

University of Kentucky

UKnowledge

---

Theses and Dissertations--Molecular and Cellular Biochemistry

Molecular and Cellular Biochemistry

---


2021

## Substrate Trafficking Within the Type VII Secretion Systems of Pathogenic Mycobacteria

Zachary A. Williamson

*University of Kentucky*, zachary.williamson@uky.edu

Author ORCID Identifier:

 <https://orcid.org/0000-0002-7070-5693>

Digital Object Identifier: <https://doi.org/10.13023/etd.2021.262>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

### Recommended Citation

Williamson, Zachary A., "Substrate Trafficking Within the Type VII Secretion Systems of Pathogenic Mycobacteria" (2021). *Theses and Dissertations--Molecular and Cellular Biochemistry*. 55.  
[https://uknowledge.uky.edu/biochem\\_etds/55](https://uknowledge.uky.edu/biochem_etds/55)

This Doctoral Dissertation is brought to you for free and open access by the Molecular and Cellular Biochemistry at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Molecular and Cellular Biochemistry by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Zachary A. Williamson, Student

Dr. Konstantin Korotkov, Major Professor

Dr. Trevor Creamer, Director of Graduate Studies

SUBSTRATE TRAFFICKING WITHIN THE TYPE VII  
SECRETION SYSTEMS OF PATHOGENIC MYCOBACTERIA

---

DISSERTATION

---

A dissertation submitted in partial fulfilment of the  
requirements for the degree of Doctor of Philosophy  
in the College of Medicine  
at the University of Kentucky

By  
Zachary Alan Williamson

Lexington, Kentucky

Director: Dr. Konstantin Korotkov, Associate Professor of  
Molecular and Cellular Biochemistry

Lexington, Kentucky

2021

Copyright © Zachary Alan Williamson 2021  
<https://orcid.org/0000-0002-7070-5693>

## ABSTRACT OF DISSERTATION

### SUBSTRATE TRAFFICKING WITHIN THE TYPE VII SECRETION SYSTEMS OF PATHOGENIC MYCOBACTERIA

Tuberculosis (TB), primarily caused by infection of *Mycobacterium tuberculosis* (*Mtb*) in the lungs, is the deadliest infectious bacterial disease killing 1.5 million people annually. A major determinant of virulence is active secretion through three specialized type VII secretion (ESX) systems; ESX-1, ESX-3, and ESX-5. A large group of substrates exported by the ESX systems is the PE (Proline-Glutamine) and PPE (Proline-Proline-Glutamate) families of proteins, which are highly expanded in the pathogenic species of *Mycobacteria* and encompass over 7% of *Mtb*'s genome coding capacity. PE and PPE proteins interact together to form PE-PPE heterodimers, and are secreted through specific ESX systems. Despite this massive expansion and the implication of a few select members in key virulence processes, most family members have still undefined functions. This can be partially attributed to previously reported difficulties of working with the purified proteins *in vitro* and a poor understanding of how heterodimer pairs are trafficked within the mycobacterial cell to their cognate ESX system. Each ESX system that secretes PE-PPE heterodimers encodes a unique copy of the chaperone protein, EspG. The work contained here aims to elucidate the mechanism of PE-PPE heterodimer recognition by EspG for the ESX-3 and ESX-5 systems. Structural analysis of ESX-3-specific PE5-PPE4-EspG<sub>3</sub> heterotrimer shows that EspG<sub>3</sub> and EspG<sub>5</sub> employ unique binding modes to their cognate PE-PPE heterodimers, which presents unique interfaces of the highly conserved PPE proteins to each EspG. The ESX-5-specific PPE proteins are variable at the hydrophobic (hh) motif, which is shielded from solvent upon binding of EspG. Structural analysis of selected hh mutants in the context of the PE25-PPE41-EspG<sub>5</sub> suggested plasticity within the PPE-binding region of EspG<sub>5</sub> to allow it to bind the various ESX-5-specific PPE proteins. Taken together, these results improve our understanding of trafficking of an important group of ESX substrates, setting the stage for more targeted studies of individual PE and PPE proteins to determine the still unknown functions of most family members. This may prove to be a fruitful avenue of therapeutic development to lower the burden of the global public health emergency caused by TB.

**KEYWORDS:** *Mycobacterium tuberculosis*, X-ray crystallography, type VII secretion system, chaperone, protein-protein interactions, mycobacteria

Zachary Alan Williamson  
*(name of student)*

04/28/2021  
Date

SUBSTRATE TRAFFICKING WITHIN THE TYPE VII  
SECRETION SYSTEMS OF PATHOGENIC MYCOBACTERIA

By

Zachary Alan Williamson

Konstantin Korotkov  
Director of Dissertation

Trevor Creamer  
Director of Graduate Studies

04/28/2021  
Date

DEDICATION

*To my parents, for their unending support*

## Acknowledgments

The work presented in this dissertation would not have been possible without support from numerous people. First, I would like to thank my mentor, Dr. Konstantin Korotkov. Thank you for the support, mentorship, and guidance that has helped me become the scientist I am today.

I would also like to thank my dissertation committee: Dr. Craig Vander Kooi, Dr. David Rodgers, and Dr. Sarah D’Orazio. During my PhD training you have all challenged me to reach my potential, and I am truly grateful for that. I would also like to thank my Outsider Examiner, Dr. Yinan Wei, for her time and effort in reviewing my dissertation.

I would also like to thank my fellow students, the faculty, and the staff of the Biochemistry Department. The community created by everybody in the department makes it a great place to learn and grow as a graduate student. I would like to thank Dr. Trevor Creamer for being a wonderful servant to the students as the DGS of the department and Dr. Rebecca Dutch for her leadership as Department Chair, and for her support. I am also incredibly grateful to the administrative staff of the department; Rachel Putty, Brenda Woods, Tonya Simon, and Phil Dickinson; for their efforts to keep the department running smoothly.

To the members of the Korotkov and Korotkova labs, thank you for creating a wonderful work environment. A special thanks to Dr. Natalia Korotkova for her support and providing valuable feedback on my experiments, manuscripts, and seminars over the years. I am also incredibly grateful to Dr. Becky Edgar for helping me learn to trust myself in the lab, and for teaching me to talk like a true New Zealander. To the current members of the labs; Dr. Catherine Chaton, Dr. Jeffrey Rush, Dr. Mizan Rahman, Dr. Prakash Parajuli, Sveta Zamakhaeva, and Cameron Kenner; you have made the lab and wonderful place to be every day and I have greatly missed you all during my work from home spell.

I would like to thank my family and friends. Without them and their support I would not have made it through my PhD training. Thank you to Dr. Michael Orr, especially for your support during my first year. Thank you to Dr. Chelsea Barrett for always reminding me that I’ve got this. Thank you to Dr. Brittani Price for always needing a cup of coffee at the same time as me. To all my other friends thank you for your support and friendship. To my parents, Gary and Laurie, you have been incredibly supportive for my entire life and I would not have made it to where I am without you. I’d also like to thank the rest of my family, my aunts, uncles, cousins, and Grandma Ruthie. Thank you to the extended Smith and Noel families for welcoming me into your families and giving me an extra home in Kentucky.

And lastly to my partner Caroline. You are truly incredible. Thank you for your patience, for picking me up when I am down, and for celebrating my triumphs.



## Table of Contents

<b>Acknowledgments</b>	<b>iii</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>Chapter 1. Introduction.</b>	<b>1</b>
1.1. Tuberculosis – the ancient, deadly infectious disease.	1
1.2. The lifecycle of the <i>Mtb</i> bacterium.	2
1.3. The type VII secretion system and mycobacteria	5
1.3.1. Protein secretion in mycobacteria.	5
1.3.2. Genetic organization and evolution of the ESX systems.	7
1.3.3. The core components of the ESX systems.	9
1.3.4. Substrates of the ESX systems.	11
1.4. The PE and PPE proteins.	14
1.4.1. Classification and genetic organization of PE and PPE proteins.	14
1.4.2. PE and PPE interactions with other ESX proteins.	16
1.5. Functions of the ESX systems.	18
1.6. Conclusions and dissertation overview.	20
<b>Chapter 2. Materials and Methods.</b>	<b>24</b>
2.1. Bacterial Strains and growth conditions.	24
2.2. Bacterial Transformation.	24
2.3. Expression and purification of recombinant proteins.	24
2.3.1. PE5-PPE4-EspG <sub>3</sub> heterotrimers.	24
2.3.2. PE25-PPE41-EspG <sub>5</sub> heterotrimers.	25
2.4. Crystallization, data collection, and structure solution.	26
2.4.1. PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG <sub>5mm</sub> .	26
2.4.2. PE25 <sub>mt</sub> -PPE41 <sub>mt</sub> -EspG <sub>5mm</sub> .	26
2.5. Size-exclusion chromatography with multi-angle light scattering (SEC-MALS).	27
2.6. Small angle X-ray scattering (SAXS) data comparison and <i>ab initio</i> model reconstruction.	27
2.7. Sequence Analysis.	27
2.8. Structural Analysis.	28
2.9. Data Availability.	28
<b>Chapter 3. Structural plasticity of EspG<sub>5</sub> accommodates variety seen in ESX-5-specific PPE proteins.</b>	<b>29</b>
3.1. Introduction.	29
3.2. Results.	30
3.2.1. Structure of PE25 <sub>mt</sub> -PPE41 <sup>A124L</sup> <sub>mt</sub> -EspG <sub>5mm</sub> .	30
3.2.2. Structural plasticity of EspG <sub>5</sub> accommodates variety in hh motifs.	31
3.2.3. Limited tolerance of bulkier hh motif residues in PPE41-EspG <sub>5</sub> interface.	32
3.3. Discussion.	34
<b>Chapter 4. PE5–PPE4–EspG<sub>3</sub> heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems.</b>	<b>43</b>

4.1. Introduction.	43
4.2. Results.	44
4.2.1. EspG <sub>3</sub> forms a complex with PE <sub>5</sub> –PPE <sub>4</sub> , and binding is conserved across species.	44
4.2.2. Overall structure of PE <sub>5<sub>mt</sub></sub> –PPE <sub>4<sub>mt</sub></sub> –EspG <sub>3<sub>mm</sub></sub> .	45
4.2.3. Interface between PPE <sub>4<sub>mt</sub></sub> and EspG <sub>3<sub>mm</sub></sub> .	46
4.2.4. Mutations cause disruptions in the PPE <sub>4</sub> –EspG <sub>3</sub> interface.	46
4.2.5. Structure of EspG <sub>3</sub> in and out of heterotrimer complex.	47
4.2.6. Comparison of ESX-3 and ESX-5 PE–PPE–EspG heterotrimers.	48
4.3. Discussion.	49
<b>Chapter 5. Discussion.</b>	<b>70</b>
5.1. Summary of this dissertation.	70
5.2. Recognition of PPE by EspG and the ESX-1 heterotrimer mystery.	70
5.3. Recognition of the PE-PPE-EspG heterotrimer by the core ESX machinery.	72
5.4. Function of the conserved PE and PPE domains.	76
5.5. Concluding remarks.	77
<b>References</b>	<b>79</b>
<b>Vita</b>	<b>101</b>

## List of Tables

Table 3.1. Data collection and refinement statistics of ESX-5 mutant heterotrimer. ....	36
Table 3.2. Structural deviations between PPE41 <sup>LL</sup> mutant heterotrimer and previously published ESX-5 heterotrimers. ....	37
Table 3.3. Packing density of hh motif residues. ....	38
4.1. Summary of constructs utilized for crystallization experiments and final outcomes.	52
Table 4.2. Data collection and refinement statistics of ESX-3 heterotrimer. ....	53
Table 4.3. Structural variations in copies of PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG <sub>3mm</sub> structure in RMSD (Å). ....	54
Table 4.4. Structural variations between individual components of ESX-3 heterotrimer and the previously published ESX-5 heterotrimers in RMSD (Å). ....	55
Table 4.5. Summary of analysis of PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG <sub>3mt</sub> interactions in vitro. ....	56

## List of Figures

Figure 1.1. Genetic organization of ESX clusters in <i>Mtb</i> . .....	22
Figure 1.2. Structures of selected ESX substrates. ....	23
Figure 3.1. Variety in hh motifs in ESX-5-specific PPE proteins. ....	39
Figure 3.2. Structure of PE25 <sub>mt</sub> -PPE41 <sup>LL</sup> <sub>mt</sub> -EspG5 <sub>mm</sub> and comparison to wild-type heterotrimer. ....	40
Figure 3.3. Interactions between the hh motif of PPE proteins and EspG5. ....	41
Figure 3.4. Sequence tolerance of hh motifs in selected ESX-5 PPE proteins. ....	42
Figure 4.1. Solution characterization of the PE5–PPE4–EspG3 heterotrimer. ....	57
Figure 4.2. PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> dimer is bound by EspG3 from various mycobacterial species. ....	58
Figure 4.3. Crystal structure of the PE5 <sub>mt</sub> –PPE4 <sub>mt</sub> –EspG3 <sub>mm</sub> heterotrimer and selected interactions in PPE4 <sub>mt</sub> -EspG3 <sub>mm</sub> interface. ....	59
Figure 4.4. Comparison of PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG3 <sub>mm</sub> crystal structure and PE5 <sub>ms</sub> -PPE4 <sub>ms</sub> -EspG3 <sub>ms</sub> SAXS data. ....	60
Figure 4.5. Interface between PPE4 <sub>mt</sub> and EspG3 <sub>mm</sub> . ....	61
Figure 4.6. Sequence alignment of <i>M. tuberculosis</i> ESX-3-specific PPE genes. ....	62
Figure 4.7. Sequence alignment of selected EspG3's shows interacting residues are conserved. ....	64
Figure 4.8. Co-purification of selected PPE4 <sub>mt</sub> and EspG3 <sub>mt</sub> mutants with their wild-type partners. ....	65
Figure 4.9. EspG3 exists in multiple structural forms. ....	66
Figure 4.10. PE5 <sub>mt</sub> –PPE4 <sub>mt</sub> interacts with EspG3 <sub>mm</sub> chaperone in a unique mode compared with ESX-5 PE–PPE dimers. ....	67
Figure 4.11. Comparison of interfaces in the ESX-3- and ESX-5-specific PPE-EspG complexes. ....	69

## Chapter 1. Introduction.

### 1.1. Tuberculosis – the ancient, deadly infectious disease.

Tuberculosis (TB) is currently a leading cause of death worldwide, being the deadliest bacterial disease. Nearly 10 million people annually fall ill with TB, with 1.5 million ultimately succumbing to the disease (1). The primary causative agent of TB is *Mycobacterium tuberculosis* (*Mtb*), which has been infecting humans for tens of thousands of years, predating the Neolithic period (2,3). In fact, the interactions between the bacterium and the human host drove the evolution of *Mtb* (4), and some estimates have roughly one third of the world's population as infected with *Mtb*. Despite discovery of *Mtb* in 1882 by Robert Koch (5), it is still a major global public health concern all these years later and has been considered an emergency for the past 25 years (6).

Currently TB diagnosis is followed by the implementation of well-studied treatment regimens. The exact treatment regimen implemented depends on numerous factors, but most importantly on the drug susceptibility of the disease-causing strain of *Mtb* within the patient. The World Health Organization recommendation for drug-susceptible strain infection is a six month treatment plan of two months of daily dosing of four medications; isoniazid, rifampicin, pyrazinamide, and ethambutol; followed by continuing the daily dosing of isoniazid and rifampicin for an additional four months (7). The latest data from treatment outcomes of drug-susceptible TB is that 85% of patients had a successful treatment (1). Despite this high success rate the World Health Organization estimated that in 2019 about 2.9 million cases of TB were not reported, or about 29% of the estimated 10 million global cases (1). The historical taboo associated with TB, along with current treatment options involving lengthy daily dosing of numerous therapeutics, and with many of these therapeutics having harmful side effects (8), has led to low adherence. Low adherence has aided in the rise and spread of drug-resistant strains of *Mtb*. With rifampicin and isoniazid being the most commonly used first line therapeutics, resistance to them is often first acquired, and strains with resistance to both are known as multi-drug resistant TB (MDR-TB) (9). Treatment of MDR-TB infections is extremely prolonged, often lasting longer than 12 months, and contains regimens with daily dosing of more than five separate therapeutics (9). Often times these second-line therapeutics have considerably worse side

effects than standard anti-TB therapeutics (10). New therapeutics that can shorten the required treatment and can be effective against all strains of *Mtb* are gravely needed.

An effective vaccine against *Mtb* also does not exist. A strain of a related pathogen, *Mycobacterium bovis* BCG, is commonly used as a vaccine, however this is only done in areas where TB is endemic and to protect children from severe forms of *Mtb* infection, such as tubercular meningitis (11-13). *M. bovis* BCG vaccination does not protect patients from developing pulmonary TB later in life (14). *M. bovis* BCG was created by the serial passage of a virulent *M. bovis* strain that ultimately resulted in its attenuation and was carried out by Albert Calmette and Camille Guérin at the Institut Pasteur of Lille between 1908 and 1921 (11). Calmette and Guérin first tested the *M. bovis* BCG strain as a vaccine in humans in 1921, before sending out the newly created vaccine to the rest of the world in 1924 (11). The main cause for attenuation in *M. bovis* BCG can be attributed to the deletion in region of difference 1 (RD1) (15,16). This deletion lies within ESX-1, one of the type VII secretion systems found within mycobacteria, and deletes the *eccCb1*, *pe35*, *ppe68*, *esxB*, *esxA*, and *espI* genes, and these deletions make ESX-1 non-functional (16). Ultimately the *M. bovis* BCG vaccine is not sufficient to fight the global spread of TB. With TB persisting as a leading cause of death worldwide, and the continued rise of drug-resistant strains of *Mtb* it will be nearly impossible to eradicate *Mtb* and TB without the aid of effective vaccines.

