definitively secreted, (2) does blocking secretion of one protein block secretion of other proteins, (3) which secreted proteins interact with each other, (4) which secreted proteins are dependent on chaperones, (5) by what basis are they targeted for secretion, (6) what are the direct functions of secreted proteins when they get outside the *Mtb* cell, and (7) how and why are certain secreted proteins such EspB proteolytically processed during secretion?

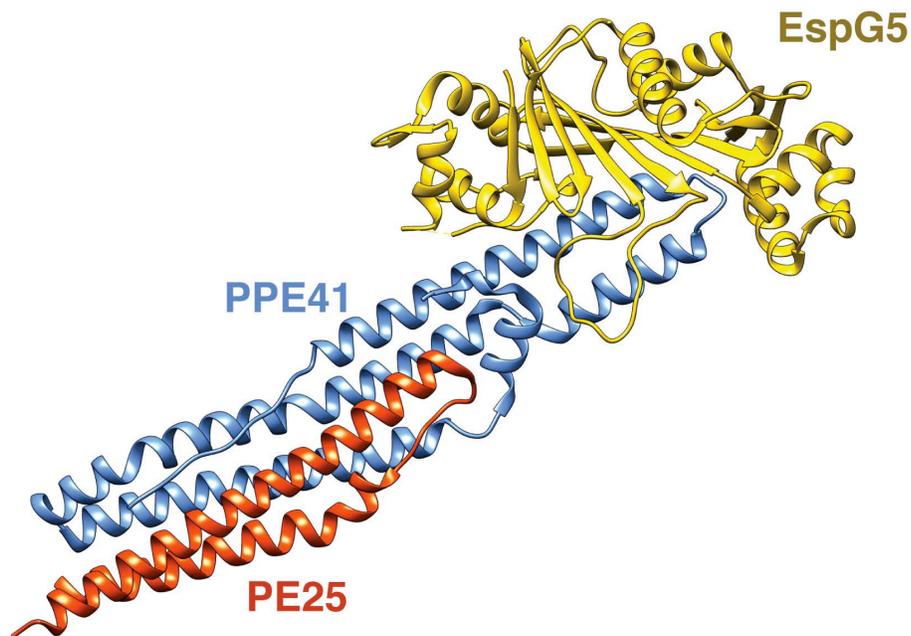
### **1.3.9 Type VII secretion chaperones**

In many secretion systems, chaperones are required to prevent premature self-polymerization, guide substrates, regulate secretion hierarchy, and keep substrates in a secretion-competent state (Korotkova et al., 2014). ESX systems have one such chaperone, EspG, which is required for virulence but is not necessary for EsxAB secretion (Bottai et al., 2010). Instead, EspG5 interacts in a 1:1:1 stoichiometry with PE/PPE proteins secreted through the ESX-5 system and is required for their secretion/stability (Daleke et al., 2012b; Teutschbein et al., 2009). There is some controversy as to whether or not EspG5 itself is secreted (Bottai et al., 2010; Chen et al., 2011; Daleke et al., 2012b).

An exciting breakthrough came with crystal structures of the EspG5/PE25/PPE41 ternary complex (Ekiert and Cox, 2014; Korotkova et al., 2014). The structure and accompanying functional analyses added support for EspG5's role as a chaperone. EspG5

binds to one end of the elongated PE25/PPE41 heterodimer, leaving the secretion signals on the other end of the complex free to interact with the ESX secretion machinery. Mutating residues hypothesized to confer chaperone-substrate specificity surprisingly does not abrogate binding, indicating that sequence variability in the chaperone is not solely responsible for dictating substrate recognition (Korotkova et al., 2014). EspG5 plays a critical role in keeping aggregation-prone PE/PPE substrates in solution; a specific set of hydrophobic residues on the PPE tip promote aggregation (Korotkova et al., 2014).

The role of EspG-like chaperones in mediating stability of other secreted proteins from the ESX-1 locus remains to be investigated, but EspD is required for EspAC stability (Chen et al., 2011). Korotkova et al. noted EspG's similarity to the SycE/SycO chaperones from the *Yersinia pestis* type III secretion system, which mask aggregation-prone regions on hydrophobic membrane domains necessary for effector localization within the host (Korotkova et al., 2014).



**Figure 1.9. PE25/PPE41 in complex with the EspG5 chaperone, PDB: 4KXR.** (Korotkova et al., 2014)

### 1.3.10 ESX-1 involvement in DNA conjugation

Finally, we discuss ESX-1's intriguing role in DNA-conjugation. An ESX-1 gene cluster closely resembling those of pathogenic bacteria is present in the non-pathogenic fast-growing species *M. smegmatis*. However in this organism, ESX-1 is involved in DNA conjugation rather than virulence (Converse and Cox, 2005). The DNA transfer phenomenon was first observed to occur between different isolates of *M. smegmatis* in the 1970s (Tokunaga et al., 1973). Years later, Parsons and Derbyshire revived the research by creating a collection of chromosomally marked antibiotic resistant strains that were used to determine the conditions under which DNA transfer occurs (Parsons et al., 1998). They observed that extensive segments of donor chromosome were unidirectionally transferred from mc<sup>2</sup>155 donor strains to MKD8 recipient strains (Parsons et al., 1998). The genetic factors that distinguish donor and recipient strains are not entirely clear. Transfer requires extended cell-cell contact, is insensitive to DNase treatment, and does not involve mycobacteriophage transduction. Transfer also requires multiple cis-acting DNA sequences termed *bom* which are found scattered around the genome. The conjugal process also involves gap repair, indicating that the DNA transferred is linear (Wang et al., 2003). The mechanism is clearly different from classic Hfr-mediated chromosomal transfer (Wang et al., 2005), creating a diversely blended genome comprised of DNA from both parental strains in a manner that is reminiscent of meiotic recombination seen in sexual organisms (Gray et al., 2013).

To identify trans-acting factors involved in the conjugation process, transposon insertion mutagenesis experiments were carried out, first with the donor strain. It was discovered that strains with transposon insertions in ESX-1 ORFs display increased rates of chromosomal transfer (Flint et al., 2004). This finding was opposite to what would be expected if the donor transfers DNA directly through the ESX-1 secretion system. It was hypothesized that ESX-1 secretes factors that reduce conjugation, perhaps coating the surface with proteins that interfere with cell-cell contacts (Flint et al., 2004). Strikingly, transforming a plasmid encoding the *Mtb* ESX-1 virulence gene cluster into an *M.*

*smegmatis* ESX-1 mutant restores DNA conjugation to wild type levels, indicating that the two systems are functionally equivalent (Flint et al., 2004).

The story became more complicated when the transposon mutagenesis experiment was carried out with the recipient strain. As seen with the donor, transposons inserted into ESX-1 ORFs impacted conjugation rates. However, the mutant strains exhibit *reduced* rates of chromosomal transfer, indicating that ESX-1 has a different role in facilitating conjugation depending on whether it is carried by the donor or recipient strain (Coros et al., 2008). While its precise role in conjugation remains unclear, the *M. smegmatis* ESX-1 system offers an valuable opportunity to study ESX function in an organism that is less cumbersome to manipulate than *Mtb*. Importantly, any attempt to explain ESX-1's role in *Mtb* virulence must be compatible with its conservation in non-pathogenic species like *M. smegmatis* as well as its involvement in DNA conjugation.

T7SS-mediated DNA conjugation also appears to occur between slow-growing pathogenic mycobacteria. A recent study identified a plasmid named pRAW that encodes a predicted T7SS along with components of a type IV secretion system (T4SS) (Ummels et al., 2014). Intriguingly, pRAW also encodes the T4SS components VirB4, VirD4, and VirB8, but lacks the outer membrane proteins typically associated with T4SS. The T7SS encoded by pRAW is most closely related to the ESX-5 system. The plasmid was found to mediate conjugal DNA transfer between slow growing mycobacteria, and transfer is dependent on both VirB4 and EccC. It was speculated that either the two secretion systems function separately (for example, T4SS transfers DNA, while T7SS mediates cell-cell contacts possibly via biofilms), or that they function jointly as a single entity brought together through an undefined mechanism (Ummels et al., 2014).

## 1.4 Research scope

The research here investigates the relationship between one of the ESX-1 apparatus proteins, MycP1, and the secreted protein that it processes, EspB. Using mass spectrometry and N-terminal sequencing, MycP1 is shown to cleave two EspB variants from *Mtb* and *M. smegmatis* at a conserved cleavage site *in vitro*. The x-ray crystallographic structure of *M. smegmatis* MycP1 reveals a subtilisin-like fold, and a comparative analysis with other subtilisin-like enzymes highlights MycP1's unique structural features. The crystallized construct includes the putative N-terminal extension previously thought to be removed by autoproteolysis, an event believed to be required for MycP1 to mature from an inactive to an active state. However, our results demonstrate this is unlikely to be the case, as the N-terminal extension shows no structural similarity to previously characterized protease propeptides, instead wrapping intimately around the catalytic domain. There is no indication that the extension is removed, and the protease is active with the N-terminal extension intact. A series of MycP1 truncation mutants suggests the N-terminal extension plays a role in stability, perhaps in a manner analogous to stabilizing metal ions found in other subtilisin-like enzymes. Finally, one of the identified EspB cleavage sites is modelled into the MycP1 active site to generate a plausible mode of cleavage site recognition.

Chapter 3 takes a closer look at the secreted protein, EspB. Its x-ray crystallographic structure is determined, revealing a fused PE/PPE-homology domain with a structured ESX secretion signal that is similar to those seen in other ESX-exported proteins, indicating EspB is secreted through the same pathway as other model substrates. Strikingly, the PE/PPE-homology domain of EspB mediates oligomerization, forming a barrel-shaped heptamer with a central pore. Two-dimensional electron microscopy averages are obtained for full-length and truncated EspB variants. A comparison of the two sets of averages hint that MycP1 cleavage modifies the ultrastructure of EspB. An electron microscopy reconstruction of the EspB multimer is generated, and a model of the oligomer is fit to the density using symmetric Rosetta docking. The final model is supported by mass spectrometry-based detection of chemically cross-linked peptides from adjacent subunits. These findings provide insight into how the specialized Esp proteins encoded within the

ESX-1 locus are targeted for secretion, and for the first time indicate an oligomerization-dependent role for Esp virulence factors.

In Chapter 4, the structure of one of the predicted periplasmic ESX apparatus proteins, EccB1 is determined. EccB1 is one of three ESX membrane apparatus proteins predicted to have globular domains located in the periplasm-like space. EccB adopts a novel elongated chair-like fold with 2-fold internal pseudo-symmetry that can be divided into four  $\alpha/\beta$  subdomains and a central beta sheet. These  $\alpha/\beta$  units are structurally homologous to the Gram-positive phage lysin protein PlyCb, suggesting a related cell-wall binding function.

# Chapter 2: Structure of a Mycosin Protease

## 2.1 Introduction

MycP1 is essential for ESX-1 secretion in pathogenic *Mtb* and for efficient DNA conjugation in the avirulent saprophyte *M. smegmatis* (Coros et al., 2008; Ohol et al., 2010). MycP1 is one of six conserved components found in all five T7SS. Although a recent compositional analysis of the ESX-5 T7SS inner membrane complex did not identify MycP5 as part of the assembly (Houben et al., 2012b), knocking out MycP1 abrogates ESX-1 protein secretion, indicating its presence is essential (Converse and Cox, 2005; Ohol et al., 2010).

Proteins exported by the T7SS have been detected in culture filtrates and in cell wall fractions as lower molecular weight/proteolytically processed species when compared to those which are not exported (Abdallah et al., 2009; Daleke et al., 2011; McLaughlin et al., 2007; Ohol et al., 2010; Xu et al., 2007). For example, LipY, a substrate of the ESX-5 system, is cleaved between its N-terminal PE domain and its C-terminal lipase domain following secretion (Daleke et al., 2011). Mycosins may function in removing these PE/PPE “leader” domains following (or concomitantly with) secretion (Daleke et al., 2011). Ohol and coworkers provided the first evidence of MycP1<sub>tb</sub> proteolytic activity against a T7SS substrate by complementing a  $\Delta mycP1_{tb}$  knockout strain with a mutant in which the predicted serine nucleophile of the enzyme was changed to alanine (MycP1<sub>tb</sub>-S332A)

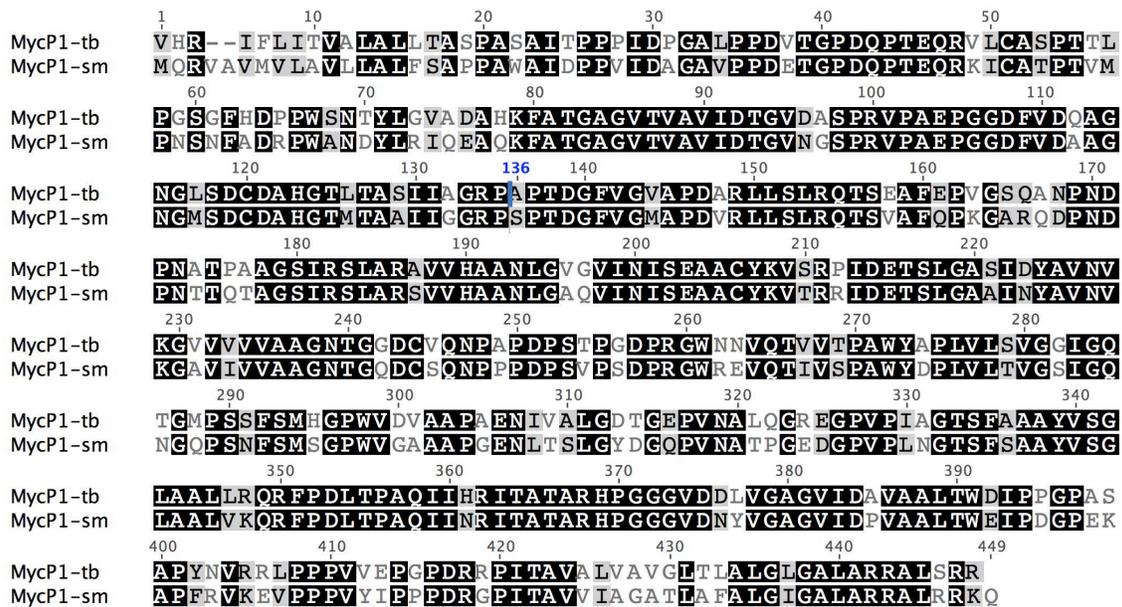
(Ohol et al., 2010). This catalytically impaired mutant resulted in an altered cleavage pattern of the C-terminal portion of the secreted protein EspB<sub>tb</sub> that suggested native MycP1<sub>tb</sub> mediates cleavage at two sites (Ohol et al., 2010). Intriguingly, the mutant strain also showed a 2-fold increase in secretion of known ESX-1 substrates as well as reduced virulence in mice, leading to the hypothesis that MycP1 protease activity negatively regulates secretion via EspB (Ohol et al., 2010). While several groups have documented secretion-coupled cleavage of EspB's C-terminus (McLaughlin et al., 2007; Ohol et al., 2010; Xu et al., 2007), data showing the exact residue at which proteolysis occurs has not been provided.

Proteases are often synthesized with propeptide segments that act as intramolecular chaperones to promote folding, as well as regulate activity to prevent untimely and potentially destructive proteolysis in the incorrect cellular compartment (Khan and James, 1998). These propeptides range from short dipeptides to independent domains, and are often proteolytically processed and removed as the enzyme matures (Khan and James, 1998). Sequence analysis (Brown et al., 2000; Dave et al., 2002) of the MycP family indicates they have an N-terminal Sec signal sequence followed by a 40 amino acid stretch that, despite showing no sequence homology to propeptides of other subtilisin-like proteases, has previously been proposed to operate as a propeptide (Dave et al., 2002; Ohol et al., 2010). Following this is a predicted subtilisin domain which is attached to a predicted C-terminal transmembrane anchor by a ~ 30 amino acid proline-rich linker.

It is unclear what maturation events lead to MycP activity. In one study, MycP1<sub>tb</sub> expressed in infected macrophages was detected at a lower apparent molecular weight after six weeks when compared to samples taken from initial cell cultures; it was suggested this electrophoretic shift was related to loss of the putative propeptide (Dave et al., 2002). Recently, Ohol et al reported that a recombinant maltose binding protein (MBP) fusion construct of *M. smegmatis* MycP1<sub>sm-[24-407]</sub> is initially inactive (Ohol et al., 2010). However, extended sample aging with Factor Xa resulted in a shortened species as determined by SDS-PAGE relative to the MBP fusion as well as an apparent gain in promiscuous proteolytic activity against a range of fluorogenic tetrapeptide substrates

(Ohol et al., 2010). The authors suggested the shortened species was a result of loss of the proposed prodomain.

Toward clarifying these issues pertaining to MycP1 maturation in relation to subsequent proteolytic activity, we present the first crystallographic analysis of a mycosin protease, including the proposed propeptide region (which we will herein refer to as the “N-terminal extension”). MycP1<sub>sm</sub> from *M. smegmatis* shares 72% sequence identity to its counterpart in *Mtb* (**Figure 2.1**), making it a suitable model system for studying mycosins of pathogenic mycobacteria. Our structure shows that the N-terminal extension is not a conventional subtilisin foldase, but rather an extended proline-rich, disulfide-tethered appendage that interacts extensively with the subtilisin core. The N-terminal extension of MycP1 does not occupy the active site cleft in a manner typical of previously characterized propeptides, suggesting regulation of the catalytic domain through a novel mechanism. Further, we show that MycP1<sub>sm</sub> with the N-terminal extension cleaves the C-terminal region of purified EspB from both *M. smegmatis* and *Mtb* and determine the conserved cut site using mass spectrometry and N-terminal sequencing.



**Figure 2.1.** Pairwise sequence alignment of *Mtb* and *M. smegmatis* MycP1 orthologues.

## 2.2 Methods

### 2.2.1 Expression and purification

The full MycP1<sub>sm</sub> gene was synthesized with codons optimized for expression in *E. coli* (Biobasic). Residues 24-407 were cloned into pET28a(+) using the restriction-free method (van den Ent and Lowe, 2006) and transformed into *E. coli* BL21 codon-plus RIPL competent cells (Stratagene) for use in protein expression. BL21 *E. coli* cells expressing native His-tagged (MGSSHHHHHHSSGLVPRGSH) MycP1<sub>sm</sub>-[24-407] were grown as described (Graslund et al., 2008), and SeMet derivative MycP1<sub>sm</sub>-[24-407] were grown in minimal media as described (Larrson, 2009). Induction was initiated by adding 1 mM IPTG at OD<sub>600</sub> = 0.6 followed by growth for 20 hours. After harvesting the cells by centrifugation, the pellets were resuspended in 2x lysis buffer (100mM HEPES pH 7.5, 1M NaCl, 20 mM imidazole, and 1 mM TCEP) and stored at -80 °C until further use. When needed, the pellet was thawed and diluted 3x with IMAC buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 0.5 mM TCEP) (16) and lysed using an Avisten C5 homogenizer. Lysate was spun at 40,000 RPM for 1 hour and the supernatant loaded onto a 1 ml HisPur IMAC column (Thermo, exae#88225) using an AKTA purifier FPLC system, washed with 10 CV of 25 mM imidazole, and eluted with 8 column volumes 500 mM imidazole without subsequent cleavage of the His-tag. Eluate was immediately concentrated and injected onto a Superdex 75 XK 26/60 column equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl, and 1 mM TCEP, and the fractions containing the centre of the main UV peak immediately concentrated to 20 mg/ml and set up for crystallization. The N-terminal deletion constructs were created using the Quick Change protocol (Stratagene). Small scale protein purification of the MycP1 truncation mutants were carried out as previously described (Graslund et al., 2008).

### 2.2.2 Crystallization and x-ray crystallographic analysis

MycP1<sub>sm</sub>[24-407] crystals grew from 0.2M NaF, 25% PEG 3350 and reached maximum size after 1 week. Crystals were harvested and briefly washed in mother liquor containing

23% glycerol and rapidly cooled in liquid nitrogen before screening using a Rigaku RU200 rotating anode and Mar345db detector. SAD data was collected at the Canadian Light Source CMCF2 at a peak wavelength of 0.97895 Å and processed with imosflm and SCALA (Battye et al., 2011; Evans, 2006). Phenix AutoSol (Adams et al.) was used for phasing, with six selenium atoms found in the substructure solution (FOM = 0.37). The model was built and refined using Phenix and Coot (Emsley and Cowtan, 2004). Data collection and refinement statistics are given in Table 1. The resultant structure served as a molecular replacement search model to solve a second, higher resolution dataset using crystals grown in 0.1M Bis-Tris propane pH 6.77, 0.18M sodium thiocyanate, and 26% PEG 3350, which were cryoprotected as described above. No significant model changes were observed despite the differing space groups (I222 vs. P212121). The structure was analyzed and figures generated using UCSF Chimera (Pettersen et al., 2004). Active site cleft sizes for MycP1<sub>sm</sub>[24-407] and thermitase (1THM) were determined with DogSiteScorer (Volkamer et al., 2012).

### 2.2.3 Cleavage assay

Full length EspB<sub>tb</sub>[1-460] (Rv3881c) and EspB<sub>sm</sub>[1-520] (MSMEG\_0076) were cloned into pET28a(+) and N-terminally His-tagged (MGSSHHHHHSSGLVPRGSH). Both variants were expressed in *E. coli* BL21 codon-plus RIPL cells grown and processed as described above for MycP1<sub>sm</sub>[24-407]. Cells were harvested and lysed in 100 mM HEPES pH 7.5, 500 mM NaCl and purified over HisPur Ni-NTA resin (Thermo, #88223) and further purified using a Superdex 75 column with 20mM HEPES 7.5, 150 mM NaCl as the elution buffer. MycP1<sub>sm</sub>[24-407] used for activity assays was purified as above except the buffer used in all stages was 20 mM MES pH 6.0, 50 mM NaCl. MycP1<sub>sm</sub>[24-407] and EspB variants were incubated separately (for controls) or mixed together in 20 mM MES pH 6.0, 50 mM NaCl at a concentration of 0.2 µg/µl in 50 µl reaction volumes and incubated at 37 degrees. 10 µl aliquots were withdrawn at the indicated time points. The samples were analyzed using sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

#### **2.2.4 Determination of EspB cleavage products**

Protein bands were excised and in-gel digested with TrypsinGold (Promega) as described (Shevchenko et al., 1996). Eluted peptides were acidified with 0.5% formic acid and purified using C18 Stage tips (Rappsilber et al., 2007) before analysis on a capillary liquid chromatography system (LC Packings, Dionex) coupled to a quadrupole time-of-flight mass spectrometer (QSTAR XL, Applied Biosystems, operated by the UBC Centre for Blood Research Mass Spectrometry Suite). Peptides were separated on a column packed with Magic C18 resin (Michrom Bioresources) using a 0–80% gradient of organic phase over 90 min. MS data were acquired automatically using the Analyst QS software, v1.1 (Applied Biosystems). Peptides were identified from a protein database containing the *Mtb* proteome database (Uniprot) with appended *Mycobacterium smegmatis* MTB48, common laboratory contaminant sequences and a reverse sequence decoy database spectrum-to-sequence, using Mascot v.2.3 (Matrix Science). Search parameters included mass tolerances of 200 ppm for the parent ion and 0.4 Da for the fragment ions, trypsin or semi-trypsin as cleavage specificity with up to two missed cleavages, and carboxyamidomethylation of cysteine residues (+57.02 Da).

#### **2.2.5 Edman sequencing**

Separated EspB cleavage fragments were transferred from SDS-PAGE gel to a PVDF membrane and stained with 0.1% (w/v) Coomassie brilliant blue R-250, 40% methanol, 10% acetic acid for 5 minutes. The membrane was destained in 40% methanol/10% acetic acid and then rinsed for 30 seconds in 90% methanol, 5% acetic acid. The membrane was dried and EspB cleavage products were cut out. The first six amino acids were determined by Edman sequencing at the Advanced Protein Technology Centre within the Hospital for Sick Children, Toronto.

#### **2.2.6 Peptide docking**

MycP1 was aligned with subtilisin BPN' (2SNI) and the CI-2 chain merged into the MycP1 structure. Residues PVGTIVTMEYRIDR were changed to the identified EspB<sub>Sm</sub> cut

sequence DPSLGKPASAGGGGG and the remainder of CI-2 deleted. Hydrogen atoms and missing sidechains were added to the model which served as the input for FlexPepDock (London et al., 2011). 100 low resolution and 100 high resolution structures were generated and the peptide with the lowest energy/RMSD was analyzed.

### **2.2.7 Differential scanning fluorimetry**

DSF Experiments were carried out using 0.2 mg/ml protein in 100 mM MES, pH 6.0, 50 mM NaCl with Sypro Orange (5x final concentration, Invitrogen Cat. #S6650). The mixture was monitored in MicroAmp Fast optical reaction plates (Applied Biosystems Cat. #4346906) using a 25  $\mu$ l assay volume and an Applied Biosystems StepOnePlus RT-PCR system set to ROX (Ex: 488 nm, Em: 620 nm). Control experiments with no protein were carried out to ensure the observed  $T_m$  values were not dependent on the buffer.

### **2.2.8 Circular dichroism**

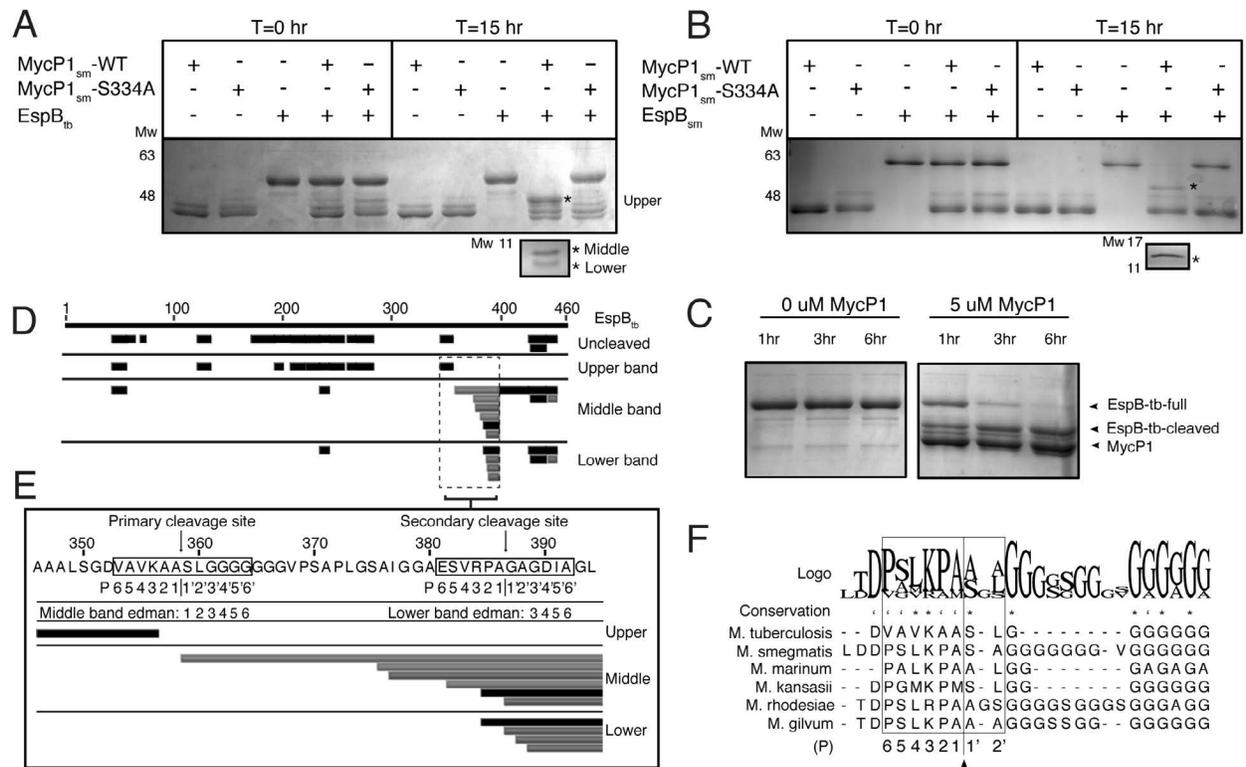
CD spectra were acquired using 5  $\mu$ M protein in 10 mM TRIS pH 7.5, 30 mM NaCl, 1 mM EDTA over wavelengths spanning 195 nm to 275 nm.

## **2.3 Results**

### **2.3.1 MycP1 cleaves EspB at two sites**

We first analyzed MycP1<sub>sm</sub> (residues 24-407, with predicted N-terminal signal sequence and C-terminal transmembrane helix deleted) by mass spectrometry ( $\pm$  3 Da resolution) and measured the molecular weight to be that of the expected construct, indicating the proposed propeptide is stably attached (data not shown). To confirm that MycP1<sub>sm</sub>-[24-407] cleaves *Mtb* EspB (EspB<sub>tb</sub>) *in vitro* as previously show (Ohol et al., 2010), we purified EspB<sub>tb</sub> and mixed it with MycP1<sub>sm</sub> as described in the methods followed by observation of cleavage products using SDS-PAGE analysis. EspB<sub>tb</sub> (consistently running as an ~55 kDa band) migrated at lower molecular weight (~48 kDa) after overnight incubation with MycP1<sub>sm</sub>[24-407] at 37°C; this truncated band was not observed when EspB<sub>tb</sub> was

incubated with the MycP1<sub>sm</sub>[24-407]-S334A catalytic mutant (**Fig. 2.2a**). A doublet of two additional closely spaced lower molecular-weight EspB<sub>tb</sub> bands were also produced (~10 kDa). We further tested and confirmed MycP1<sub>sm</sub> activity against purified EspB from *M. smegmatis* (EspB<sub>sm</sub>)(**Fig. 2.2b**). In contrast to the ~10 kDa doublet band observed for EspB<sub>tb</sub>, a single ~12 kDa band was produced when EspB<sub>sm</sub> was cleaved. At an equimolar ratio under the assay conditions tested, MycP1<sub>sm</sub>[24-407] required over six hours incubation to completely cleave either EspB variant (**Fig. 2.2c**). To determine where MycP1<sub>sm</sub>[24-407] cleaves EspB<sub>tb</sub>, the cleavage products, along with full-length EspB<sub>tb</sub> as a control, were excised from the gel and subjected to in-gel digestion by trypsin followed by LC-ESI-MS/MS analysis. Peptides from the large molecular weight band mapped to a discrete N-terminal region (Residues 1-356), while peptides from the smaller two bands mapped primarily to the C-terminal region (358-460), with the exception of two possible contaminating peptides (**Fig. 2.2d**). Perhaps the EspB<sub>tb</sub> variant has an extra cleavage site as a redundant measure to ensure cleavage occurs; EspB processing may be particularly important for the pathogen. A similar peptide distribution was also observed in the LC-MS/MS analysis of the EspB<sub>sm</sub> cleavage products (data not shown). Thus, MycP1<sub>sm</sub>[24-407] with its N-terminal extension intact cleaves two distinct EspB homologues in the C-terminal region.



**Figure 2.2 Identification of the MycP1 cleavage site of EspB** (a) SDS-PAGE gel showing MycP1<sub>sm</sub>[24-407] cleavage of EspB<sub>tb</sub>, with the three EspB<sub>tb</sub> cleavage products indicated by asterisks. (b) SDS-PAGE gel showing MycP1<sub>sm</sub>[24-407] cleavage of EspB<sub>sm</sub>, two cleavage products indicated by asterisks. (c) SDS-PAGE gel showing MycP1<sub>sm</sub>[24-407]-dependent EspB degradation over 6 hours. (d) Summary of LC-MS/MS analysis of uncleaved EspB<sub>tb</sub> in comparison to the three EspB<sub>tb</sub> cleavage products. Tryptic peptides (black) and semi-tryptic peptides in (grey) are mapped against the full EspB<sub>tb</sub> sequence (residues 1-460), indicated by the numbered upper black bar. (e) EspB<sub>tb</sub> sequence surrounding the proposed cleavage sites with LC-ESI-MS/MS, Edman sequencing results, and proposed P residues mapped. (f) Multiple sequence alignment of EspB homologues with the residues involved in conferring specificity in subtilisin-like enzymes indicated by a box.

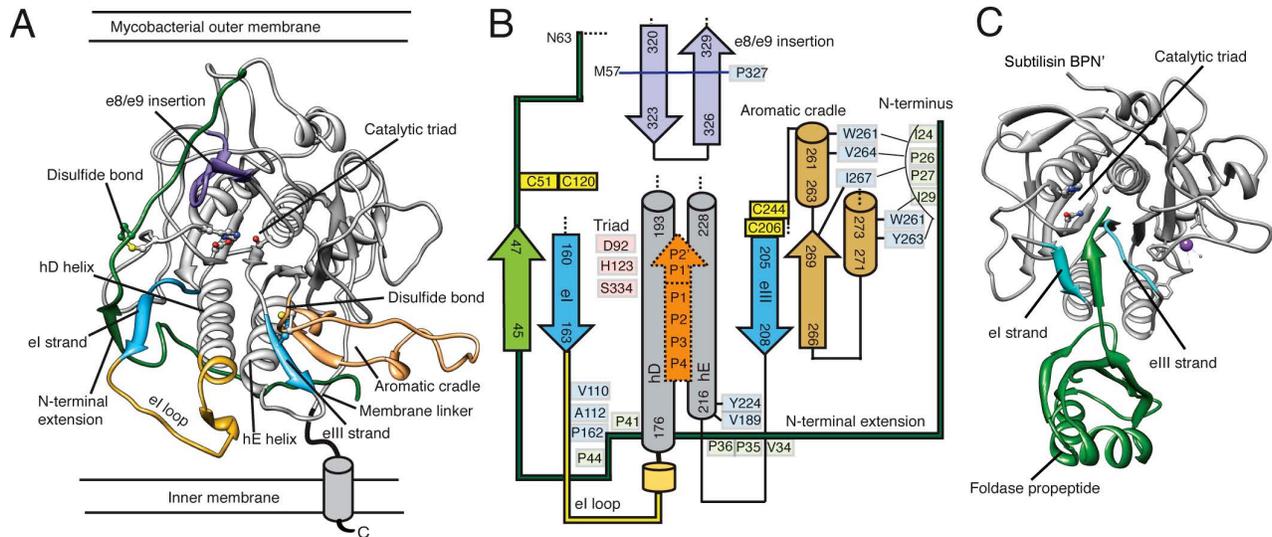
To precisely locate the MycP1 cleavage site, we repeated the spectrum-to-sequence assignment searches to include semi-tryptic peptides, i.e. peptides that had been cut by trypsin at one terminus only, which would suggest the remaining terminus resulted from MycP1 protease activity. Analysis of the middle (~10 kDa) EspB<sub>tb</sub> cleavage product revealed the most N-terminal semi-tryptic peptide matching to EspB<sub>tb</sub> at residues 359-394, indicating a non-tryptic cleavage between Ala358 and Ser359

(<sup>354</sup>AVKAA<sup>358</sup>|<sup>359</sup>SLGGG<sup>363</sup>, **Fig. 2.2e**). Edman sequencing independently confirmed this site as the N terminus of the C-terminal EspB<sub>Tb</sub> fragment, strongly suggesting MycP1 cleavage at this site. No semi-tryptic peptide could be identified for *M. smegmatis* EspB<sub>Sm</sub> but Edman sequencing identified the N terminus of the C terminal cleavage product at (<sup>402</sup>SLKPA<sup>406</sup>|<sup>407</sup>SAGGG<sup>411</sup>). For both EspB homologues, the predicted molecular weights of fragments resulting from cleavage at these sites correlated well to their observed apparent molecular weights on the gel. Pairwise sequence alignment of EspB<sub>Tb</sub> and EspB<sub>Sm</sub> indicates this cleavage site is at a homologous location, immediately preceding a polyglycine stretch (**Fig. 2.2f**). Multiple sequence alignment indicates the cleavage site is similar across species (**Fig. 2.2f**). According to protease nomenclature, substrate residues upstream of the cleaved scissile bond are denoted as P1-P6 in the sequence and residues downstream of the scissile bond are termed P1'-P6' (Siezen and Leunissen, 1997). The binding of the P1 side chain to the corresponding S1 pocket of the enzyme is often deemed most critical in aiding in the optimal positioning of the adjacent substrate peptide carbonyl for nucleophilic attack and thus greatest conservation is typically observed at this position. The P2', P2 and P4 positions are also typically key specificity determinants in subtilisin-like protease substrate recognition. Thus it is notable that residues comprising these positions in our identified EspB<sub>Tb</sub> and EspB<sub>Sm</sub> cleavage sites are identical or similar in chemical nature to each other (P4:[L/V], P3:[K], P2:[P/A], P1:[A], and P2': [L/A]). Interestingly, a sequence resembling these cut sites (<sup>384</sup>VRPA<sup>387</sup>) is found 29 residues downstream in the identified EspB<sub>Tb</sub> sequence and may indicate a second cleavage site, which would explain the doublet band we observe when EspB<sub>Tb</sub> is cleaved. Duplex EspB<sub>Tb</sub> cleavage has also been observed previously in the cellular context (7). Indeed, LC-MS/MS analysis of the lower EspB<sub>Tb</sub> cleavage product identified an N-terminus at Gly387, and N-terminal sequencing supports cleavage at this second VRPA cut site (**Fig. 2.2e**). Both EspB<sub>Tb</sub> sites we identified at Ala358 and Ala386 are within the vicinity but differ from previously proposed cut sites suggested from experiments using

cultures of *Mtb* and *M. marinum* cells secreting EspB (McLaughlin et al., 2007; Ohol et al., 2010; Xu et al., 2007).

### 2.3.2 Overall architecture of MycP1

To further understand MycP1's substrate specificity as well as the possible role of the N-terminal extension as a regulatory propeptide, we sought to determine its crystal structure. MycP1<sub>sm</sub>[24-407] crystallized in the orthorhombic space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and the structure was determined to 1.86 Å resolution (Table 1, PDB ID code 4J94). As predicted from sequence analysis, residues 62-390 of MycP1<sub>sm</sub>[24-407] adopts a subtilisin-like  $\alpha/\beta$  fold centered on a central 7-stranded parallel  $\beta$ -sheet (**Fig. 2.3a**), with a DALI (Holm and Rosenstrom, 2010) search yielding subtilisin BPN' as its closest structural homologue (RMSD of 0.965 Å over 186 common CA atoms) (**Fig. 2.3c**). Thus, the canonical subtilisin nomenclature will be used to describe the structure (Siezen and Leunissen, 1997). The N-terminal extension of MycP1<sub>sm</sub>, residues 24-61, proposed earlier to form the putative propeptide (Dave et al., 2002; Ohol et al., 2010), is observed to be structurally and functionally distinct from typical subtilisin propeptides (Siezen and Leunissen, 1997) in that it does not occupy the active site nor have a characteristic foldase domain, which acts as a intramolecular chaperone for subtilisin-like enzymes (Shinde and Inouye, 1995). MycP1<sub>sm</sub>[24-407] also contains several other unique insertions localized around the active site (**Fig. 2.3b**). Part of the C-terminal proline-rich membrane linker (390-403) was resolved and is found hugging the subtilisin core in an extended conformation, passing adjacent to the protein N-terminus. Coulombic electrostatic surface maps indicate MycP1<sub>sm</sub>[24-407] is negatively charged over the majority of its surface, including the active site cleft. In contrast to other subtilisin-like proteases (Siezen and Leunissen, 1997), MycP1<sub>sm</sub>[24-407] does not appear to bind calcium ions to promote structural integrity.



**Figure 2.3 Overall structure of *M. smegmatis* MycP1<sub>sm</sub>[24-407].** (a) Crystal structure of *M. smegmatis* MycP1<sub>sm</sub>[24-407] displayed as a ribbon representation, with key features labeled and highlighted by colour. (b) Schematic of active site region illustrating how the unique MycP1 secondary structural features and selected conserved residues interact. Green = N-terminal extension, blue = eI/eIII substrate docking strands, yellow = eI loop, brown = aromatic cradle, purple = e8/e9 insertion, orange = substrate location. (c) For comparison, the structure of subtilisin BPN' with its propeptide foldase domain (1SBP).

**Table 2.1. Data collection and refinement statistics for MycP1 structures**

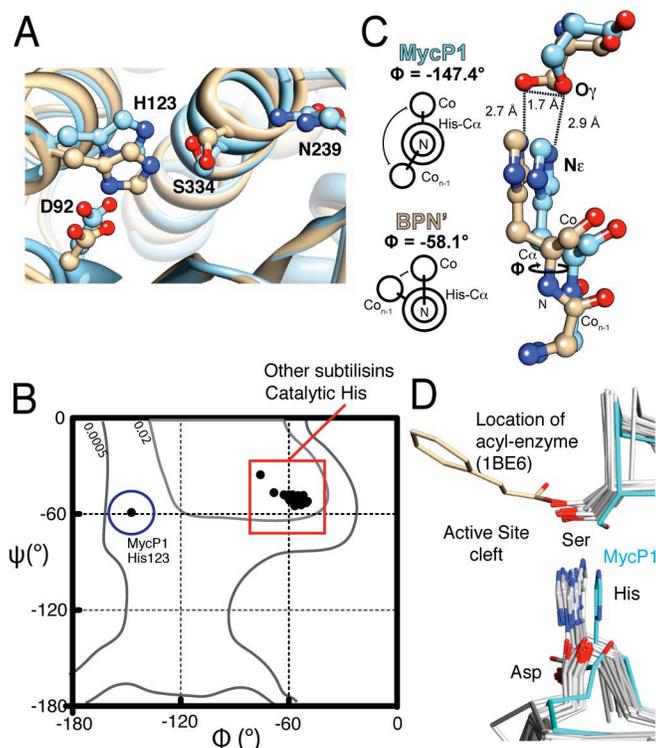
	4KPG	4J94
<b>Crystal parameters</b>		
Phasing method	SeMet SAD (peak)	MR
Space group	I 2 2 2	P 21 21 21
Cell dimensions		
<i>a, b, c</i> (Å)	81.00, 94.08, 119.60	56.49, 77.29, 84.93
$\alpha, \beta, \gamma$ (°)	90, 90, 90	90, 90, 90
<b>Data collection</b>		
Resolution (Å)	47.04-2.15	37.22-1.86
Completeness (%)	99.7 (99.0) <sup>a</sup>	86.3 (73.03)
Redundancy	4.8 (4.5)	2.0 (2.0)
Wavelength	0.97895	0.97895
<i>f' / f''</i>	8.25/5.13	
<i>R</i> <sub>merge</sub>	0.036 (0.553)	0.067 (0.179)
<i>I</i> / $\sigma$ <i>I</i>	10.4 (2.7)	7.50 (3.95)
<b>Refinement</b>		
Resolution (Å)	47.04-2.15	37.22-1.86
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>b</sup>	0.1690/0.2043	0.1553/0.1913
No. reflections	25,225	27,671 (2,296)
No. atoms		
Protein	2,945	2,738
Ligand/ion	0	0
Water	184	421
<i>B</i> -factors (Å <sup>2</sup> )		
Protein	36.9	15.2
Water	41.9	31.8
Root mean square deviations		
Bond lengths (Å)	0.007	0.006
Bond angles (°)	1.103	1.11
Ramachandran values		
Favored (%)	97.4	98
Allowed (%)	100	100

<sup>a</sup> Values in parentheses are for the highest resolution shell.

<sup>b</sup> Fraction of reflections in test set = 0.1.

### 2.3.3 MycP1's catalytic triad

In structures derived from two different space groups, the positioning of the MycP1<sub>sm</sub>[24-407] catalytic triad is unusual – namely, the side chain conformation of the nucleophile Ser334 is flipped and facing away from the presumed scissile bond location (**Fig. 2.4a**). Moreover, the phi angle ( $\varphi$ ) for His123 in our MycP1<sub>sm</sub>[24-407] structure is  $-147^\circ$ , compared to  $-60^\circ$  in subtilisin BPN'. A survey of subtilisin-like enzymes of known structure reveal both  $\varphi/\psi$  torsion angles for the catalytic histidine are conserved in all other cases ( $\varphi = -58^\circ \pm 5.8$  (S.D.),  $N = 22$ ), making the MycP1's backbone atypical in this region (**Fig. 2.4b**). The increased planarity of this peptide bond alters the back bone scaffold, and in combination with other global structural changes, His123 C $\alpha$  and its imidazole side chain are positioned  $\sim 1.7$  Å further from the active site cleft in comparison to subtilisin BPN' (**Fig. 2.4c and 2.4d**). The other two catalytic triad members, Ser334 and Asp92, maintain their hydrogen bonds with His123 (Ser334-O $\gamma$ -His123-N $\epsilon$  and Asp92-O $\delta$ 1-His123-N $\delta$ , respectively), resulting in all three triad residues occupying positions further from the presumed scissile bond position, with perhaps the most important change being the Ser334 side chain adopting a conformation that in our structure appears suboptimal for nucleophilic attack and a conformational change would be required to properly position the catalytic machinery. The oxyanion hole residue N239 appears similarly positioned as in other subtilisin-like enzymes.

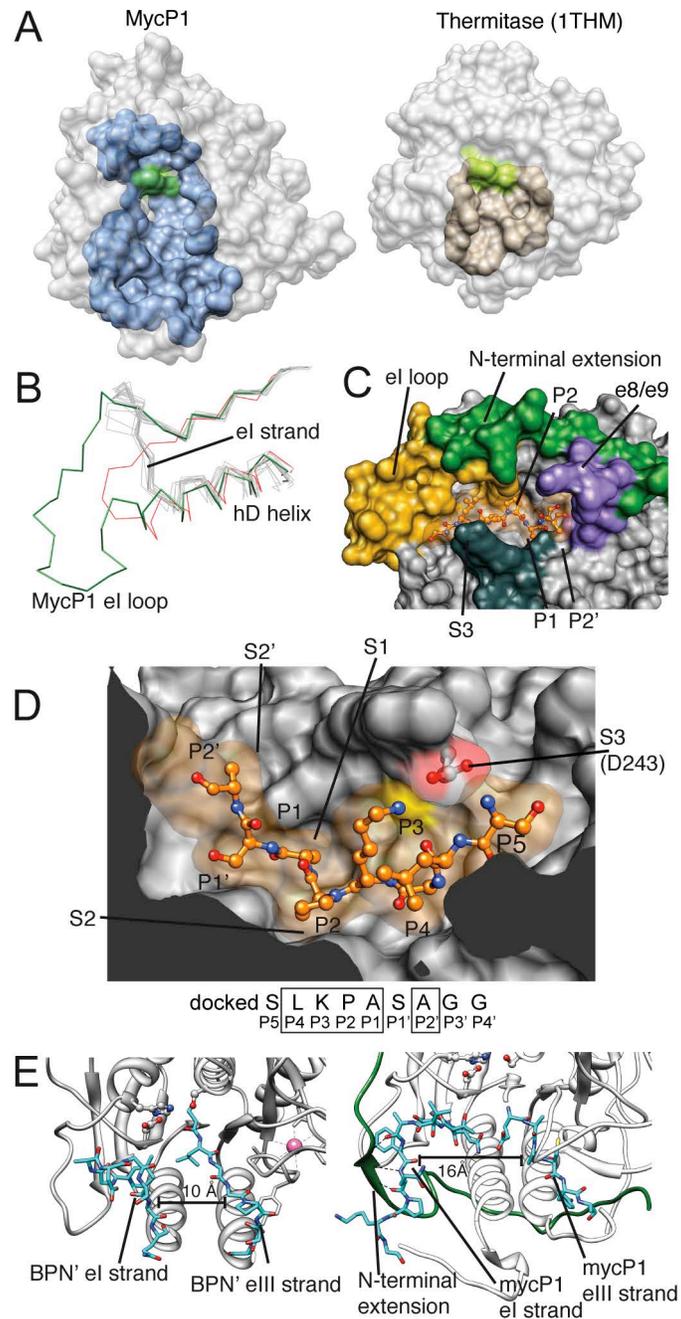


**Figure 2.4. MycP1's catalytic triad.** (a) Structural overlay of MycP1<sub>sm</sub>[24-407] (blue) and subtilisin BPN' (2SNI, tan) with catalytic triads displayed. (b) Ramachandran plot indicating  $\phi/\psi$  torsion angles for MycP1 and 22 subtilisin-like enzymes. (c) Differing  $\phi$  angles and His/Ser geometries for MycP1/subtilisin BPN' (d) Structural alignment of subtilisin-like enzymes with catalytic triad displayed. MycP1's triad is highlighted in cyan. An acyl-enzyme structure of subtilisin Carlsberg (1BE6) indicates the serine location and orientation required for nucleophilic attack.

### 2.3.4 Active Site Cleft

The predicted MycP1<sub>sm</sub>[24-407] catalytic cleft is unusually large (volume = 1083 Å<sup>3</sup>) in comparison to typical subtilisin-like enzymes of known structure (e.g. thermitase = 373 Å<sup>3</sup>) (Volkamer et al., 2012) (Fig. 2.5a). This is in part due to a unique 18 residue insertion after the e1 substrate binding strand (termed here the e1 loop; Fig. 2.5b) that creates a wide, elongated groove via its interaction with the N-terminal extension (Fig. 2.5c). A second mycosin-specific insertion, the e8/e9 beta strands, juts over the cleft to form a "lid" over the catalytic triad. To locate putative subsites (denoted by S) we used Rosetta FlexPepDock (London et al., 2011) to model the newly identified *M. smegmatis*

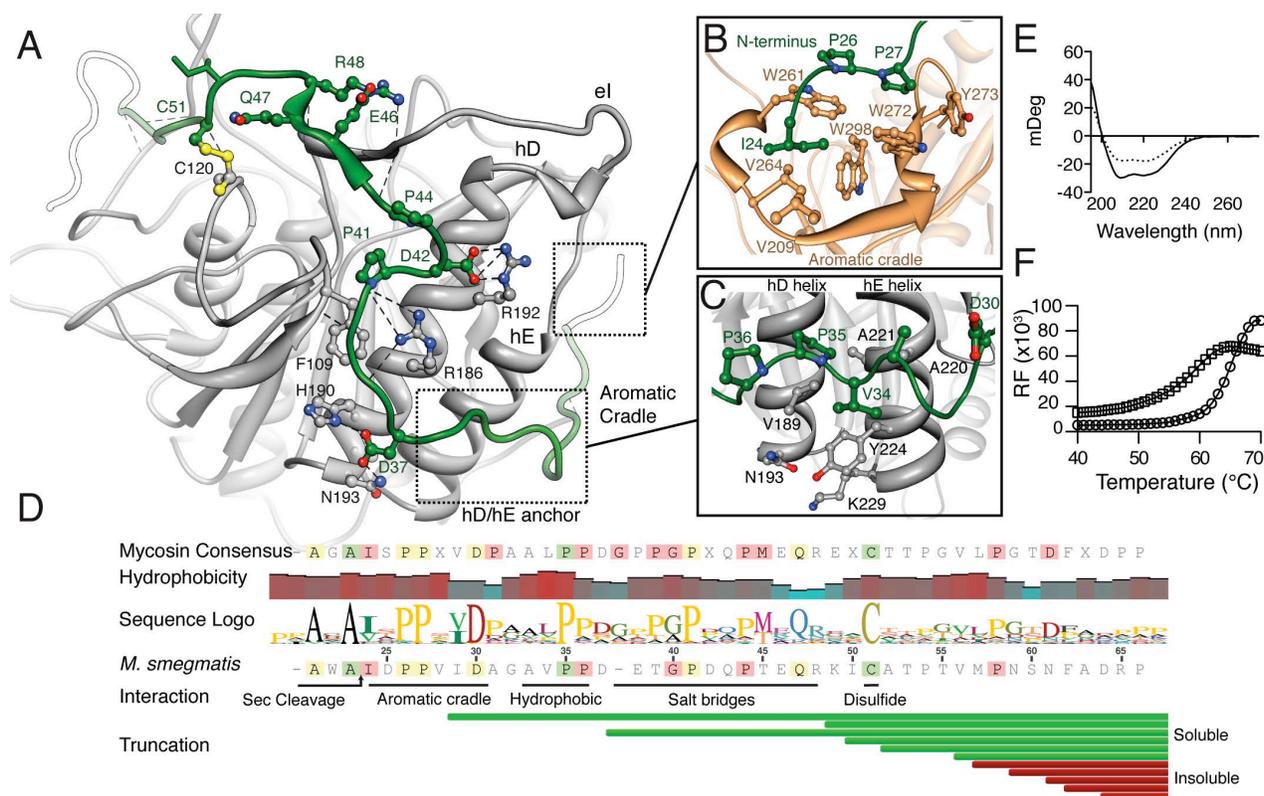
recognition motif (SLKPASAGG) into the cleft of MycP1<sub>sm</sub>[24-407] (**Fig. 2.5d**). As mentioned above, P1-P4 and P2' are typically the most important determinants for specificity in subtilisin-like enzymes. The conserved P1 Ala and P2 Pro side chains are complementary to their respective subsites and contact Ala204 (C $\beta$ ) and Thr156 (C $\gamma$ ), respectively. MycP1<sub>sm</sub>[24-407] appears to have a strong preference for Lys/Arg at P3, unusual in that this side chain is typically solvated and not a major contributor to specificity in most subtilisin-like enzymes (Siezen and Leunissen, 1997). However, MycP1 has evolved a disulfide-stabilized loop that protrudes over top of the cleft to scaffold Asp243, creating a negatively charged S3 site that is optimally positioned to interact with the positively charged P3 side chain. The S2' site is very similar to that of subtilisin BPN' - an apolar pocket created by Phe292 that could accommodate a hydrophobic residue, and is consistent with the Leu/Ala identified in our cleavage site. Although P4 is [LV] and should reasonably bind to a hydrophobic pocket, an obvious S4 subsite was not evident in the docking experiment. Of final note, the eI/eIII substrate-docking strands that typically orient the peptide for substrate for catalysis (Siezen and Leunissen, 1997) are separated by ~15 Å in the MycP1<sub>sm</sub>[24-407] structure (**Fig. 2.5e**) and appear incapable of forming such a triple-stranded ES complex without a conformational change/induced fit mechanism that would bring these active site elements into sufficiently close proximity. Taken together, the MycP1<sub>sm</sub> S2', S1, S2, and S3 subsites in the structure appear complementary to the P2', P1, P2, and P3 residues identified in the EspB<sub>sm</sub> cleavage site, however it is possible additional conformational changes may be needed to optimize the beta-strand interactions and formation of downstream subsites involved in binding substrate, and subtle repositioning of catalytic groups to promote optimal catalysis.