## 1.2. The lifecycle of the *Mtb* bacterium.

*Mtb* is an intracellular pathogen that primarily spreads in aerosolized droplets from a patient with active TB disease to an uninfected patient. *Mtb* is most often phagocytosed by macrophages, where it sets up its replicative niche and modulates the host immune responses. The primary site of infection is within the lungs, and extrapulmonary TB is less than 10% of all cases (17). Inside the lungs *Mtb* targets the lower lung and the alveolar macrophages. The lower lung tends to have fewer competing microbes than the upper respiratory tract (18), which allows *Mtb* to focus solely on infecting the host and not on competing with other microbes nor dealing with the host immune responses to the competing microbes. *Mtb* contains a surface lipid, phthiocerol dimycocerosate (PDIM), to hide their pathogen-activated molecular patterns (PAMPs), which normally would recruit bactericidal macrophages(19) Instead, *Mtb* utilizes another surface lipid, phenolic

glycolipid (PGL), to recruit growth-permissive macrophages (19). This would be futile in the upper respiratory tract that is filled with other microbes that are recruiting bactericidal macrophages, and *Mtb* would simply be collateral damage. Further evidence for the lower respiratory tract being the preferred target for *Mtb* infection comes from studies looking at transmission in confined spaces (20,21), and these studies were supported by a rabbit infection study that showed aerosol droplet size negatively correlated with infection, as the most infectious droplets contained one to three mycobacteria while droplets containing thousands of mycobacteria caused little to no infection (22).

Once within the macrophage *Mtb* has to avoid destruction by these professional phagocytes. *Mtb*'s first action is to prevent phagosome maturation and its fusion with the lysosome (23) and the subsequent acidification and maturation of the compartment. *Mtb*-containing phagosomes maintain characteristics of early phagosomes such as limited acidification (24,25), maintenance of a small GTPase associated with early endosomal events (26-28), and access to the rapid recycling pathway (26,29,30). Modulation of phagosomal maturation is accomplished through a variety of methods. Numerous cell wall lipids have been implicated in stunting phagosomal maturation (31-35), along with two secreted proteins, SapM (36) and PknG (37). Despite their best efforts, *Mtb* can be found within acidified phagolysosomes, which typically reduces *Mtb*'s replicative potential and often kills the mycobacteria (38,39). Yet it seems that lysosomal fusion is not a death sentence for *Mtb*, as co-infection with *Coxiella burnetii*, a bacterium that occupies acidified vacuole and readily fuses with other vacuoles, drives *Mtb* to co-localize with *C. burnetii* into the acidified vacuole replicating the lysosomal fusion event, but does not rapidly kill the mycobacteria and may not affect the mycobacterium's replication potency (40). Also, if opsonized or in later stages of infection when anti-*Mtb* antibodies are present, *Mtb* can enter macrophages through Fc-receptors and reside within phagolysosomes and even replicate within these compartments (41,42). Survival within the macrophage is crucial for *Mtb* to establish the next part of its infectious cycle, the granuloma.

Granulomas at their most basic are described as a compact, organized aggregate of mature macrophages in response to a continual stimulus (43-45). The first description of granulomas was done in *Mtb* infected lungs in 1679 and was called a tubercle, predating the description of *Mtb* by Koch by over 200 years (46). The tubercle and TB were so

intimately connected that it is where TB gets its name (47), and even today TB remains the leading cause of granuloma formation (46). Granulomas are incredibly heterogeneous in terms of cellular composition, oxygenation levels, inflammatory milieu, and bacterial burden not only across different host but also within a single host (48). Transport of *Mtb*-infected macrophages across the alveolar epithelium into the lung tissue is the first step in the initiation of the granuloma (49) and is made possible because of the recruitment of the correct growth-permissive macrophages by *Mtb*'s surface lipid PGL when the mycobacterium first enters the host (19). New macrophages are recruited, in part by substrates of ESX-1 (50,51), and in combination with an innate immune response this is sufficient to initiate the formation of the granuloma (52). *Mtb* exploits the granuloma for growth and dissemination within the host, and granulomas are supportive of high bacterial burdens (53). Residing within the granuloma are *Mtb*-infected macrophages which are actively dying, and these dying macrophages are actively recruiting new macrophages into the granuloma to allow phagocytosis of the dying cell by the newly arriving macrophages (54). Multiple newly arriving macrophages will phagocytose a single dying *Mtb*-infected macrophage, thus spreading *Mtb* in multiple new host cells to continue its replication cycle (54). The direct attraction of macrophages to phagocytose dying *Mtb*-infected macrophages is dependent on a functional ESX-1 system (54). EsxA, a substrate of ESX-1, may also drive the death of *Mtb*-infected macrophages, as it induces apoptosis through numerous pathways in cultured cells, some of which may be present in the context of the granuloma (55-59). The center of the mature granuloma is a hypoxic environment full of dying macrophages, many of which have undergone necrotic death (60). Interestingly, the necrotic release of *Mtb* may be beneficial as the extracellular milieu is very permissive for growth (61,62). The granuloma therefore sustains *Mtb* as not only are newly arriving macrophages regularly infiltrating and phagocytosing dying *Mtb*-infected macrophages, but also the extracellular environment supports the growth of released *Mtb* from *Mtb*-infected macrophages that undergo necrosis. When a person with active TB disease coughs, sneezes, or spits the *Mtb* from their lungs can be aerosolized and expelled into the environment where an uninfected person can inhale the droplet and allow *Mtb* to spread in a new host.



### 1.3. The type VII secretion system and mycobacteria

#### 1.3.1. Protein secretion in mycobacteria.

*Mtb* and the rest of the mycobacterial species have their own family, Mycobacteriaceae, within the Actinobacteria phylum, or the high GC Gram-positive bacteria, which contain a high proportion of guanine-cytosine base pairings in their genome. Mycobacteria can be split into two different subgroups, depending on their rate of replication. The so-called fast-growing mycobacteria include most of the non-pathogenic, saprophytic mycobacteria, such as *Mycobacterium smegmatis*, although some species, such as *Mycobacteroides abscessus* can cause disease, and are considered the more ancestral of the mycobacteria (63). Most of the pathogenic mycobacteria, such as *Mtb* and the human pathogen *Mycobacterium leprae*, belong to the so-called slow-growing group (63). All mycobacteria contain a complex cell envelope with characteristic long chain fatty acids, known as mycolic acids, attached to peptidoglycan-arabinogalactan network (64-66). This cell envelope creates an exceptionally strong permeability barrier for mycobacteria; however, the organization of the cell envelope was unclear until 2007. Two studies identified evidence of an outer membrane bilayer that exists in the cell envelope, that would become known as the mycomembrane (67,68). The mycomembrane presents an export challenge for mycobacteria similar to the one present in Gram-negative bacteria. None of the specialized secretion systems found in Gram-negative bacteria have been identified in mycobacterial species, and in fact the identification of mycomembrane transporters has thus far eluded researchers (69,70)., This limits the current knowledge of protein secretion in mycobacteria to mechanisms and systems of transport across the plasma membrane, with still unclear modes of transport across the mycomembrane.

Numerous pathways across the inner membrane of mycobacteria have been identified including the conserved general secretion (Sec) pathway, twin-arginine translocase (Tat) pathway, and the type VII secretion system. The Sec pathway is a highly conserved and essential export pathway ubiquitously found within all bacteria (71,72). The SecYEG complex forms the main complex through which Sec export occurs. SecY forms the main pore of the channel, which is stabilized by SecE, and the efficiency of transport is improved with SecG (73-75). Proteins are translocated across the channel in an unfolded state and contain a conserved signal peptide in their N-terminus, which is characterized by

a positive N-terminal domain, a hydrophobic central domain, and an uncharged C-terminal domain with a cleavage site (76,77). Export through SecYEG is a posttranslational process, however SecYEG also works to insert integral membrane proteins into the plasma membrane in a co-translational manner with the aid of SRP and FtsY (78-80). The chaperone SecA is also associated with the SecYEG channel and acts to bring proteins to the channel and use its intrinsic ATPase activity to drive export (81). Mycobacteria, including *Mtb*, along with other actinomycetes and a small subset of Gram-positive bacteria have an additional SecA paralog known as SecA2 (72,82). In mycobacteria, SecA1 is the SecA paralog with the highest similarity to the *E. coli* SecA protein. A large variety of proteins are secreted in an SecA1-dependent manner through the SecYEG channel and have roles in a large variety of cellular functions such as cell wall synthesis and virulence (72). SecA2 overexpression does not allow for *secA1* to be deleted, suggesting that SecA2 interacts in a unique manner and similarly SecA1 overexpression does not rescue a  $\Delta secA2$  phenotype (82). SecA2 utilizes the same SecYEG channel in mycobacteria (83-85). SecA2 is not essential for growth *in vitro* yet does function in virulence in both macrophage and animal models of mycobacterial infection (72,86-90). Overall, numerous vital proteins are secreted in a Sec-dependent manner, whether with the aid of SecA1 or SecA2, that not only allow mycobacteria to grow but also play roles in the virulence of pathogenic species.

Another conserved secretion pathway in mycobacteria is the Tat pathway. The Tat pathway in mycobacteria functions similarly to the Tat systems studied in other bacteria and encompasses the integral membrane proteins TatA, TatB, and TatC (91-94). TatA and TatB are small homologous proteins with a single transmembrane domain, while TatC is a large integral membrane protein (95). TatBC binds substrates and recruits TatA homooligomers to the complex for translocation across the membrane with energy from the proton motor force (96-98). The exact mechanism of transport is currently unknown. Some models propose TatA oligomers form a channel for substrates to translocate through, while others propose that TatA destabilizes the membrane near the TatBC-substrate complex to allow for TatC to drive translocation across the plasma membrane without a channel (99,100). The signal peptide of Tat substrates shares many similarities with the Sec signal peptide such as a positively charged N-terminus followed by a hydrophobic domain and a neutral polar C-terminal domain. The Tat signal peptide does contain a key difference in

the presence of a Tat motif, with characteristic twin arginines followed by another amino acid and two hydrophobic amino acids (101). Most Tat systems are dispensable, at least when bacteria are grown in standard laboratory media (102). This is not the case for *Mtb*, as it is one of the few species where Tat is essential (94), and even in other mycobacteria, such as *M. smegmatis*, the Tat pathway is not essential (92,93). The Tat pathway represents an alternative pathway for secretion of proteins across the plasma membrane.

The major export system in mycobacteria is the type VII secretion systems, also known as the ESX systems. The ESX systems will be the primary focus of the rest of this section. All mycobacterial species encode for at least one ESX system, with numerous unique systems being found within pathogenic mycobacteria, such as *Mtb* which contains five ESX systems. Genetically the systems are organized in clusters which encode all the core components that make up the inner membrane transport machinery, conserved accessory proteins, and a few select substrates. A more detailed description of the ESX systems including genetic organization, core components, substrate properties, and individual system functions will be the focus of the rest of this chapter.

### 1.3.2. Genetic organization and evolution of the ESX systems.

The ESX systems are genetically organized in loci/clusters that contain the conserved core components *eccB*, *eccC*, *eccD*, *eccE*, and *mycP*, a pair of *esx* genes, the accessory genes *espG* and *eccA*, and a *pe* and *ppe* pair (Figure 1.1.). The naming of the genes follows a standardized nomenclature, for example the *eccD* gene from ESX-1 is known as *eccD<sub>1</sub>* (103). Each component, along with functions of the various systems will be discussed further in the following subsections. The precise arrangement of the genes has changed over the evolution of the systems.

ESX-4 is the most ancestral of the ESX systems and similar systems are present in non-*Mycobacterium* genera such as *Gordonia*, *Nocardia*, and non-mycolic acid producing *Actinobacteria* orders such as *Streptomycetales* and *Pseudonocardiales* (104). The more distantly related phylum *Firmicutes* also has ESX-like systems, however only *eccC* and *esxA* homologs exist in these systems (105-107). These ESX-like systems in *Firmicutes* species have been shown to be functional for the secretion of their *esxA* homologs, but are likely mechanistically distinct from any *Actinobacteria* ESX systems, such as the ones found within in *Mtb* (105,108-113). The presence of orthologous ESX-4 systems in a

variety of actinobacterial species, many which do not contain a mycomembrane, suggests that it plays a role only in inner membrane translocation, and this could carry to the other ESX systems found within mycobacterial species. Functional ESX-4 evidence has only recently been identified, including intracellular survival in *M. abscessus* (114) and conjugation in *M. smegmatis* (115,116). The lack of the *eccE<sub>4</sub>* gene in *Mtb* (Figure 1.1.) could explain why ESX-4 mediated secretion has not been observed using *Mtb* systems, however recent work has identified a co-dependence on intact ESX-4 in order to secrete the toxin CpnT (117). Also missing from ESX-4 are *eccA<sub>4</sub>*, *espG<sub>4</sub>*, *pe*, and *ppe* genes. As these four genes have hypothesized interconnections, which will be discussed in a future subsection, it is unsurprising that all four are absent together.

ESX-3 is present in nearly every sequenced mycobacterial species (104) and contains all conserved core components, along with the substrates *esxG* and *esxH*, and the PE-PPE-associated genes; *eccA<sub>3</sub>*, *espG<sub>3</sub>*, *pe5*, and *ppe4*. ESX-3 is essential for *in vitro* growth of *Mtb* (118-120).

ESX-1 is present in almost all mycobacteria, only a few of the most ancestral species, such as *M. abscessus*, contain only ESX-4 and ESX-3 (104). A few mycobacterial species such as *Mycobacterium avium*, *Mycobacterium ulcerans*, and *Mycobacterium xenopi* had ESX-1 present in their genomes at one point, but it has since been deleted from their genomes. The vaccine strain *M. bovis* BCG strain has a main attenuation deletion, RD1, that is located within the ESX-1 cluster. The presence of ESX-1 in most mycobacteria, including pathogenic and non-pathogenic species, suggests a non-virulence primary function, and perhaps virulence effectors have evolved more recently in the pathogenic species (104). Both *eccE<sub>1</sub>* and *mycP<sub>1</sub>* are often inverted, along with numerous gene insertions not shown in Figure 1.1. The *eccC* gene is unique in ESX-1 as it is split into two different genes, *eccCa<sub>1</sub>* and *eccCb<sub>1</sub>*, yet the two gene products interact with each other to function similar to other EccC paralogs from *Mtb*.

The most recently evolved systems are ESX-2 and ESX-5 and are only found within the slow-growing mycobacteria such as *Mtb*. There is no evidence that either cluster causes the slow-growing phenotype as ESX-2 is absent in numerous slow-growing mycobacteria, such as *M. leprae* and *M. marinum* (104). Furthermore, the deletion of ESX-5 does not increase the growth rate of *Mtb* or *M. marinum* (121,122). ESX-5 is unique in that it

contains multiple *pe* and *ppe* genes within the cluster (Figure 1.1.) and is hypothesized to secrete most of the known PE and PPE proteins (123). The evolution of ESX-5 is closely linked with the expansion of the *pe* and *ppe* genes (124). The expansion of the *pe* and *ppe* genes will be discussed further in 1.4.

### 1.3.3. The core components of the ESX systems.

The core ESX machinery can be defined as the complex containing EccB, EccC, EccD, EccE, and MycP. Most studies do not capture MycP during purification of the complex (125-128), however work on MycP shows it does interact with the other components and is hypothesized to play a role in stabilizing the complex (129), and therefore is considered a component of the core complex. The first structures of the core complex were negative-stain electron microscopy reconstructions from the ESX-5 systems of *M. marinum* and *M. bovis* BCG, and revealed that the machinery exists solely within the plasma membrane (126) and it is still unknown what facilitates transport of ESX substrates across the mycomembrane. Recently, cryo-EM structures of ESX-3 machinery from *M. smegmatis* have been published (127,128). These structures both identified a core unit of EccB<sub>3</sub>C<sub>3</sub>D<sub>3</sub>E<sub>3</sub> in a 1:1:2:1 ratio that forms a stable dimer of protomers (127,128). The original ESX-5 negative-stain structure contains six-fold symmetry and three of the ESX-3 dimers can be modelled within its envelope (127). Famelis *et al.*, (127) propose this hexameric model of the ESX-3 machinery as the functional unit (127). The functional model contains an inner pore created by transmembrane helices from EccC<sub>3</sub> that is roughly 25 Å in diameter, which is wide enough to secrete the folded EsxGH heterodimer (127). Poweleit *et al.* propose two models in their publication (128). The first largely agrees with Famelis *et al.* (127), however Poweleit *et al.* (128) indicate it is not clear how EccC<sub>3</sub>'s ATPase activity would be activated based upon the current understanding of its mechanism of action (127,128). In Poweleit *et al.*'s second model for secretion the core protomer dimerizes and an inner pore is created by EccD<sub>3</sub> (128). This EccD<sub>3</sub>-pore is also wide enough to secrete the folded EsxGH, a known ESX-3 substrate. However, because of the hydrophobicity of the transmembrane helices from EccD<sub>3</sub> a unique, non-water mediated secretion mechanism would be required, which has not been observed in other bacterial secretion systems (128). These structures provide insight into the mechanism of secretion via ESX, and recently two structures of the ESX-5 core complex have been published or

made available as a preprint on bioRxiv (130,131) and no structure exists for the ESX-1 core machinery. The exact mechanism of secretion through the core ESX machinery remains elusive, despite the available structural information of the core complex. The pathway of translocation of substrates across the mycomembrane also remains unknown.

EccB is required for proper secretion (132) and interacts with MycP to tether MycP to the core complex (133). EccB<sub>1</sub> is 51 kDa protein with a small N-terminal soluble domain of about 40 amino acids and a large C-terminal soluble domain of about 400 amino acids connected by a single helix transmembrane domain, and all paralogs are roughly organized the same. The C-terminal domain of EccB<sub>1</sub> has been structurally characterized and revealed a single elongated fold in a distorted propeller shape, which includes a quasi 2-fold symmetry (134). A central core domain composing a six stranded  $\beta$ -sheet is flanked on either side by two repeat domains (134). The structure of EccB<sub>1</sub>'s C-terminal domain was not significantly similar to any other known structure (134). EccB<sub>1</sub> also has evidence of ATPase activity, but the localization of the large soluble domain in the periplasm does not indicate a clear functional role of the reported ATPase activity (135). The recent structures of the ESX-3 core machinery provided insight on EccB function as both reported structures reported dimerization of the protomers that was stabilized and/or facilitated by the dimerization of EccB<sub>3</sub>'s periplasmic domain (127,128).