**Figure 2.5. Active site cleft of MycP1.** (a) MycP1 and thermitase (1THM) active site cleft regions, as predicted by DogSiteScorer (27). The catalytic triad is coloured green. (b) The el loop and hD helix are unusually long in comparison to nearly all other subtilisin-like enzymes; the structural alignment<sup>2</sup> of subtilisin-like enzymes highlights how these features are typically conserved in length and position. The outliers are C5a peptidase (red, 3EIF), and MycP1 (green). Note how the el substrate docking strand is unusually positioned. (c) The N-terminal segment (green) and el loop (gold) interact to form one side of the cleft, the docked EspB<sub>SM</sub> peptide (orange) indicates the cleft. (d) A docked peptide corresponding to the EspB<sub>SM</sub> recognition sequence is shown (orange) with P and S sites indicated. (e) MycP1<sub>SM</sub> el/eIII substrate docking strand spacing (left) compared with subtilisin BPN' (right).

### 2.3.5 N-terminal extension

The unusual N-terminal extension of MycP1 (residues 24-61) follows a securely anchored extended path, wrapping along the surface of the subtilisin domain (**Fig. 2.6**). At the extreme N-terminal end, a diproline motif (residues 24-29) sits in an aromatic cradle of MycP1<sub>sm</sub>[24-407] formed by conserved residues W272, W261, W298, and Y273. This cradle is the result of an insertion (between the e5/e6 strands) composed of two short 310 helices flanking a beta strand, that effectively creates an outcropping from the catalytic domain (**Fig. 2.6b**). The two proline side chains of the N-terminal motif stack favourably onto the tryptophan indole rings (**Fig. 2.6b**). This region of the N-terminal extension adopts a polyproline II helix, with characteristic 120° spacing between residues imposed by the conformational rigidity of the proline residues (Ball et al., 2005). Its interaction with the aromatic cradle is reminiscent of proline recognition motifs (PRM) of the WW family (Zarrinpar and Lim, 2000). The N-terminal extension passes underneath the hD/hE active site helices where a second diproline motif forms another polyproline II helix and is anchored by interactions with a patch of hydrophobic amino acids (**Fig. 2.6c**). The N-terminal extension is further bound through a series of electrostatic interactions, beta sheet interactions with the e1 loop and tethering by the C51-C120 disulfide bond (**Fig. 2.6a**). Generally, the N-terminal extension is rich in prolines, which appear required to facilitate its extended path around the subtilisin domain of MycP1<sub>sm</sub>[24-407]. We carried out a BLAST search to find MycP1 homologues and used the identified sequences to generate a multiple sequence alignment with the program MUSCLE. A sequence logo conservation plot is shown for the N-terminal extension, with purification results for a series of N-terminal truncation mutants mapped below (**Fig 2.6d**). We observed MycP1<sub>sm</sub>[24-407] was expressed and folded (albeit with reduced solubility) even with up to 33 of 40 N-terminal extension residues removed, suggesting its primary function is unrelated to folding but may contribute to stability (**Figs. 2.6d and 2.6e**). Differential scanning fluorimetry comparing MycP1<sub>sm</sub>[24-407] and MycP1<sub>sm</sub>[49-407] showed a 7 °C difference in melting temperature ( $T_m = 64\text{ °C}$  vs.  $57\text{ °C}$ ), indicating the N-terminal extension contributes to overall stability (**Fig. 2.6f**).



**Figure 2.6. Interactions between the N-terminal extension and subtilisin domain.** (a) Overview of the propeptide path emphasizing electrostatic interaction, beta sheet formation with the el loop, and the C51-C121 disulfide bond. (b) The N-terminal region of MycP1 adopts a polyproline helix that binds to a PRM-like aromatic cradle. The IDPP stretch interacts with a triple-tryptophan arrangement, where the proline rings pack parallel to two of the tryptophan residues. I24 and I29 pack into adjacent hydrophobic pockets. (c) The N-terminal extension forms a second polyproline II helix that binds to the base of the hD and hE helices. D30 is strictly conserved and faces the solvent. (d) N-terminal extension sequence conservation with the mycosin consensus. A hydrophobicity plot shows four distinct hydrophobic patches on the N-segment (sliding window average= 3). The N-terminal portion is more highly conserved across all mycosins than C-terminal portion. Results of expression of N-terminal extension truncations shown as soluble (green bars) or insoluble (red bars). (e) Circular dichroism of MycP1<sub>sm</sub>[24-407] (solid line) and the N-terminal segment truncation mutant MycP1<sub>sm</sub>[49-407] (dashed line). (f) Differential scanning fluorimetry experiment of MycP1<sub>sm</sub>[24-407] (circles) and MycP1<sub>sm</sub>[49-407] (squares).

## 2.4 Discussion

MycP1<sub>sm</sub>[24-407] crystallized with a native catalytic triad that failed to hydrolyze the proposed pro-domain, indicating this full-length enzyme is highly stable over the course of

the purification procedure and crystallization experiment. In contrast to subtilisin pro-domains, which form a typical foldase fold that occupies the active site within reach of the Ser-His-Asp catalytic triad, the structurally distinct MycP1 N-terminal extension is not bound within the cleft and is more than 13 Å away from the S334 nucleophile. This rules out the possibility of autoproteolytic cleavage of this region in cis, barring conformational rearrangement. Crystallization of wild type subtilisin-like enzymes with their uncleaved pro-domains is rare (a catalytically impaired form is typically required), the only other example to our knowledge is the proprotein convertase PCSK9 (Cunningham et al., 2007). Further, our mass spectrometry data showing MycP1, with its N-terminal extension intact, is able to cleave EspB (albeit slowly) at a [L/V]K[P/A]A | S[L/A]GG site adds further support that this appendage is not a propeptide. Thus, it is possible that the N-terminal extension may not be cleaved but may modulate proteolysis in response to an appropriate trigger. This mode of operation may be comparable to the intracellular subtilisin proteases (ISPs), which have a short N-terminal extension that operates through a combined active site blocking/catalytic triad rearrangement mechanism (Vevodova et al., 2010). Another possibility is that cleavage of the N-terminal extension may occur in a cellular context but the appendage could remain anchored to the subtilisin domain through the disulfide bond (e.g. as with chymotrypsin (Khan and James, 1998)), which could help explain previous reports of activation resulting from addition of Factor Xa protease (Ohol et al., 2010). We have shown that the majority of the N-terminal extension is not required for folding, but does contribute to overall stability. Given that the extension interacts with several features surrounding the active site (the e1 strand (residues 159-163), aromatic cradle (residues 261-274), the C51-C120 disulfide bond, and e8/e9 extension (residues 320-328), we suspect a role in modulating catalysis or conferring specificity. Another possibility is that the N-terminal extension is serving as a placeholder in obstructing binding surfaces, and if removed, these surfaces would have the potential to target substrates or other components of the secretion system, perhaps through the same proline recognition motifs which anchor the uniquely pro-rich N-terminal extension in the structure.

As MycP1<sub>sm</sub>[24-407] inefficient at cleaving EspB *in vitro*, we hypothesize the MycP1 structure presented here is an isolated T7SS component in its pre-assembly form and could still represent a zymogen with “leaky” activity. Given that the previous *in vitro* fluorogenic substrate profiling data indicated an uncharacterized form of MycP1 is capable of catalysis with broad specificity (Ohol et al., 2010), it is unsurprising that MycP1 is synthesized with the intrinsic capacity to prevent rapid premature proteolysis, which could cause damaging non-specific cleavage. It is well understood that nucleophilic cleavage of peptide bonds requires a precisely defined geometric arrangement (Dodson and Wlodawer, 1998). Our analysis of MycP1’s native catalytic triad allows us to propose one possible mechanism by which a subtle conformational change, for example through N-terminal segment cleavage or binding to a cofactor or binding partner, could twist the His123  $\phi$  torsion angle and Ser334 side chain into positions more similar to that of catalytically competent proteases. MycP1 may have extrinsically low activity in the absence of unknown cofactors including the lipidic membrane environment, other ESX components such as the predicted EccBCDE ATPase/translocon complex, or as-of-yet uncharacterized components of the extracellular portion of the secretion apparatus. Perhaps an undefined temporal trigger mechanism may ensure proteolysis only occurs as ESX substrates are crossing the membrane.

Due to its essential action in virulence and relatively accessible nature, MycP1 is an attractive extracellular anti-TB drug target. Additionally exciting are the mycosin’s potential as an engineerable factor in rational design of more effective TB vaccines due to its influence on processing and presumably the downstream function of highly antigenic/virulence-associated EspB and the PE/PPE protein families (Ohol et al., 2010). Indeed, manipulation of T7S systems have driven major advancements in mycobacterial immunology (Pym et al., 2003; Sweeney et al., 2011), and we have provided the first structural template for engineering vaccine strains that perturb this protease’s influence on regulation, secretion, and processing of secreted TB antigens.

# Chapter 3: Structure of EspB and Insights Into its Export Mechanism

## 3.1 Introduction

Despite its central role in pathogenesis, the structure of the ESX apparatus remains poorly characterized. The core inner membrane export machinery consists of an EccC ATPase, a membrane-bound MycP protease that processes secreted proteins (Ohol et al., 2010), and several other transmembrane apparatus proteins (Abdallah et al., 2007) (**Fig. 3.1a**). Each gene cluster also encodes a more variable set of components such as chaperones, ATPases, transcription factors, and crucially, a specific set of secreted proteins (Houben et al., 2013).

These secreted proteins are the focus of intense study due to their external location in the bacterial cell wall and growth media, placing them in position to manipulate the surrounding environment. However, their molecular functions remain an issue of some contention. Previous hypotheses suggest ESX-secreted proteins may possess pore-forming activity (de Jonge et al., 2007; Hsu et al., 2003; Smith et al., 2008), chelate metals, (Ilghari et al., 2011), modulate extracellular signalling pathways (Pathak et al., 2007), or are structural components that form the extracellular portion of the secretion apparatus itself (Pallen, 2002). In any case, each ESX system is associated with a specific set of secreted proteins that likely contribute to the specialized nature of each system. A close comparison of unique secreted proteins across ESX systems will help explain how, for example, ESX-3 functions in iron acquisition while ESX-1 is associated with pathogenesis

and DNA conjugation. Along these lines, ESX-1 secretes a number of proteins not encoded by the simpler ESX loci of Gram-positive bacteria. These proteins are also absent in ESX-2 through ESX-5 loci of mycobacteria, many of which are essential for ESX-1 virulence phenotypes in *Mtb* and conjugal DNA transfer in *M. smegmatis*. These include EspA (Fortune et al., 2005), EspC (Millington et al., 2011), EspE (Sani et al., 2010), EspF (Sani et al., 2010), EspJ (Champion et al., 2014), EspK (Champion et al., 2014; Sani et al., 2010), and EspB (Gao et al., 2004; McLaughlin et al., 2007; Xu et al., 2007) (**Fig. 3.1b, red asterisks**).

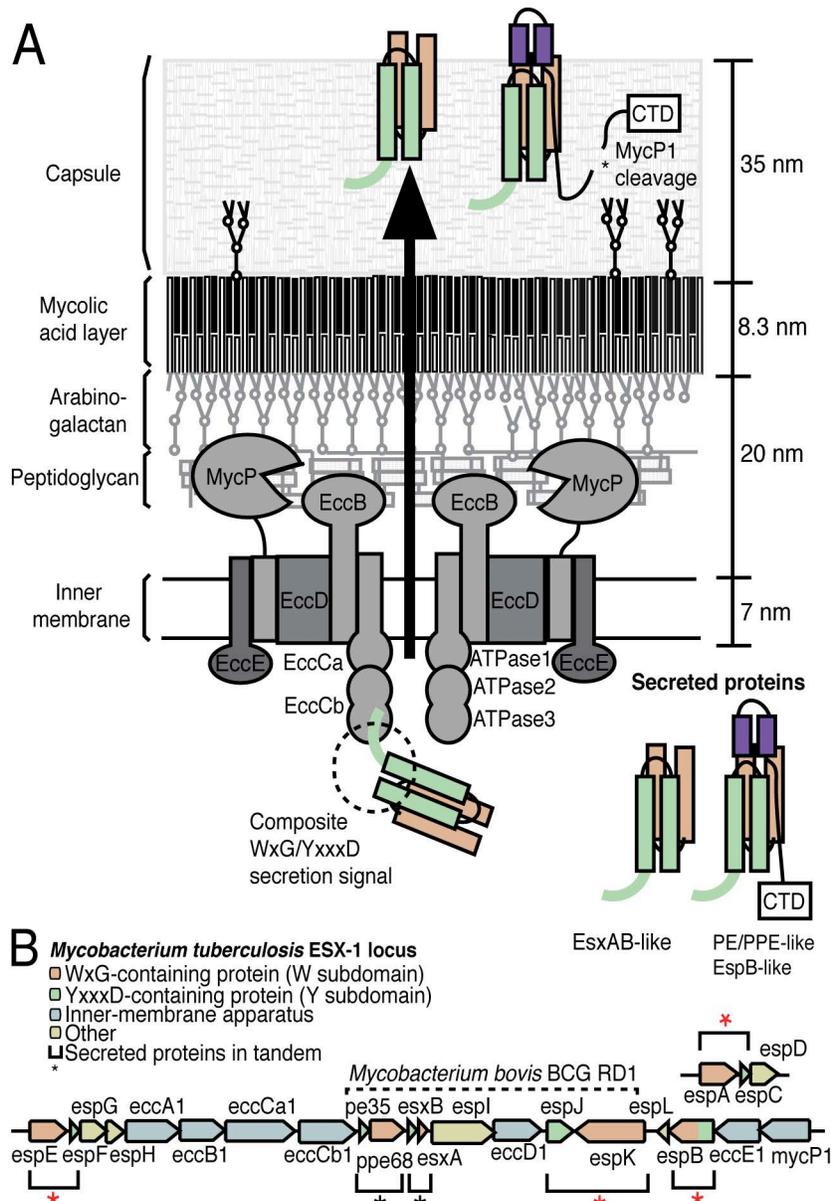
EspB in particular is directly involved in ESX-1 membranolytic function, and strains lacking this secreted protein are as attenuated for virulence as those lacking the entire secretion apparatus (Gao et al., 2004). Knocking out EspB in the related pathogen *Mycobacterium marinum* abrogates red blood cell hemolysis (Gao et al., 2004), eliminates macrophage cytotoxicity (Gao et al., 2004), inhibits intracellular growth and phagosome maturation (McLaughlin et al., 2007; Xu et al., 2007), and completely attenuates virulence in zebrafish infection models (Gao et al., 2004). A transposon mutant of EspB also decreases *M. smegmatis* conjugation efficiency by 1000-fold, suggesting a role in conjugal transfer as well (Coros et al., 2008). EspB has a C-terminal domain (CTD) that is processed by the membrane-bound protease, mycosin-1 (MycP1) following secretion (McLaughlin et al., 2007; Ohol et al., 2010; Solomonson et al., 2013; Wagner et al., 2013a; Xu et al., 2007) (**Fig. 3.1a**). The purpose of this cleavage event is unknown but is required for full pathogenesis of *Mtb* and has been reported to modulate the quantity of protein secreted by ESX-1 (Ohol et al., 2010) and/or regulate phospholipid binding by EspB (Chen et al., 2013b).

Despite this dual role in regulating and mediating ESX-1 virulence, structural data is lacking for EspB or any other secreted Esp protein in the ESX-1 locus. Moreover, the mechanism by which Esp proteins are secreted through ESX-1 is not entirely clear. Recent work has culminated in the identification of a general 'secretion signal' that targets cytoplasmic ESX substrates, including EspB, to the secretion apparatus for export across the inner membrane (Champion et al., 2006; Daleke et al., 2012a) (**Fig. 3.1a**). The structure of this secretion signal has primarily been studied in the simplest ESX-secreted

substrates of the WXG100 family such as EsxA and EsxB (Arbing et al., 2013; Arbing et al., 2010; Ilghari et al., 2011; Poulsen et al., 2014; Renshaw et al., 2005).

WXG100 family members are secreted as homo- or heterodimeric complexes that adopt elongated, antiparallel helical bundles, where a WxG motif is present in at least one member of the heterodimer (labeled “W subdomain” and colored orange in **Fig. 3.1 and 3.2**). This WxG motif forms one half of a bipartite composite secretion signal (Sysoeva et al., 2014) that is thought to target the heterodimeric complex to EccC for export (Champion et al., 2006) (**Fig. 3.1a**). The second member of the heterodimer contributes the other half of this composite secretion signal: a flexible C-terminal ‘export arm’ that harbours a characteristic Tyr-x-x-x-[Asp/Glu] (YxxxD) motif and at least one additional conserved hydrophobic residue seven positions downstream from the YxxxD motif (Poulsen et al., 2014) (this member is labeled “Y subdomain” and colored green in **Fig. 3.1 and 3.2**). Previous NMR (Renshaw et al., 2005), crystallographic (Arbing et al., 2013; Poulsen et al., 2014), and two-hybrid studies (Champion et al., 2006) indicate this dynamic export arm is stabilized such that residues align on the same face to interact with the EccC ATPase, representing a key phase of the export cycle and a potential determinant of ESX substrate specificity (Champion et al., 2009).

The final class of ESX-secreted protein relevant to the work presented here are the PE and PPE protein subfamilies, named for the characteristic Pro-Glu or Pro-Pro-Glu motifs located in the N-terminal regions of the respective variants (Sampson, 2011). Like EspB, PE and PPE proteins also possess CTDs, which encode for low complexity repetitive structures hypothesized to promote antigenic variation at the cell surface. Crystallographic analysis of the PE25/PPE41 complex from ESX-5 revealed a helical bundle reminiscent of the EsxAB heterodimer (Strong et al., 2006) (**Fig. 3.2b**).



**Figure 3.1. Overview of ESX-1 secretion.** (a) Schematic summary of inner membrane secretion complex, secreted proteins, and mycobacterial cell wall composition and dimensions (schematic adapted from references (Brennan and Crick, 2007; Houben et al., 2012b; Sani et al., 2010)). (b) The *Mtb* ESX-1 virulence locus, with tandemly-organized secreted proteins that together possess the full composite secretion signal marked by asterisks.

Here, we describe the first crystal structure of EspB from *M. smegmatis* (EspB<sub>sm</sub>), which reveals structural elements common to both the EsxAB and PE/PPE family of proteins, with a structured view of the composite secretion signal that clearly classifies it

as a member of the WXG100 superfamily. EspB<sub>sm</sub> adopts a distinct homomeric helical bundle and an additional customized appendage that, like the PE/PPE heterodimers, replaces one of the dual signal sequences present in canonical ESX substrates. Our light scattering and negative stain electron microscopy data further show that EspB from *Mtb* (EspB<sub>tb</sub>) has a propensity to oligomerize, revealing the potential for EspB to serve as structural subunits in the construction of cell wall-associated architectures of mycobacteria as widely hypothesized (Champion et al., 2009; Fortune et al., 2005).

## 3.2 Methods

### 3.2.1 Purification, crystallization, and structure determination of EspB

EspB coding sequences from *M. smegmatis* (MSMEG\_0076) and *Mtb* (Rv3881c) were cloned into pET28a(+), in frame with an N-terminal cleavable histidine affinity tag with the sequence MGSSHHHHHHHHHHSSGLVPRGSH. Plasmids were transformed into *E. coli* BL21 codon-plus (DE3)-RIPL-competent cells (Stratagene). Induction was initiated with 1 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside at A<sub>600</sub> = 0.6 followed by growth for 20 hours at 20 °C. Cells expressing Se-Met derivative EspB<sub>sm</sub>-[1-292] were grown as previously described (Larrson, 2009). Prior to purification, cells were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, and protease inhibitor tablets (Roche cOmplete) and lysed with a C5 homogenizer (Avisten). Lysate was spun at 45,000 rpm for 1 h, and the supernatant was passed over HisPur Ni-NTA resin (Thermo, 88222) and washed with buffer containing 25 mM imidazole and finally eluted in buffer containing 250 mM imidazole. Eluate was concentrated and further purified on a Superose 6 10/300 GL column equilibrated with 20 mM HEPES pH 7.5, 150 mM imidazole. EspB<sub>tb</sub> protein corresponding to the oligomeric peak was concentrated and flash-frozen at -80 prior to EM, light scattering, and analytical ultracentrifuge analyses. The EspB<sub>sm</sub> variants were digested overnight with thrombin and purified on a Superdex 200 column prior to crystallization.

Diffraction Se-Met derivative P3221 crystals grew using the EspB<sub>Sm</sub>-[1-292] construct at 80 mg/ml in 100 mM HEPES pH 6.8, 540 mM MgCl<sub>2</sub>, 15% PEG 6000, and reached maximum size in 5 days. Crystals of the P65 space group were obtained using the EspB<sub>Sm</sub> full-length construct at 20 mg/ml in 0.1 M Bis-tris propane pH 6.5, 200 mM MgCl<sub>2</sub>, and PEG 3350, reaching a maximum size after 1 month. The P3221 crystals were flash frozen directly, while the P65 crystals were cryoprotected in 20% glycerol prior to flash freezing. The data collected at the Canadian Light Source Canadian Macromolecular Crystallography Facility (CMCF) was integrated, scaled, and merged with Mosflm/Pointless (Bailey, 1994) and the structure was solved, built, and refined with the Phenix package (Adams et al., 2010) and Coot (Emsley and Cowtan, 2004). PHENIX Autosol located 13 selenium atoms in the substructure solution (figure of merit = 0.42 using data from 3.2 – 61.8 Å). The P3221 structure served as a molecular replacement search model for solving the P65 native dataset using PHASER (McCoy et al., 2007). The structures were analyzed and figures generated using UCSF Chimera (Pettersen et al., 2004).

### **3.2.2 Protein fragment complementation assay**

The coding sequence for EspB<sub>Sm</sub> residues 70-100 was cloned into pUAB100, in-frame with the a C-terminal murine dihydrofolate reductase F[1,2] fragment, and coding sequence for EccCb1<sub>Sm</sub> residues 341-617 was cloned into pUAB200, in frame with a C-terminal murine dihydrofolate reductase F[3] fragment, and the assay was carried out as described (Singh et al., 2006). Plasmids were co-transformed into *M. smegmatis mc*<sup>2</sup><sub>155</sub> by electroporation and plated on 7H10 media containing 25 µg/ml kanamycin and 50 µg/ml hygromycin. Overnight cultures grown from glycerol stocks were adjusted to OD = 1.0 and 4 µl of cells were spotted on 7H11 plates containing 25 µg/ml kanamycin and 50 µg/ml hygromycin, in the absence or presence of 5 µg/ml trimethoprim.

### **3.2.3 Multi-angle light scattering**

EspB<sub>tb</sub>-[1-460] samples (10 mg/ml, volumes of 100  $\mu$ l) were loaded onto a Superose 6 10/300 GL column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl at 25°C with a flow rate of (0.5 ml/min) followed by light scattering/refractive index measurements made with a miniDAWN TREOS detector coupled to a Optilab T-rEX differential refractometer following chromatographic separation. The ASTRA software package (Wyatt Technologies) was used to analyze the data and determine molar mass/polydispersity.

### **3.2.4 Analytical ultracentrifuge**

Analytical ultracentrifuge experiments were carried out at 20 °C using a Beckman XL-I analytical ultracentrifuge, an An-60 Ti rotor, and absorbance optics. EspB<sub>tb</sub>-[1-460] samples were loaded at 1.0 mg/ml and spun at 47,000 RPM in two-channel carbon filled epon centerpieces. The data were analyzed with Sedfit (Schuck, 2000).

### **3.2.5 Negative stain electron microscopy**

EspB<sub>tb</sub> was diluted to 0.01 mg/ml and prepared for EM as described previously (Ohi et al., 2004) and visualized using a Tecnai Spirit transmission electron microscope (FEI) operated at with an accelerating voltage of 120kV. Images were taken at a nominal magnification of 49,000x using an FEI Eagle 4K x 4K charge-coupled device (CCD) camera at a defocus value of -1.2 $\mu$ m. For image processing, 2 x 2 image pixels were averaged for a 4.7Å pixel size. For 2D analysis, individual particle images were selected using Boxer (Ludtke et al., 1999). The particles were next subjected to reference-free alignment and sorted into classes by K-means classification using algorithms in the SPIDER image processing suite (Frank et al., 1996). Particle images in each class were averaged to generate EspB 2D class averages.

To produce the EspB 3D reconstruction, a representative 2D class average corresponding to a side view of the EspB heptamer was rotationally extruded about the symmetry axis to form an initial 3D model using John Rubinstein's program

"build\_fspace\_v2\_00" (<https://sites.google.com/site/rubinsteingroup/3-d-fourier-space>). Single particle images belonging to class averages corresponding to both top and side views of the EspB complex (1,455 particles) were used to refine the initial model using EMAN2. The final resolution was determined using the Fourier shell correlation (FSC) function using the 0.5 FSC criterion.

### **3.2.6 Oligomeric modeling**

Homology models were constructed in RosettaCM (Song et al., 2013) using the EspBsm P3221 crystal structure as a template and an alignment from hhsuite (Soding, 2005). Non-terminal gaps in the sequence alignment were rebuilt from backbone fragments, and the whole structure was refined with a physically realistic forcefield combined with pairwise distance restraints derived from the homologue. A total of 25 homology models were generated (due to high similarity between the template and target sequence, convergence was quite good), and the lowest-energy model was selected.

Using this selected model, we ran Rosetta's symmetric docking protocol (Soding, 2005) guided by density data. Models were randomly oriented and a C7 symmetric system was built. Docking was carried out in two stages. In the first stage moves consist of rigid body perturbations and "slide-into-contact" moves, and structures are evaluated with a course energy function where sidechains are replaced with single interaction centers. In the second stage, sidechains are built, and moves consist of rigid-body perturbation followed by sidechain optimization; structures are evaluated with Rosetta's all-atom energy function. In both stages, an additional fit-to-density term ("elec\_dens\_fast") was included in model-building with a weight of 2.

Due to the low resolution of the data, no backbone optimization against the experimental density data was performed. In total, 25,000 docked models were generated. The lowest-energy 500 were selected and clustered, yielding four distinct clusters. The centroids of each cluster were then compared to crosslinking data.

### 3.2.7 Crosslinking and mass spectrometry

EspBtb[1-348] and EspBtb[1-460] samples were crosslinked with 1 mM cyanur-biotin-dimercapto-propionyl-succinimide (CBDPS-H8/D8; Creative Molecules Inc.) and incubated for 30 min at 25 °C. The reaction mixtures were quenched with 10 mM ammonium bicarbonate at 25 °C. Crosslinked proteins were digested with either trypsin for 18 hours at 37 °C (Promega, Madison, WI) and/or proteinase K (Worthington, Lakewood, NJ) for 60 min at 37 °C, both at 1:15 (w:w) enzyme:substrate ratios. Trypsin and proteinase K digestions were inhibited by the addition of AEBSF to a final concentration of 10 mM. Trypsin and Proteinase K in-gel digests of dimer and pooled higher molecular weight bands were also prepared. CBDPS crosslinked peptides were enriched on monomeric avidin beads (Thermo Scientific, Rockford, IL), eluted from the beads with 0.1% FA 50% acetonitrile, concentrated by lyophilization, reduced with 10 mM DTT for 10 min at 25 °C, and finally acidified with formic acid. Mass spectrometric analysis was carried out with a nano-HPLC system (Easy-nLC II, ThermoFisher Scientific, Bremen, Germany) coupled to the ESI-source of an LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific, Bremen, Germany) (Petrotchenko et al., 2014a). Samples were injected onto a 100 µm ID, 360 µm OD trap column packed with Magic C18AQ (Bruker-Michrom, Auburn, CA), 100 Å, 5 µm pore size (prepared in-house) and desalted by washing for 15 min with 0.1% formic acid (FA) before peptides were separated with a 70 minute gradient (0-60 min: 4-40% B, 60-62 min: 40-80% B, 62-70 min: 80% B, with solvent B: 90% acetonitrile, 10% water, 0.1% FA) on a 75 µm ID, 360 µm OD analytical column packed (in-house) with Magic C18AQ, 100 Å, 5 µm pore size with IntegraFrit (New Objective Inc. Woburn, MA) and equilibrated with 95% solvent A (2% acetonitrile, 98% water, 0.1% FA). MS data were acquired with Xcalibur (ver. 2.1.0.1140) Mass Tags and Dynamic Exclusion enabled in global data dependent settings. MS scans and MSMS scans were acquired in the Orbitrap mass analyzer at 60000 and 30000 resolution respectively. MSMS fragmentation was performed by CID activation at normalized collision energy of 35%. Data analysis was performed using DXMSMS Match of ICC-CLASS. To determine intra- or inter-protein origin of the identified crosslinks crosslinking

was performed with equimolar ratio of  $^{14}\text{N}$  and  $^{15}\text{N}$ -metabolically labeled EspB (Taverner et al., 2002). Data were analyzed using 14N15N DXMSMS Match program (Petrotchenko et al., 2014b).

## 3.3 Results

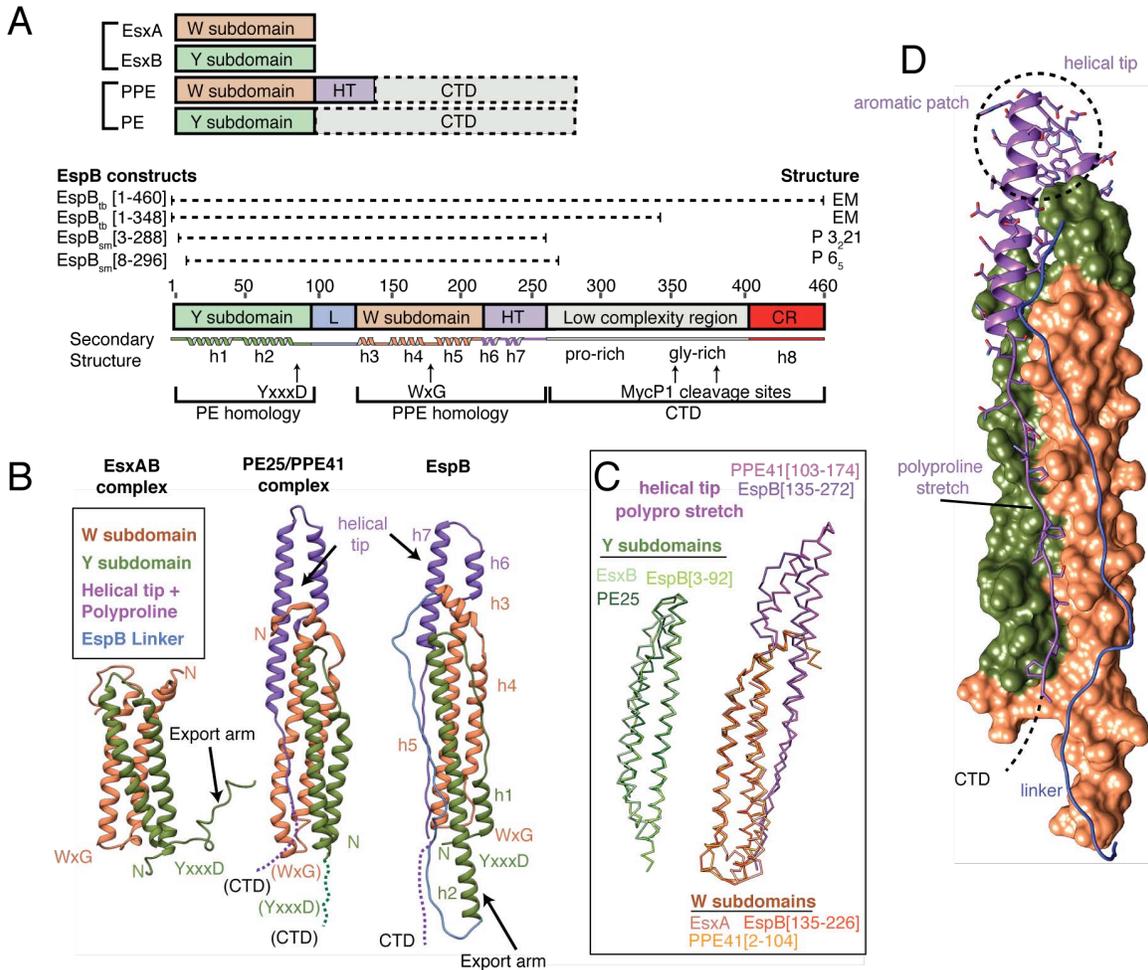
### 3.3.1 Crystal structure of EspB

The x-ray crystallographic structure of the N-terminal helical domain of EspB from *M. smegmatis* was determined in two crystal forms, one encompassing residues 3-288, (space group P3221, PDB code 4WJ1) and a second crystal form encompassing residues 8-296, (space group P65, PDB code 4WJ2 (**Table 3.1**). Due to predicted disorder, the CTD was not included in the construct used to grow the P3221 form (**Fig. 3.2a**), and this CTD was also not observed in the P65 form even though it was included in the construct, likely due to proteolytic degradation or disorder within the large crystal solvent channels of this crystal form (~80% solvent). Although highly similar in overall architecture (0.69 Å backbone RMSD), the two structures display significant differences in functionally relevant regions (discussed below). Unless noted, we will describe structural features based on the P3221 form, as this structure was refined to higher resolution (2.4 Å).

EspB<sub>Sm</sub> (residues 3-288, **Fig. 3.2a**) is composed of long, antiparallel coiled-coil helices characteristic of a PE/PPE heterodimer (**Fig. 3.2b, right**). Unlike other ESX-associated proteins, EspB<sub>Sm</sub>[3-288] is not a heterodimer but rather adopts a PE/PPE-like fold due to an apparent genetic fusion of heterodimeric members into a single open reading frame by a 35-residue linker (**Fig. 3.2a**). A side-by-side comparison of EspB<sub>Sm</sub>[3-288] with the crystal structures of the ESX-secreted substrates EsxAB and PE25/PPE41 reveals the modular nature of ESX-secreted proteins (**Fig. 3.2b and 2c**). The helix-turn-helix domains of PE25, EsxB, and the N-terminal domain of EspB<sub>Sm</sub> residues 1-92 are roughly superimposable, each with a characteristic YxxxD motif at the turn (this subdomain is labeled “Y subdomain” and colored green in **Fig. 3.1 and 3.2**.) EspB<sub>Sm</sub> residues 135-226, EsxA, PPE41 residues 2-104 are also roughly superimposable, each with

a WxG motif at the turn of the second helix-turn-helix subdomain (labeled “W subdomain” and coloured orange in **Fig. 3.1 and 3.2**). In EspB<sub>Sm</sub>[3-288], the Y and W subdomains form an extensive interface of non-covalent interactions, and the linker connecting them adopts an extended conformation that anchors loosely to the helical bundle through van der Waals interactions (**Fig. 3.2d**).

Compared to PE25/PPE41, EspB<sub>Sm</sub>[3-288] has a shorter h1 helix, with residues in this position instead adopting an extended conformation (**Fig. 3.2b**). EspB<sub>Sm</sub>[3-288] and PE25/PPE41 possess features that are not found in EsxAB. Notably, helix 3, helix 6 and helix 7 of EspB<sub>Sm</sub>[3-288](residues 231-269) create a “helical tip” that is rich in solvent-exposed hydrophobic/aromatic/acidic residues (**Fig. 3.2d**, purple). This helical tip is slightly smaller in EspB<sub>Sm</sub>[3-288] than in PE25/PPE41. The tip ends in an extended polyproline stretch (residues 270-288) that sandwiches between the Y and W subdomains through hydrophobic interactions provided by prolines and aliphatic side chains (**Fig. 3.2d**). Sequence comparisons suggest that in EspB<sub>Sm</sub> and many PPE-family proteins, this helical tip is connected via the polyproline stretch to the CTD of varying sequence (**Fig. 3.2d**).



**Figure 3.2. Crystal structure of *M. smegmatis* EspB compared to model ESX substrates EsxAB and PE25/PPE41. (a) Summary of ESX-secreted proteins discussed in this paper. *W subdomain* indicates presence of WxG motif, *Y subdomain* indicates presence of YxxxD motif, *HT* - helical tip, *L* - linker, *CR* - conserved region of C-terminal domain (b) Structures of representative members of Esx and PE/PPE-families for comparison with EspB, colour-coded by feature (c) Structural superposition of isolated Y and W subdomains across ESX-secreted protein classes demonstrates conserved modularity. (d) Surface view emphasizing helical tip.**

**Table 3.1. Data collection and refinement statistics for the EspB structures**

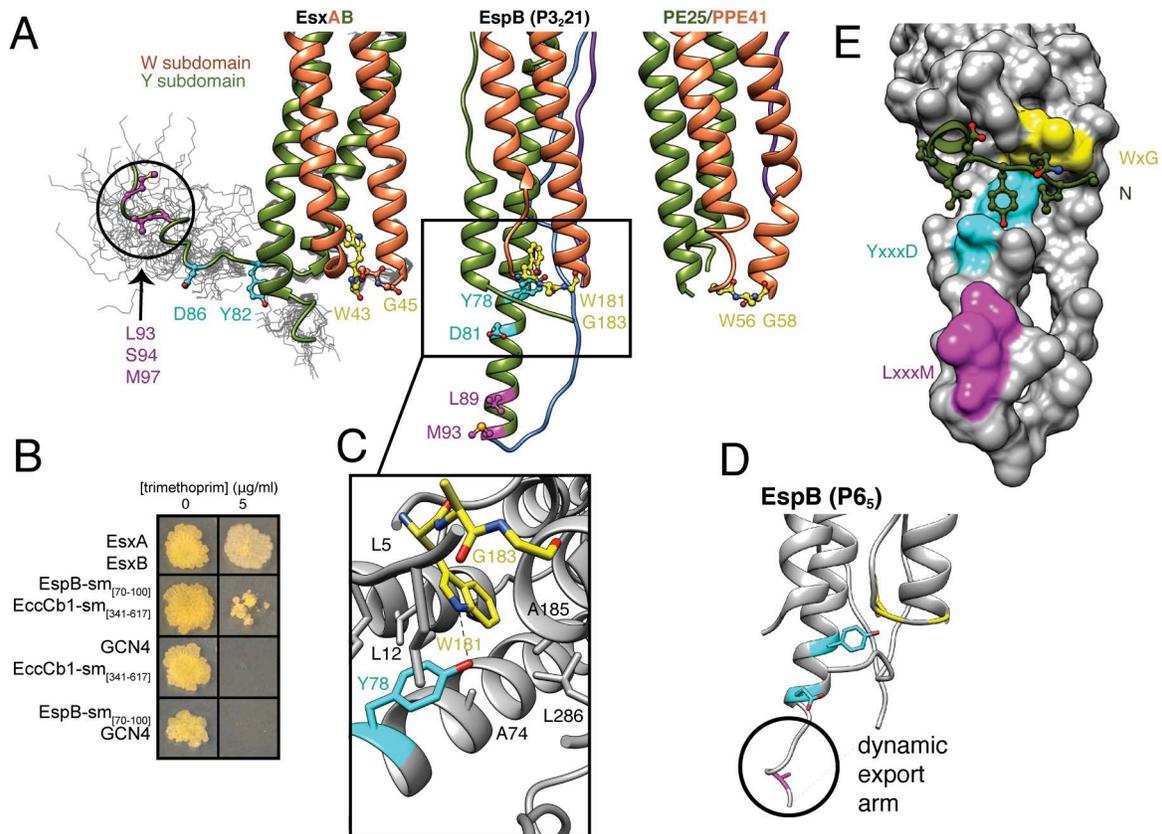
	<i>M. smegmatis</i> EspB [3–288] Se-Met SAD	<i>M. smegmatis</i> EspB [8–296] Native MR
Data Collection		
Space group	P 3 <sub>2</sub> 2 1	P 6 <sub>5</sub>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	125.75, 125.75, 75.06	148.29, 148.29, 49.69
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120	90, 90, 120
Resolution (Å) <sup>a</sup>	61.8–2.41 (2.51–2.41)	48.5–2.80 (2.95–2.8)
CC <sub>1/2</sub>	0.999 (0.972)	0.997 (0.845)
<i>R</i> <sub>pim</sub>	0.026 (0.137)	0.054 (0.309)
<i>R</i> <sub>meas</sub>	0.121 (0.627)	0.145 (0.852)
<i>I</i> / $\sigma$ <i>I</i>	22.8 (5.6)	12.9 (2.3)
Completeness (%)	100.0 (100.0)	97.6 (99.3)
Redundancy	21.8 (20.7)	7.1 (7.4)
Refinement		
Resolution (Å)	61.8–2.41	48.5–2.8
No. of reflections	26,584	15,287
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>b</sup>	0.1755/0.2011	0.2378/0.2718
No. of atoms		
Protein	2162	2039
Ligand/ion	0	0
Water	152	41
B factors (Å <sup>2</sup> )		
Protein	48	57
Ligand/ion	0	0
Water	51	64
Rms deviations		
Bond lengths (Å)	0.008	0.010
Bond angles (°)	1.033	1.015
Ramachandran		
% Favored	99	96
% Allowed	1	4
% Outliers	0	0
Molprobit		
Clashscore	6	10

<sup>a</sup>Values in parentheses represent the highest-resolution shell.

<sup>b</sup>10% of reflections were excluded from refinement and used to calculate *R*<sub>free</sub>.

### 3.3.2 Structure of EspB's secretion signal

The EspB<sub>Sm</sub> crystal structure also reveals the unique relative disposition of multiple characteristic ESX sequence motifs including a structured view of the helical “export arm”. **Fig. 3.3a, middle**). In this work, we define the export arm as EspB<sub>Sm</sub> residues EspB 78-93, corresponding to the appendage that is dynamic in the EsxAB NMR structure (**Fig 3.3a, left**) and lacking electron density in the PE25/PPE41 structure (**Fig 3.3a, right**). The prevalent view is that residues from the export arm align on the same helical face upon binding the EccCb1 ATPase, targeting ESX-secreted substrates for secretion (Callahan et al., 2009; Champion et al., 2006; Daleke et al., 2012a). In keeping with this hypothesis, residues within the export arm of EspB are identical in sequence to those shown to be essential for EsxB<sub>Tb</sub> interaction with EccCb1 (Champion et al., 2006), and either mutation of Y78 in the YxxxD motif (Daleke et al., 2012a) or deletion of the EccCb1 ATPase (McLaughlin et al., 2007) similarly disrupts EspB secretion. In the EspB<sub>Sm</sub> P3221 structure, this arm takes the form of a fully stabilized helix where export arm residues Y78, D81, L89, S90, and M93 indeed align on the same helical face to interact with a symmetry-related molecule in the P3221 crystal, representing a disposition that we suggest mimics the EccCb ATPase-bound state (**Fig. 3.3a, middle** and **Fig. S1a**). The alpha helical pattern is similar to that seen in recent Esx structures (Arbing et al., 2013; Poulsen et al., 2014), in particular to the *Streptococcus agalactiae* homodimeric EsxA structure determined by Poulsen and colleagues. Notably, the portion of this arm containing L89/S90/M93 was not resolved in the electron density of the P65 structure of EspB<sub>Sm</sub>, which has no such stabilizing crystal contacts, indicating that in the unbound conformation, a portion of EspB<sub>Sm</sub>'s export arm adopts a similar dynamic state as EsxB (**Fig. 3.3d**).



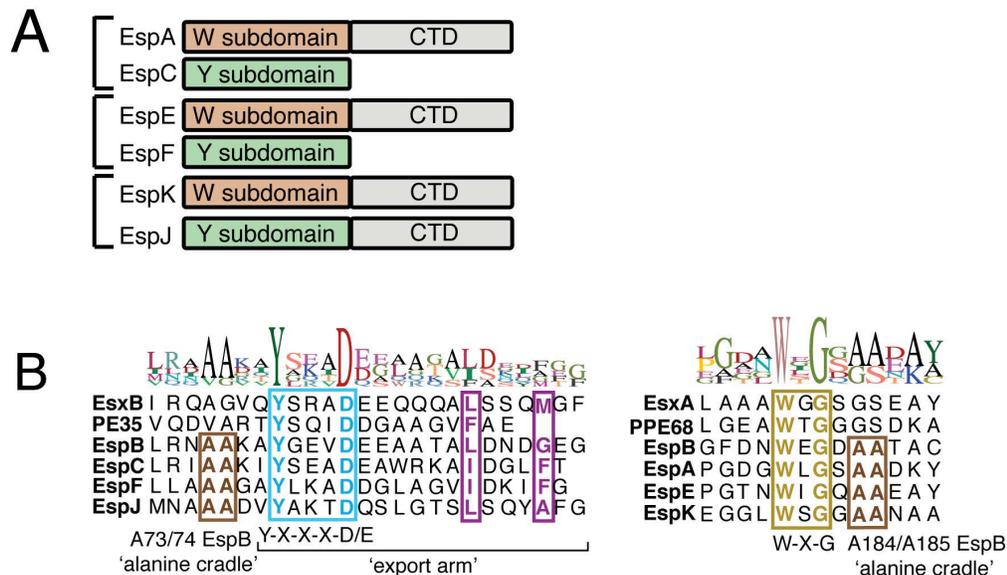
**Figure 3.3. Structure of EspB<sub>sm</sub>'s stabilized bipartite composite secretion signal.** (a) Comparison of secretion signal structures across the major ESX-secreted protein classes. 28 structures of the EsxA/B NMR ensemble (1WA8) and PE25/PPE41 (2G38) crystal structures are shown. (b) *M. smegmatis* murine dihydrofolate reductase protein fragment complementation assay demonstrates *in vivo* interaction between EspB<sub>sm</sub>[70-100] 'export arm' and the second FtsK-homology domain of EccCb1<sub>sm</sub>, residues [341-617] (c) Compared to EsxA/B, EspB has a flipped WxG motif that allows the W180 indole nitrogen to form a direct hydrogen bond with the Y78 side chain hydroxyl. The interaction is stabilized by conserved hydrophobic residues that form an 'alanine cradle' (d) Dynamic export arm in the P65 structure (e) The composite secretion signal creates a continuous hydrophobic surface that interacts with the N-terminal portion of the construct in the P3221 structure

To determine if EspB<sub>sm</sub>'s export arm interacts with EccCb1<sub>sm</sub> *in vivo*, we carried out a murine dihydrofolate reductase (mDHFR) protein fragment complementation assay (Singh et al., 2006). EspB<sub>sm</sub> residues 70 to 100 which encompass the 'export arm' were fused to the mDHFR-F[1,2] fragment. The second FtsK-like ATPase domain of EccCb1, residues 341 to 617, was fused to the mDHFR-F[3] fragment. If the fused proteins interact, the F[1,2] and F[3] halves of mDHFR associate and enzyme activity is reconstituted, allowing

*M. smegmatis* to survive on media containing the antibiotic trimethoprim. A strain expressing known interacting proteins, EsxB-F[1,2] and EsxA-F[3], served as a positive control, and growth was observed (**Fig. 3.3b**). For negative controls, EspB<sub>Sm</sub>-F[1,2] and EccCb1-F[3] were co-expressed with GCN4-F[3] and GCN4-F[1,2], respectively. GCN4 is a protein not associated with ESX-1 and no interaction is expected, and indeed, no growth for these strains was observed (**Fig. 3.3b**). The experimental strain expressing EspB<sub>Sm</sub>-F[1,2] and EccCb1-F[3] did exhibit growth indicating an interaction, but to a lesser extent than the positive control (**Fig. 3.3b**). The strength of interaction is difficult to evaluate from this assay. The structural and *in vivo* interaction data are thus compelling evidence that EspB and EsxB are secreted through ESX-1 by a similar export mechanism involving recognition of a conserved stabilized helix by the EccCb1 ATPase.

Recent secretion/mutagenesis experiments with the *Bacillus subtilis* homodimeric ESX substrate YukE suggests the WxG motif works in conjunction with the YxxxD motif adjacent to the stabilized export arm (Sysoeva et al., 2014). The WxG motif has also been shown to be indispensable for EspA secretion (Chen et al., 2013a) and this is presumably the case as well for EspB. Remarkably, in the EspB<sub>Sm</sub> crystal structure, Y78 from the YxxxD motif and W181 from the WxG motif are seen interacting directly (**Fig. 3.3c**). A survey of Esx protein structures deposited in the PDB indicates this is the first observation in which both elements of the Esx bipartite signal sequence are in direct contact. In our EspB<sub>Sm</sub> structure, the WxG motif is uniquely flipped towards the YxxxD motif within the export arm such that the Y78 side chain hydroxyl forms a direct hydrogen bond with the W180-N $\epsilon$  side chain indole nitrogen. Furthermore, the flip of the WxG is dependent on G182 within the motif adopting dihedral angles that can only be assumed by glycine (**Fig. 3.3c**). The interaction between Y78 and W180 is specifically stabilized by surrounding hydrophobic residues - primarily through a conserved "alanine cradle" formed by side chains of L12, A73, and A74 which interact with the Y78 aromatic ring, and L12, A184, A185, L286 which interact with the indole ring atoms of W181 (**Fig. 3.3c**). These surrounding hydrophobic residues appear to stabilize a configuration in which the Y78 and W181 side chains form a perpendicular, aromatic surface that is observed in both

crystal forms, independent of crystal contact influence (**Fig. 3.3d**). When structures from the two EspB<sub>Sm</sub> crystal forms are compared, this aromatic surface is variably bound to hydrophobic side chains of Y6 in P3221 structure or L294 in the P65 structure, suggesting this surface may have indiscriminately bound the adjacent free N and C terminal regions of the constructs that were crystallized (**Fig. 3.3d**). In a physiological context, it is possible that rather than binding these termini, the stabilized export arm and WxG loop may operate as a unified protein-protein interaction surface that is poised to interact with the EccCb ATPase (Sysoeva et al., 2014).



**Figure 3.4. Secretion signals in other Esp proteins** (a) Presence of secretion signal motifs in other Esp proteins of the ESX-1/EspACD locus. (b) Multiple sequence alignment of the bipartite composite signal sequence-containing proteins in the *Mtb* ESX-1/EspACD locus.

We next investigated if this composite secretion signal is a general feature of other Esp proteins in the ESX-1/EspACD loci. Using the structural homology server PHYRE2 (Kelley and Sternberg, 2009) we established that EspC, EspF and EspJ each have predicted Y subdomains, and EspA, EspE, and EspK each have predicted W subdomains fused to CTDs (**Fig. 3.4a**). Furthermore, the residues forming the alanine cradle that stabilize the interaction between the WxG and YxxxD motifs in the EspB<sub>Sm</sub> structure are conserved

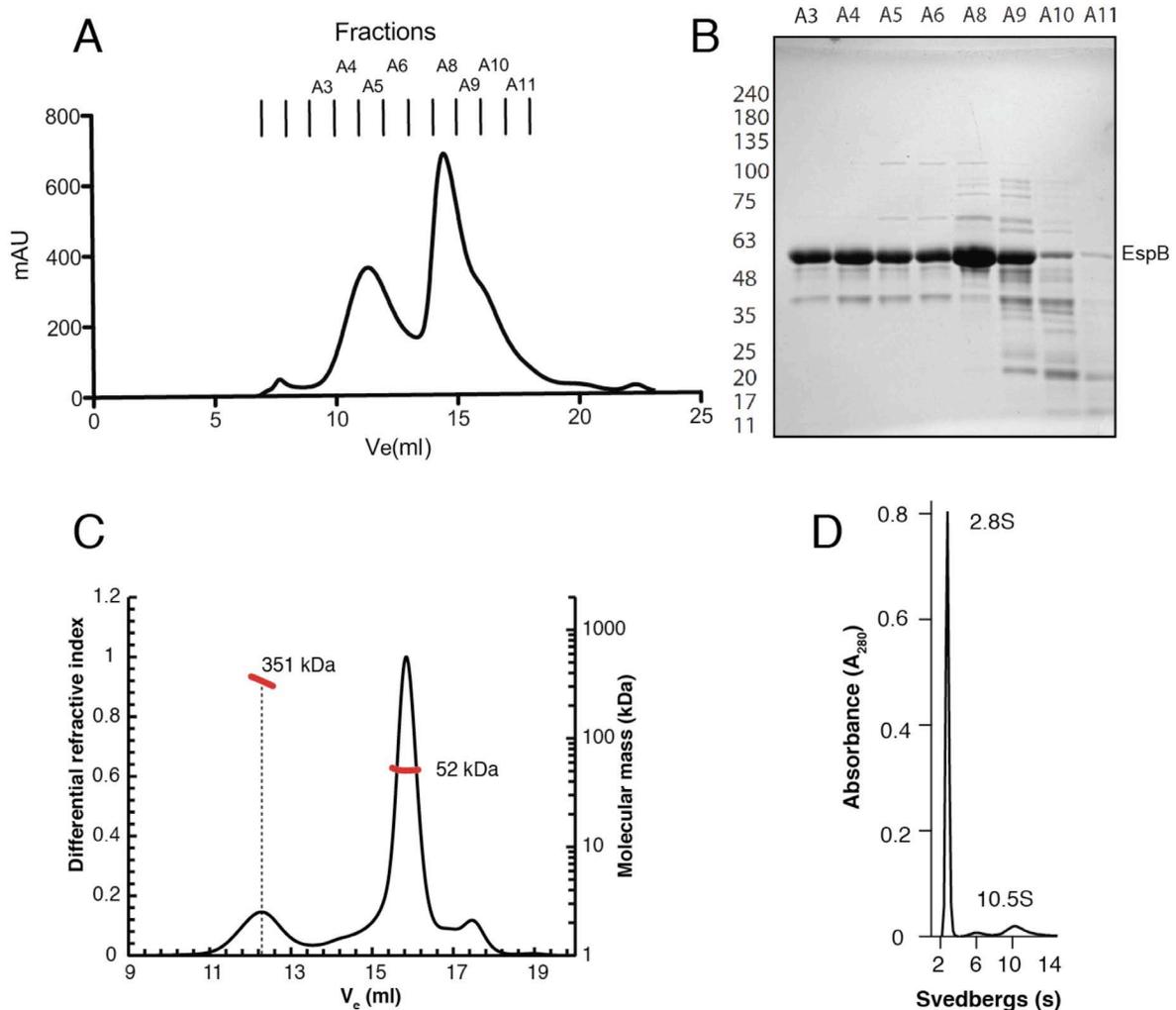
across these Esp proteins (**Fig. 3.4b**). As EspA and EspC, EspE and EspF, and EspJ and EspK are tandemly arrayed in a similar manner as known ESX heterodimers (**Fig. 3.1a**), it is tempting to speculate that these pairs also associate with similarly positioned export features as EspB, and that these other Esp proteins likely belong to the WXG100 superfamily.