EccC has two transmembrane helices located N-terminal to a domain of unknown function, which is followed by three ATPase domains. Interestingly, in ESX-1 EccC is split into two separate proteins, EccCa<sub>1</sub> containing the transmembrane helices, the DUF and the first ATPase domain and EccCb<sub>1</sub> which contains the other two ATPase domains. EccCa<sub>1</sub> and EccCb<sub>1</sub> interact to form a functional EccC (136). The hypothesized function of EccC is to recognize substrates and provide the energy required for their transport. EccC is required for proper secretion (136), and specifically its ATPase activity is required (127,137). Work on an EccC homolog suggests that it binds a specific sequence on a substrate that modulates EccC multimerization, and ultimately its activity (138). It has also been hypothesized that the third ATPase domain may specifically recognize cognate substrates as variety exists in numerous loops of *Mtb* EccC paralogs (139). Both ESX-3 core machinery structure publications proposed models of secretion through a channel partially formed by EccC<sub>3</sub>'s transmembrane helices and the pore's opening could be linked

to ATPase activity of EccC's ATPase domains through other members of the core machinery (127,128).

EccD has a small N-terminal domain followed by 11 transmembrane helices. It was originally thought that EccD could form the main pore in the plasma membrane as it has the most transmembrane helices of any conserved component. The N-terminal domain of EccD<sub>1</sub> was structurally characterized and revealed a ubiquitin-like fold (134). The structural characterization utilized two crystal forms of EccD<sub>1</sub> and in both forms EccD<sub>1</sub> dimerized, this along with the extensive nature of the dimerization interface suggested that EccD<sub>1</sub> is a natural homodimer (134). The homodimerization of EccD could be a tethering of two separate channels or perhaps the formation of one large pore containing all 22 transmembrane helices (134). With the core structures from ESX-3 it is now unlikely that EccD creates the central pore for the ESX system, instead EccD could be considered as the scaffold on which the core machinery is built as it interacts with every other conserved component, and in each ESX-3 core machinery structure EccD is dimerized along the N-terminal domain interface (127,128).

EccE contains an N-terminal domain with two transmembrane helices and a C-terminal soluble domain, that exists in the cytoplasm. To date no biochemical analysis has been published on EccE. A single structural study on the topology of the N-terminal transmembrane helices domain of EccE<sub>1</sub> was recently published (140), which agrees with the models of EccE<sub>3</sub> present in both ESX-3 core machinery structures (127,128). Additionally, EccE<sub>1</sub> has been shown to be essential for *Mtb* virulence *ex vivo*, and acts to stabilize the other members of the core machinery (141).

#### 1.3.4. Substrates of the ESX systems.

The substrates of the ESX systems can be classified into three major groups: the Esx proteins, the Esp proteins, and the PE and PPE proteins. The Esx and Esp proteins will be described in more detail within this subsection, while a more in-depth description of the PE and PPE proteins will follow in section 1.4. All of these proteins belong to the Pfam clan EsxAB (Pfam CL0352, (142)). It is hypothesized that the substrates are secreted as folded proteins with a structural motif of a four-helix bundle. This four-helix bundle is often created by the heterodimerization of two substrates, such as EsxA and EsxB (143) or PE25 and PPE41 (144), but can also be formed by a single protein, such as EspB (145,146).

Each substrate pair also contains a conserved ESX secretion signal of YXXXD/E (147,148). The signal was first identified as the C-terminal tail of EsxB, and shown to interact with EccC, but the critical residues of EsxB could not be identified (148). The critical residues of tyrosine (Y) and a negatively charged amino acid (D/E) separated by three residues are also found in PE25 (147). The ESX secretion signal is found within all PE proteins, EsxB and its homologs, and many of the Esp proteins (147). However, this signal is only a general signal for all of the ESX systems and does not confer any system specificity.

The Esx proteins are small proteins belonging to the WxG100 family (Pfam PF06013) within the larger EsxAB clan (Pfam CL0352), characterized by a tryptophan and glycine separated by a single residue in the turn of a helix-turn-helix structure of about 100 amino acids. Each ESX loci contains a pair of *esx* genes which form a heterodimer, that is a characteristic four helix bundle. There is a total of eleven tandem pairs of *esx* genes, five of which are located within the ESX-1 through ESX-5 loci, and the other six at other places within the genome (149). The *esx* genes located outside the ESX loci are closely related to ones within the ESX loci, and likely arose from recent duplication events (105). Functionally Esx proteins can serve a wide variety of functions, as EsxA and EsxB play important roles in ESX-1-related functions such as phagosomal rupture, host membrane pore formation, T-cell modulation, and macrophage recruitment in the granuloma (50,51,55-59,150-154). EsxG and EsxH serve very different functions in iron homeostasis (155). It seems there is no unifying function of Esx proteins. The Esx proteins are best represented by EsxA and EsxB of ESX-1, as they are the most well studied of the family. EsxA and EsxB are also known as ESAT-6 and CFP-10, respectively. EsxA was the first identified substrate of an ESX system (156), and in fact is responsible for the naming of the ESX systems. EsxB was discovered shortly after EsxA (157), and it forms a 1:1 heterodimeric complex with EsxA (158). The structure of the EsxAB heterodimer was the first of an ESX substrate (143), and with the solution structure of EsxGH from ESX-3 appears to be standard for Esx proteins (159). The Esx proteins are likely the most archaic substrate of the ESX systems as they are found in all ESX systems, not only in mycobacteria but also in other species (107). The more distantly related *Firmicutes* do not always contain a pair of *esx* genes, and instead only have a single *esx* gene. Yet there is



evidence that these single Esx proteins from *Firmicutes* systems homodimerize to form the classical helical bundle structure of ESX substrates and is secreted as a homodimer (160,161).

The Esp proteins are unique to ESX-1 and are not found in other ESX systems. The *espB*, *espE*, *espF*, *espH*, *espI*, *espJ*, *espK*, and *espL* genes are all located with the ESX-1 genetic loci (not shown in Figure 1.1.). In addition, an operon of *espACD* is located about 260 kbp upstream of the ESX-1 loci (162). EspG<sub>1</sub> is not a secreted protein and was originally mischaracterized since it is within the ESX-1 loci and does not share sequence similarity with its EspG paralogs in other ESX systems. It has kept its name for historical reasons, and simply follows the standard nomenclature for ESX homologs in multiple systems (103). EspG will be discussed further in later sections, and the rest of this section's discussion will focus on the other Esp proteins. The rest of the Esp proteins all belong to the EsxAB clan (Pfam CL0352) and are hypothesized to have the characteristic helix-turn-helix structural motif. Some Esp proteins have been shown to be secreted (162-168), while some have been shown to play roles in the secretion of other ESX-1 substrates (169-173). EspB is perhaps the most well studied Esp protein. Initially, it has been structurally characterized by two different groups and revealed that EspB's N-terminal domain shares similarities with the PE-PPE heterodimers (145,146). In solution EspB forms a heptameric complex in a donut-ring shape (145,146), and just recently a high resolution model of this structure was published (174). EspB is within all mycobacteria that contain an ESX-1 system, except for *M. leprae* (175). EspB functions in membrane lysis and is essential for macrophage cytotoxicity and intracellular mycobacterial growth (166-168,176). Loss of *espB* attenuates *M. marinum* in zebrafish and prevents conjugation in *M. smegmatis* (166,177). Interestingly EspB is required for EsxAB secretion, and likewise EsxA and EsxB are required for EspB secretion (168). Upon secretion EspB is cleaved by MycP, and the importance of this cleavage to the mature EspB in ESX-1 function and/or regulation is still under active investigation (129,167,168,176,178,179). While much is left to be uncovered about the Esp proteins and why they only exist in ESX-1, they highlight the interdependence of ESX substrates on each other for proper secretion.

## 1.4. The PE and PPE proteins.

### 1.4.1. Classification and genetic organization of PE and PPE proteins.

The PE and PPE families of proteins are a major class of substrates for the ESX systems, and they were discovered with the sequencing of the *Mtb* genome in 1998 (149). Remarkably, at the time of sequencing these two families encompassed 10% of the *Mtb* genome (149), although current analysis has them at about 7% of the genome (180). PE and PPE genes are also unique to mycobacterial species and are yet to be found in any other genera (124). These families are named for conserved Pro-Glu (PE) or Pro-Pro-Glu (PPE) residues within their conserved N-terminal domains, and both families can have expanded C-terminal domains. The conserved PE N-terminal domain is roughly 110 amino acids in length and forms a helix-turn-helix structure (PFAM: PF00934). The C-terminal portion of the PE domain also contains the conserved YxxxD/E ESX secretion signal, which is required for proper secretion (147). The conserved PPE domain is about 180 amino acids in length and is mostly helical in structure (PFAM: PF00823). PPE proteins do not contain the ESX secretion signal, and only contain the WxG motif between the second and third alpha helix of all PPE proteins (147,181). As it is currently understood, PE and PPE proteins form heterodimers in a helical bundle structure, and it is this folded structure that is secreted through the ESX systems (106). This PE-PPE helical bundle is extremely similar to the bundle seen in other ESX substrates, such as the EsxAB heterodimer, except that the PPE portion of the helical bundle has an extended helical tip (Figure 1.2.). Functionally most PE and PPE proteins are dependent on cell envelope or extracellular localization, and numerous studies have highlighted the conserved PE and PPE domains as the drivers of this localization (180,182-189). Along with their genetic linkage to the ESX systems, it has been experimentally shown that many PE and PPE proteins rely on ESX systems for proper localization (124,155,185,190,191).

PE and PPE proteins can be found within pathogenic and non-pathogenic mycobacterial species, however they are extremely expanded within pathogen species, such as *Mtb*. For example, there are only 2 *pe* and 2 *ppe* genes in the non-pathogenic *M. smegmatis*, while *Mtb* has 99 *pe* genes and 69 *ppe* genes (124,149). The expansion of *pe* and *ppe* genes in pathogenic mycobacteria has suggested an important role in virulence for the PE and PPE proteins, yet their prevalence in non-pathogenic species also suggests they

originally carried a more central role to mycobacterial metabolism and that some still do (124,192). The expansion of *pe* and *ppe* genes is also closely linked to the expansion of the ESX systems (124). Expansion of *pe* and *ppe* genes are also tightly linked with pathogenicity, as *Mtb* and the other members of the Mycobacterium tuberculosis Complex, which includes closely related species that cause Tb in humans or other animals, contain the most copies of *pe* and *ppe* genes (193).

It is also important to note that it is unclear exactly which PE and PPE proteins interact together to form heterodimers. While some *pe* and *ppe* genes are located in operons with only one possible pairing, like *pe5-ppe4* and *pe25-ppe41*, others are located in uneven groupings, like *ppe25-pe19* in the ESX-5 cluster, where it is unclear which *pe* and *ppe* genes encode functional PE-PPE dimers. Some *pe* genes also exist in operons either with only additional *pe* genes and even some exist by themselves without any adjacent *pe* or *ppe* genes (144). There exists an uneven amount of each group, with 99 *pe* genes and 69 *ppe* genes (149). Since both PE and PPE groups can have extended C-terminal domains, each subgroup can be further divided based upon the features of the C-terminal domain.

PE proteins can be sub-classified into two distinct groups: PE and PE-PGRS. The PE-PGRS sub-classification is for *pe* genes with expanded C-terminal domains that contain polymorphic GC-rich sequences (PGRS). The PE sub-classification contains all other *pe* genes and they can contain either no expanded C-terminal domain or a C-terminal domain with no distinctive features. Most *pe* genes belong to the PE-PGRS sub-classification, including 65 of the 99 genes in *Mtb* (123,124,149). The PGRS domain contains the repeated motif of GGAGGX, which can be repeated over 30 times (194). The size of PE-PGRS proteins varies wildly from as small as 175 amino acids in length, to as large as 1900 amino acids, with the average PE-PGRS protein being 550 amino acids (194). Due to the extended size of the GC-rich repeats within the genes and the high hydrophobicity of the amino acid repeats, along with the long length of many PGRS domains it has been incredibly difficult to study the PE-PGRS proteins exact functions (123). Originally the PE-PGRS proteins were thought to be involved in immune evasion strategies, although that has been under increasing scrutiny in recent years (195-199).

PPE proteins are sub-classified into three groups: PPE, PPE-PPW PPE-SVP, and PPE-MPTR. PPE-PPW proteins contain a PXXPXXW motif that is located about 10 to 30

amino acids from the C-terminus (123). PPE-SVP proteins contain a conserved SVP motif within their C-terminal domain and are the largest of the PPE groupings (123). The PPE-SVP proteins average between 350 and 468 amino acids, other than two truncated genes (123). PPE-MPTR proteins have C-terminal domains that contain major polymorphic tandem repeats (MPTR) (123). These repeats vary around a NXGXGNXG motif and have widely variable sizes with some members being over 3700 amino acids in length (123). The PPE grouping contains all other *ppe* genes. A vast majority of PPE proteins are hypothesized to be dependent on ESX-5 for secretion including both PPE-SVP and PPE-MPTR groups (121,191,200,201).

The first PE-PPE structure was of PE25-PPE41 and was published in 2006 (144). To date no other PE-PPE heterodimer structure has been solved without the presence of the cytosolic chaperone EspG. Full discussion on EspG and its function in context of PE-PPE heterodimers will follow in the next subsection. The lack of structures of other PE-PPE heterodimers highlights the difficulty associated with the study of these two families. It has been incredibly challenging to study individual PE and PPE proteins and most do not have assigned functions. The limited PE and PPE proteins with known functions mostly have C-terminal domains with conserved features, such as LipY (185). There are some that are known to function in broader pathways, such as PE5-PPE4 in iron metabolism (155) and PE22-PPE36 and PPE62 in heme utilization (183), but their exact roles in these pathways is unclear. There is also thought that many *pe* and *ppe* genes are redundant or co-dependent on each other, which further complicates knockout experiments in mycobacteria (123).

#### 1.4.2. PE and PPE interactions with other ESX proteins.

A discussion of PE and PPE proteins is incomplete without mention of their cytosolic chaperone EspG. A copy of the *espG* gene is located within each ESX locus that also contains a *pe* and *ppe* gene, with only ESX-4 in *Mtb* lacking an *espG* gene (Figure 1.1.). It was originally not thought to be a conserved component since sequence homology of *espG* paralogs across the ESX systems is low, however significant homology exists and in fact each structurally characterized EspG has the same overall fold (103,202). EspG specifically interacts with PPE proteins, and recent structural studies of PE-PPE-EspG heterotrimers has uncovered this interaction in ESX-5-specific heterotrimers (203-206).

The first two structures were of the PE25-PPE41-EspG<sub>5</sub> heterotrimer, which is also the only PE-PPE heterodimer to be structurally characterized without EspG present (144,204,205). Analysis of the PPE41-EspG<sub>5</sub> interface highlighted that the tip of PPE41, which includes the helix-turn-helix motif between  $\alpha 4$  and  $\alpha 5$ , is completely shielded by interactions with EspG<sub>5</sub> (204,205). Also, of note is that the structure of the PE25-PPE41 heterodimer is conserved between the apo- and EspG<sub>5</sub>-bound structures (204,205). It is hypothesized that binding of EspG does not alter the conformation of PE-PPE heterodimers. EspG aids in the folding and/or stability of PE-PPE heterodimers, partially through preventing self-aggregation of the heterodimers (205). Currently PE25-PPE41 is the only known PE-PPE heterodimer to stably express in an *E. coli* expression system without the presence of its EspG chaperone (144,204,205). In the loop between  $\alpha 4$  and  $\alpha 5$  of PPE proteins exists an hh motif, characterized by two hydrophobic amino acids. PPE41's hh motif is Ala-Leu, and mutations to bulkier hh motifs reduced protein yield compared to WT (205).

EspG only binds cognate PE-PPE heterodimers (203,205). However, the molecular mechanism for this selectivity is unclear. Currently the only structural information available on PE-PPE-EspG heterotrimers comes from the ESX-5 system with PE25-PPE41-EspG<sub>5</sub> and PE8-PPE15-EspG<sub>5</sub> being the only structurally characterized heterotrimers (204-206). Analysis of the EspG-binding region on PPE proteins showed a high level of conservation at a sequence level regardless of which ESX system the PPE is associated with (205). Even mutational studies on PE25-PPE41 to alter the EspG-binding region of PPE41 to match ESX-1- and ESX-3-specific PPE proteins were unable to block PPE41-EspG<sub>5</sub> interaction (205). EspG recognition of cognate PPE proteins is still unclear.

The interaction between PPE and EspG ranges from the low nanomolar range to low hundred nanomolar range depending on the PE-PPE heterodimer and methodology used (204-206). In spite of this strong affinity, the PE-PPE heterodimers are secreted through their cognate ESX systems, while EspG is maintained within the cytosol (203). An extra factor must uncouple PE-PPE heterodimers from EspG for efficient secretion, and the primary suspect is EccA. EccA is a conserved component of the ESX systems but is not a part of the conserved secretion machinery that sits within the plasma membrane. EccA is also not encoded in ESX systems that do not encode PE and PPE proteins, such as ESX-

4 (Figure 1.1.). Interestingly, an *EccA5* knockout accumulated EspG-bound PPE proteins in the cytosol (203). EccA has an N-terminal domain with six tetratricopeptide repeat (TPR) motifs and a C-terminal ATPases associated with various cellular activities (AAA) domain. The N-terminal domain of *EccA1* has been structurally characterized and it forms a canonical super-helix structure with a unique  $\beta$ -finger insertion in the concave face (207). Six tandem TPR repeat motifs form a right-handed super-helix that has a pitch of about 60 Å and a width of about 40 Å (207). The  $\beta$ -finger insert is between TPR motifs 2 and 3 and resides within the concave face of the super-helix (207). TPR domains are well defined protein-protein interaction domains, and although they usually mediate interactions with the concave face of their super-helix, there is evidence of proteins utilizing the convex face or similar  $\beta$ -finger insertions to mediate interactions (208-212). EccA interacts with both PPE and EspG proteins in yeast-two hybrid experiments, and this is likely mediated by the N-terminal TPR domain (204,213). The C-terminal domain of EccA can function as an ATPase (214,215), although there is no published evidence of ATPase activity of any full-length EccA paralog. It is plausible that EccA's interacting partner would need to be present to stimulate its ATPase activity.