### 3.3.3 Biophysical analysis of EspB quaternary structure

While our structural studies of the N-terminal domain of *M. smegmatis* EspB<sub>sm</sub> provides insight into the molecular basis of ATPase-mediated export, subsequent analysis of our full-length *Mtb* EspB (EspB<sub>tb</sub>-[1-460]) revealed that EspB has the capability of forming higher-order oligomeric states. Immediate clues of EspB oligomerization arose during purification, with protein eluting on a gel filtration column at multiple retention times, suggesting that EspB can self-associate in a concentration-dependent manner (**Fig. 3.5a,b**). To more accurately determine the subunit stoichiometry of the higher molecular species, we carried out size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis. The SEC-MALS experiment detected a monodisperse peak of molecular weight 52 kDa (corresponding to mass of a single EspB<sub>tb</sub>-[1-460] monomer) as well as a polydisperse high-molecular weight peak with a molecular mass of 351 kDa at its centre, equivalent to seven EspB copies (**Fig. 3.5c**). In agreement with the SEC-MALS data, our analytical ultracentrifugation sedimentation velocity experiments identified a low molecular weight 2.1S species ( $44 \pm 5$  kDa) and a high-molecular weight species with a sedimentation coefficient of 10.5S ( $317 \pm 41$  kDa) (**Fig. 3.5d**).

To characterize the structural properties of the EspB<sub>tb</sub>-[1-460] oligomer, we next subjected the peak fraction from our gel filtration run to negative stain electron microscopy analysis. Raw images show ring-shaped structures that were evenly distributed across the grid (**Fig. 3.6a**). Indeed, over 90% of the two dimensional (2D) class averages obtained from reference-free classification of 2,733 particles show an overall circular ring shape with several displaying unfused or incomplete rings (**Fig. 3.6b,d**). To determine which part of EspB was responsible for ring formation, we generated a truncated construct

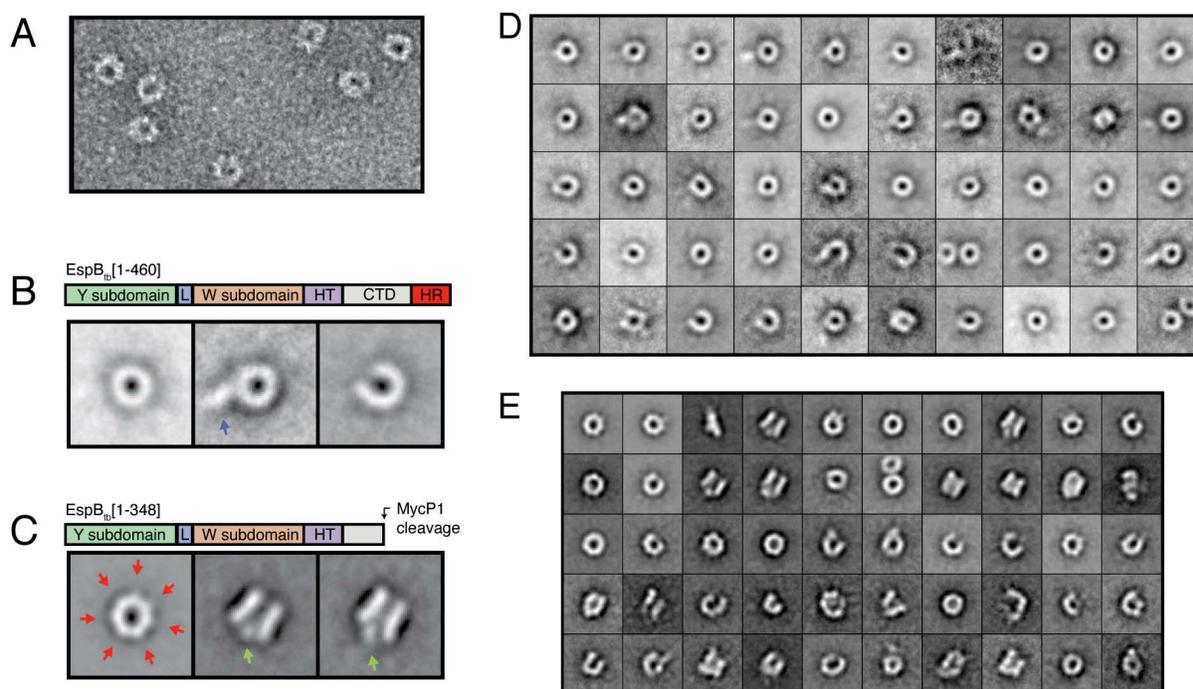
EspBtb[1-348] that approximates the previously determined position of the mycosin protease (MycP1) cleavage site and thus mimics the physiologically processed form (Solomonson et al., 2013). This truncation mutant, which encompasses the N-terminal PE/PPE homology domain and a portion of the low-complexity region of the CTD, still produced ring-shaped structures, indicating that it is the PE/PPE homology region of EspB forming the ring (**Fig. 3.6c,e**).



**Figure 3.5. Purification and size determination of EspBtb[1-460].** (a) EspB elutes with a bimodal distribution on a Superose 6 column. (b) SDS-PAGE analysis confirms EspB as the major species in the higher-order peak. (c) SEC-MALS analysis of EspBtb[1-460] (d) Analytical ultracentrifuge analysis of EspBtb[1-460].

However, the truncated EspBtb-[1-348] shows a more diverse set of orientations compared to full length EspBtb-[1-460], with 20% of the class averages depicting a tubular shape that appears to represent a side view as its width is consistent with that estimated from the top view averages (**Fig. 3.6c**). Even with no imposed symmetry, projection averaging of the “ring” views clearly showed seven distinct “nodes” (**Fig. 3.6c, red arrows**), which is fully consistent with our SEC-MALS and analytical ultracentrifugation data. Notably, a distinct appendage was seen protruding from the ring’s core in some of the top view EspBtb-[1-460] averages (**Fig. 3.6b**, blue arrowhead). This structure was absent in the EspBtb-[1-348] averages, suggesting that this region corresponds to EspB residues 350-460. However the EspBtb-[1-348] averages did reveal some unidentified density located at one end of the putative side view, which may correspond to the residual 50 CTD amino acids (**Fig. 3.6c**, green arrowhead).

We next determined the 3D reconstruction of the EspBtb-[1-348] multimer using an approach that involves generating an initial model by rotating a side view class average 360° about the Z-axis and iteratively refining this initial model against untilted data containing top and side views. The final map, with 7-fold symmetry imposed, has a resolution of 30 Å as measured by the Fourier shell correlation (FSC) function using the 0.5 FSC criterion (EM Data Bank accession code EMD-6120) (**Fig. 3.7a**). This EM structure shows that seven copies of EspBtb-[1-348]’s PE/PPE-homology domain associate to form an overall cylindrical shape of 100 Å by 80 Å with a central pore of 50 Å in diameter.



**Figure 3.6. Electron microscopy characterization of EspB** (a) Representative negative stain EM micrograph of full length EspB<sub>tb</sub>[1-460] (b) 2D class averages of EspB<sub>tb</sub>[1-460] and (c) truncated EspB<sub>tb</sub>[1-348]. Red arrows draw attention to distinct nodes in the unsymmetrized averages that provide evidence for heptameric stoichiometry. The blue arrow indicates appendage protruding from the ring's core found only in EspB<sub>tb</sub>[1-460] averages. Green arrows indicate unexplained density in the side view. (d) Complete reference-free 2D averages of EspB<sub>tb</sub>[1-460] (50 averages, 2733 particles) and (e) EspB<sub>tb</sub>[1-348] (50 averages, 3257 particles).

### 3.3.4 Data-guided computational model of EspB oligomer

To determine the molecular details of how the monomeric EspB structure assembles into a heptameric complex, we examined the P3221 and P65 EspB<sub>sm</sub> crystal lattices for potential interaction interfaces. The P3221 crystal exhibits a remarkable hexagonal lattice in which EspB monomers interlock in a herringbone pattern formed by tight side-by-side stacking at the hexagonal edges and tip to tail contacts between export arm residues and the helical tip. However, the potential oligomeric structures generated by these symmetries do not match the dimensions estimated from the EspB<sub>tb</sub> EM structure. We next carried out chemical crosslinking coupled to mass-spectrometry (CXMS) using the isotopically coded cleavable affinity-purifiable crosslinker cyanur-biotin-dimercapto-

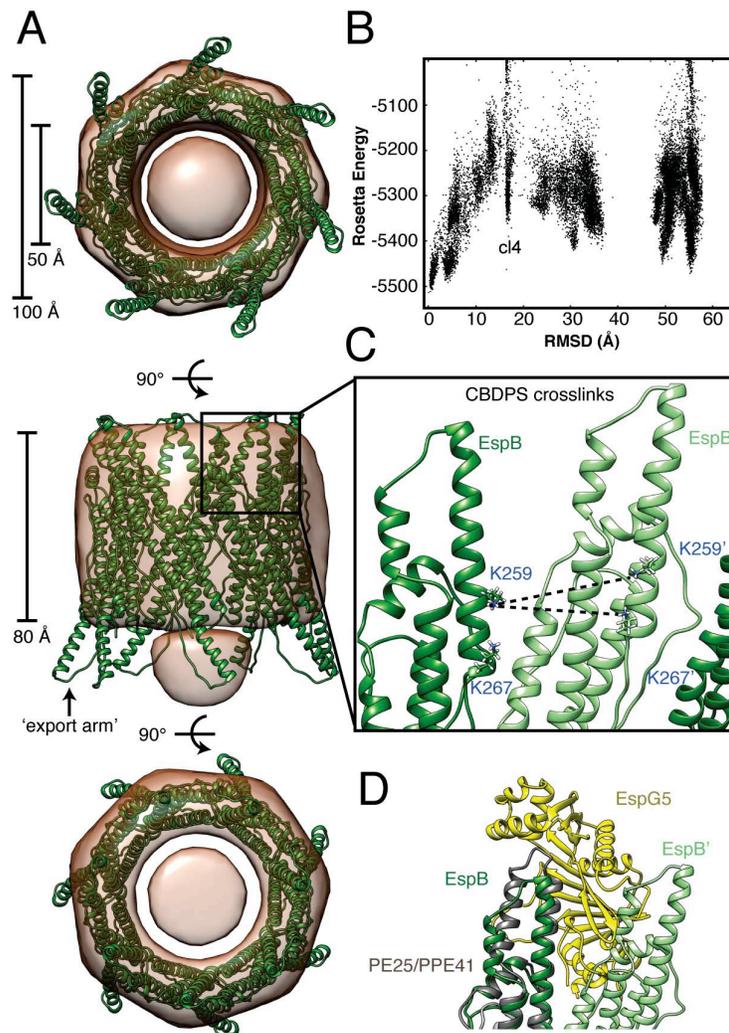
propionyl-succinimide (CBDPS). In multiple independent experiments, we identified CBDPS crosslinks between K259-K259' and K259-K267' peptides (**Table 3.2**), indicating these residues are located within 25 Å from one another. Our CXMS data definitively shows that intermolecular interactions based on crystal contacts are different from the ones mediating EspB oligomerization.

**Table 3.2. Identified inter-protein CBDPS crosslinks of EspB<sub>tb</sub>-[1-348] and EspB<sub>tb</sub>-FL.** L, light isotopic form of the crosslink. Δ, mass difference between theoretical and observed crosslink masses. Peptide sequences are shown with preceding and following residues in the protein sequence in parentheses. Fragment ion, intensity scores and CID-cleavage products number reflect crosslink MS/MS spectrum match quality. Tr, trypsin, PK, proteinase K. Reconfirmation of crosslinks as intra- or inter-protein origin with <sup>14</sup>N/<sup>15</sup>N crosslinking experiments is indicated.

Deconvoluted mass (MH <sup>+</sup> ) (L)	Δ, ppm	Charge	Residue numbers (1)	Crosslinked residue (1)	Peptide sequence (1)	Residue numbers (2)	Crosslinked residue (2)	Peptide sequence (2)	Fragment ion score	Intensity score	CID-cleavage products	Modification	Digestion	<sup>14</sup> N/ <sup>15</sup> N Inter/Intra
<b>EspB<sub>tb</sub>-[1-348]</b>														
3157.43919	-0.1	2	257-267	259	(R)SEKVLTEYNNK(A)	257-267	259	(R)SEKVLTEYNNK(A)	85	42	4	CBDPS	Tr	Inter
1242.55321	0.7	2	259-261	259	(E)KVL(T)	265-267	267	(Y)NNK(A)	188	9	4	CBDPS	PK	Inter
1543.74107	1.2	2	257-262	259	(R)SEKVLT(E)	259-261	259	(E)KVL(T)	107	16	4	CBDPS	PK	Inter
1644.79063	0	2	257-262	259	(R)SEKVLT(E)	259-262	259	(E)KVL(T)	150	18	4	CBDPS	PK	Inter
<b>EspB<sub>tb</sub>-FL</b>														
3173.43207	0.3	4	257-267	259	(R)SEKVLTEYNNK(A)	257-267	259	(R)SEKVLTEYNNK(A)	185	50	4	CBDPS +10x	Tr	Inter

We decided to carry out *in silico* Rosetta-based symmetric modeling with an EspB<sub>tb</sub>[11-286] homology model using an approach we have previously developed (Bergeron et al., 2013). Simulations with a stoichiometric constraint of seven resulted in model convergence to several low energy clusters (**Fig. 3.7b**). The fourth-lowest energy cluster was selected as the final model as it satisfies constraints from our CXMS data (density-fitted model PDB code: 3J83, **Fig. 3.7c**). The calculated model-map correlation was 0.9, indicating good agreement between the heptameric EspB<sub>tb</sub> model and the EM reconstruction (**Fig. 3.7a**). Regions of EspB lying outside the EM density correspond to the dynamic export arm, which, according to the P65 EspB crystal structure, are unlikely to be

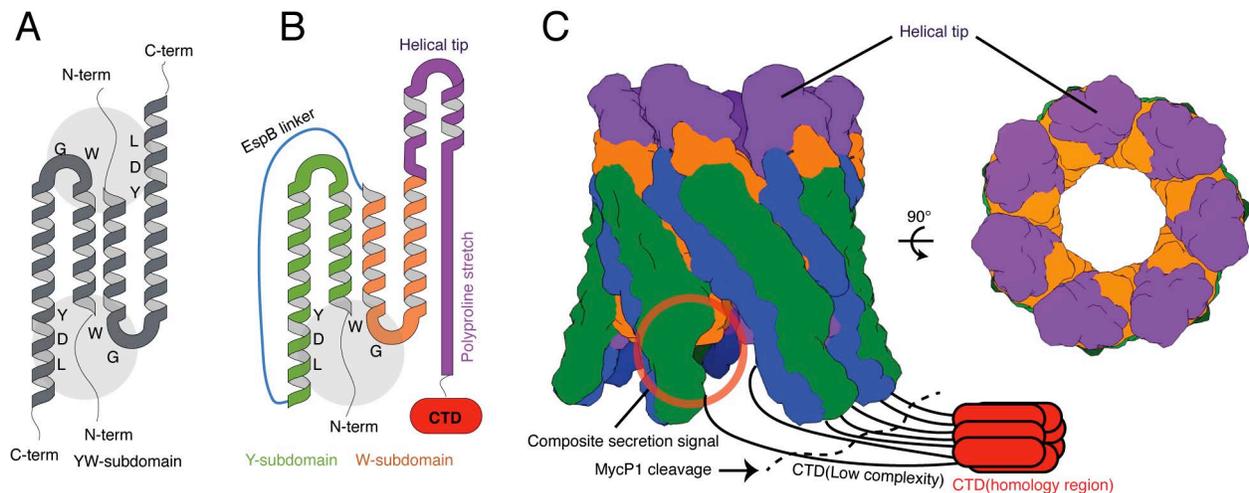
well-ordered in the EM reconstruction. The model places the C-terminal region of the EspB<sub>tb</sub>'s helical domain in close proximity to the unexplained density located at one end of the multimer, which is consistent with this density corresponding to the residual 50 residues of the CTD. While further experiments will be necessary to refine the exact orientation and molecular interactions involved, this model provides a framework for delineating how EspB<sub>tb</sub> oligomerizes in the context of the ESX secretion apparatus.



**Figure 3.7. Heptameric EspB<sub>tb</sub> model fit to negative stain 3D reconstruction density with location of mass spectrometry-derived cross-links mapped to adjacent subunits.** (a) EspB negative stain 3D reconstruction with heptameric EspB<sub>tb</sub> model fit to the EM density, approximate dimensions indicated (b) Top Rosetta energies for EspB<sub>tb</sub> heptameric model (c) Locations of CBDPS-crosslinked lysine residues mapped to the proposed model. (d) Overlay of the PE25/PPE41/EspG5 ternary chaperone complex (PDB: 4KXR) on the heptameric EspB model.

### 3.4 Discussion

The EspB crystal structure provides new insight into the common structural characteristics shared by ESX substrates, as well as the elevated complexity of ESX-secreted proteins found in mycomembrane-encapsulated bacteria. Ancestral ESX-secreted proteins in Gram-positive bacteria likely originated as symmetric homodimers with WxG and YxxxD motifs located at both ends, and this is still observed in homodimeric substrates such as *B. subtilis* YukE (Sysoeva et al., 2014) (**Fig. 3.8a**). However, organisms such as *Mtb* encode more complicated heterodimeric ESX substrates, with each “monomer” possessing either a WxG motif or a YxxxD motif, but not necessarily both. This arrangement places the composite signal sequence at just one end of the helical bundle (**Fig. 3.8b**). In the case of EspB, a linker has merged the heterodimer into a single ORF. Furthermore, in EspB and other PPE proteins of mycobacteria, the canonical motifs on one end of the helical bundle have been phased out and replaced by a ‘helical tip’ (**Fig. 3.8b**).



**Figure 3.8 Schematic summary of ESX-secreted protein specialization in mycobacteria.** (a) Ancestral ESX substrates were homodimeric and displayed essential export motifs on both sides of the helical bundle with N/C termini on separate sides. (b) Mycobacterial ESX proteins such as EspB which are located in the cell walls of these organisms have become increasingly asymmetric, losing one of the bipartite signal elements and gaining a helical appendage (c) Proposed model of EspB quaternary structure, placing the helical tip and secretion signals on opposite sides.

The EspB structure reveals an unexpected hydrogen bonded interaction between the defining tyrosine and tryptophan side chains in the YxxxD and WxG motifs positioned in our structure on the same helical face as conserved acidic and hydrophobic residues of the stabilized export arm. We note the similarity of the stabilized export arm to the signal peptides of the secretory pathway, which also adopt a helical conformation when bound to lipids or translocon components (Briggs et al., 1986; Chou and Gierasch, 2005). We suspect the unique conformation of the WxG and YxxxD motifs observed in our crystal structure represents a functional disposition that is adopted by all ESX substrates at some critical point during their translocation. The conserved alanine cradle serves to stabilize this direct interaction, perhaps favouring a secretion-competent conformation of the composite signal. Our analysis also shows that the ESX-1-associated EspA, EspE, EspK have WxG motifs and EspC, EspF, EspJ have YxxxD motifs, and all have residues of the 'alanine cradle' that potentially support their direct interaction. We predict EspAC, EspEF, and EspJK are heterodimeric complexes that associate to ensure the presence of both secretion motifs in the complex with a similar structural configuration we observe in EspB, perhaps allowing some or all of these unique ESX-1 substrates to be secreted through the same pathway as the well-studied EsxAB heterodimer. Thus, we suggest these proteins be included in the WXG100 family, resulting in a total of five heterodimeric WXG100 pairs as well as one fused homomer (EspB) in the ESX-1 locus, which in pairwise combination, possess the complete composite secretion signal (**3.4b**).

Of particular interest are the mycobacteria-specific features of ESX-secreted proteins, such as the helical tip we observe in EspB and PPE proteins (**Fig. 3.9b**). The helical tip is relatively hydrophobic, raising the possibility of its involvement in traversing the mycomembrane or perhaps in facilitating interaction with an unidentified component within the mycomembrane that aids this process. This appendage may also have a role in oligomerization, particularly in light of recent studies that show EspG chaperones bind directly to the helical tip of PPE41 from the PE25/PPE41 complex, preventing self-aggregation (Daleke et al., 2012b; Ekiert and Cox, 2014; Korotkova et al., 2014). When the PE25/PPE41/EspG5 ternary complex (4KXR) is aligned to EspB in the context of our heptameric model, the EspG5 chaperone is situated at an identical interface as the cross-

linked lysines identified by our CXMS experiments (**Fig. 3.5d**). This raises the possibility that EspB may also have a yet to be identified chaperone that prevents premature self-association in the *Mtb* cytoplasm.

In the context of the ESX system as a whole, the finding that EspB<sub>tb</sub> can oligomerize is particularly exciting as it indicates the helical domains of ESX-secreted proteins may serve a structural role in addition to, or as part of, their role in facilitating export. Our proposed model places EspB's helical tips on one side of the multimer, opposite the composite secretion signals and CTD (**Fig. 3.8c**). This creates an arrangement of seven closely spaced hydrophobic helical tips, which in combination with the pore generated by the multimer, may contribute to a mycomembrane-puncturing event that facilitates EspB's path across the mycobacterial cell wall and/or carry out ESX-1-mediated phenotypes such as phagosome permeabilization within *Mtb*-infected macrophages (de Jonge et al., 2007; Hsu et al., 2003; Smith et al., 2008) or conjugal transfer of DNA in *M. smegmatis*. Does EspB promote these activities in isolation or rather in concert with other components of ESX-1, such as the remaining five WXG100 pairs encoded by the locus and/or the inner membrane apparatus? It is currently thought that the inner membrane-spanning portion of the ESX apparatus forms a large molecular weight complex of probable symmetric composition (Houben et al., 2012b), and EspB could contribute to the extracellular portion of this, perhaps anchoring to a tip of a membrane-spanning filament. That ESX-1's secreted proteins are extracellular components of the secretion apparatus is supported by the co-localization of apparatus and secreted components to the same pole of the bacteria (Carlsson et al., 2009; Wirth et al., 2011) and is further supported by the finding that many Esx/Esp proteins are co-dependent for secretion (Champion et al., 2009; Fortune et al., 2005). Moreover, expression of EspB with its endogenous promoter has been shown to be important in recovering the wild-type ESX phenotype, suggesting stoichiometric expression levels of EspB is key (Xu et al., 2007). It should perhaps not be ruled out that *in vivo*, EspB may form structures of a higher order than cyclic heptamer to extend its span to the extracellular space, for example through helical assembly. The extent to which ESX-

secreted proteins structures resemble the filamentous protein building blocks from other secretion systems has been noted (Pallen, 2002).

Finally, what might our data reveal about the function of MycP1-mediated cleavage of EspB? By comparing 2D averages of the full-length and truncated EspB<sub>tb</sub> constructs, we speculate that MycP1 proteolysis serves to alter the structure of the EspB<sub>tb</sub> multimer. In the top view averages for EspB<sub>tb</sub>[1-460], we observed a region of density extending from the ring that we propose to be EspB<sub>tb</sub>'s CTD; this density was lacking in the truncated construct (**Fig. 3.6b, blue arrow**). A previous circular dichroism experiment showed the CTD is unstructured (Wagner et al., 2013a), but the EM data here suggests the CTD may become structured in the context of the multimer, perhaps through self-association of the CTD's conserved region (see **Fig. 3.2a**, red). In contrast, the EM averages for EspB<sub>tb</sub>[1-348], a construct that resembles processed EspB, exhibited a region of density located at the mouth of the pore in the side view averages where no such density was found in averages of the full length protein. If EspB indeed contributes to the secretion system channel through which other ESX substrates are also secreted, perhaps the region of the protein corresponding to this density shifts from the perimeter of the ring to occlude the centre of the pore following MycP1 cleavage. This hypothesis would explain the observation that in the absence of MycP1 cleavage, hypersecretion ensues (Ohol et al., 2010), as no "plug" is situated in the pore. Alternatively, perhaps this density corresponds to a domain that mediates interaction with other secretion system components, allowing assembly to be regulated by MycP1 processing.

We have shown that EspB possesses a WXG100 bipartite secretion signal that accounts for its localization to the *Mtb* cell wall. Once exported to the cell wall, EspB appears to be equipped to operate as a structural building block, which we propose to underlie its involvement in ESX-1-mediated macrophage killing that is essential for *Mtb* disease progression. Further characterization of key *Mtb* secreted proteins such as EspB will likely accelerate TB vaccine development - a field where major advancements have previously depended on genetic manipulation of the ESX loci and their associated secreted antigens.

# Chapter 4: Structure of EccB1

## 4.1 Introduction

There are five conserved membrane proteins associated with the inner membrane of the type VII secretion apparatus: EccA, EccB, EccC, EccD, EccE, and MycP (Bitter et al., 2009a) (**Fig. 4.1a**). These components are thought to form a complex that exports cytosolic proteins across the multi-layered envelope of mycobacteria to the outer capsular layer and secretion media (Houben et al., 2012b). The structure and function of the apparatus complex is not well characterized, and mechanism by which it passages secreted proteins through these unique mycobacterial cell wall barriers is unclear. These barriers include the inner membrane, peptidoglycan, arabinogalactan, mycolic acid, and capsular layers (**Fig. 4.1a**).