#### 1.5. Functions of the ESX systems.

ESX-1 was the first discovered ESX system (136). Since its discovery it has been perhaps the most studied of all the ESX systems. ESX-1 secretes the two most well studied ESX substrates, *EsxA* and *EsxB*, which were the first discovered substrates and important antigens for immune responses (136,156). Loss of ESX-1 function was also identified as a key contributor to *M. bovis* BCG attenuation (15). *Mtb* and *M. marinum* are also attenuated in their natural hosts without a functional ESX-1 system (16,136,166,216). There are numerous virulence functions that have been assigned to ESX-1 including: phagosomal escape into the cytosol, inhibiting T-cell responses, inhibiting phagosomal maturation and acidification, autophagy, and host cell-death (150-154,217). ESX-1 also highlights the interdependence of substrates on each other for proper secretion. Deletion of both *esxA* and *esxB* blocks the secretion of the ESX-1 substrate *EspA* (162). Since *M. bovis* BCG lacks a functional ESX-1 system it cannot escape the phagosome and enter the host cytosol, and this may have important immune response side effects. *M. bovis* BCG does not induce some immune signaling pathways because of its inability to access the host cytosol and

this could be the reason for why it is not effective at preventing pulmonary TB (218,219). ESX-1 is also found in non-pathogenic species. In these non-pathogenic species ESX-1 is required for conjugation and functions to negatively regulate the process in donor cells, while its activity is required for recipient cells (177,220,221).

ESX-3 is essential for the *in vitro* growth of *Mtb* (118,119). The entire genetic loci containing ESX-3 is under the regulation of transcription factors Zur, and IdeR (222,223), being down-regulated in the presence of zinc and iron, respectively. The ESX-3 locus is also induced during manganese limitation through MntR (224), indicating a role for ESX-3 in broad divalent cation homeostasis processes. This has direct virulence implications, especially because of the iron homeostasis roles of ESX-3. The availability of aqueous ferric iron ( $\text{Fe}^{3+}$ ) is incredibly scarce within human serum at about  $10^{-24}$  M (225). Humans use a number of proteins to sequester and store iron. Pathogenic bacteria have to find a way to uptake and utilize iron from the host. *Mtb* has two pathways to do this, either via its siderophores mycobactin or carboxymycobactin, or in a siderophore-independent uptake and utilization of heme-bound iron sources. ESX-3 effectors have roles in iron uptake, as loss of secretion prevents mycobacteria from properly utilizing mycobactin (120,155,226). Interestingly, ESX-3 is not essential in the non-pathogenic *M. smegmatis* (118), although this could be due to redundant iron, zinc, and/or manganese homeostasis regulators that have been lost through genetic reduction in *Mtb*. ESX-3 also modulates the host response to the *Mtb*-containing phagosome through the EsxGH heterodimer. The action of EsxGH is iron-independent, as a  $\Delta\text{esxH}$  strain is attenuated *in vivo*, while a  $\Delta\text{pe5-pep4}$  strain is not, despite both strains displaying a similar iron phenotype (155). The connection between iron or other divalent cations homeostasis and the modulation of the host response to the *Mtb*-contained phagosome could be a mechanism to signal the proper spatial and/or temporal timing to the mycobacteria.

ESX-5 functions to modulate the host immune responses and nutrient uptake (121,122,200,227) and also potentially in the biogenesis and/or homeostasis of the cell envelope (200,228). ESX-5 is the most recently evolved ESX system, and its emergence seems to correlate with the expansion of the PE and PPE families. Most PE and PPE are expected to be secreted through ESX-5 as all PPE-SVP, PPE-MPTR, and PE-PGRS genes are ESX-5-substrates (121,191,200,201). In fact, the only *pe* genes not hypothesized to be

secreted via ESX-5 are located within an ESX cluster or are closely associated with a specific class of *ppe* genes, the *ppe-ppw* genes, which are all hypothesized to be secreted via ESX-3 (121,200,201,229-231).

As mentioned earlier, evidence of a functional ESX-4 has only recently been established. *M. abscessus* intracellular survival depends on functional ESX-4 (114). *M. abscessus* is an emerging opportunistic pathogen (232-240), and the environmental host of *M. abscessus* remains unknown (241). *M. abscessus* can survive within free-living amoeba (242,243) and its virulence increases upon co-culturing with amoebae (244). Importantly, the *M. abscessus* ESX-4 system contains an *eccE<sub>4</sub>* gene (245). The *M. abscessus* ESX-4 system showed similar functions to the *Mtb* ESX-1, such as blockage of phagosomal acidification and phagosomal membrane permeation (114). As *M. abscessus* lacks an ESX-1 system (104,245), perhaps ESX-4 has replaced this function. The ESX-4 system from *M. smegmatis* has also recently been implicated in conjugation (115,116). Functional ESX-4 is required in recipient *M. smegmatis* strains for conjugation to occur as loss of either *eccC<sub>4</sub>* or *eccD<sub>4</sub>* abolished conjugation (116). Interestingly, ESX-4 function was dispensable for donor strains (116). In fact, donor strains actually signal recipient strains to express ESX-4 genes through SigM (115). Therefore, ESX-4 and ESX-1 have non-redundant roles in conjugation (116).

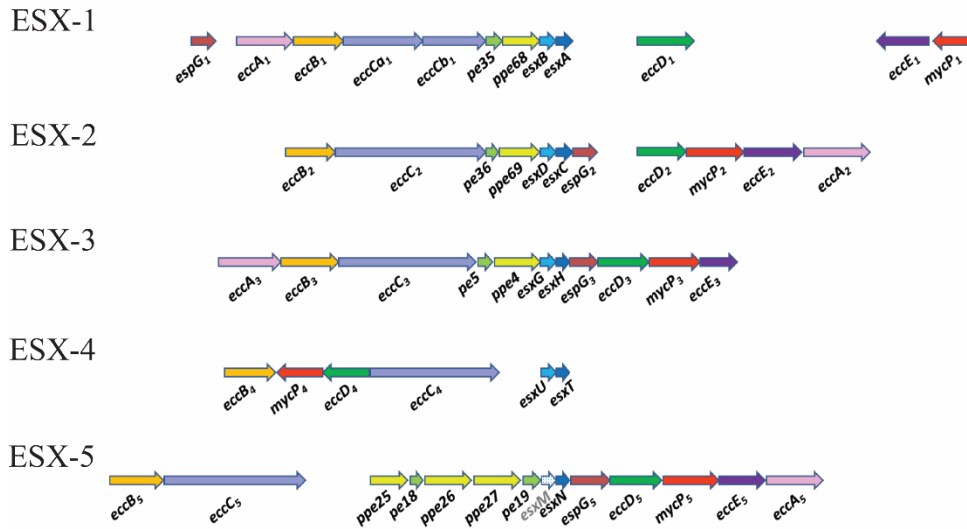
Currently there is no evidence that ESX-2 secretes any substrates and whatever function, if any, it serves remains unclear.

#### 1.6. Conclusions and dissertation overview.

The PE and PPE proteins are uniquely mycobacterial proteins that are a large substrate class of the ESX systems. Each of the three active ESX systems in *Mtb* secrete their own unique subset of PE-PPE heterodimers, and there is no cross-talk between any of the systems. PE-PPE heterodimers interact with a cytosolic chaperone, EspG, that is unique to its cognate ESX system and aids in the folding and/or stability of the heterodimers and keeps them in a secretion-competent state. It is poorly understood what the molecular mechanisms that dictate EspG binding and what determines system specificity. The work in this dissertation aims to close the gap by providing a more complete picture of the PE-PPE heterodimer interaction with EspG. Chapter 3 examines the ESX-5 system and how its chaperone, EspG<sub>5</sub>, interacts with the large variety of ESX-

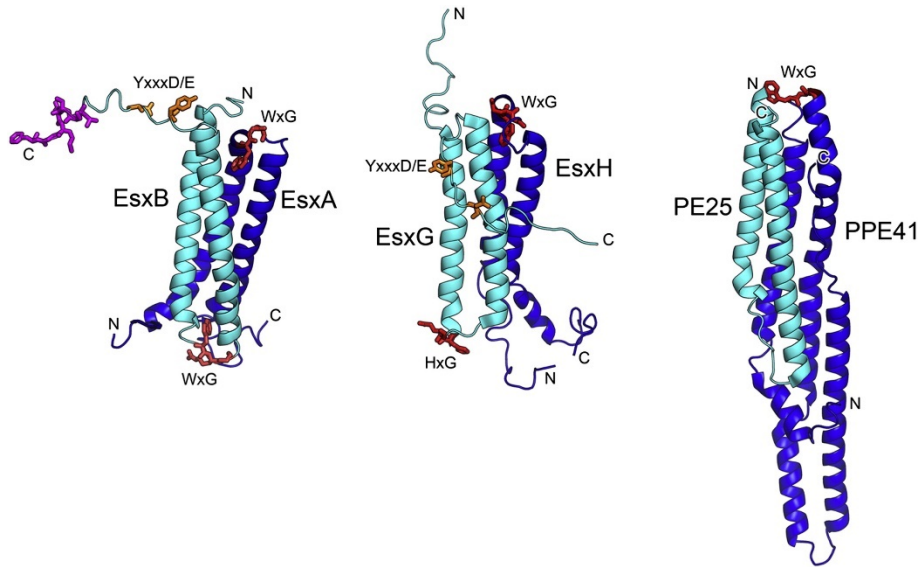


5-specific PPE proteins. Chapter 4 presents the structure of the PE5-PPE4-EspG<sub>3</sub> heterotrimer, the first from the ESX-3 system, and analyzes how EspG differentiates between cognate and non-cognate PPE proteins. Chapter 5 provides some discussion on the results from Chapters 3 and 4, and how they fit in the field, along with some exciting new avenues of future research.



**Figure 1.1. Genetic organization of ESX clusters in *Mtb*.**

Genetic organization of ESX clusters in *Mtb*. Only conserved core components, the *ecc* genes and *mycP*, selected other core components, *espG* and *eccA*, and selected substrates, *esx*, *pe*, and *ppe*, genes are shown. Paralogous genes are color coded the same color. The *esp* genes from ESX-1 are excluded.



**Figure 1.2. Structures of selected ESX substrates.**

This figure is reprinted from (105) with permission. The structures of the EsxAB heterodimer (143), EsxGH heterodimer (159), and PE25-PPE41 heterodimer (144) are presented from left to right. The WXG motifs on EsxA, EsxB, EsxH, and PPE41 are highlighted. EsxG has a HXG variant of the WXG motif. The YXXXD/E ESX secretion signal motifs (147) in EsxAB and EsxGH are also highlighted. This signal is disordered in the PE25-PPE41 structure. EspB is not shown, but its N-terminal domain has the same fold as the PE25-PPE41 heterodimer (145,146).

## Chapter 2. Materials and Methods.

### 2.1. Bacterial Strains and growth conditions.

*Escherichia coli* Rosetta2 (DE3) strains were grown in Luria-Bertani Miller (LB) liquid medium (IBI Scientific) or on solid LB agar (LB liquid with 1.5 % w/v agar) at 37° C. When grown in liquid medium, cells were agitated at 200 rpm. Antibiotics were added, as needed, at the following concentrations: chloramphenicol at 10 µg/ml, streptomycin at 50 µg/ml, and kanamycin at 50 µg/ml.

### 2.2. Bacterial Transformation.

Competent *E. coli* Rosetta2 (DE3) cells were thawed from -80° C on ice for 10 m in 50 µl aliquots. Plasmids were added to cells at 1 µl per 100 mg/ml. If multiple constructs were transformed into the same cell, they were added at the same time. The cell-plasmid mixture was incubated on ice for an additional 20 min. Cells were heat shocked at 42 °C for 60 s and placed immediately back on ice for an additional 5 m. Afterwards, 500 µl of Super Optimal Broth (SOB, 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, and 10 mM MgSO<sub>4</sub>) was added and the cells were grown at 37 °C with agitation for 50 min. Cells were plated on LB agar plates with appropriate antibiotics and 2% glucose and allowed to grow at 37 °C for 20 h. A single colony was selected and grown in liquid media at 37 °C for 20 h. Selected colonies were then stored in liquid LB supplemented with 15% glycerol at -80 °C until needed for expression and purification.

### 2.3. Expression and purification of recombinant proteins.

#### 2.3.1. PE5-PPE4-EspG<sub>3</sub> heterotrimers.

DNA sequences of full-length PE5 and residues 1-180 of PPE4 from the *Mycobacterium tuberculosis* genome were optimized for *E. coli* expression and obtained from Invitrogen. Constructs were inserted into a pRSF-NT vector (246) using NcoI and HindIII restriction sites. When expressed, PE5 has an N-terminal His<sub>6</sub> tag that is cleavable by TEV protease. Plasmids containing EspG<sub>3</sub> from *Mycobacterium marinum* and *Mycobacterium tuberculosis* were created by the Korotkov Lab and are described in (205). Plasmids containing EspG<sub>3</sub> from *Mycobacterium smegmatis*, *Mycobacterium hassiacum*, and *Mycobacterium kansasii* were also created by the Korotkov Lab as

described in (247). Mutations in PPE4<sub>mt</sub>, EspG<sub>3mt</sub>, and EspG<sub>3mm</sub> were introduced with Gibson assembly mutagenesis (SGI-DNA).

*E. coli* strains containing the appropriate PE5<sub>mt</sub>-PPE4<sub>mt</sub> and EspG<sub>3</sub> plasmids were induced at an A<sub>600nm</sub> of 0.5 to 0.8 with 0.5 mM of isopropyl β-D-thiogalactopyranoside. After induction strains continued growing at 16 °C for 20 h. Cells were harvested by centrifugation at 5000 rpm. Cells were resuspended in a 1:10 ratio of Lysis Buffer (300 mM NaCl, 20 mM Tris-HCl, pH 8.0, and 10 mM imidazole) with the addition of 1:100 ratio of Halt protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA). The cells were lysed using an EmulsiFlex-C5 homogenizer (Avestin, Ottawa, Canada). Soluble lysate was separated from insoluble lysate by centrifugation at 18000 rpm for 50 m. Soluble lysate was purified over a nickel-nitrilotriacetic resin in a two-step purification. For the first step after loading the soluble lysate, the resin is washed with Lysis Buffer and then the heterotrimer is eluted with Elution Buffer (300 mM NaCl, 20 mM Tris-HCl, pH 8.0, and 250 mM imidazole). The His<sub>6</sub> tag is cleaved by addition of TEV Protease in a 1:20 ratio at 4° C for 20 h, which is performed while dialyzing the sample against Wash Buffer (300 mM NaCl and 20 mM Tris-HCl, pH 8.0). Cleaved sample is then reapplied to the nickel-nitrilotriacetic resin, and is this time washed with Wash Buffer (300 mM NaCl and 20 mM Tris-HCl, pH 8.0) before being eluted with Elution Buffer. Samples from the first and second steps were analyzed by SDS-PAGE for cleavage of His<sub>6</sub>-tag and homogeneity of samples from the second step. Flow-through and wash fractions from the second step were pooled and concentrated for size-exclusion chromatography over a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences, Marlborough, MA) that was pre-equilibrated in Buffer SE (100 mM NaCl and 20 mM HEPES, pH 7.5). Samples from the size-exclusion chromatography run were analyzed by SDS-PAGE for purity and then fractions were pooled and concentrated before either being utilized immediately or flash frozen in liquid N<sub>2</sub> and stored at -80° C.

### 2.3.2. PE25-PPE41-EspG<sub>5</sub> heterotrimers.

All PE25<sub>mt</sub>-PPE41<sub>mt</sub> constructs were obtained from lab storage and their creation is outlined in (205). An EspG<sub>5mm</sub> construct was created by the Korotkov lab. The protocol outlined in 2.3.1. was followed for expression and purification of all PE25-PPE41-EspG<sub>5</sub> heterotrimers.

## 2.4. Crystallization, data collection, and structure solution.

### 2.4.1. *PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>5mm</sub>*.

Purified protein was concentrated to 4.2 mg/ml. Initial screening was done with the MCSG crystallization suite (Anatrace, Maumee, OH) and was set up with the mosquito crystal crystallization robot (SPT Labtech, Melbourn, England) in 200 nL drops with mother liquor to protein ratios of 1:3, 1:1, and 3:1. The initial screen produced P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> crystals grown in 200 mM NH<sub>4</sub> tartrate and 20% PEG 3350. Optimization around three other hits from the initial screen that all contained NaCl as the precipitant and various buffers ranging from pH 5.5 to pH 8.0, led to the growth of I422 crystals, which were grown in 2.0 M NaCl and 100 mM Bis-Tris, pH 6.5. Crystals of each form were harvested and transferred to cryoprotectant solution containing the crystallization mother liquor supplemented with either 20% (P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>) or 25% (I422) glycerol and flash cooled in liquid N<sub>2</sub>. Data was collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID Beamline at the Advanced Photon Source, Argonne National Laboratory. Data were processed using XDS and XSCALE (248). Molecular replacement using Phaser (249) was used to solve structures from both crystal forms. First, the model of the PE25<sub>mt</sub>-PPE41<sub>mt</sub> heterodimer from 4KXR (205) and the model of EspG<sub>3mm</sub> from 5DLB (202) were used as search models for the I422 data set. Later, an early model of the I422 data set was used as a search model for the P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> data set. Starting models of both P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> and I422 data sets were iteratively rebuilt and refined using Coot and phenix.refine (250,251). The final model of both crystal forms was refined in phenix.refine (251), with the P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> model using non-crystallographic symmetry restraints. The final models were assessed using Coot (250) and the MolProbity server (252) for quality.

### 2.4.2. *PE25<sub>mt</sub>-PPE41<sub>mt</sub>-EspG<sub>5mm</sub>*.

Purified PE25<sub>mt</sub>-PPE41<sub>mt</sub>-EspG<sub>5mm</sub> was concentrated to 3.41 mg/ml and initial screening was performed using MCSG (Anatrace, Maumee, OH). Screens were set up with the mosquito crystal crystallization robot (SPT Labtech, Melbourn, England) as outlined in 2.4.1. Diffraction quality crystals were grown in the original MCSG screening plates, with the well containing 100 mM sodium citrate pH 5.6, 10% (w/v) PEG 4000, and 10% (v/v) isopropanol. Crystals were harvested and transferred to a cryoprotectant solution containing the crystallization mother liquor supplement with 30% glycerol and then flash-

cooled in liquid N<sub>2</sub>. Data was collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID Beamline at the Advanced Photon Source, Argonne National Laboratory. Data were processed using XDS and XSCALE (248). Molecular replacement was performed using Phaser (249) with WT PE25<sub>mt</sub>-PPE41<sub>mt</sub>-EspG5<sub>mt</sub> (PDB: 4KXR, (205)) as the search model. The starting model was iteratively rebuilt and refined using Coot and phenix.refine (250,251). The final model was refined in phenix.refine (251). The final model was assessed using Coot (250) and the MolProbity server (252) for quality.

## 2.5. Size-exclusion chromatography with multi-angle light scattering (SEC-MALS).

Proteins were expressed and purified as described in 2.3.1. and then passed over an AKTA pure with an inline Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences), miniDAWN TREOS, and Optilab T-rEX (Wyatt Technologies, Santa Barbara, CA). The system was equilibrated and run with in Buffer SE (100 mM NaCl and 20 mM HEPES, pH 7.5). The samples were loaded at a volume of 500 µl at a concentration of 2-4 mg/ml, and the system was run at 0.5 ml/min. Analysis of light scattering data was performed using Astra (Wyatt Technologies). Molecular mass determination was done by analyzing peaks at one-half their maximum. Graphics from these results were prepared using Prism (GraphPad Software, La Jolla, CA).

## 2.6 Small angle X-ray scattering (SAXS) data comparison and *ab initio* model reconstruction.

PE5<sub>Ms</sub>-PPE4<sub>Ms</sub>-EspG3<sub>Ms</sub> heterotrimer SAXS data (SASDDX2) (202) was compared with a single copy of the mixed PE5<sub>mt</sub>- PPE4<sub>mt</sub>-EspG3<sub>mm</sub> heterotrimer structure (PDB code 6UUJ) using CRY SOL (253). *Ab initio* reconstruction of the envelope was completed using GASBOR (254). Monomeric symmetry was used as a constraint for GASBOR. Twenty *ab initio* models were generated and averaged using the DAMAVER software package (255). DAMSEL rejected only one model.

## 2.7. Sequence Analysis.

Sequence alignments were performed using the EMBL-EBI analysis tools, specifically the Clustal Omega program (256) via the EMBL-EBI webpage. Rendering of sequence analysis was done with the ESPript server via its webpage (257).

## 2.8. Structural Analysis.

Structural figures were generated using PyMOL (<http://www.pymol.org>). Electrostatic surface potentials were calculated using the APBS Electrostatics plugin in PyMOL (258). Structural alignments were performed using the Dali server (259). Packing analysis was performed with the Voronoia software (260) using the ProtOr radii definition (261) and grid spacing of 0.05 Å while reporting all atoms. Sequence tolerance was performed on ROSIE (262) with the `sequence_tolerance` protocol (262-264) with the generalized protocol and the default options (264).

## 2.9. Data Availability.

The coordinates and structure factors were deposited in the Protein Data Bank with accession codes 6UUJ (PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub>, P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> data), and 6VHR (PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub>, I422 data). The hh mutant highlighted in Chapter 3 was crystalized and solved by the Korotkov Lab and is also deposited in the Protein Data Bank with accession code 6VJ5.



### **Chapter 3. Structural plasticity of EspG<sub>5</sub> accommodates variety seen in ESX-5-specific PPE proteins.**

#### 3.1. Introduction.

Binding of EspG to PE-PPE heterodimers is required for proper folding and/or stability of the PE-PPE heterodimers (203,205). Each ESX system secretes its own unique subset of PE-PPE heterodimers and each system encodes its own unique copy of EspG. EspG only binds to the PE-PPE heterodimers from its cognate ESX system (203,205). Structures of ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimers showed that EspG<sub>5</sub> makes extensive contacts with the PPE protein, and not with PE (204-206). Currently there are three different structures of two different heterotrimers from the ESX-5 system: two PE25-PPE41-EspG<sub>5</sub> structures that represent the same crystal form (204,205), and one PE5-PPE18-EspG<sub>5</sub> structure (206). We hypothesize that all ESX-5-specific PPE proteins bind to EspG<sub>5</sub> in the same manner as these structures.

ESX-5 is hypothesized to secrete the most PE-PPE heterodimers, and currently 46 PPE proteins are believed to be exported via ESX-5 (123). There is variety within the EspG<sub>5</sub>-binding domain of PPE proteins, as defined by the PPE41-EspG<sub>5</sub> interface. Of note is the hh motif within this binding domain that is located in a turn between two alpha helices of PPE and is buried within the PPE-EspG<sub>5</sub>. The hh motif is simply two hydrophobic residues and is highly variable within ESX-5-specific PPE proteins. The hh motifs of PPE41 and PPE18 are AL and VL, respectively. Bulkier hh motifs, such as FF and WF, can be found within ESX-5-specific PPE proteins. The PPE-EspG<sub>5</sub> interface would require some level of plasticity to allow for the various substitutions found within the interface, particularly with the substitutions found within the hh motif. Ekiert and Cox (204) hypothesized that this was accomplished by underpacking of the interface in the PPE-EspG<sub>5</sub> interface of PPE proteins that contained smaller hh motifs, such as PPE41 and its motif of AL.

In this chapter we attempted to crystallize various PPE41 mutants in the context of the PE25-PPE41-EspG<sub>5</sub> heterotrimer to investigate how the bulkier hh motifs found within the ESX-5-specific PPE proteins affect the PPE41-EspG<sub>5</sub> interface. Here we present the structure of PE25<sub>mt</sub>-PPE41<sup>A124L</sup><sub>mt</sub>-EspG<sub>5mm</sub>, which contains an hh motif of LL. Our results suggest that EspG<sub>5</sub> is dynamic and the PPE-EspG<sub>5</sub> interface is flexible to accommodate the

various amino acids found within the numerous ESX-5-specific PPE proteins, particularly within the hh motif.

## 3.2. Results.

### 3.2.1. Structure of PE25<sub>mt</sub>-PPE41<sup>A124L</sup><sub>mt</sub>-EspG<sub>5mm</sub>.

To understand how EspG<sub>5</sub> accommodates the variety of hh motifs present in ESX-5-specific PPE proteins, high-resolution structures were needed. Ideally the structure of unique PE-PPE heterodimers in complex with EspG<sub>5</sub> would have been solved, but the uncertainty in defining PE-PPE interacting pairs lead us to take the defined PE25-PPE41 heterodimer and make mutations in PPE41's hh motif. As mentioned earlier PPE41 has an AL hh motif, and we made mutations to create PE25-PPE41 heterodimers with LL, LF, FF, or WF hh motifs. These motifs represent 56.5% of the hh motifs found in ESX-5-specific PPE proteins (Figure 3.1.). We also made the decision to co-express the hh motif mutants with *Mycobacterium marinum* EspG<sub>5</sub> due to stability and/or degradation issues we noticed in our previous work with *M. tuberculosis* EspG<sub>5</sub> (205). We also crystallized the mixed heterotrimer of WT PE25<sub>mt</sub>-PPE41<sub>mt</sub>-EspG<sub>5mm</sub>, and while it was successful the quality of the data and subsequent model was worse than the lab's previous full *Mtb* WT heterotrimer (4KXR, (205)). Therefore, the mixed heterotrimer was not used for any comparison analysis.

We were able to obtain diffraction quality crystals of only the LL hh motif mutant, and the crystals diffracted to 2.4 Å (Table 3.1). The mutations in the hh motif do not affect the overall fold of any protein in the heterodimer, as the heterotrimer is similar to the previously reported ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimers, including the WT PE25-PPE41-EspG<sub>5</sub> heterotrimers, and crystallized in the same crystal form (204,205) (Figure 3.2A).

The PE25-PPE41 heterodimer is a compact, elongated alpha-helical bundle. PPE41 encompasses the entire heterodimer, while PE25 is located at one end and forms a four helical bundle with  $\alpha 2$  and  $\alpha 3$  of PPE41. The extension of PPE41, containing  $\alpha 4$  and  $\alpha 5$ , interacts with EspG<sub>5</sub>. Importantly, the C-termini of both PE25 and PPE41 are located distal from EspG<sub>5</sub>, as many PE and PPE proteins have expanded C-terminal domains it is unlikely that they would sterically clash or interact with EspG<sub>5</sub>. The YXXXD/E motif on PE25 is

also distal to the PPE41-EspG<sub>5</sub> interaction and is available for interaction with the core ESX machinery.

The mutant heterotrimer was aligned to both available models of the WT heterotrimer and to the PE8-PPE15-EspG<sub>5</sub> model (206) and each component aligns well with RMSD less than 0.4 Å for both PE25 and PPE41 to both models and 0.7 Å for EspG<sub>5</sub> to both models (Table 3.2). The PPE41-EspG<sub>5</sub> interface buries 3210 Å<sup>2</sup> of solvent-accessible surface area, as calculated by the PISA server (265), and is similar to the interface area of the previous WT models of 2880 Å<sup>2</sup> (205) and 3500 Å<sup>2</sup> (204). Despite the overall similarity of the PE25-PPE41<sup>A124L</sup>-EspG<sub>5</sub> heterotrimer to the previously reported ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimers, we noticed differences between them in their PPE-EspG<sub>5</sub> interfaces. We next investigated these PPE-EspG<sub>5</sub> interfaces to understand the changes caused by the mutation of the WT hh motif to LL.

### 3.2.2. Structural plasticity of EspG<sub>5</sub> accommodates variety in hh motifs.

Accommodation of bulkier hh motifs by EspG<sub>5</sub> could be accomplished by either an under-packing of the PPE41<sup>A124L</sup>-EspG<sub>5</sub> interface to allow the bulkier residues to fit in the already defined interface, or alternatively by structural flexibility of EspG<sub>5</sub> to adjust to fit the variety of hh motifs present in ESX-5-specific PPE proteins. We compared our interface of the mutant PPE41<sup>A124L</sup>-EspG<sub>5</sub> to the WT PPE41-EspG<sub>5</sub> interface to see if there is any structural change in EspG<sub>5</sub>. To best compare the structural differences in EspG<sub>5</sub> in the two different interfaces, we utilized the PPE41-based structural alignment. Upon initial visual inspection of the alignment, we noticed a deviation in the EspG<sub>5</sub> proteins in the C-terminal helical bundle (Figure 3.2B). We next used Gesamt (266) in the CCP4 suite (267) to analyze the deviations on a per-residue basis with an alignment of 4KXR and 6VJ5's EspG<sub>5</sub> (Figure 3.2C). The deviations matched our visual inspection, as the C-terminal helical bundle residues largely deviated by more than the overall rmsd of 0.7 Å. We hypothesize that this helical bundle structurally adjusts to accommodate the variations in hh motifs of the ESX-5-specific PPE proteins.

We next took a closer inspection on the hh motif and its interactions with EspG<sub>5</sub> to investigate which residues were altered by the mutation from Ala to Leu in PPE41<sup>A124L</sup> by comparing the 4KXR and 6VJ5 models. We utilized the MolProbity all-atom contacts (268) feature in Coot (250) to identify which residues of EspG<sub>5</sub> are interacting with hh

motif of PPE41 in the 4KXR model. It identified Q187<sup>EspG5</sup>, F191<sup>EspG5</sup>, and L216<sup>EspG5</sup> as interacting with A124<sup>PPE41</sup>; and M179<sup>EspG5</sup>, L180<sup>EspG5</sup>, T183<sup>EspG5</sup>, L254<sup>EspG5</sup>, and L268<sup>EspG5</sup> as interacting with L125<sup>PPE41</sup>. As L125<sup>PPE41</sup> is present in both 4KXR and 6VJ5, we did not expect many alterations in the residues interacting with it in EspG<sub>5</sub>, and in fact all of the residues are in mostly the same position and orientation with only minor differences (Figure 3.3A). A124<sup>PPE41</sup> is mutated to L124<sup>PPE41</sup> in the 6VJ5 model, and we expected this to cause alterations in the EspG<sub>5</sub> residues. L216<sup>EspG5</sup> and Q187<sup>EspG5</sup> are mostly unaffected by the mutation, with Q187<sup>EspG5</sup> having some slight differences, largely due to the difference in the backbone of EspG<sub>5</sub> differing between the two models in this position (Figure 3.3B). F191<sup>EspG5</sup> is most altered as it is shifted about 2.3 Å away from the mutated L124<sup>PPE41</sup> (Figure 3.3B). This makes sense as the mutation from Ala to Leu introduces a longer amino acid and F191<sup>EspG5</sup> is interacting with the wild-type Ala residue in the space the mutated Leu takes up. We hypothesize that this shift in F191<sup>EspG5</sup> is the cause of the perturbation in the helical bundle of EspG<sub>5</sub>.

### 3.2.3. Limited tolerance of bulkier hh motif residues in PPE41-EspG<sub>5</sub> interface.

To support our hypothesis that EspG<sub>5</sub> is dynamic and adjusts to accommodate the bulkier hh motifs that can be found in ESX-5-specific PPE proteins we investigated the packing and sequence tolerance of the hh motif in the context of available ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimers. Currently there are four structural models: two models of WT PE25-PPE41-EspG<sub>5</sub> heterotrimers (4KXR (205)) and 4W4L (204)), our new model of a heterotrimer with a mutated PPE41<sup>A124L</sup> (6VJ5), and a PE8-PPE15-EspG<sub>5</sub> heterotrimer (5XFS (206)). We took all of the available models and ran them in the MolProbity webserver to add hydrogens at electron-cloud positions (252) to fully analyze packing and interactions in the PPE-EspG<sub>5</sub> interface. We first visually inspected the interactions between each PPE protein's hh motif with EspG<sub>5</sub>. Each interface appears to be well packed with no “extra space” available to accommodate substitutions of bulkier amino acids at either hh motif position (Figure 3.3A-B). We also analyzed the interface using Voronoia (260) to quantify the packing of the hh motif of each model. The hh motifs of the PPE proteins are either AL (4KXR and 4W4L), LL (6VJ5), or VL (5XFS). We first examined the first position of the hh motif. PPE41 is unique to all other ESX-5-specific PPE proteins in that it has an Ala in the first hh motif. PPE15 contains a Val which is only found in 20%

of the ESX-5-specific PPE proteins (Figure 3.1B.). Our mutated PPE41<sup>A124L</sup> contains Leu, which is the most common of first position among ESX-5-specific PPE proteins, with 37% containing it (Figure 3.1B.). Overall, the two Ala containing models, 4KXR and 4W4L, have their first positions slightly “underpacked” having packing densities of 0.59 and 0.62, respectively (Table 3.3). The bulkier substitutions of Val (5FXS) or Leu (6VJ5) increasing the residue’s overall packing to 0.71 and 0.73, respectively, which is mostly attributed to the packing density of the gamma or delta carbons being 1.0, respectively (Table 3.3). About 24% of the ESX-5-specific PPE proteins contain the bulkier Phe or Trp at this first position (Figure 3.1B.), and either of these residues fitting within this interface without any kind of rearrangement is unlikely. We also analyzed the second position of the hh motif, which all four models contain a Leu at this position. ESX-5-specific PPE proteins contain either a Leu or Phe at this position, with Leu appearing in 41% of PPE proteins (Figure 3.1C.). Interestingly all four models pack similarly at this position with residue packing densities of 0.81, 0.81, 0.78, and 0.80 for 4KXR, 4W4L, 5XFS, and 6VJ5 respectively. Again, the last carbons, the delta carbons, are completely packed with packing densities of 1.0 for all four models. Thus, it is hard to imagine a Phe substitution taking place at this position and fitting in the observed interfaces without any rearrangement from either the PPE proteins or EspG<sub>5</sub>.

We also utilized the `sequence_tolerance` protocol on the ROSIE web server (262-264) to determine if substitutions in the hh motif could be tolerated in any of the known PE-PPE-EspG<sub>5</sub> heterotrimer structures. For the first position in the hh motif in PPE41 (residue 124), Ala in the WT 4KXR and 4W4L and Leu in the mutant 6VJ5, there is some tolerance as most of the standard amino acids are hits in the `sequence_tolerance` output (Figure 4.4A-C). However, both WT heterotrimers only had Glu account for at least 50% of possible outcomes (Figure 3.4A-B). This is similar to the mutant 6VJS, except both Glu and Leu make up about 80% of the possible outcomes (Figure 3.4B). Interestingly none of the other amino acids found in ESX-5 hh motifs (Figure 3.1B) are represented well in the outcomes (Figure 3.4A-C). The first hh motif position, residue 125, for PPE8 in the 5XFS heterotrimer is more tolerate with about 20% each of outcomes being Ile, Leu, and Trp (Figure 3.4D). The second position of the hh motif, Leu in all heterotrimers and residue 124 in 4KXR, 4W4L, and 6VJ5, and residue 125 in 5XFS, only accepts a Leu as nearly

every sequence tolerance outcome contains this residue (Figure 3.4). Over half of the expected ESX-5-specific PPE proteins contain a Phe in this second hh motif position (Figure 3.1C.), and therefore structural rearrangement beyond what is allowed within the sequence\_tolerance protocol must occur for EspG<sub>5</sub> to bind these PPE proteins.

### 3.3. Discussion.

In this work we present a structure of a mutated PPE41<sup>A124L</sup> in the context of the PE25-PPE41-EspG<sub>5</sub> heterotrimer. We opted to use the defined PE25-PPE41 heterodimer to make mutations within the hh motif and assess the impacts they had on the PPE-EspG<sub>5</sub> interface because of the difficulty of assigning PE-PPE heterodimers. We also attempted to solve the structures of numerous hh motif mutants, including the motifs LF, FF, and FF but we were only successful in solving the structure of the LL mutant. Overall, the mutant heterotrimer has the same overall fold and shape as the previously solved WT PE25-PPE41-EspG<sub>5</sub> heterotrimers (204,205) and the PE5-PPE18-EspG<sub>5</sub> heterotrimer (206). This suggests that that changes within the hh motif do not alter the binding mode of ESX-5-specific PPE proteins to the cognate chaperone EspG<sub>5</sub>. The hh motif is of interest because it is hypothesized to be a large contributor to the binding energy of the PPE-EspG<sub>5</sub> interface and be a driving factor to aggregation that causes PE-PPE heterodimers to be insoluble in the absence of EspG. The mutations we tested here made the PE25-PPE41 heterodimer insoluble in the absence of EspG<sub>5</sub> during protein purification, and ultimately prevented the ability to analyze the binding affinity changes that occur with substitutions in the hh motif. The structure presented here is the fourth ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimer structure, two WT PE25-PPE41-EspG<sub>5</sub> (204,205), one PE5-PPE18-EspG<sub>5</sub> (206), and the mutated PE25-PPE41<sup>A124L</sup>-EspG<sub>5</sub>. All of these structures have the same basic shape with PE-PPE forming an elongated helical bundle with one being solely made up of a helix-turn-helix from PPE that is bound by EspG<sub>5</sub>, suggesting that this is a conserved binding mode and all ESX-5-specific PPE proteins bind EspG<sub>5</sub> in the same manner.