While the architecture of the complete T7SS assembly is unknown, structures of some of the individual soluble domains are becoming available, and many have general-purpose folds and are structurally homologous to proteins from other systems. For example, EccA has an AAA+ ATPase domain with six tetratricopeptide repeats (Wagner et al., 2013b), and EccC exhibits sequence homology with the FtsK/SpolIII-like family of AAA+ ATPases (Gey Van Pittius et al., 2001; Pallen, 2002), which are involved in DNA translocation and chromosomal segregation (Aussel et al., 2002; Massey et al., 2006). The structure of the MycP protease revealed a subtilisin-like domain that processes proteins as they are secreted (Solomonson et al., 2013; Sun et al., 2013; Wagner et al., 2013a), and EccD is a multi-spanning integral membrane protein with 10 predicted TM helices and a ubiquitin-like cytosolic domain. However aside from the MycP protease, structures of the

apparatus components predicted to be in direct proximity to the outer cell wall barriers, namely EccB and EccE, are lacking.

EccB and EccE do not exhibit sequence homology to any structurally characterized proteins. Structural data may provide clues as to how these proteins contribute to T7SS export and will make steps toward understanding architecture and function of the assembled apparatus. In this chapter, we report the structural determination and analysis of the soluble domain of EccB1 from the *Mycobacterium smegmatis* ESX-1 system.

## 4.2 Methods

### 4.2.1 Purification of EccB1

The DNA encoding residues 66-479 from EccB1 (EccB1<sub>sm</sub>-66-479) from the *M. smegmatis* ESX-1 system was amplified from *M. smegmatis mc*<sup>2</sup> 155 genomic DNA by PCR and cloned into a modified pET28a(+) vector, in-frame with an N-terminal thrombin-cleavable histidine tag with the sequence MGSSHHHHHHHHSSGLVPRGSH. The plasmid was transformed into *Escherichia coli* BL21-CodonPlus (DE3)-RIPL competent cells. The protein was expressed in LB medium containing 25 µg/ml kanamycin and 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C overnight following induction at an OD<sub>600</sub> of 0.6. Cells were harvested by centrifugation and lysed using an Avestin C-5 homogenizer in buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, DNase, and cOmplete protease inhibitor tablets (Roche). Lysate was centrifuged at 45,000 RPM at 4°C for 1 hour. The supernatant was passed over HisPur Ni-NTA affinity resin (Thermo #88223). Protein was eluted with buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, and 250 mM imidazole. Eluate was concentrated and injected on a Superdex 200 HR 10/30 column (GE Healthcare) equilibrated in gel filtration buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, and the fraction containing EccB1<sub>sm</sub>-66-497 was pooled.

### 4.2.2 Crystallization and structure determination of EccB1

EccB1<sub>66-497</sub> protein was concentrated to 45 mg ml<sup>-1</sup> in gel filtration buffer. Plate-like crystals were grown by sitting drop vapour diffusion with protein mixed at a 1:1 ratio with 15% w/v PEG8000, 100 mM imidazole, pH 6.7, at a temperature of 20°C. The crystals were cryoprotected by soaking in 1.6x mother liquor with 20% glycerol and then flash-cooled in liquid nitrogen.

A native data set was collected at the Canadian Light Source Macromolecular Crystallography Facility using the 08B1-1 insertion device beamline and processed using iMOSFLM (Battye et al., 2011). The data were merged and scaled with AIMLESS (Bailey, 1994). Shortly after collection of our dataset, the *Mtb* EccB1 structure was deposited in the PDB (4KK7; unpublished) by Korotkov et al. This served as a molecular replacement search model for phasing our *Mycobacterium smegmatis* EccB1 native dataset using the Phaser software package (McCoy et al., 2007). Two molecules were located in the asymmetric unit. The structure was built and refined with Coot (Emsley et al., 2010), and PHENIX (Adams et al., 2010). Structure validation statistics were calculated with MolProbity (Chen et al., 2010). Regions of the molecule that are part of the construct but could not be built include chain A N-terminal residues 66-74, residues 119-127, and C-terminal residues 466-479. In chain B, N-terminal residues 66-75, residues 122-128, and C-terminal residues 461-479 could not be built.

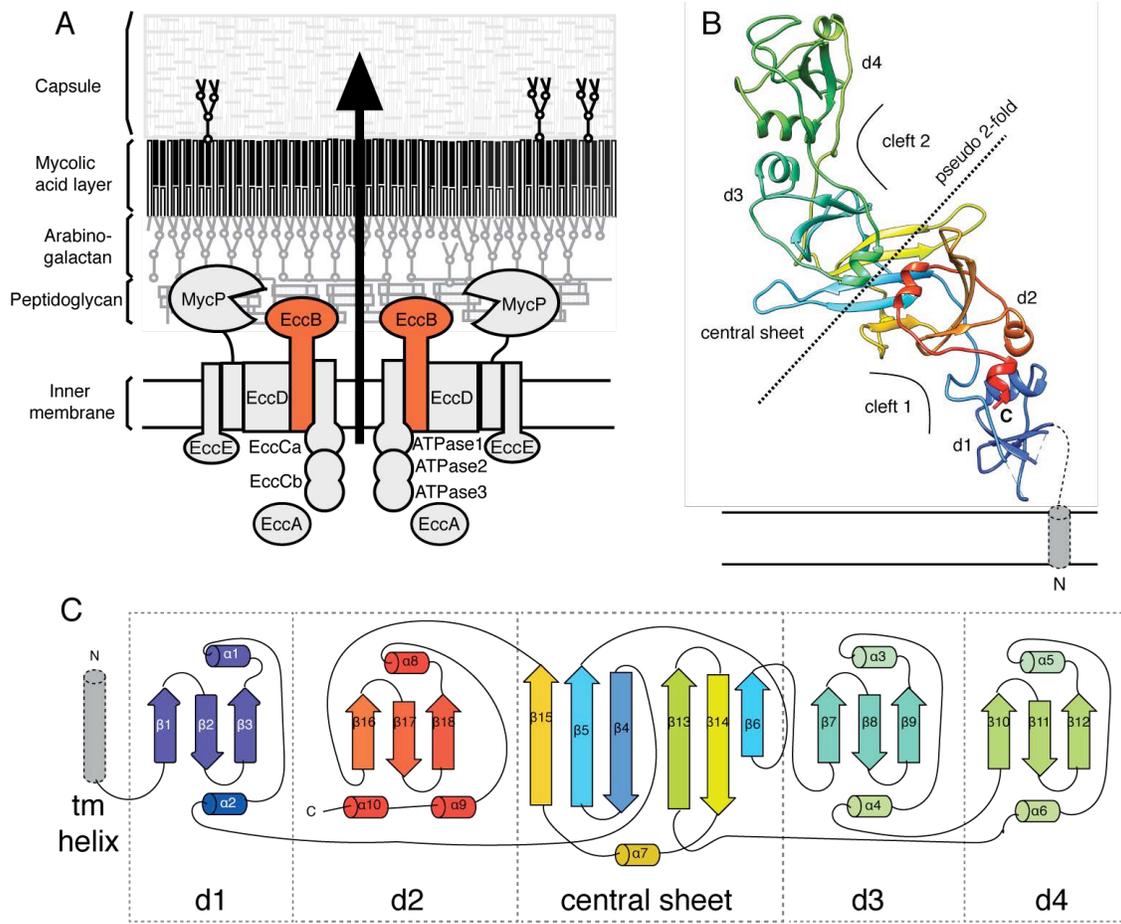
Structural homologues of EccB1<sub>sm</sub>[66-479] were identified using DALI (Holm and Rosenstrom, 2010). The structure was analyzed and figures generated with UCSF Chimera (Goddard et al., 2007). For the conservation analysis, homologues were located with hhblits (Remmert et al., 2012) and the results were mapped to the structure in Chimera. The surface electrostatics were calculated with PDB2PQR (Dolinsky et al., 2004) and the Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al., 2001).

## 4.3 Results

### 4.3.1 Overall structure of EccB

Transmembrane hidden markov analysis (Krogh et al., 2001) of *Mycobacterium smegmatis* EccB1 predict residues 1-39 to be cytoplasmic, residues 40-62 to form a single-spanning transmembrane helix, and residues 63-479 to be soluble and located outside the cell. We purified a stable construct spanning residues 66-479 of this soluble C-terminal region which crystallized in the orthorhombic space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, herein referred to as EccB1<sub>sm</sub>[66-479]. The 2.3 Å resolution structure was solved by molecular replacement using *Mtb* EccB1<sub>tb</sub>[74-458] (PDB code 4KK7; unpublished) as a search model and refined to R/R<sub>free</sub> = 0.21/0.25. As described in the methods, the N-termini, C-termini, and 2 loops are lacking electron density.

EccB1<sub>sm</sub>[66-479] adopts a ‘chair-shaped’ structure with internal 2-fold pseudo-symmetry (**Fig. 4.1b**). The curvature of the protein results in two distinct clefts (**Fig. 4.1b**, labelled cleft 1 and cleft 2). The C-terminus of EccB is located in close proximity to where the predicted N-terminal transmembrane helix would attach. In contrast to the recently deposited EccB1<sub>tb</sub> structure, EccB1<sub>sm</sub> crystallized with two molecules in the asymmetric unit related by a twofold axis of symmetry, resulting in an X-shaped unit. The two copies in the asymmetric unit have an 0.64 Å RMSD over 346 C $\alpha$  atoms. The PISA interface analysis tool (Krissinel and Henrick, 2007) calculated the buried surface area at the asymmetric unit dimer interface to be ~1337.5 Å<sup>2</sup> with 26 hydrogen bonds. However, the protein also migrates as a monomer on a gel filtration column. Thus, the biological relevance of the crystallographic dimer is unclear.



**Figure 4.1. Structure of *Mycobacterium smegmatis* EccB1<sub>sm</sub>[66-479].** (a). Schematic diagram of the ESX-1 type VII secretion membrane complex & mycobacterial outer membrane. EccB is highlighted in orange. (b). Structure of EccB1 coloured by N/C rainbow. The central beta sheet and four  $\alpha/\beta$  subdomains (labelled d1 - d4) are indicated (c). Secondary structure diagram.

Table 4.1. Data collection and refinement statistics for EccB1

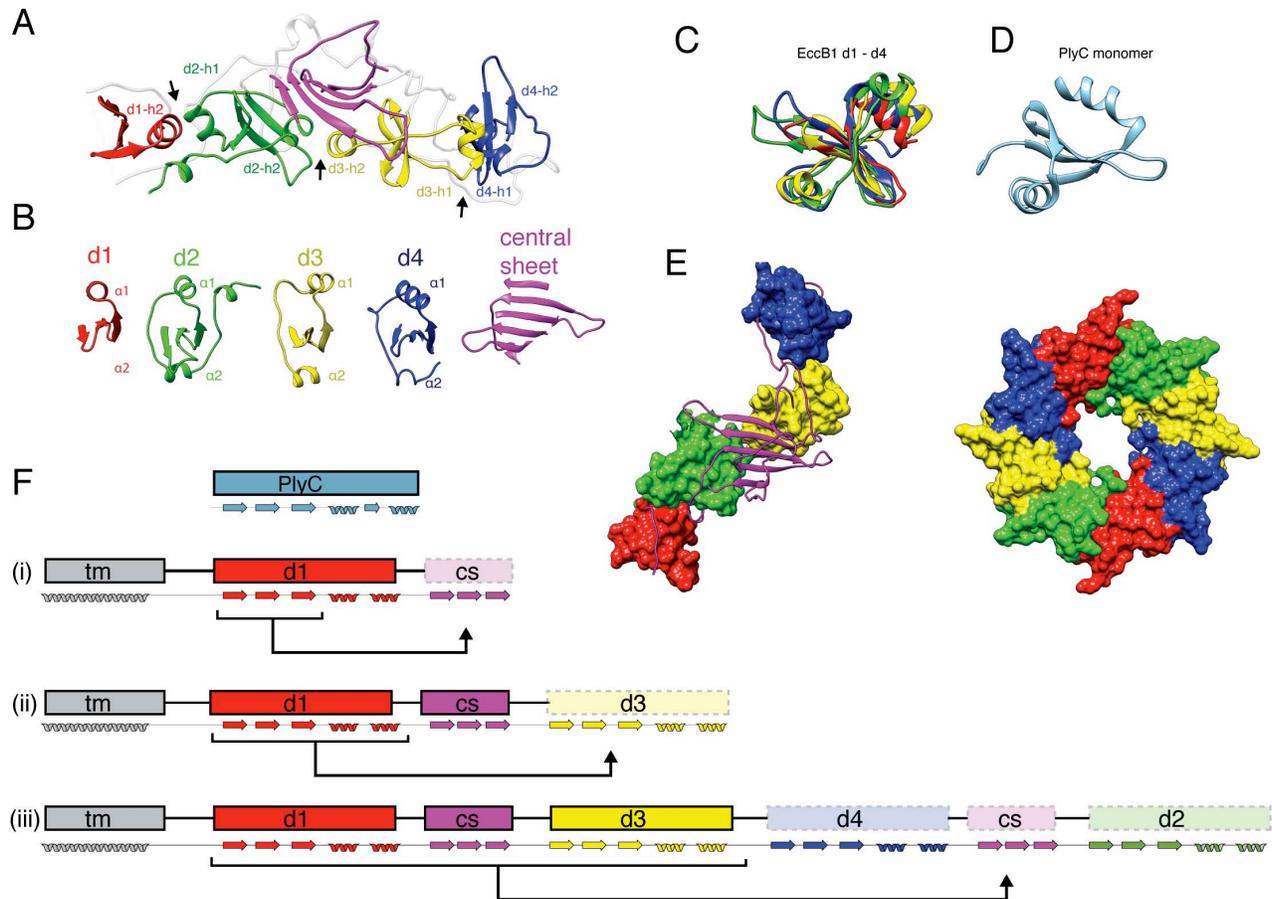
<b>Data collection</b>	<b><i>M. smegmatis</i> EccB1 [66-432]</b>
Space group	P 21 21 21
Wavelength	0.97895
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	62.19 115.02 125.42
$\alpha$ , $\beta$ , $\gamma$ (°)	90 90 90
Resolution (Å) <sup>a</sup>	44.16 – 2.3
$R_{\text{merge}}$	0.03742 (0.2362)
$I / \sigma I$	8.8 (2.3)
Completeness (%)	97.66 (99.05)
Redundancy	1.9 (2.0)
<b>Refinement</b>	
Resolution (Å)	44.16 – 2.3
No. reflections	87140
$R_{\text{work}} / R_{\text{free}}^{\text{b}}$	0.1817/0.2351
No. atoms	
Protein	5322
Ligand/ion	0
Water	414
Average B-factor (Å <sup>2</sup> )	39.1
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.22
Ramachandran	
% Favored	97
% Outliers	0.4
Molprobit	
Clashscore <sup>c</sup>	4.90

<sup>a</sup> Values in parentheses are for highest-resolution shell.

<sup>b</sup> 5% of reflections were excluded from refinement and used to calculate  $R_{\text{free}}$

### 4.3.2 Pseudosymmetric modularity of EccB

EccB1's 2-fold pseudo-symmetry is generated by four  $\alpha/\beta$  subdomains (labelled d1 through d4) centred around a 6-stranded beta sheet (**Fig. 4.1b and 4.1c**). The order in which these subdomains connect in the 3D tertiary structure does not strictly correspond to their locations in the polypeptide chain primary sequence. For example, while the d1 and d2 subdomains are adjacent in the tertiary structure, they are formed by N and C terminal regions of the primary sequence of the crystallized construct (**Fig. 4.1c**). Each of the four  $\alpha/\beta$  subdomains in EccB1<sub>sm</sub>[66-479] are comprised of three-stranded antiparallel beta sheets capped on both sides by short alpha helices (**Fig. 4.2a and 4.2b**). The two alpha helical caps are oriented roughly 90° from one another and are the primary mediators of inter-subdomain contacts (**Fig. 4.2a**, arrows). The four domains can be overlaid reasonably well, despite no detectable similarity in amino acid sequence (**Fig. 4.2c**).



**Figure 4.2. Pseudo-symmetric modularity of EccB tertiary structure.** (a) Core secondary structure elements from the five EccB subdomains highlighted by color. The rest of the polypeptide connecting these subdomains in grey. (b) Side-by-side comparison and (c) Overlay of the four EccB  $\alpha/\beta$  domain repeats. (d) Monomeric structure of PlyCb (4F87) (McGowan et al., 2012), the top DALI search result (Z-score 3.19). (e). Comparison of EccB's PlyCb homology domains with the PlyCb octamer. (f) Proposed evolution of EccB modularity.

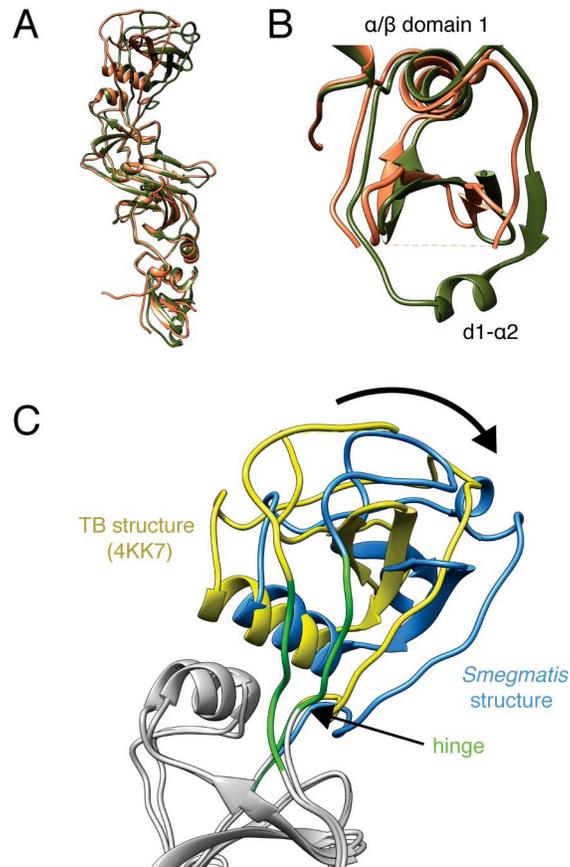
A structural similarity search of the PDB using the full EccB<sub>1sm</sub> as a search model in the DALI structural homology server (Holm and Rosenstrom, 2010) returned no significant hits. However, when an individual  $\alpha/\beta$  subdomain (the d2 subdomain from chain A) was queried, the results showed statistically significant structural homology to phage lysin PlyCb (Z-score 3.19) (PDB code 4F87) (**Fig 4.2d**). PlyCb and its partner PlyCa comprise a two-component phage lysin that ruptures the cell wall of *Streptococcus* species (McGowan et al., 2012). PlyCb is responsible for binding to the bacterial cell wall such that PlyCa, is positioned hydrolyze peptidoglycan.

By taking into account EccB1's modularity as observed in the crystal structure, and by examining how each modular unit is connected in the primary sequence of the polypeptide chain, it is possible to infer how EccB's subdomains might have arisen from a single  $\alpha/\beta$  unit through three sequential duplication/fusion events followed by genetic drift and sequence divergence. In the first event, the beta strands from the first ancestral  $\alpha/\beta$  subdomain could have duplicated to form half of the central strand (**Fig. 4.2f, i**). In the second event, the ancestral  $\alpha/\beta$  subdomain duplicated in its entirety to form a second  $\alpha/\beta$  domain (**Fig. 4.2f, ii**). In the third event, the whole unit duplicated to generate the full 2-fold pseudo-symmetric protein (**Fig. 4.2f, iii**). Note that events (i) and (ii) are not mutually exclusive and thus interchangeable in time; either could have been predecessors to the final event (iii). Genetic drift then led to loss of sequence similarity across the four subdomains. The central sheet appears to act as the "glue" that helps hold together halves at the pseudo-2-fold via alternating  $\beta$ -strand intercalation and a disulfide bond between the  $\beta_4$  and  $\beta_{13}$  strands (**see Fig. 4.1c**).

These observations suggest EccB's fold may have arisen through a series of duplication and fusion events followed by genetic drift to the point where each modular unit is unrelated in sequence but highly similar in tertiary fold. Some other examples of this type of pseudo-symmetric evolution are the trefoil protein (Broom et al., 2012), beta propeller domains (Chaudhuri et al., 2008), and gamma crystallin (Wolynes, 1996). These duplication events may have extended the protein's reach within the cell wall, perhaps providing surfaces that interact with other T7SS components, for example through the two clefts generated by the concave curvature of the 5 subdomains. An intriguing possibility, based on the DALI search results, is that a Gram-positive phage lysin PlyCb-like protein may have been the progenitor of EccB's  $\alpha/\beta$  subdomains. Gram-positive bacteria could have adapted this fold for T7SS-mediated export rather than the phage protein's lytic function. A similar proposal was put forth for the Gram-negative type VI secretion system, which has several components that are structurally homologous to bacteriophage tail proteins (Leiman et al., 2009; Pell et al., 2009).

### 4.3.3 Comparison with EccB-tb structure

There are a few notable differences between the EccB1<sub>tb</sub>[74-458] structure recently deposited by Korotkov et al (PDB: 4KK7) and the *M. smegmatis* structure determined here, despite these proteins having relatively high sequence identity (65%) (**Fig. 4.3a**). First, the  $\alpha$ 2 helix of the first  $\alpha/\beta$  subdomain was not resolved in the electron density (**Fig. 4.3b**). Second, there is a subtle change in position of the d4  $\alpha/\beta$  subdomain relative to the rest of the protein. To quantify this change, we analyzed the structures using the DynDom protein domain motion analysis server (Hayward and Lee, 2002) by threading the *M. smegmatis* sequence over the *Mtb* structure and evaluated the structures as separate conformers. DynDom calculated that the d4  $\alpha/\beta$  subdomain, formed by EccB1<sub>sm</sub> residues 244-333, is offset by a rotation angle of 18.7° and a translation of 3.7 Å (**Fig. 4.3b**). Residues 243-248 and 333-334 were identified as hinge residues that provide conformational flexibility for the differing positions. This difference is perhaps too large to be accounted for by sequence divergence alone, which at 65% would predict an RMSD of <1 Å across the entire sequence (Aloy et al., 2003), and may instead reflect an influence of crystal contacts on protein conformation. In the EccB1<sub>sm</sub> structure, the most extensive crystal contact from a symmetry-related molecule that could influence domain 4 positioning is generated by the  $-x+3/2, -y+1, z-1/2$  symmetry operator, creating a buried surface area of 804 Å<sup>2</sup>. On the other hand, in the EccB1<sub>tb</sub> structure there is significantly more packing around the d4 domain with symmetry-related molecules generated by  $x-1, y, z$  and  $x-1/2, -y+1/2, -z+1$  operators, which contribute 889 Å<sup>2</sup> and 664 Å<sup>2</sup> buried surface area, respectively. This subtle difference in the relative position of the d4 domain in the two structures may or may not be indicative of a conformational change in EccB that occurs during the T7SS export cycle or when bound to other components.

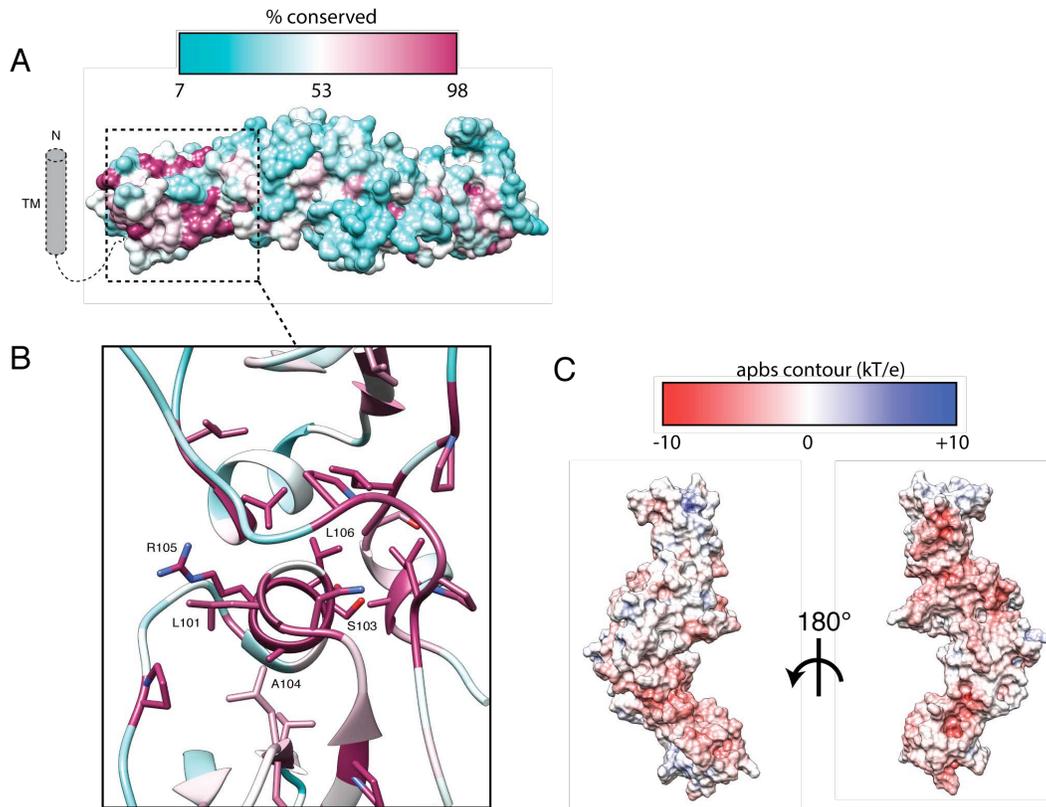


**Figure 4.3. Structural differences between *Mtb* and *Msmeag* EccB** (a). Superposition of *Mycobacterium smegmatis* and *Mtb* (4KK7) structures. (b). Absence of  $\alpha 2$  helix in domain 1. (c). Difference in domain 4 position seen in the *Mtb* and *M. smegmatis* crystal structures, relative to the rest of the protein.