The structure presented here reveals EspG<sub>5</sub> has some structural plasticity to accommodate the bulkier LL motif present in the PPE41 mutant. The C-terminal helical bundle of EspG<sub>5</sub> shifts to allow the bulkier LL motif to fit within the PPE-EspG<sub>5</sub> interface. Most of the perturbances of the EspG<sub>5</sub> structure are minor, with the exception of F191<sup>EspG5</sup> (Figure 3.3B.). Because Leu contains an additional alkyl group compared to Ala, it makes

sense that the changes are localized on the EspG<sub>5</sub>-side of the interface. Bulkier Phe and Trp make up a smaller percentage of residues in the first position, it poses the question if the changes within EspG<sub>5</sub> are broader to accommodate the added bulk.

Ekiert and Cox hypothesized that the PPE-EspG<sub>5</sub> interface was underpacked in their PE25-PPE41-EspG<sub>5</sub> structure to allow for substitutions of bulkier amino acids found within ESX-5-specific PPE proteins (204). Our analysis of the structure through the Voronoia software (260) and using the `sequence_tolerance` protocol on ROSIE (262-264) does not support this. The Voronoia analysis showed that all structures we analyzed; the two WT PE25-PPE41-EspG<sub>5</sub> heterotrimers, the PE5-PPE18-EspG<sub>5</sub> heterotrimer, and the newly present mutant PE25-PPE41<sup>A124L</sup>-EspG<sub>5</sub>; all have similar packing levels within the hh motifs. The available space for the bulkier Phe or Trp substitutions to reside did not seem plausible without some level of rearrangement from EspG<sub>5</sub>. The `sequence_tolerance` results show some tolerance available for the first position of the hh motif, however the residues found in ESX-5-specific PPE proteins; Ale, Leu, Val, Phe, and Trp; were not enriched. This first position allowing other residues to fit makes some sense because this location is close to the surface of the interface. The second hh motif position showed almost zero tolerance for anything other than Leu in all four models. Therefore, for the bulkier Phe amino acid, which is present in over half of ESX-5-specific PPE proteins, to fit in the PPE-EspG<sub>5</sub> interface some broader structural rearrangement which is beyond what is allowed by the `sequence_tolerance` protocol must occur. The second hh motif position is buried deeper within the PPE-EspG<sub>5</sub> interface.

In conclusion, we presented a mutant PE25-PPE41<sup>A124L</sup>-EspG<sub>5</sub> heterotrimer structure. This structure revealed that bulkier substitutions within the hh motif of ESX-5-specific PPE proteins cause structural changes within the PPE-EspG<sub>5</sub> interface. This suggests that EspG<sub>5</sub> has structural plasticity and can accommodate the various hh motifs found in ESX-5-specific PPE proteins.

**Table 3.1. Data collection and refinement statistics of ESX-5 mutant heterotrimer.**

	<b>PE25<sub>mt</sub>-PPE41<sub>mt</sub><sup>LL</sup>-EspG<sub>3mm</sub> (PDB ID 6VJ5)</b>
<b>Data Collection</b>	
Wavelength (Å)	0.9791
Space group	P6 <sub>1</sub> 22
Cell Dimensions:	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	139.02, 139.02, 170.57
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 120
Resolution (Å)	45.5 – 2.4 (2.46 – 2.4) <sup>a</sup>
R <sub>sym</sub>	0.154 (0.964)
R <sub>pim</sub>	0.067 (0.674)
CC <sub>1/2</sub> <sup>b</sup>	97.5 (73.3)
<i>I</i> / <i>σ</i>	6.73 (2.01)
Completeness (%)	99.7 (99.4)
Multiplicity	4.8 (4.9)
<b>Refinement</b>	
Resolution (Å)	45.5 – 2.4
No. reflections (total/free)	38512/1981
R <sub>work</sub> /R <sub>free</sub>	21.00/25.38
Number of atoms:	
Protein	4230
Ligand/ion	2
Water	184
<i>B</i> -factors:	
Protein	55.08
Water	51.78
All atoms	58.72
Wilson <i>B</i>	48.83
R.m.s. deviations:	
Bond lengths (Å)	0.004
Bond angles (°)	0.59
Ramachandran distribution <sup>c</sup> (%)	
Favored	97.58
Allowed	2.23
Outliers	0.19
Rotamer outliers <sup>c</sup> (%)	0.00
Clashscore <sup>d</sup>	2.15
MolProbity Score <sup>e</sup>	1.07

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

<sup>b</sup> CC<sub>1/2</sub> correlation coefficient is defined in (269) and was calculated with XSCALE (248).

<sup>c</sup> Calculated with the MolProbity server (<http://molprobity.biochem.duke.edu>) (252).

<sup>d</sup> Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

<sup>e</sup> MolProbity Score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution (252).

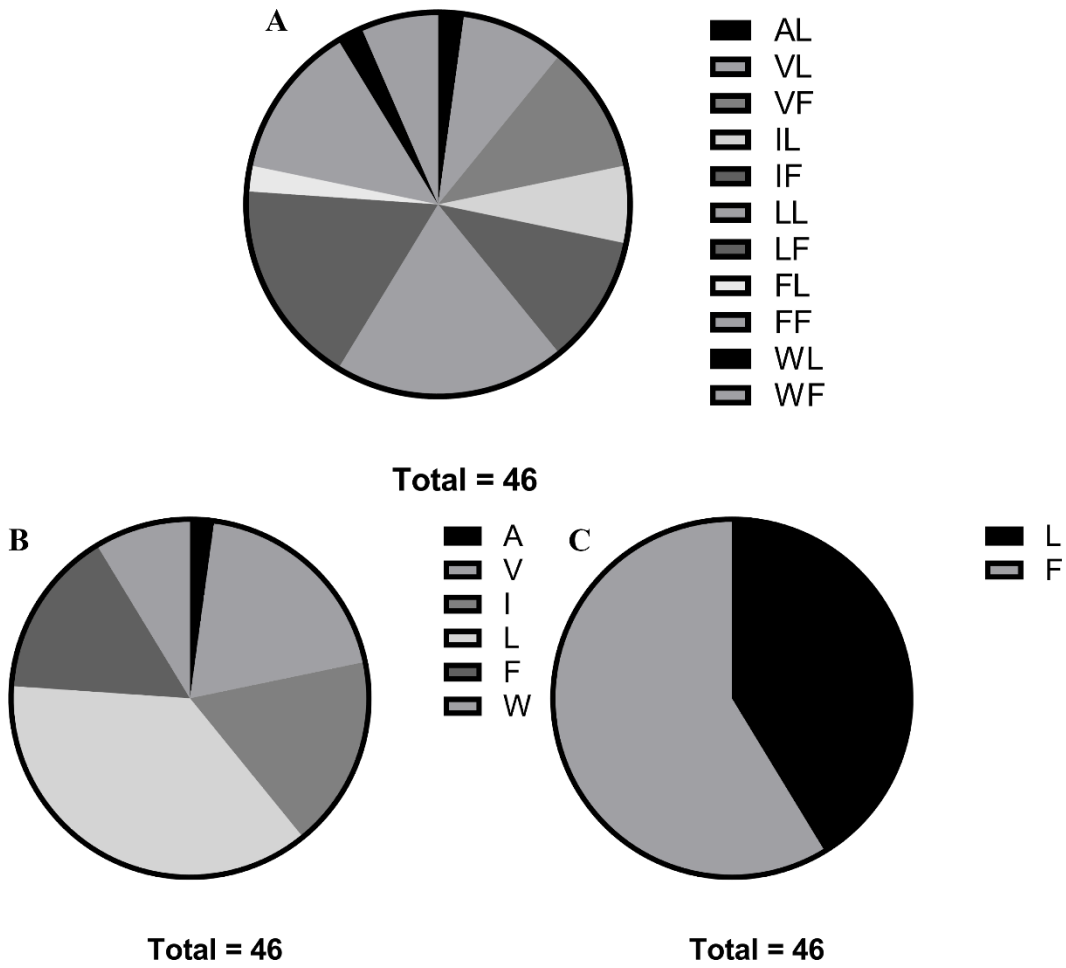


**Table 3.2. Structural deviations between PPE41<sup>LL</sup> mutant heterotrimer and previously published ESX-5 heterotrimers.**

		4KXR	4W4L	5XFS
6VJ5	PE25	0.3	0.3	1.1
	PPE41 <sup>A124L</sup>	0.3	0.3	2.2
	EspG <sub>5</sub>	0.7	0.7	0.6

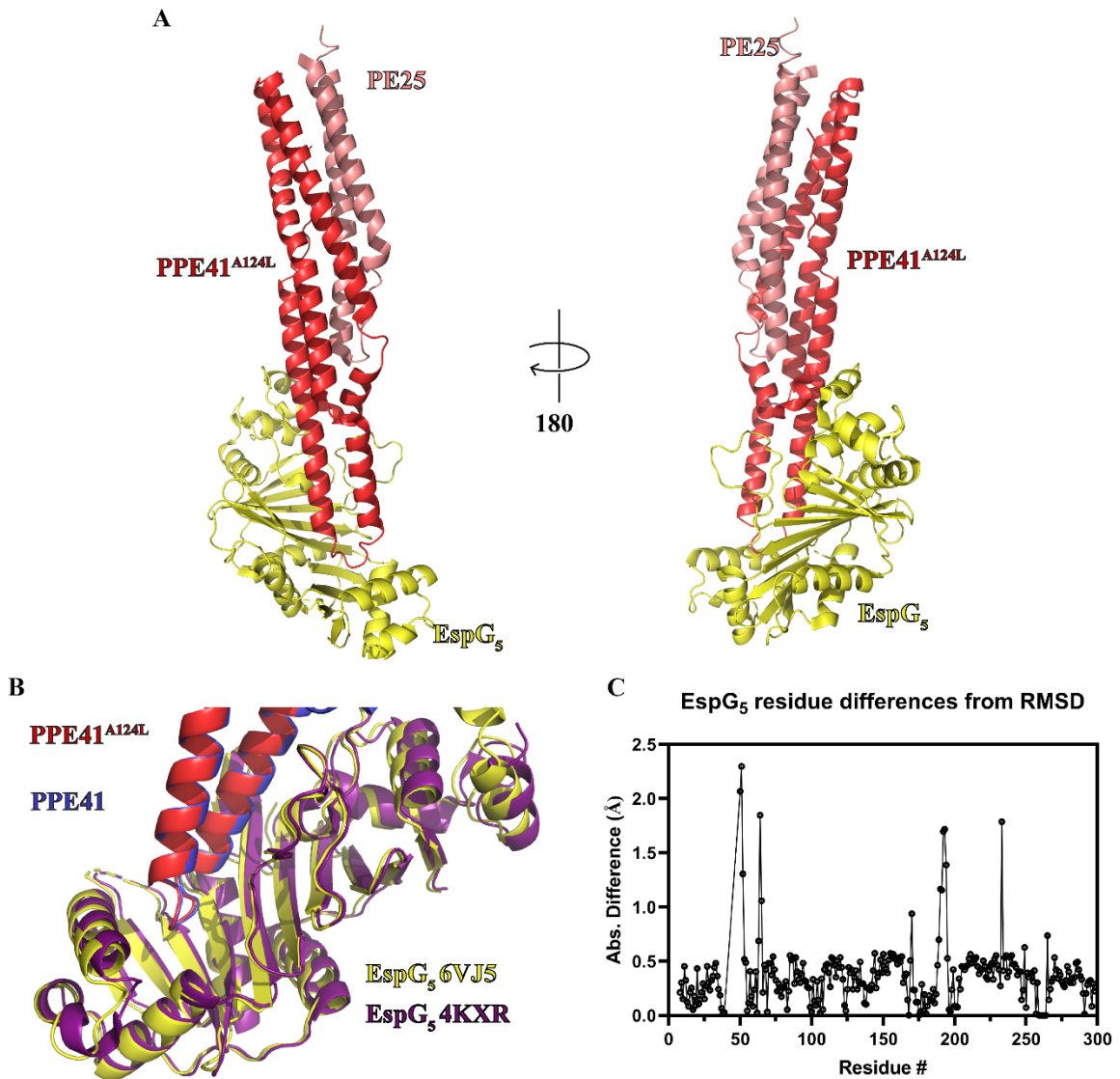
**Table 3.3. Packing density of hh motif residues.**

	4KXR	4W4L	5XFS	6VJ5
h1 residue	Ala	Ala	Val	Leu
h1 PD	0.59	0.62	0.71	0.73
$\beta$ carbon PD	0.491	0.513		
$\gamma$ carbon PD			1/1	
$\delta$ carbon PD				1/1
h2 residue	Leu	Leu	Leu	Leu
h2 PD	0.81	0.81	0.78	0.80
$\delta$ carbon PD	1/1	1/1	1/1	1/1



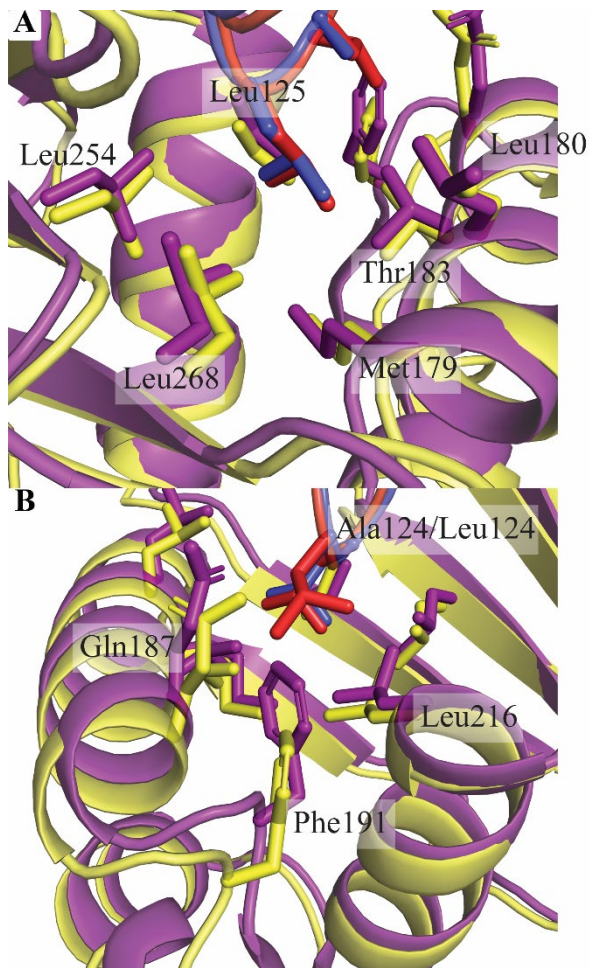
**Figure 3.1. Variety in hh motifs in ESX-5-specific PPE proteins.**

*A*, The representation of each of the eleven unique hh motifs found in ESX-5-specific PPE proteins. *B*, The variety in the first position of the hh motif among the ESX-5-specific PPE proteins. *C*, The variety in the second position of the hh motif among the ESX-5-specific PPE proteins.



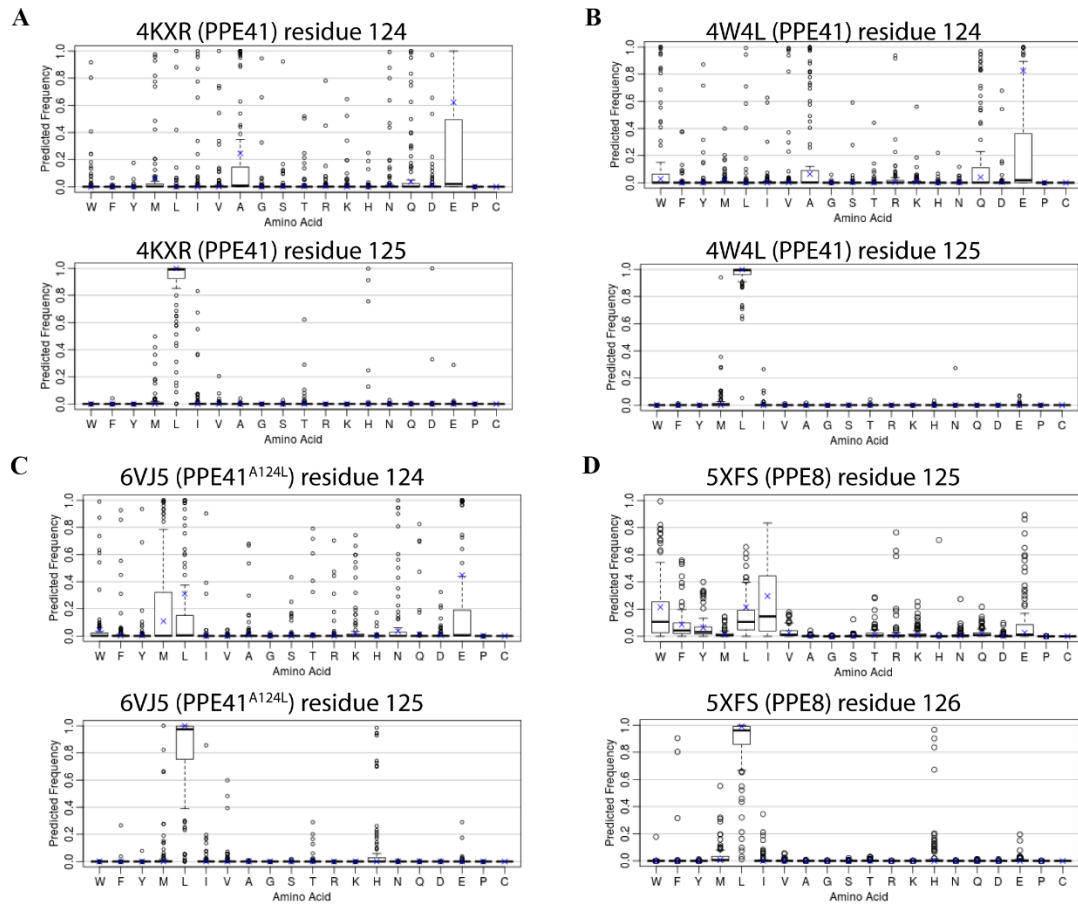
**Figure 3.2. Structure of PE25<sub>mt</sub>-PPE41<sup>LL</sup><sub>mt</sub>-EspG<sub>5mm</sub> and comparison to wild-type heterotrimer.**

*A*, Cartoon representation of the PE25-PPE41<sup>A124L</sup>-EspG<sub>5mm</sub> in two views related by a rotation of 180°. *B*, Comparison of 4KXR and 6VJ5 trimer PPE-based alignment structures highlighting differences in C-terminal helical bundle. *C*, rmsd deviations in angstrom between equivalent EspG<sub>5</sub> residues when aligning EspG<sub>5</sub>.



**Figure 3.3. Interactions between the hh motif of PPE proteins and EspG5.**

View of hh motif on PPE41-based alignment of 6VJ5 and 4KXR. *A*, Leu125 of PPE41 (4XKR in blue and 6VJ5 in red) interacts with Met179, Leu180, Thr183, Leu254, and Leu268 of EspG5. There are minimal structure changes in the EspG5 residues between the 4XKR (purple) and 6VJ5 (yellow). *B*, Ala124 (4XKR, blue) and Leu124 (6VJ5, red) from PPE41 interacts with Gln187, Phe191, and Leu216 of EspG5. The mutation of Ala124Leu in 6VJ5 is likely the cause of the 2.3 Å shift in Phe191 this causes the perturbation in the alpha helix that Gln187 is in.



**Figure 3.4. Sequence tolerance of hh motifs in selected ESX-5 PPE proteins.**

Boxplot representation of the distributions of tolerated amino acid substitutions in the hh motifs of *A*, 4KXR; *B*, 4W4L; *C*, 6VJ5; and *D*, 5XFS as calculated within the ROSIE server (262).

## **Chapter 4. PE5–PPE4–EspG<sub>3</sub> heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems.**

### 4.1. Introduction.

This chapter was originally published in the Journal of Biological Chemistry. Zachary A. Williamson, Catherine T. Chaton, William A. Ciocca, Natalia Korotkova and Konstantin V. Korotkov. PE5–PPE4–EspG<sub>3</sub> heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. J Biol Chem. 2020; 295:12706-12715. © Williamson, *et al.* The introduction of the original publication has been modified for this chapter.

Each ESX system secretes a unique subset of PE–PPE heterodimers, and therefore each encodes an EspG that binds to only its corresponding heterodimers (203,205). The first structural insight into the EspG and PE–PPE interaction was revealed by analysis of the structure of the PE25–PPE41–EspG<sub>5</sub> complex, a heterotrimer from ESX-5 (204,205). EspG<sub>5</sub> interacts solely with PPE41 at the tip distal to the PE25 interaction and aids in preventing PE–PPE heterodimer aggregation in part by shielding a conserved hydrophobic tip on the PPE proteins, known as the hh motif (205). The additional structure of the ESX-5–related, PE8–PPE15–EspG<sub>5</sub> heterotrimer, revealed similar interactions of the substrate PE–PPE dimer with the EspG<sub>5</sub> chaperone (206). Despite high conservation among PPE proteins in the identified EspG<sub>5</sub>-binding region from PPE41, three residues vary depending on whether the PPE protein is secreted by ESX-1, ESX-3, or ESX-5 (205). Alteration of any or all of these positions in the ESX-5–dependent PPE41 did not disrupt PPE41-EspG<sub>5</sub> binding (205). Based on this observation it has been suggested that structural elements outside of the EspG-binding region differentiate the ESX-5–specific PPE proteins from their ESX-1 and ESX-3 homologs to bind EspG<sub>5</sub> (205).

The work in this chapter was initiated to understand the how each EspG from the different ESX systems specifically recognizes its unique subset of cognate PE–PPE heterodimers. Here we present the structure of PE5–PPE4–EspG<sub>3</sub> from ESX-3. This structure reveals a novel binding mode of PE–PPE proteins with the EspG chaperone and suggests the molecular mechanism by which the PE–PPE dimers are specifically targeted by cognate chaperones.

## 4.2. Results.

### 4.2.1. *EspG<sub>3</sub>* forms a complex with PE5–PPE4, and binding is conserved across species.

To understand the mechanism for the specificity of PE–PPE recognition by cognate chaperones, a high-resolution structure of a heterotrimer produced by the ESX systems, other than ESX-5, was needed. Our efforts have focused on optimizing the ESX-3 PE–PPE–EspG heterotrimer for X-ray structural studies. Constructs of full-length PE5 (Rv0285), the conserved N-terminal PPE domain of PPE4 (Rv0286, residues 1–181), in a complex with the cognate full-length EspG<sub>3</sub> (Rv0289) from *Mycobacterium tuberculosis* (Figure 4.1A) never formed high-resolution diffraction quality crystals, despite our best efforts. The difficulty could be due to some heterogeneity in the processing of EspG<sub>3mt</sub> within the *Escherichia coli* cell, as seen by the double band in Figure 4.1B and Figure 4.2A. Numerous variations of PE5–PPE4–EspG<sub>3</sub> constructs were screened utilizing multiple mycobacterial species, different fusion approaches, and even mixing PE5–PPE4 dimers with EspG<sub>3</sub> chaperones from different species (Table 4.1). This latter approach was inspired by the work done on the *Plasmodium* aldolase–thrombospondin–related anonymous protein complex (270) and in the end, produced the best crystals for further diffraction experiments. To ensure that the mixed heterotrimers behaved the same in solution as the WT heterotrimer, a size-exclusion chromatography with multiangle light scattering (MALS) experiment was performed on both the WT PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mt</sub> heterotrimer and the mixed PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mm</sub> heterotrimer that contained the *Mycobacterium marinum* EspG<sub>3</sub> (MMAR\_0548) with 78% sequence identity to EspG<sub>3mt</sub> (Figure 4.1B–C). Both heterotrimers form a 1:1:1 complex with experimental molecular masses of 56.2 kDa (Figure 1B) for the full *M. tuberculosis* heterotrimer (theoretical heterotrimer molecular mass of 58.8 kDa) and 54.6 kDa (Figure. 1C) for the mixed heterotrimer with the *M. marinum* EspG<sub>3</sub> (theoretical heterotrimer molecular mass of 58.1 kDa). Co-purification assays were run with both *M. tuberculosis* and *M. marinum* EspG<sub>3</sub> with the *M. tuberculosis* PE4–PPE5, along with EspG<sub>3s</sub> from *Mycobacterium smegmatis* (MSMEG\_0622), *Mycobacterium hassiacum* (MHAS\_04631), and *Mycobacterium kansasii* (MKAN\_17015). Because of the His<sub>6</sub> tag only being present on PE5<sub>mt</sub>, EspG<sub>3</sub> copurification required interaction with the PE5<sub>mt</sub>–PPE4<sub>mt</sub> heterodimer. Across all species that were tested, EspG<sub>3</sub> co-purified with the PE5<sub>mt</sub>–PPE4<sub>mt</sub> heterodimer



(Figure 4.2A–E). The binding of different EspG<sub>3S</sub> to the same PE–PPE heterodimer suggests a common protein–protein recognition mechanism within the ESX-3 family.

#### 4.2.2. Overall structure of PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub>–EspG<sub>3<sub>mm</sub></sub>.

The PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub>–EspG<sub>3<sub>mm</sub></sub> heterotrimer was able to form diffraction quality crystals, and two different crystal forms were observed that diffracted to 3.3 Å (I422) and 3.0 Å (P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>) (Table 4.2). The final refinement and data statistics are shown in Table 4.2. Overall, there is little structural variation between the individual proteins across the copies present in the two crystal forms (Table 4.3).

For all structural analysis and comparisons, the first copy of the PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub>–EspG<sub>3<sub>mm</sub></sub> heterotrimer from the higher resolution P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form was used because it diffracted at a higher resolution and has the lowest *B*-factors from the noncrystallographic copies in the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> form. EspG<sub>3<sub>mm</sub></sub> interacts solely with the tip of PPE<sub>4<sub>mt</sub></sub> (Figure 4.3), similar to EspG<sub>5</sub> in the previously solved ESX-5 heterotrimers (204–206). However, the orientation of PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub> relative to EspG<sub>3<sub>mm</sub></sub> is dramatically different from what was observed for either ESX-5 heterotrimer, and the differences between them will be described in later sections. The YXXX(D/E) motif for ESX secretion of PE<sub>5<sub>mt</sub></sub> is accessible for interactions with the rest of the ESX machinery because it is located distal to the EspG<sub>3<sub>mm</sub></sub> interaction (203). In both crystal forms, this secretion motif is disordered, similar to the motif in PE<sub>8<sub>mt</sub></sub> from the PE<sub>8<sub>mt</sub></sub>–PPE<sub>15<sub>mt</sub></sub>–EspG<sub>5<sub>mt</sub></sub> heterotrimer (206). The individual components of the PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub>–EspG<sub>3<sub>mm</sub></sub> heterotrimer align well to the individual components of the previously reported ESX-5 heterotrimers, both PE<sub>25<sub>mt</sub></sub>–PPE<sub>41<sub>mt</sub></sub>–EspG<sub>5<sub>mt</sub></sub> (4KXR and 4W4L) and PE<sub>8<sub>mt</sub></sub>–PPE<sub>15<sub>mt</sub></sub>–EspG<sub>5<sub>mt</sub></sub> (5XFS), with only moderate variations (Table 4.4).

In a previous study on EspG structures (202), a small-angle Xray scattering (SAXS) experiment was done on the PE<sub>5</sub>–PPE<sub>4</sub>–EspG<sub>3</sub> heterotrimer from *M. smegmatis*. Comparisons between this SAXS analysis and our crystal structure were performed to see whether the solution-based characterization of the heterotrimer matched the X-ray-based characterization. We ran CRY SOL (253) on our crystal structure compared with the experimental scattering data from the *M. smegmatis* heterotrimer. The overall  $\chi^2$  is 2.53, which is acceptable given that the heterotrimers are from different species with only 54.0–73.8% sequence identity across the different components (Figure 4.4). The main

differences are in the extreme high- and low-resolution areas, likely arising from differences in the primary structure between the two samples and from aggregation in the SAXS sample, respectively. Therefore, we are confident that the crystal structure is an appropriate model of the ESX-3 heterotrimer because it exists in solution.

#### 4.2.3. Interface between PPE4<sub>mt</sub> and EspG3<sub>mm</sub>.

The interface between EspG3<sub>mm</sub> and PPE4<sub>mt</sub> contains numerous hydrophobic interactions, multiple hydrogen bonds, and two salt bridges centered around Glu140 of PPE4<sub>mt</sub> (Figure 4.3B–F). Overall, the interface buries 3,121 Å<sup>2</sup> of solvent-accessible surface area, as calculated by the PISA server (265), and has the shape correlation S<sub>c</sub> value of 0.664 (271). The interface is comprised of 30 total residues from PPE4<sub>mt</sub> and 49 residues from EspG3<sub>mm</sub> (Figure 4.5). The tip of PPE4<sub>mt</sub> containing the ends of α4 and α5, and the loop between them is inserted into a groove on EspG<sub>3</sub> composed of its central β sheet and C-terminal helical bundle. This bundle shields the hydrophobic tip of PPE4<sub>mt</sub>, including the hh motif of Phe128-Phe129, from solvent access. The tip of PPE4<sub>mt</sub> is interacting with EspG<sub>3</sub> in such a way that the complex is unlikely to disengage at the ESX secretion machinery without structural rearrangement of the chaperone.

#### 4.2.4. Mutations cause disruptions in the PPE4–EspG<sub>3</sub> interface.

To probe the interface of the crystal structure and test the importance of interacting residues, we made several mutations on both PPE4 and EspG<sub>3</sub> sides of the interface and opted to use the cognate PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG3<sub>mt</sub> heterotrimer to test our mutations. The PISA output (265) of the interface was analyzed along with sequence alignments of the current known ESX-3 PPE proteins (Figure 4.6) and alignments of the EspG<sub>3</sub> used in this study (4.7) to select which residues in the interface would be mutated. PPE4<sub>mt</sub> is well-conserved along the interface among ESX-3-specific PPE proteins (4.6), and we targeted strictly conserved residues in the interface. We selected N127 and N132 because they contain buried hydrogen bonds, F128 and F129 because they are the hh motif and contribute a large amount of solvation energy to the interface according to PISA (265), and E140 because it is part of the salt bridges in the interface. We ran co-purification pulldown assays with mutated PPE4<sub>mt</sub> and EspG3<sub>mt</sub> (Table 4.5). As described earlier, EspG3<sub>mt</sub> is only co-purified with the PE5–PPE4 heterodimer if it forms a complex. The introduction of charges into the buried hydrogen bonds with N127D and N132E was unable to break the

PPE4<sub>mt</sub>-EspG<sub>3mt</sub> interaction, and neither was the charge reversal of E140R, because all three mutations co-purify with EspG<sub>3mt</sub> (Figure 4.8A). This suggests that disruption of any of these single positions is not sufficient to abolish PPE4<sub>mt</sub>-EspG<sub>3mt</sub> interaction. Conversely, the introduction of charged residues into the hh motif with F128R or F129E did disrupt the interface and prevented EspG<sub>3mt</sub> from being co-purified (Figure 4.8A), because it interrupts with the hydrophobic environment deep within the EspG<sub>3mt</sub>-binding pocket. The interface of EspG<sub>3mt</sub> is also well-conserved among the various EspG<sub>3</sub>s tested in this study (Figure 4.7), and again, we targeted strictly conserved residues. We selected R208 and E212 because they contain buried hydrogen bonds, R87 and R102 because they form the salt bridge within the interface, and S231 because it sits at the top of the groove of EspG<sub>3</sub> and could sterically block entrance into the pocket. Neither single mutation of the salt bridge, R87E or R102E, was able to prevent co-purification of EspG<sub>3mt</sub> (Figure 4.8B). Also, the introduction of a charged residue with R208E was unable to prevent the interaction (Figure 4.8B). In contrast, E212R was sufficient to prevent co-purification, as well as S231Y (Figure 4.8B), because both prevent the hydrophobic tip of PPE4<sub>mt</sub> from interacting with the binding pocket of EspG<sub>3mt</sub> either by charge repulsion or steric hindrance. Thus, our mutations on both PPE4<sub>mt</sub> and EspG<sub>3mt</sub> highlight the importance of the hydrophobic environment deep within the PPE4<sub>mt</sub>-EspG<sub>3mt</sub> interface.

#### 4.2.5. Structure of EspG<sub>3</sub> in and out of heterotrimer complex.

Our structure is the first of EspG<sub>3</sub> solved in complex with a cognate PE-PPE dimer, and thus we wanted to compare it with the previously solved unbound EspG<sub>3</sub> structures. In total, there are six available EspG<sub>3</sub> structures, four of EspG<sub>3ms</sub> (PDB codes 4L4W, 4RCL, 5SXL, and 4W4J (202,204)), one EspG<sub>3mt</sub> (PDB code 4W4I (204)), and one EspG<sub>3mm</sub> (PDB code 5DLB (202)). These six structures can be classified into two different forms, an “open” form and a “closed” form. The differentiation between these two forms is the orientation of the C-terminal helical bundle relative to the core  $\beta$ -sheet. The EspG<sub>3mm</sub> structure (PDB code 5DLB) is representative of the open form, and one of the EspG<sub>3ms</sub> structures (PDB code 4RCL) is representative of the closed form. Analysis of EspG<sub>3mm</sub> as it exists in the PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> heterotrimer was done relative to these two representative structures. The overall alignment of the representative structures to the bound EspG<sub>3mm</sub> was good with RMSDs of 2.1 and 1.9 Å for the open and closed forms,

respectively (Figure 4.9A). Inspection of these alignments show the majority of differences to be within the arrangement of the C-terminal helical bundles, with the bound form of EspG<sub>3mm</sub> being in close to the orientation found in the closed form (Figure 4.9B-C). The bound EspG<sub>3mm</sub> cannot be any closer to the closed form orientation because the C-terminal helical bundle makes contact with PPE4<sub>mt</sub>. We hypothesized that this C-terminal helical bundle is dynamic and closes on cognate PPE proteins upon interaction. A comparison between the bound EspG<sub>3mm</sub> structure and the open EspG<sub>3mm</sub> was performed with the DynDom server to test this hypothesis (272). DynDom identified a moving domain within the structures that was located in the C-terminal helical bundle (Figure 4.9D). DynDom's analysis also performed a whole structure alignment that agreed with the previous Dali alignment in Figure 4.9A-B. DynDom performed alignments between the fixed domains (residues 11–168 and 189–279) and the moving domains (residues 168–188), which resulted in much better alignments with RMSDs of 1.76 and 0.86 Å, respectively. Therefore, the moving domain, the C-terminal helical bundle, is essentially structurally identical between PPE4<sub>mt</sub>-bound EspG<sub>3mm</sub> and the open EspG<sub>3mm</sub> and its rotation of 30.2° and translation of 0.8 Å is moderately perturbing the fixed domain. Because the moving domain making extensive contact with PPE4<sub>mt</sub> and PPE4<sub>mt</sub> would sterically clash with the current orientation of the C-terminal helical bundle, the movement from the closed to the open orientation could be significant in releasing the secreted PE–PPE dimers from the chaperone at the secretion machinery.

#### 4.2.6. Comparison of ESX-3 and ESX-5 PE–PPE–EspG heterotrimers.

A vastly different binding mode is observed when comparing the ESX-3-specific PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mm</sub> heterotrimer to the previously published ESX-5-specific heterotrimers. As mentioned earlier, there is good agreement when comparing individual components of the ESX-3-specific heterotrimer to the available ESX-5-specific heterotrimers (Table 4.4). The difference between the two sets of heterotrimers became apparent when they were aligned via EspG (Figure 4.10A-B and Figure 4.11) (259). Our results focused on comparisons with the PE25<sub>mt</sub>–PPE41<sub>mt</sub>–EspG<sub>5mt</sub> (PDB code 4KXR) heterotrimer, but the same differences were present with the PE8<sub>mt</sub>–PPE15<sub>mt</sub>–EspG<sub>5mt</sub> (PDB code 5XFS) heterotrimer. The interaction angle of the different PE–PPE heterodimer with EspG is drastically different between the two heterotrimers, with a 30° angle

difference (Figure 4.9B). Another difference lies within the hh motif loops of PPE25<sub>mt</sub> ( $\alpha$ 4- $\alpha$ 5 loop) and PPE4<sub>mt</sub> ( $\alpha$ 5- $\alpha$ 6 loop) (Figure 4.10C). In PPE25<sub>mt</sub>, this loop is seven residues long and undertakes a compact conformation that is not altered during EspG<sub>5mt</sub> binding (205). In contrast, in PPE4<sub>mt</sub>, this loop is nine residues long and has an extended conformation. This difference was rapidly apparent when PPE25<sub>mt</sub> and PPE4<sub>mt</sub> were aligned (Figure 4.10C).