#### 4.3.4 EccB surface conservation and electrostatics

To determine which EccB1 residues might be functionally important, we queried the *M. smegmatis* EccB sequence on the hhblits homology detection server (Remmert et al., 2012). The top 60 hits, mostly from mycobacteria and closely related species, were used to generate a multiple sequence alignment and the percent conservation level was mapped on the EccB1<sub>sm</sub> structure using UCSF Chimera (Goddard et al., 2007) (**Fig. 4.4a**). The most highly conserved residues are shown in purple, the most poorly conserved in teal. EccB1 exhibits the highest density of conserved residues in the region closest to where the predicted transmembrane-spanning helix would attach, in particular near the

interface between the d1 and d2  $\alpha/\beta$  subdomains (**Fig. 4.4a**). Many of these conserved residues are facing the solvent in cleft 1 (see **Fig. 4.1b**). A strictly conserved Ser-Ala-Arg-Leu (SARL) motif is observed on the  $\alpha 2$  helix. Ser103 and Ala104 appear to contribute to inter-subdomain contacts, and Arg105 and Leu106 are surface-exposed (**Fig. 4.4b**). Arg105 is notable in that it is largely buried except for the guanidinium group which pokes out adjacent to the Leu of the motif. On the face opposite the surface formed by Arg105/Leu106 is a tri-proline arrangement formed by conserved residues Pro134, Pro136, and Pro451. Together, these two strips of conserved solvent-exposed residues, located in close proximity to the predicted site of membrane attachment, are strong candidates for protein-protein interaction surfaces, and could for example, anchor to the EccD integral membrane protein, which is hypothesized to form an inner-membrane channel. In general, the protein is mostly acidic or neutral, as measured by Adaptive Poisson-Boltzmann calculations (Baker et al., 2001) (**Fig. 4.4c**).



**Figure 4.4. EccB conservation and surface electrostatics.** (a). Conserved residues mapped to the EccB1<sub>sm</sub> surface. A distinct conservational bias is seen at the d1-d2 interface, near the membrane spanning attachment. (b). Close-up view of the d1-d2 interface. (c). APBS surface electrostatics.

## 4.4 Conclusion

In conclusion, we have solved the structure of EccB1 from the ESX-1 type VII secretion system complex. An analysis of its structural modularity allows us to propose an evolutionary scenario. The four EccB1  $\alpha/\beta$  subdomains are structurally homologous to the Gram-positive phage lysin protein, PlyCb, suggesting a shared ancestry. EccB1 has a conserved residue patch in close proximity to where the single-spanning transmembrane helix is predicted to attach, representing a candidate surface for interaction with other components of the T7SS inner membrane complex.

# Chapter 5: Concluding Chapter

## 5.1 Summary

In this thesis, I have investigated three components from the type VII secretion system with the hope that structural information about these individual pieces could lead to insight into how the system functions as a whole. The MycP1 structure revealed a subtilisin-like fold with unique structural features I predict are adaptations that allow the protease to work in concert with the rest of the system, perhaps by mediating interactions with other proteins or by regulating activity. The MycP1 structure and cleavage site determination also provided insight into how it cleaves its target, EspB. The EspB study showed for the first time a very interesting attribute of the PE/PPE fold: the ability to self-associate. I predict this to be a general feature of the PE/PPE fold that is central to its trafficking function across the mycolic acid layer. The EspB structure also contributes to our understanding of the common structural requirements of secreted proteins across ESX systems, and I suspect the structures of other ESX-1 Esp proteins will to some extent resemble those of EspB, EsxAB, and the PE/PPE proteins. Finally, the EccB1 structure revealed that it is constructed from four  $\alpha/\beta$  subdomains that display structural homology to a phage lysin protein.

## 5.2 The two faces of MycP1

There are several logical routes to further investigation of MycP1. First, its dual role as a required ESX component as well as a protease has not been explained (David, 2010). Deletion of MycP1 abrogates secretion, suggesting it is bona fide apparatus component (Ohol et al., 2010). However, MycP5 was not detected in the purified ESX-5 complex (Houben et al., 2012b); perhaps it is loosely attached and was lost during purification.

Secondly, we know that MycP1 processes EspB, but to what extent is this protease a general “signal peptidase” of ESX systems? How many other secreted proteins might it process? Some ESX-5 substrates like LipY are also cleaved during secretion, and it was hypothesized this is carried out by the MycP5 protease from that system (Daleke et al., 2011); MycP5 is indeed essential to ESX-5 function system and for growth of pathogenic mycobacteria (Ates et al., 2015). We now know the common attribute of EspB and LipY is their dependence on a PE/PPE domain for transport. Perhaps PE/PPE domains are the vehicles required for transfer across the mycolic layer and the unique C-terminal domains carry out functions outside the cell; the mycosins serve to liberate the CTDs from their cognate PE/PPE transporters following translocation.

Inquiries into MycP1’s relevance as a potential drug target are ongoing. As a MycP1 catalytic mutant significantly reduces *Mtb* virulence in mouse models, it is hoped that a compound which inhibits MycP1 activity may similarly curb *Mtb* virulence (Ohol et al., 2010). Recent efforts to discover such a compound have included 4D fingerprinting (Hamza et al., 2014), the use of pentapeptide boronic acid inhibitors (Frasinyuk et al., 2014), and substrate profiling (Wagner et al., 2013a). Along with identifying an inhibitory compound, there are additional requirements for targeting MycP1: drugs must be delivered to the site of infection and they must traverse the mycomembrane.

In the quest for new preventative measures to combat TB, the MycP1/EspB system could also be exploited to engineer new *Mtb* strains that may lead to novel vaccines. As previously suggested, a MycP1 catalytic mutant *Mtb* strain that hyper-secretes ESX substrates may provide enhanced protective immunity through stimulation of cytosolic surveillance pathways (Ohol et al., 2010). Or perhaps the substrates themselves could be

engineered to perturb the infection cycle to favour of the host immune system. For example, point mutations in key ESX-1 substrates like EspA have deleterious effects on secretion and virulence; importantly, these effects are different than what is observed when the corresponding gene is completely disrupted (Chen et al., 2013a). Therefore, vaccines based on such point mutant strains may have a protective capacity that is different from, and hopefully better than, BCG. These ESX-1 point mutant strains have the potential to be more protective than BCG because a larger variety and quantity of ESX-1 antigens would be sampled by the immune system, coupled to the pathogen's exposure to cytosolic surveillance pathways as a result of the ESX-1-mediated cytosolic access, resulting in a stronger immune response. However, such novel vaccines must also be attenuated to the extent that they do not cause disease in vaccinated subjects. The structural data presented here could be used to guide further rational mutagenesis efforts aimed at perturbing the functions of EspB and MycP1, ideally resulting in a strain of *Mtb* that is attenuated for virulence but not to the extent that secretion is completely abrogated and the stimulatory effect of ESX-1 on the immune system lost. For example, point mutations in the MycP1 N-terminal extension or at EspB's PE/PPE interface could have a desirable destabilizing effect.

### 5.3 The PE/PPE fold

One of the central questions remaining in T7SS research: what is the molecular function of the PE/PPE fold? As discussed in **section 3.4**, the PE/PPE and EsxA/B families probably have a shared ancestry, but why have mycobacteria evolved this PE/PPE fold as part of its secreted protein repertoire? How can we elucidate the function of the PE/PPE fold *in vivo*? Proteins with this fold are generally difficult to express and purify; however, EspB is an exception. EspB should be regarded, along with PE25/PPE41, as a model ESX substrate that has the potential to provide key insight into the general function of the PE/PPE fold. For example, Chen et al. have already exploited this to discover EspB binds to specific phospholipids (Chen et al., 2013b). Do other PE/PPE proteins bind specific lipids as well? To better understand the function of this fold, perhaps we should focus on devising

labeling/imaging strategies to observe ESX-1-secreted proteins *in vivo*. For example, super-resolution microscopy combined with cryo-electron tomography could be used to locate, observe, and structurally characterize these secreted proteins *in situ*. Such approaches have recently led to breakthroughs in our understanding of the structure and function of the type VI secretion system (T6SS) (Basler et al., 2012).

## 5.4 Visualization of the T7SS apparatus

Full understanding of the T7SS system will require structural characterization of the assembled apparatus. Toward this goal, I can suggest three possible strategies. First, endogenous T7SS apparatus could be purified from a source under conditions in which it is naturally expressed. I argue the best candidate system for this is *M. smegmatis* ESX-1. *M. smegmatis* assembles ESX-1 under well-characterized conditions, can be grown in large quantities in biosafety level 1 facilities, and has an ESX-1 gene cluster that is nearly identical to that of *Mtb*, allowing inferences to be made about ESX-1 as it functions in the pathogen. To ensure the system is expressed and functional, robust methods to detect various apparatus components and secreted proteins are needed. For example, Houben et al. have made progress toward purifying the ESX-5 membrane complex of *M. marinum* by designing a suite of antibodies to detect its constituent proteins (Houben et al., 2012b). Tagging ESX-1 components with fluorescent proteins may also help facilitate structural studies, and these can be added with disrupting ESX-1 function (Carlsson et al., 2009; Wirth et al., 2011). A potential disadvantage of this approach might be low protein yield. Perhaps this could be addressed by over-expressing of regulatory transcription factors that increase ESX-1 protein quantity. Such an approach has been applied to increase yields of the *Salmonella* SPI1 type III secretion system by over-expressing the transcription factor HilA prior to purification of the system by differential centrifugation in a caesium chloride gradient (Eichelberg and Galan, 1999). Transcription factors that may increase expression of ESX-1 include EspR (Blasco et al., 2011; Raghavan et al., 2008) and the WhiB6/PhoP regulon (Frigui et al., 2008; Solans et al., 2014).

A second possible strategy for obtaining T7SS apparatus protein for structural studies is through heterologous expression of the complex in *E. coli* followed by tandem affinity purification and single particle cryo-EM analysis. Such an approach recently led to the first structural characterization of the type IV secretion system (T4SS) inner membrane complex (Low et al., 2014). These authors identified a plasmid-encoded T4SS under control of a single promoter which was replaced with a high-expression/inducible T7 promoter. They engineered the individual membrane components to have unique affinity tags, and after a series of affinity purification steps, were able to enrich sufficient quantities of homogeneously assembled membrane complex for single particle cryo-EM analysis. The tags were also used to approximate the positions of the constituent apparatus components in low-resolution EM maps using nanogold labeling. Could a similar approach be exploited to obtain T7SS apparatus protein for EM studies? It is difficult to identify T7SS gene clusters in which all the genes appear to be under control of a single promoter that could easily be replaced with T7. While rare for T7SS gene clusters, one such operon appears to encompass ESX-3, and another is found on the conjugation-related pRAW plasmid described in section **1.3.10** (Ummels et al., 2014). Alternatively, if a given ESX gene cluster is comprised of multiple operons, these operons could be cloned individually into separate plasmids or engineered as an artificial operon. It may also be worth identifying, cloning, and carrying out expression trials for a wide variety of ESX systems from non-mycobacterial actinobacteria. From a genetic perspective, some of these systems appear to be less complex than mycobacterial systems and are controlled by just one or a few promoters, reducing the technical challenges associated with engineering expression constructs. It may also be beneficial to utilize Gram-positive heterologous expression systems such as those developed for *Rhodococcus jostii* or *M. smegmatis*; these expression systems could provide a native-like cell wall environment better-suited for ESX system assembly.

Finally, it is possible that much could be learned from observing the T7SS structure *in situ* by cryo-electron tomography. Advances in sample preparation techniques (Villa et al., 2014), subtomogram averaging (Briggs, 2013), and correlated cryo-fluorescence spectroscopy (Chang et al., 2014; Schorb and Briggs, 2014) have made it easier to obtain

quantitative structural data for multi-protein complexes in their native context. It is now possible to differentiate between the various assembly states of large molecular machines in eukaryotic cells (Asano et al., 2015). Through these methods, perhaps the goal of making a “movie” of *Mtb* T7SS assembly and function as the infection progresses in eukaryotic cells will be achievable.

# Bibliography

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

Abdallah, A.M., Savage, N.D.L., van Zon, M., Wilson, L., Vandenbroucke-Grauls, C., van der Wel, N.N., Ottenhoff, T.H.M., and Bitter, W. (2008). The ESX-5 Secretion System of *Mycobacterium marinum* Modulates the Macrophage Response. *Journal of Immunology* 181, 7166-7175.

Abdallah, A.M., Verboom, T., Weerdenburg, E.M., van Pittius, N.C.G., Mahasha, P.W., Jimenez, C., Parra, M., Cadieux, N., Brennan, M.J., Appelmelk, B., *et al.* (2009). PPE and PE\_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Molecular Microbiology* 73, 329-340.

Abouzeid, C., Smith, I., Grange, J., Steele, J., and Rook, G. (1986). Subdivision of Daughter Strains of *Bacille Calmette-Guerin* (Bcg) According to Secreted Protein-Patterns. *Journal of General Microbiology* 132, 3047-3053.

Abouzeid, C., Smith, I., Grange, J.M., Ratliff, T.L., Steele, J., and Rook, G.A.W. (1988). The Secreted Antigens of *Mycobacterium-Tuberculosis* and Their Relationship to Those Recognized by the Available Antibodies. *Journal of General Microbiology* 134, 531-538.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* 66, 213-221.

Aloy, P., Ceulemans, H., Stark, A., and Russell, R.B. (2003). The Relationship Between Sequence and Interaction Divergence in Proteins. *J Mol Biol* 332, 989-998.

Andersen, P. (1994). Effective vaccination of mice against *Mycobacterium tuberculosis* infection with a soluble mixture of secreted mycobacterial proteins. *Infect Immun* 62, 2536-2544.

Andersen, P., Askgaard, D., Ljungqvist, L., Bennedsen, J., and Heron, I. (1991a). Proteins released from *Mycobacterium tuberculosis* during growth. *Infect Immun* 59, 1905-1910.

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

Andersen, P., and Heron, I. (1993). Specificity of a protective memory immune response against *Mycobacterium tuberculosis*. *Infect Immun* 61, 844-851.

Arbing, M.A., Chan, S., Harris, L., Kuo, E., Zhou, T.T., Ahn, C.J., Nguyen, L., He, Q., Lu, J., Menchavez, P.T., *et al.* (2013). Heterologous Expression of Mycobacterial Esx Complexes in *Escherichia coli* for Structural Studies Is Facilitated by the Use of Maltose Binding Protein Fusions. *PLoS One* 8, e81753.

Arbing, M.A., Kaufmann, M., Phan, T., Chan, S., Cascio, D., and Eisenberg, D. (2010). The crystal structure of the *Mycobacterium tuberculosis* Rv3019c-Rv3020c ESX complex reveals a domain-swapped heterotetramer. *Protein Science* 19, 1692-1703.

Armstrong, J.A., and Hart, P.D. (1975). Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. *J Exp Med* 142, 1-16.

Asano, S., Fukuda, Y., Beck, F., Aufderheide, A., Foerster, F., Danev, R., and Baumeister, W. (2015). A molecular census of 26S proteasomes in intact neurons. *Science* 347, 439-442.

Ates, L.S., Ummels, R., Commandeur, S., van der Weerd, R., Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S.R., Abdallah, A.M., *et al.* (2015). Essential Role of

the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLoS genetics* *11*, e1005190.

Aussel, L., Barre, F.-X., Aroyo, M., Stasiak, A., Stasiak, A.Z., and Sherratt, D. (2002). FtsK Is a DNA Motor Protein that Activates Chromosome Dimer Resolution by Switching the Catalytic State of the XerC and XerD Recombinases. *Cell* *108*, 195-205.

Bach, H., Papavinasasundaram, K.G., Wong, D., Hmama, Z., and Av-Gay, Y. (2008). Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. *Cell Host & Microbe* *3*, 316-322.

Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D-Biological Crystallography* *50*, 760-763.

Baird, P.N., Hall, L.M., and Coates, A.R. (1989). Cloning and sequence analysis of the 10 kDa antigen gene of Mycobacterium tuberculosis. *J Gen Microbiol* *135*, 931-939.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* *98*, 10037-10041.

Ball, L.J., Kuhne, R., Schneider-Mergener, J., and Oschkinat, H. (2005). Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew Chem Int Ed Engl* *44*, 2852-2869.

Barnes, P.F., Mehra, V., Hirschfield, G.R., Fong, S.J., Abou-Zeid, C., Rook, G.A., Hunter, S.W., Brennan, P.J., and Modlin, R.L. (1989). Characterization of T cell antigens associated with the cell wall protein-peptidoglycan complex of Mycobacterium tuberculosis. *J Immunol* *143*, 2656-2662.

Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012). Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* *483*, 182-186.

- Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67, 271-281.
- Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., and Small, P.M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284, 1520-1523.
- Bergeron, J.R., Worrall, L.J., Sgourakis, N.G., DiMaio, F., Pfuetzner, R.A., Felise, H.B., Vuckovic, M., Yu, A.C., Miller, S.I., Baker, D., *et al.* (2013). A refined model of the prototypical Salmonella SPI-1 T3SS basal body reveals the molecular basis for its assembly. *PLoS Pathog* 9, e1003307.
- Berthet, F.X., Rasmussen, P.B., Rosenkrands, 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-Uk* 144, 3195-3203.
- Bitter, W., Houben, E.N., Bottai, D., Brodin, P., Brown, E.J., Cox, J.S., Derbyshire, K., Fortune, S.M., Gao, L.Y., Liu, J., *et al.* (2009a). Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog* 5, e1000507.
- Bitter, W., Houben, E.N., Luirink, J., and Appelmelk, B.J. (2009b). Type VII secretion in mycobacteria: classification in line with cell envelope structure. *Trends Microbiol* 17, 337-338.
- Blasco, B., Chen, J.M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., Pojer, F., and Cole, S.T. (2011). Virulence regulator EspR of Mycobacterium tuberculosis is a nucleoid-associated protein. *PLoS Pathog* 8, e1002621.
- Bottai, D., Majlessi, L., Simeone, R., Frigui, W., Laurent, C., Lenormand, P., Chen, J., Rosenkrands, I., Huerre, M., Leclerc, C., *et al.* (2010). ESAT-6 Secretion-Independent Impact of ESX-1 Genes espF and espG(1) on Virulence of Mycobacterium tuberculosis. *Journal of Infectious Diseases* 203, 1155-1164.

Brennan, P.J., and Crick, D.C. (2007). The cell-wall core of *Mycobacterium tuberculosis* in the context of drug discovery. *Current Topics in Medicinal Chemistry* 7, 475-488.

Briggs, J.A.G. (2013). Structural biology in situ - the potential of subtomogram averaging. *Current Opinion in Structural Biology* 23, 261-267.

Briggs, M.S., Cornell, D.G., Dluhy, R.A., and Gierasch, L.M. (1986). Conformations of signal peptides induced by lipids suggest initial steps in protein export. *Science* 233, 206-208.

Brites, D., and Gagneux, S. (2015). Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunological Reviews* 264, 6-24.

Brodin, P., de Jonge, M.I., Majlessi, L., Leclerc, C., Nilges, M., Cole, S.T., and Brosch, R. (2005). Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of *Mycobacterium tuberculosis*, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. *Journal of Biological Chemistry* 280, 33953-33959.

Brodin, P., Majlessi, L., Marsollier, L., de Jonge, M.I., Bottai, D., Demangel, C., Hinds, J., Neyrolles, O., Butcher, P.D., Leclerc, C., *et al.* (2006). Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infection and Immunity* 74, 88-98.

Broom, A., Doxey, A.C., Lobsanov, Y.D., Berthin, L.G., Rose, D.R., Howell, P.L., McConkey, B.J., and Meiering, E.M. (2012). Modular Evolution and the Origins of Symmetry: Reconstruction of a Three-Fold Symmetric Globular Protein. *Structure* 20, 161-171.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., *et al.* (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99, 3684-3689.

Brown, G.D., Dave, J.A., Gey van Pittius, N., 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, 147-155.

Callahan, B., Nguyen, K., Collins, A., Valdes, K., Caplow, M., Crossman, D.K., Steyn, A.J.C., Eisele, L., and Derbyshire, K.M. (2009). Conservation of Structure and Protein-Protein Interactions Mediated by the Secreted Mycobacterial Proteins EsxA, EsxB, and EspA. *Journal of Bacteriology* 192, 326-335.

Cambier, C.J., Falkow, S., and Ramakrishnan, L. (2014). Host Evasion and Exploitation Schemes of *Mycobacterium tuberculosis*. *Cell* 159, 1497-1509.

Carlsson, F., Joshi, S.A., Rangell, L., and Brown, E.J. (2009). Polar Localization of Virulence-Related Esx-1 Secretion in Mycobacteria. *Plos Pathogens* 5.

Carlsson, F., Kim, J., Dumitru, C., Barck, K.H., Carano, R.A.D., Sun, M., Diehl, L., and Brown, E.J. (2010). Host-Detrimental Role of Esx-1-Mediated Inflammasome Activation in Mycobacterial Infection. *Plos Pathogens* 6.

Cascioferro, A., Delogu, G., Colone, M., Sali, M., Stringaro, A., Arancia, G., Fadda, G., Palu, G., and Manganelli, R. (2007). PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. *Molecular Microbiology* 66, 1536-1547.

Champion, M.M., Williams, E.A., Kennedy, G.M., and Champion, P.A.D. (2012). Direct Detection of Bacterial Protein Secretion Using Whole Colony Proteomics. *Molecular & Cellular Proteomics* 11, 596-604.

Champion, M.M., Williams, E.A., Pinapati, R.S., and Champion, P.A. (2014). Correlation of Phenotypic Profiles Using Targeted Proteomics Identifies Mycobacterial Esx-1 Substrates. *J Proteome Res.*

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* 73, 950-962.

Champion, P.A.D., and Cox, J.S. (2007). Protein secretion systems in Mycobacteria. *Cellular Microbiology* 9, 1376-1384.

Champion, P.A.D., 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, 1632-1636.

Chang, Y.-W., Chen, S., Tocheva, E.I., Treuner-Lange, A., Loebach, S., Sogaard-Andersen, L., and Jensen, G.J. (2014). Correlated cryogenic photoactivated localization microscopy and cryo-electron tomography. *Nature Methods* 11, 737-U179.

Chao, J., Wong, D., Zheng, X., Poirier, V., Bach, H., Hmama, Z., and Av-Gay, Y. (2010). Protein kinase and phosphatase signaling in *Mycobacterium tuberculosis* physiology and pathogenesis. *Biochim Biophys Acta* 1804, 620-627.

Chaudhuri, I., Soding, J., and Lupas, A.N. (2008). Evolution of the beta-propeller fold. *Proteins-Structure Function and Bioinformatics* 71, 795-803.

Chen, J.M., Boy-Roettger, S., Dhar, N., Sweeney, N., Buxton, R.S., Pojer, F., Rosenkrands, I., and Cole, S.T. (2011). EspD Is Critical for the Virulence-Mediating ESX-1 Secretion System in *Mycobacterium tuberculosis*. *Journal of Bacteriology* 194, 884-893.

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

- Chen, J.M., Zhang, M., Rybniker, J., Boy-Röttger, S., Dhar, N., Pojer, F., and Cole, S.T. (2013b). Mycobacterium tuberculosis EspB binds phospholipids and mediates EsxA-independent virulence. *Molecular Microbiology* 89, 1154-1166.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12-21.
- Chou, Y.T., and Gierasch, L.M. (2005). The conformation of a signal peptide bound by Escherichia coli preprotein translocase SecA. *J Biol Chem* 280, 32753-32760.
- Choudhary, R.K., Mukhopadhyay, S., Chakhaiyar, P., Sharma, N., Murthy, K.J.R., Katoch, V.M., and Hasnain, S.E. (2003). PPE antigen Rv2430c of Mycobacterium tuberculosis induces a strong B-cell response. *Infection and Immunity* 71, 6338-6343.
- Clark, M., Riben, P., and Nowgeseic, E. (2002). The association of housing density, isolation and tuberculosis in Canadian First Nations communities. *International Journal of Epidemiology* 31, 940-945.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., *et al.* (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393, 537-+.
- Collins, F.M., Lamb, J.R., and Young, D.B. (1988). Biological activity of protein antigens isolated from Mycobacterium tuberculosis culture filtrate. *Infect Immun* 56, 1260-1266.
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., *et al.* (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics* 45, 1176-U1311.

Comas, I., and Gagneux, S. (2009). The Past and Future of Tuberculosis Research. *Plos Pathogens* 5.

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, 1238-1245.

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, 794-808.

Cunningham, D., Danley, D.E., Geoghegan, K.F., Griffor, M.C., Hawkins, J.L., Subashi, T.A., Varghese, A.H., Ammirati, M.J., Culp, J.S., Hoth, L.R., *et al.* (2007). Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. *Nature Structural & Molecular Biology* 14, 413-419.

Daleke, M.H., Cascioferro, A., de Punder, K., Ummels, R., Abdallah, A.M., van der Wel, N., Peters, P.J., Luirink, J., Manganeli, R., and Bitter, W. (2011). Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) Protein Domains Target LipY Lipases of Pathogenic *Mycobacteria* to the Cell Surface via the ESX-5 Pathway. *Journal of Biological Chemistry* 286, 19024-19034.

Daleke, M.H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C.M., Luirink, J., and Bitter, W. (2012a). General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A* 109, 11342-11347.

Daleke, M.H., van der Woude, A.D., Parret, A.H., Ummels, R., de Groot, A.M., Watson, D., Piersma, S.R., Jimenez, C.R., Luirink, J., Bitter, W., *et al.* (2012b). Specific Chaperones for the Type VII Protein Secretion Pathway. *J Biol Chem* 287, 31939-31947.

Dave, J.A., van Pittius, N.C.G., 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. *Bmc Microbiology* 2.

David, R. (2010). Bacterial Secretion the Two Faces of Mycp1. *Nature Reviews Microbiology* 8.

de Jonge, M.I., Pehau-Arnaudet, G., Fretz, M.M., Romain, F., Bottai, D., Brodin, P., Honore, N., Marchal, G., Jiskoot, W., England, P., *et al.* (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, 6028-6034.

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

Derrick, S.C., and Morris, S.L. (2007). The ESAT6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression. *Cellular Microbiology* 9, 1547-1555.

Desjardins, M., Houde, M., and Gagnon, E. (2005). Phagocytosis: the convoluted way from nutrition to adaptive immunity. *Immunological Reviews* 207, 158-165.

Desvaux, M., Hebraud, M., Talon, R., and Henderson, I.R. (2009a). Outer membrane translocation: numerical protein secretion nomenclature in question in mycobacteria. *Trends Microbiol* 17, 338-340.

Desvaux, M., Hebraud, M., Talon, R., and Henderson, I.R. (2009b). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* 17, 139-145.

Divangahi, M., Chen, M.J., Gan, H.X., Desjardins, D., Hickman, T.T., Lee, D.M., Fortune, S., Behar, S.M., and Remold, H.G. (2009). *Mycobacterium tuberculosis* evades macrophage defenses by inhibiting plasma membrane repair. *Nature Immunology* 10, 899-U123.

- Dodson, G., and Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends in Biochemical Sciences* 23, 347-352.
- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32, W665-667.
- Dorhoi, A., Reece, S.T., and Kaufmann, S.H.E. (2011). For better or for worse: the immune response against *Mycobacterium tuberculosis* balances pathology and protection. *Immunological Reviews* 240, 235-251.
- Eichelberg, K., and Galan, J.E. (1999). Differential regulation of *Salmonella typhimurium* type III secreted proteins by pathogenicity island 1 (SPI-1)-encoded transcriptional activators InvF and HilA. *Infection and Immunity* 67, 4099-4105.
- Ekiert, D.C., and Cox, J.S. (2014). Structure of a PE-PPE-EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. *Proceedings of the National Academy of Sciences of the United States of America* 111, 14758-14763.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* 60, 2126-2132.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62, 72-82.
- Faller, M., Niederweis, M., and Schulz, G.E. (2004). The structure of a mycobacterial outer-membrane channel. *Science* 303, 1189-1192.
- Finlay, B.B., and Falkow, S. (1997). Common themes in microbial pathogenicity revisited. *Microbiology and Molecular Biology Reviews* 61, 136-&.

Flint, J.L., Kowalski, J.C., Karnati, P.K., and Derbyshire, K.M. (2004). The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 12598-12603.

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 of the United States of America* *102*, 10676-10681.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* *116*, 190-199.

Frasinyuk, M.S., Kwiatkowski, S., Wagner, J.M., Evans, T.J., Reed, R.W., Korotkov, K.V., and Watt, D.S. (2014). Pentapeptide boronic acid inhibitors of *Mycobacterium tuberculosis* MycP(1) protease. *Bioorganic & Medicinal Chemistry Letters* *24*, 3546-3548.

Fredlund, J., and Enninga, J. (2014). Cytoplasmic access by intracellular bacterial pathogens. *Trends in Microbiology* *22*, 128-137.

Frigui, W., Bottai, D., Majlessi, L., Monot, M., Josselin, E., Brodin, P., Garnier, T., Gicquel, B., Martin, C., Leclerc, C., *et al.* (2008). Control of M-tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. *Plos Pathogens* *4*.

Ganguly, N., Giang, P.H., Basu, S.K., Mir, F.A., Siddiqui, I., and Sharma, P. (2007). *Mycobacterium tuberculosis* 6-kDa early secreted antigenic target (ESAT-6) protein downregulates lipopolysaccharide induced c-myc expression by modulating the extracellular signal regulated kinases 1/2. *Bmc Immunology* *8*.

Ganguly, N., Giang, P.H., Gupta, C., Basu, S.K., Siddiqui, I., Salunke, D.M., and Sharma, P. (2008). *Mycobacterium tuberculosis* secretory proteins CFP-10, ESAT-6 and the CFP10 : ESAT6 complex inhibit lipopolysaccharide-induced NF-kappa B transactivation by

downregulation of reactive oxidative species (ROS) production. *Immunology and Cell Biology* 86, 98-106.

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, 1677-1693.

Garces, A., Atmakuri, K., Chase, M.R., Woodworth, J.S., Krastins, B., Rothchild, A.C., Ramsdell, T.L., Lopez, M.F., Behar, S.M., Sarracino, D.A., *et al.* (2010). EspA Acts as a Critical Mediator of ESX1-Dependent Virulence in *Mycobacterium tuberculosis* by Affecting Bacterial Cell Wall Integrity. *Plos Pathogens* 6.

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 Biol* 2, RESEARCH0044.

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.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. *Journal of Structural Biology* 157, 281-287.

Graslund, S., Nordlund, P., Weigelt, J., Bray, J., Hallberg, B.M., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., *et al.* (2008). Protein production and purification. *Nature Methods* 5, 135-146.

Gray, T.A., Krywy, J.A., Harold, J., Palumbo, M.J., and Derbyshire, K.M. (2013). Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. *PLoS Biol* 11, e1001602.

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, 359-370.

Hagedorn, M., Rohde, K.H., Russell, D.G., and Soldati, T. (2009). Infection by Tubercular *Mycobacteria* Is Spread by Nonlytic Ejection from Their Amoeba Hosts. *Science* 323, 1729-1733.

Hamza, A., Wagner, J.M., Wei, N.-N., Kwiatkowski, S., Zhan, C.-G., Watt, D.S., and Korotkov, K.V. (2014). Application of the 4D Fingerprint Method with a Robust Scoring Function for Scaffold-Hopping and Drug Repurposing Strategies. *Journal of Chemical Information and Modeling* 54, 2834-2845.

Hayward, S., and Lee, R.A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* 21, 181-183.

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M., and Engelhardt, H. (2008). Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences of the United States of America* 105, 3963-3967.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38, W545-549.

Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeon, L., Abdallah, A.M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., *et al.* (2012a). ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cellular Microbiology* 14, 1287-1298.

Houben, E.N., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jimenez, C.R., Ottenhoff, T.H., Luirink, J., and Bitter, W. (2012b). Composition of the type VII secretion system membrane complex. *Mol Microbiol* 86, 472-484.

Houben, E.N., Korotkov, K.V., and Bitter, W. (2013). Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta*.

Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., *et al.* (2003). The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proceedings of the National Academy of Sciences of the United States of America* 100, 12420-12425.

Ilghari, D., Lightbody, K.L., Veverka, V., Waters, L.C., Muskett, F.W., Renshaw, P.S., and Carr, M.D. (2011). Solution Structure of the Mycobacterium tuberculosis EsxG.EsxH Complex *Journal of Biological Chemistry* 286, 29993-30002.

Kaufmann, S.H.E., and Parida, S.K. (2007). Changing funding patterns in tuberculosis. *Nature Medicine* 13, 299-303.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protocols* 4, 363-371.

Khan, A.R., and James, M.N.G. (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Science* 7, 815-836.

Kim, M.J., Wainwright, H.C., Locketz, M., Bekker, L.G., Walther, G.B., Dittrich, C., Visser, A., Wang, W., Hsu, F.F., Wiehart, U., *et al.* (2010). Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *Embo Molecular Medicine* 2, 258-274.

Kinhikar, A.G., Verma, I., Chandra, D., Singh, K.K., Weldingh, K., Andersen, P., Hsu, T.D., Jacobs, W.R., and Laal, S. (2010). Potential role for ESAT6 in dissemination of *M. tuberculosis* via human lung epithelial cells. *Molecular Microbiology* 75, 92-106.

Koo, I.C., Wang, C., Raghavan, S., Morisaki, J.H., Cox, J.S., and Brown, E.J. (2008). ESX-1-dependent cytolysis in lysosome secretion and inflammasome activation during mycobacterial infection. *Cellular Microbiology* 10, 1866-1878.

Korotkova, N., Freire, D., Phan, T.H., Ummels, R., Creekmore, C.C., Evans, T.J., Wilmanns, M., Bitter, W., Parret, A.H.A., Houben, E.N.G., *et al.* (2014). Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25–PPE41 dimer. *Molecular Microbiology*, n/a-n/a.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305, 567-580.

Kurenuma, T., Kawamura, I., Hara, H., Uchiyama, R., Daim, S., Dewamitta, S.R., Sakai, S., Tsuchiya, K., Nomura, T., and Mitsuyama, M. (2009). The RD1 Locus in the *Mycobacterium tuberculosis* Genome Contributes to Activation of Caspase-1 via Induction of Potassium Ion Efflux in Infected Macrophages. *Infection and Immunity* 77, 3992-4001.

Larson, A.M. (2009). *Protein Crystallization*, 2nd edn (International University Line, La Jolla CA).

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America* 106, 4154-4159.

Lewinsohn, D.M., Grotzke, J.E., Heinzl, A.S., Zhu, L., Owendale, P.J., Johnson, M., and Alderson, M.R. (2006). Secreted proteins from *Mycobacterium tuberculosis* gain access to the cytosolic MHC class-I antigen-processing pathway. *J Immunol* 177, 437-442.

Lewis, K.N., Liao, R.L., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., and Sherman, D.R. (2003). Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *Journal of Infectious Diseases* 187, 117-123.

Li, Y.J., Miltner, E., Wu, M., Petrofsky, M., and Bermudez, L.E. (2005). A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. *Cellular Microbiology* 7, 539-548.

Lightbody, K.L., Ilghari, D., Waters, L.C., Carey, G., Bailey, M.A., Williamson, R.A., Renshaw, P.S., and Carr, M.D. (2008). Molecular features governing the stability and specificity of functional complex formation by *Mycobacterium tuberculosis* CFP-10/ESAT-6 family proteins. *Journal of Biological Chemistry* 283, 17681-17690.

London, N., Raveh, B., Cohen, E., Fathi, G., and Schueler-Furman, O. (2011). Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. *Nucleic Acids Res* 39, W249-253.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E.V., *et al.* (2014). Structure of a type IV secretion system. *Nature* 508, 550-553.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128, 82-97.

MacGurn, J.A., and Cox, J.S. (2007). A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. *Infection and Immunity* 75, 2668-2678.

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*, 1653-1663.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., and Stover, C.K. (1996). Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent M-bovis. *Journal of Bacteriology* *178*, 1274-1282.

Manzanillo, P.S., Ayres, J.S., Watson, R.O., Collins, A.C., Souza, G., Rae, C.S., Schneider, D.S., Nakamura, K., Shiloh, M.U., and Cox, J.S. (2013). The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature* *501*, 512-+.

Manzanillo, P.S., Shiloh, M.U., Portnoy, D.A., and Cox, J.S. (2012). *Mycobacterium tuberculosis* activates the DNA-dependent cytosolic surveillance pathway within macrophages. *Cell Host Microbe* *11*, 469-480.

Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J., and Lowe, J. (2006). Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell* *23*, 457-469.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* *40*, 658-674.

McGowan, S., Buckle, A.M., Mitchell, M.S., Hoopes, J.T., Gallagher, D.T., Heselpoth, R.D., Shen, Y., Reboul, C.F., Law, R.H., Fischetti, V.A., *et al.* (2012). X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proc Natl Acad Sci U S A* *109*, 12752-12757.

McLaughlin, B., Chon, J.S., MacGurn, 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 Pathogens* *3*, 1051-1061.

Millington, K.A., Fortune, S.M., Low, J., Garces, A., Hingley-Wilson, S.M., Wickremasinghe, M., Kon, O.M., and Lalvani, A. (2011). Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Sciences*.

Mishra, K.C., De Chastellier, C., Narayana, Y., Bifani, P., Brown, A.K., Besra, G.S., Katoch, V.M., Joshi, B., Balaji, K.N., and Kremer, L. (2008). Functional role of the PE domain and immunogenicity of the *Mycobacterium tuberculosis* triacylglycerol hydrolase LipY. *Infection and Immunity* 76, 127-140.

Nair, S., Ramaswamy, P.A., Ghosh, S., Joshi, D.C., Pathak, N., Siddiqui, I., Sharma, P., Hasnain, S.E., Mande, S.C., and Mukhopadhyay, S. (2009). The PPE18 of *Mycobacterium tuberculosis* Interacts with TLR2 and Activates IL-10 Induction in Macrophage. *Journal of Immunology* 183, 6269-6281.

O'Riordan, M., Yi, C.H., Gonzales, R., Lee, K.D., and Portnoy, D.A. (2002). Innate recognition of bacteria by a macrophage cytosolic surveillance pathway. *Proceedings of the National Academy of Sciences of the United States of America* 99, 13861-13866.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* 6, 23-34.

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 & Microbe* 7, 210-220.

Okkels, L.M., and Andersen, P. (2004). Protein-protein interactions of proteins from the ESAT-6 family of *Mycobacterium tuberculosis*. *Journal of Bacteriology* 186, 2487-2491.

Pallen, M.J. (2002). The ESAT-6/WXG100 superfamily - and a new Gram-positive secretion system? *Trends in Microbiology* 10, 209-212.

Parsons, L.M., Jankowski, C.S., and Derbyshire, K.M. (1998). Conjugal transfer of chromosomal DNA in *Mycobacterium smegmatis*. *Mol Microbiol* 28, 571-582.

Pathak, S.K., Basu, S., Basu, K.K., Banerjee, A., Pathak, S., Bhattacharyya, A., Kaisho, T., Kundu, M., and Basu, J. (2007). Direct extracellular interaction between the early secreted antigen ESAT-6 of *Mycobacterium tuberculosis* and TLR2 inhibits TLR signaling in macrophages. *Nature Immunology* 8, 610-618.

Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L., and Davidson, A.R. (2009). The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proceedings of the National Academy of Sciences of the United States of America* 106, 4160-4165.

Petrochenko, E.V., Makepeace, K.A., Serpa, J.J., and Borchers, C.H. (2014a). Analysis of protein structure by cross-linking combined with mass spectrometry. *Methods Mol Biol* 1156, 447-463.

Petrochenko, E.V., Serpa, J.J., Makepeace, K.A., Brodie, N.I., and Borchers, C.H. (2014b). NN DXMSMS Match program for the automated analysis of LC/ESI-MS/MS crosslinking data from experiments using N metabolically labeled proteins. *J Proteomics* 109C, 104-110.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 1605-1612.

Philips, J.A., and Ernst, J.D. (2012). Tuberculosis Pathogenesis and Immunity. *Annual Review of Pathology: Mechanisms of Disease*, Vol 7 7, 353-384.

Poulsen, C., Panjikar, S., Holton, S.J., Wilmanns, M., and Song, Y.-H. (2014). WXG100 Protein Superfamily Consists of Three Subfamilies and Exhibits an alpha-Helical C-Terminal Conserved Residue Pattern. *Plos One* 9.

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, 709-717.

Pym, A.S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K.E., Marchal, G., Leclerc, C., and Cole, S.T. (2003). Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nature Medicine* 9, 533-539.

Raghavan, S., Manzanillo, P., Chan, K., Dovey, C., and Cox, J.S. (2008). Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature* 454, 717-U749.

Ramakrishnan, L., Federspiel, N.A., and Falkow, S. (2000). Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* 288, 1436-1439.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2, 1896-1906.

Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9, 173-175.

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., *et al.* (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *Embo Journal* 24, 2491-2498.

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, 21598-21603.

Riley, R., Pellegrini, M., and Eisenberg, D. (2008). Identifying Cognate Binding Pairs among a Large Set of Paralogs: The Case of PE/PPE Proteins of *Mycobacterium tuberculosis*. *Plos Computational Biology* 4.

Rook, G.A. (1987). Progress in the immunology of the mycobacterioses. *Clin Exp Immunol* 69, 1-9.

Rook, G.A.W., Steele, J., Barnass, S., Mace, J., and Stanford, J.L. (1986). Responsiveness to Live *Mycobacterium-Tuberculosis*, and Common Antigens, of Sonicate-Stimulated T-Cell Lines from Normal Donors. *Clinical and Experimental Immunology* 63, 105-110.

Russell, D.G. (2007). Who puts the tubercle in tuberculosis? *Nature Reviews Microbiology* 5, 39-47.

Russell, D.G. (2011). *Mycobacterium tuberculosis* and the intimate discourse of a chronic infection. *Immunological Reviews* 240, 252-268.

Russell, D.G., VanderVen, B.C., Glennie, S., Mwandumba, H., and Heyderman, R.S. (2009). The macrophage marches on its phagosome: dynamic assays of phagosome function. *Nat Rev Immunol* 9, 594-U584.

Russell, D.G., VanderVen, B.C., Lee, W., Abramovitch, R.B., Kim, M.J., Homolka, S., Niemann, S., and Rohde, K.H. (2010). *Mycobacterium tuberculosis* Wears What It Eats. *Cell Host & Microbe* 8, 68-76.

Sampson, S.L. (2011). Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin Dev Immunol* 2011, 497203.

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* 81, 305-317.

Samten, B., Wang, X., and Barnes, P.F. (2011). Immune regulatory activities of early secreted antigenic target of 6-kD protein of *Mycobacterium tuberculosis* and implications for tuberculosis vaccine design. *Tuberculosis (Edinb) 91 Suppl 1*, S114-118.

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jimenez, C.R., Daffe, M., *et al.* (2010). Direct Visualization by Cryo-EM of the Mycobacterial Capsular Layer: A Labile Structure Containing ESX-1-Secreted Proteins. *PLoS Pathog 6*, e1000794.

Schorb, M., and Briggs, J.A.G. (2014). Correlated cryo-fluorescence and cryo-electron microscopy with high spatial precision and improved sensitivity. *Ultramicroscopy 143*, 24-32.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J 78*, 1606-1619.

Schumann, G., Schleier, S., Rosenkrands, I., Nehmann, N., Halbich, S., Zipfel, P.F., de Jonge, M.I., Cole, S.T., Munder, T., and Mollmann, U. (2006). *Mycobacterium tuberculosis* secreted protein ESAT-6 interacts with the human protein syntenin-1. *Central European Journal of Biology 1*, 183-202.

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*, 6340-6344.

Seto, S., Matsumoto, S., Ohta, I., Tsujimura, K., and Koide, Y. (2009). Dissection of Rab7 localization on *Mycobacterium tuberculosis* phagosome. *Biochemical and Biophysical Research Communications 387*, 272-277.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem 68*, 850-858.

Shinde, U., and Inouye, M. (1995). Folding mediated by an intramolecular chaperone - autoprocesing pathway of the precursor resolved via a substrate assisted catalysis mechanism. *J Mol Biol* 247, 390-395.

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 of the United States of America* 106, 18792-18797.

Siezen, R.J., and Leunissen, J.A.M. (1997). Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Science* 6, 501-523.

Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., and Enninga, J. (2012). Phagosomal Rupture by *Mycobacterium tuberculosis* Results in Toxicity and Host Cell Death. *Plos Pathogens* 8.

Singh, A., Mai, D., Kumar, A., and Steyn, A.J.C. (2006). Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. *Proceedings of the National Academy of Sciences of the United States of America* 103, 11346-11351.

Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J.J., Liu, J., McDonald, K.L., Szyk, A., LaRonde-LeBlanc, N., and Gao, L.Y. (2008). Evidence for Pore Formation in Host Cell Membranes by ESX-1-Secreted ESAT-6 and Its Role in *Mycobacterium marinum* Escape from the Vacuole. *Infection and Immunity* 76, 5478-5487.

Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960.

Solans, L., Aguilo, N., Samper, S., Pawlik, A., Frigui, W., Martin, C., Brosch, R., and Gonzalo-Asensio, J. (2014). A Specific Polymorphism in *Mycobacterium tuberculosis* H37Rv Causes Differential ESAT-6 Expression and Identifies WhiB6 as a Novel ESX-1 Component. *Infection and Immunity* 82, 3446-3456.

Solomonson, M., Huesgen, P.F., Wasney, G.A., Watanabe, N., Gruninger, R.J., Prehna, G., Overall, C.M., and Strynadka, N.C. (2013). Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J Biol Chem* 288, 17782-17790.

Song, Y., DiMaio, F., Wang, R.Y., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure* 21, 1735-1742.

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*. *Infect Immun* 63, 1710-1717.

Stamm, L.M., Morisaki, J.H., Gao, L.Y., Jeng, R.L., McDonald, K.L., Roth, R., Takeshita, S., Heuser, J., Welch, M.D., and Brown, E.J. (2003). *Mycobacterium marinum* escapes from phagosomes and is propelled by actin-based motility. *Journal of Experimental Medicine* 198, 1361-1368.

Stanley, S.A., Johndrow, J.E., Manzanillo, P., and Cox, J.S. (2007). The type I IFN response to infection with *Mycobacterium tuberculosis* requires ESX-1-mediated secretion and contributes to pathogenesis. *Journal of Immunology* 178, 3143-3152.

Stanley, S.A., Raghavan, S., Hwang, W.W., and Cox, J.S. (2003). Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proceedings of the National Academy of Sciences of the United States of America* 100, 13001-13006.

Stoop, E.J.M., Schipper, T., Rosendahl Huber, S.K., Nezhinsky, A.E., Verbeek, F.J., Gurcha, S.S., Besra, G.S., Vandenbroucke-Grauls, C.M.J.E., Bitter, W., and van der Sar, A.M. (2011). Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. *Disease Models & Mechanisms*.

Strong, M., Sawaya, M.R., Wang, S.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*. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 8060-8065.

Sultana, R., Tanneeru, K., and Guruprasad, L. (2011). The PE-PPE Domain in *Mycobacterium* Reveals a Serine alpha/beta Hydrolase Fold and Function: An In-Silico Analysis. *PLoS One* *6*.

Sun, D., Liu, Q., He, Y., Wang, C., Wu, F., Tian, C., and Zang, J. (2013). The putative propeptide of MycP1 in mycobacterial type VII secretion system does not inhibit protease activity but improves protein stability. *Protein & Cell* *4*, 921-931.

Sweeney, K.A., Dao, D.N., Goldberg, M.F., Hsu, T., Venkataswamy, M.M., Henao-Tamayo, M., Ordway, D., Sellers, R.S., Jain, P., Chen, B., *et al.* (2011). A recombinant *Mycobacterium smegmatis* induces potent bactericidal immunity against *Mycobacterium tuberculosis*. *Nat Med* *17*, 1261-1268.

Sysoeva, T.A., Zepeda-Rivera, M.A., Huppert, L.A., and Burton, B.M. (2014). Dimer recognition and secretion by the ESX secretion system in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* *111*, 7653-7658.

Tan, T., Lee, W.L., Alexander, D.C., Grinstein, S., and Liu, J. (2006). The ESAT-6/CFP-10 secretion system of *Mycobacterium marinum* modulates phagosome maturation. *Cellular Microbiology* *8*, 1417-1429.

Taverner, T., Hall, N.E., O'Hair, R.A., and Simpson, R.J. (2002). Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem* *277*, 46487-46492.

Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell, B.G., and Cole, S.T. (1999). Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle and Lung Disease* *79*, 329-342.

Teutschbein, J., Schumann, G., Möllmann, 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, 253-259.

Tokunaga, T., Mizuguchi, Y., and Suga, K. (1973). Genetic recombination in mycobacteria. *J Bacteriol* 113, 1104-1111.

Ummels, R., Abdallah, M.A., Kuiper, V., Aajoud, A., Sparrius, M., Naeem, R., Spaink, H.P., van Soolingen, D., Pain, A., and Bitter, W. (2014). Identification of a Novel Conjugative Plasmid in *Mycobacteria* That Requires Both Type IV and Type VII Secretion. *Mbio* 5.

van den Ent, F., and Lowe, J. (2006). RF cloning: A restriction-free method for inserting target genes into plasmids. *J Biochem Biophys Methods* 67, 67-74.

van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., Brenner, M., and Peters, P.J. (2007). M-tuberculosis and M-leprae translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 129, 1287-1298.

van der Woude, A.D., Stoop, E.J.M., Stuess, M., Wang, S., Ummels, R., van Stempvoort, G., Piersma, S.R., Cascioferro, A., Jimenez, C.R., Houben, E.N.G., *et al.* (2014). Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. *Cellular Microbiology* 16, 280-295.

Vergne, I., Chua, J., Lee, H.H., Lucas, M., Belisle, J., and Deretic, V. (2005). Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* 102, 4033-4038.

Vevodova, J., Gamble, M., Kuenze, G., Ariza, A., Dodson, E., Jones, D.D., and Wilson, K.S. (2010). Crystal Structure of an Intracellular Subtilisin Reveals Novel Structural Features Unique to this Subtilisin Family. *Structure* 18, 744-755.

Via, L.E., Fratti, R.A., McFalone, M., Pagan-Ramos, E., and Deretic, D. (1998). Effects of cytokines on mycobacterial phagosome maturation. *Journal of Cell Science* *111*, 897-905.

Vieira, O.V., Botelho, R.J., and Grinstein, S. (2002). Phagosome maturation: aging gracefully. *Biochemical Journal* *366*, 689-704.

Villa, E., Schaffer, M., Plitzko, J.M., and Baumeister, W. (2014). Opening Windows into the Cell: Focused-Ion-Beam Milling for Cryo-Electron Tomography. *Biophysical Journal* *106*, 600A-600A.

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., and Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model* *52*, 360-372.

Volkman, H.E., Clay, H., Beery, D., Chang, J.C.W., Sherman, D.R., and Ramakrishnan, L. (2004). Tuberculous granuloma formation is enhanced by a *Mycobacterium* virulence determinant. *Plos Biology* *2*, 1946-1956.

Wagner, J.M., Evans, T.J., Chen, J., Zhu, H., Houben, E.N., Bitter, W., and Korotkov, K.V. (2013a). Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. *J Struct Biol* *184*, 115-128.

Wagner, J.M., Evans, T.J., and Korotkov, K.V. (2013b). Crystal structure of the N-terminal domain of EccA1 ATPase from the ESX-1 secretion system of *Mycobacterium tuberculosis*. *Proteins: Structure, Function, and Bioinformatics*, n/a-n/a.

Wang, J., Karnati, P.K., Takacs, C.M., Kowalski, J.C., and Derbyshire, K.M. (2005). Chromosomal DNA transfer in *Mycobacterium smegmatis* is mechanistically different from classical Hfr chromosomal DNA transfer. *Molecular Microbiology* *58*, 280-288.

Wang, J., Parsons, L.M., and Derbyshire, K.M. (2003). Unconventional conjugal DNA transfer in mycobacteria. *Nature Genetics* *34*, 80-84.

Watson, R.O., Manzanillo, P.S., and Cox, J.S. (2012). Extracellular *M. tuberculosis* DNA Targets Bacteria for Autophagy by Activating the Host DNA-Sensing Pathway. *Cell* *150*, 803-815.

Wirth, S.E., Krywy, J.A., Aldridge, B.B., Fortune, S.M., Fernandez-Suarez, M., Gray, T.A., and Derbyshire, K.M. (2011). Polar assembly and scaffolding proteins of the virulence-associated ESX-1 secretory apparatus in mycobacteria. *Molecular Microbiology* *83*, 654-664.

Wolynes, P.G. (1996). Symmetry and the energy landscapes of biomolecules. *Proceedings of the National Academy of Sciences of the United States of America* *93*, 14249-14255.

Wong, D., Bach, H., Sun, J., Hmama, Z., and Av-Gay, Y. (2011). *Mycobacterium tuberculosis* protein tyrosine phosphatase (PtpA) excludes host vacuolar-H<sup>+</sup>-ATPase to inhibit phagosome acidification. *Proc Natl Acad Sci U S A* *108*, 19371-19376.

Wong, D., Chao, J.D., and Av-Gay, Y. (2013). *Mycobacterium tuberculosis*-secreted phosphatases: from pathogenesis to targets for TB drug development. *Trends Microbiol* *21*, 100-109.

Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S.J., Edwards, N., Zhu, X.P., Fenselau, C., and Gao, L.Y. (2007). A unique *Mycobacterium* ESX-1 protein co-secreted with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. *Molecular Microbiology* *66*, 787-800.

Yamashiro, L.H., Oliveira, S.C., and Bafica, A. (2014). Innate immune sensing of nucleic acids from mycobacteria. *Microbes and Infection* *16*, 991-997.

Young, D.B., Kaufmann, S.H., Hermans, P.W., and Thole, J.E. (1992). Mycobacterial protein antigens: a compilation. *Mol Microbiol* *6*, 133-145.

Young, D.B., Perkins, M.D., Duncan, K., and Barry, C.E. (2008). Confronting the scientific obstacles to global control of tuberculosis. *Journal of Clinical Investigation* 118, 1255-1265.

Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D., and Davis, R.W. (1985). Dissection of Mycobacterium-Tuberculosis Antigens Using Recombinant DNA. *Proceedings of the National Academy of Sciences of the United States of America* 82, 2583-2587.

Zarrinpar, A., and Lim, W.A. (2000). Converging on proline: the mechanism of WW domain peptide recognition. *Nat Struct Biol* 7, 611-613.