This loop conformation also made each PPE protein incompatible with the other's binding mode. When looking at the PPE alignment in the context of the ESX-3 heterotrimer, the  $\alpha$ 4- $\alpha$ 5 loop of PPE25<sub>mt</sub> does not align over the central groove of EspG<sub>3mm</sub> and instead sterically clashes the central  $\beta$  sheet of the chaperone (Figure 4.10D). The tip of PPE41<sub>mt</sub> would have to undergo a drastically new tip confirmation to bind in the opening of EspG<sub>3mm</sub>. In the context of the ESX-5 heterotrimer, the  $\alpha$ 5- $\alpha$ 6 loop does not align with the central groove of the chaperone, and instead, PPE4<sub>mt</sub>'s hh motif sterically clashes with the C-terminal helical bundle of EspG<sub>5mt</sub> (Figure 4.10E). Also, none of the salt bridges between PPE41<sub>mt</sub> and EspG<sub>5mt</sub> are conserved in PPE4<sub>mt</sub>. Specifically, D134<sup>PPE41</sup>-K235<sup>EspG5mt</sup>, D140<sup>PPE4mt</sup>-R109<sup>EspG5mt</sup>, and D144<sup>PPE4mt</sup>-R27<sup>EspG5mt</sup>, that are all replaced with hydrophobic residues in PPE4<sub>mt</sub>: either T137<sup>PPE4mt</sup> or L138<sup>PPE4mt</sup>, V144<sup>PPE4mt</sup>, and L147<sup>PPE4mt</sup>, respectively.

### 4.3. Discussion.

In this work, we present the first structure of the PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> heterotrimer, which is from the ESX-3 system. Our structure is a mixed heterotrimer, and we presented evidence that EspG<sub>3</sub> from numerous mycobacterial species can bind the PE5<sub>mt</sub>-PPE4<sub>mt</sub> heterodimer. Conservation of the EspG<sub>3s</sub> used in this study ranged from 57 to 83% identity, yet an enrichment in conservation is observed within PPE4-interacting residues (Figure 4.9). The ability of EspG<sub>3</sub> from numerous mycobacterial species to bind PE5<sub>mt</sub>-PPE4<sub>mt</sub> suggests that the recognition mechanism is conserved within ESX systems across species. Overall, the PE5<sub>mt</sub>-PPE4<sub>mt</sub> interaction is similar to the previously reported PE-PPE-EspG heterotrimers (204-206) in that PPE4<sub>mt</sub>'s tip is solely interacting with EspG<sub>3mm</sub> and the general secretion motif of YXXX(D/E), on PE5<sub>mt</sub>, is at the distal end of the PE5<sub>mt</sub>-PPE4<sub>mt</sub> heterodimer. In all copies of PE5<sub>mt</sub>, this motif is unstructured because it is in the PE8-PPE15-EspG<sub>5</sub> heterotrimer (206), and similarly, W63<sup>PPE4mt</sup> is pointed away

from this secretion motif. This arrangement is distinct from the PE25–PPE41–EspG5 heterotrimers (204,205) and EspB, an ESX-1 substrate that has a similar structural fold to the PE–PPE heterodimers (145,146). PE8<sub>mt</sub> contains an expanded C-terminal domain, and because the secretion motif is located in the linker between the C-terminal domain and the PE domain, the orientation of the secretion motif was unclear (206). PE5<sub>mt</sub> does not have an expanded C-terminal domain and is just the conserved PE domain, yet its secretion motif is still unstructured in our heterotrimer. Therefore, the exact significance of the structural variations in the ESX secretion motif is still unclear, and further work is needed.

Our structure is the first of EspG<sub>3</sub> bound to a cognate PE–PPE heterodimer. In comparisons of the various published EspG<sub>3</sub> structures, we identified two different forms that relate to the orientation of the C-terminal helical bundle: an open form and a closed form. EspG<sub>3mm</sub>, when bound to the PE5<sub>mt</sub>–PPE4<sub>mt</sub> heterodimer, is in a conformation slightly different from the closed form because of interactions with the tip of PPE4<sub>mt</sub>. We also found that the C-terminal helical bundle is a dynamic domain and shifts between the open and closed forms via a hinge movement (Figure 4.9D). The functional significance of this domain movement could be 2-fold. First, the plasticity of the C-terminal helical bundle could allow EspG<sub>3</sub> to accommodate any variation in the ESX-3–specific PPE tips. Although the tip of ESX-3–specific PPE proteins is mostly conserved (Figure 4.6), there is some variations at the end of  $\alpha 5$  that could alter the tertiary structure and thus slightly alter the interactions with the EspG<sub>3</sub> chaperone and the PPE protein. Second, the movement of the C-terminal helical bundle could be critical to the release of the PE–PPE heterodimers at the ESX-3 secretion machinery. It is unlikely that PPE4<sub>mt</sub> could be removed from its interactions with EspG<sub>3mm</sub> without either movement of the C-terminal helical bundle or steric clashes with the C-terminal helical bundle. Movement of this helical bundle and release of PPE4<sub>mt</sub> would likely require energy input and a candidate to provide that energy is EccA. EccA is an ATPase (103) and interacts with both EspG and PPE proteins in yeast two-hybrid experiments (204,213). Recent structures of the ESX machinery from both ESX-3 (127) and ESX-5 (126) suggests overall 6-fold symmetry of the core ESX machinery within the inner membrane, and EccA could be acting not only to provide the energy required to uncouple the PE–PPE heterodimers from their EspG chaperone but also

to provide a platform for interaction with the core secretion machinery because EccA is likely hexameric when functional.

Previous studies showed that each EspG only recognizes PE–PPE heterodimers from their cognate systems (203,205). Despite the structures of two different PE–PPE–EspG heterotrimers from ESX-5 (204-206), it was still unclear how EspG<sub>5</sub> was differentiating from cognate and noncognate PE–PPE heterodimers. Our structure represents the first PE–PPE–EspG heterotrimer from ESX-3 and allows for direct comparisons between the ESX-3 and ESX-5 heterotrimers. Our structure reveals that PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub> interacts with EspG<sub>3<sub>mm</sub></sub> at a different angle of interaction than what was shown for either ESX-5 heterotrimer. This difference in interaction angle presents a different face of PPE<sub>4<sub>mt</sub></sub> to EspG<sub>3<sub>mm</sub></sub>. We hypothesize that this is a conserved feature of the ESX-3 PPE–EspG<sub>3</sub> interaction, because both characterized ESX-5 PE–PPE heterodimers (204-206) display the same face to EspG<sub>5</sub> despite 33% sequence identity between PPPE41 and PPE15. Therefore, we hypothesize that each ESX system has a unique shape complementarity between its subset of PPE proteins and their cognate EspG chaperone, and these unique shapes are likely not compatible for interaction with noncognate chaperones. Our structure is also the first of an ESX-3–specific PE–PPE heterodimer. PE<sub>5<sub>mt</sub></sub>–PPE<sub>4<sub>mt</sub></sub> shares the same global conformation as the previously solved PE–PPE heterodimers; however, it differs drastically in PPE<sub>4<sub>mt</sub></sub> in the loop between  $\alpha$ 5 and  $\alpha$ 6, which contains the hh motif. This longer, more extended loop interacts deeper in the cleft of EspG<sub>3<sub>mm</sub></sub> and is subsequently much more shielded from solvent. It is possible that the longer, extended loop conformation is a feature of ESX-3 PPE proteins and could play an essential role in EspG<sub>3</sub> recognition.

In conclusion, we presented the first structure of a PE–PPE–EspG heterotrimer from the ESX-3 system. This structure allowed us to compare the interactions of EspG<sub>3</sub> and a cognate PPE protein to the previously described EspG<sub>5</sub>–PPE interactions. We hypothesize that shape complementarity is a key feature of distinguishing cognate and noncognate PPE proteins from the EspG chaperones.

**4.1. Summary of constructs utilized for crystallization experiments and final outcomes.**

PE5 construct (all constructs contain His <sub>6</sub> purification tag)	PPE4 construct (only N-terminal PPE domain)	EspG <sub>3</sub> construct	Crystallization Outcome
<i>MSMEG_0618</i>	<i>MSMEG_0619</i>	<i>MSMEG_0622</i>	Low resolution crystals
<i>MSMEG_0618</i> with MBP fusion (two different linker lengths)	<i>MSMEG_0619</i>	<i>MSMEG_0622</i>	Low resolution crystals for both linker lengths
<i>MSMEG_0618</i> with T4L fusion (3 different forms of T4L)	<i>MSMEG_0619</i>	<i>MSMEG_0622</i>	Poor expression of trimer in all forms
<i>Rv0285</i>	<i>Rv0286</i>	<i>Rv0289</i>	Low resolution crystals
<i>Rv0285</i>	<i>Rv0286</i>	<i>MMAR_0548</i>	Two crystal forms solved



**Table 4.2. Data collection and refinement statistics of ESX-3 heterotrimer.**

	PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG <sub>3mm</sub> (PDB ID 6UUJ)	PE5 <sub>mt</sub> -PPE4 <sub>mt</sub> -EspG <sub>3mm</sub> (PDB ID 6VHR)
<b>Data Collection</b>		
Wavelength (Å)	1.000	1.000
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>I</i> 422
Cell Dimensions:		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	72.26, 158.63, 209.31	219.14, 219.14, 104.44
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	39.51 – 3.00 (3.08 – 3.00) <sup>a</sup>	35.73 – 3.30 (3.39 – 3.30)
R <sub>sym</sub>	0.131 (1.56)	0.087 (2.18)
R <sub>pim</sub>	0.070 (0.848)	0.029 (0.524)
CC <sub>1/2</sub> <sup>b</sup>	0.998 (0.590)	0.999 (0.597)
<i>I</i> / $\sigma$	9.45 (1.14)	15.70 (1.31)
Completeness (%)	99.1 (99.5)	99.8 (100)
Multiplicity	4.2 (4.4)	10.3 (9.5)
<b>Refinement</b>		
Resolution (Å)	39.51 – 3.00	35.73 – 3.30
No. reflections (total/free)	48463/2462	19335/928
R <sub>work</sub> /R <sub>free</sub>	0.266/0.303	0.248/0.266
Number of atoms:		
Protein	14535	3643
Ligand/ion	0	0
Water	4	0
<i>B</i> -factors:		
Protein	101.6	173.2
Water	70.6	
All atoms	101.6	173.2
Wilson <i>B</i>	87.9	147.3
R.m.s. deviations:		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.53	0.503
Ramachandran distribution <sup>c</sup> (%)		
Favored	96.42	91.34
Allowed	3.58	8.04
Outliers	0	0.62
Rotamer outliers <sup>c</sup> (%)	0.28	0
Clashscore <sup>d</sup>	7.02	6.94
MolProbity Score <sup>e</sup>	1.62	1.89

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

<sup>b</sup> CC1/2 correlation coefficient is defined in (269) and was calculated with XSCALE (248).

<sup>c</sup> Calculated with the MolProbity server (<http://molprobity.biochem.duke.edu>) (252).

<sup>d</sup> Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

<sup>e</sup> MolProbity Score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution (252).

**Table 4.3. Structural variations in copies of PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mm</sub> structure in RMSD (Å).**

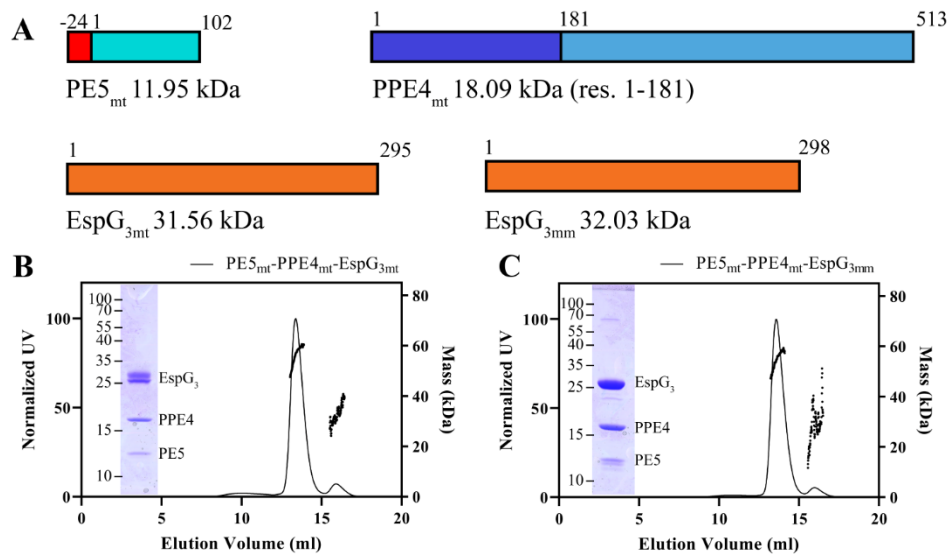
	<b>PE5<sub>mt</sub></b>	<b>PPE4<sub>mt</sub></b>	<b>EspG<sub>3mm</sub></b>
	Aligned to 6UUJ copy 1		
6UUJ copy 2	0.2	0.3	0.4
6UUJ copy 3	0.4	0.2	0.4
6UUJ copy 4	0.3	0.2	0.4
6VHR	0.4	0.5	0.6
	Aligned to 6UUJ copy 2		
6UUJ copy 1	0.2	0.3	0.4
6UUJ copy 3	0.3	0.3	0.3
6UUJ copy 4	0.3	0.3	0.4
6VHR	0.4	0.5	0.6
	Aligned to 6UUJ copy 3		
6UUJ copy 1	0.4	0.2	0.4
6UUJ copy 2	0.3	0.3	0.3
6UUJ copy 4	0.4	0.3	0.4
6VHR	0.5	0.5	0.6
	Aligned to 6UUJ copy 4		
6UUJ copy 1	0.3	0.2	0.4
6UUJ copy 2	0.3	0.3	0.4
6UUJ copy 3	0.3	0.3	0.4
6VHR	0.4	0.5	0.7

**Table 4.4. Structural variations between individual components of ESX-3 heterotrimer and the previously published ESX-5 heterotrimers in RMSD (Å).**

	PE25 <sub>mt</sub> -PPE41 <sub>mt</sub> -EspG5 <sub>mt</sub> (4KXR)	PE25 <sub>mt</sub> -PPE41 <sub>mt</sub> -EspG5 <sub>mt</sub> (4W4L)	PE8 <sub>mt</sub> -PPE15 <sub>mt</sub> -EspG5 <sub>mt</sub> (5XFS)
PE5 <sub>mt</sub>	2.3	2.3	1.4
PPE4 <sub>mt</sub>	3.3	3.4	2.5
EspG <sub>3mm</sub>	2.4	2.7	2.3

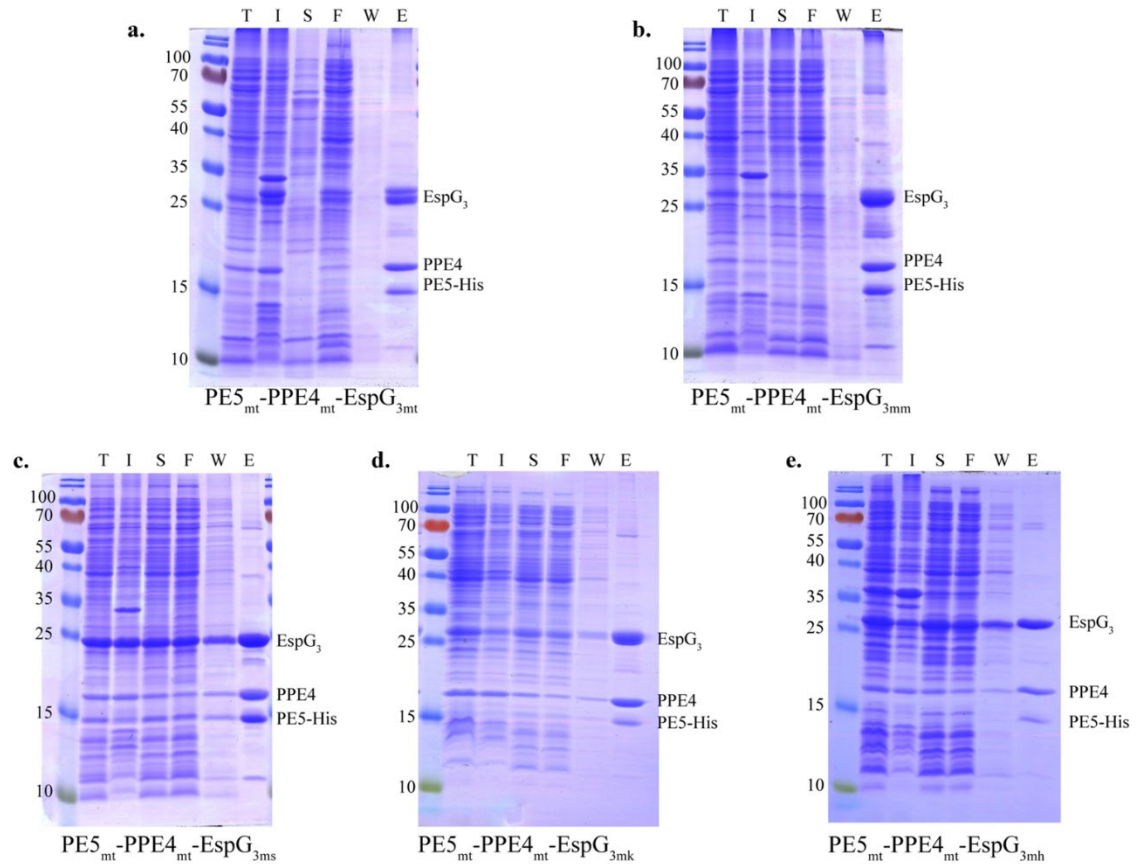
**Table 4.5. Summary of analysis of PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mt</sub> interactions in vitro.**

<b>PPE4 mutations</b>	
N127D	+
F128R	-
F129E	-
N132E	+
E140R	+
<b>EspG<sub>3</sub> mutations</b>	
R87E	+
R102E	+
R208E	+
E212R	-
S231Y	-



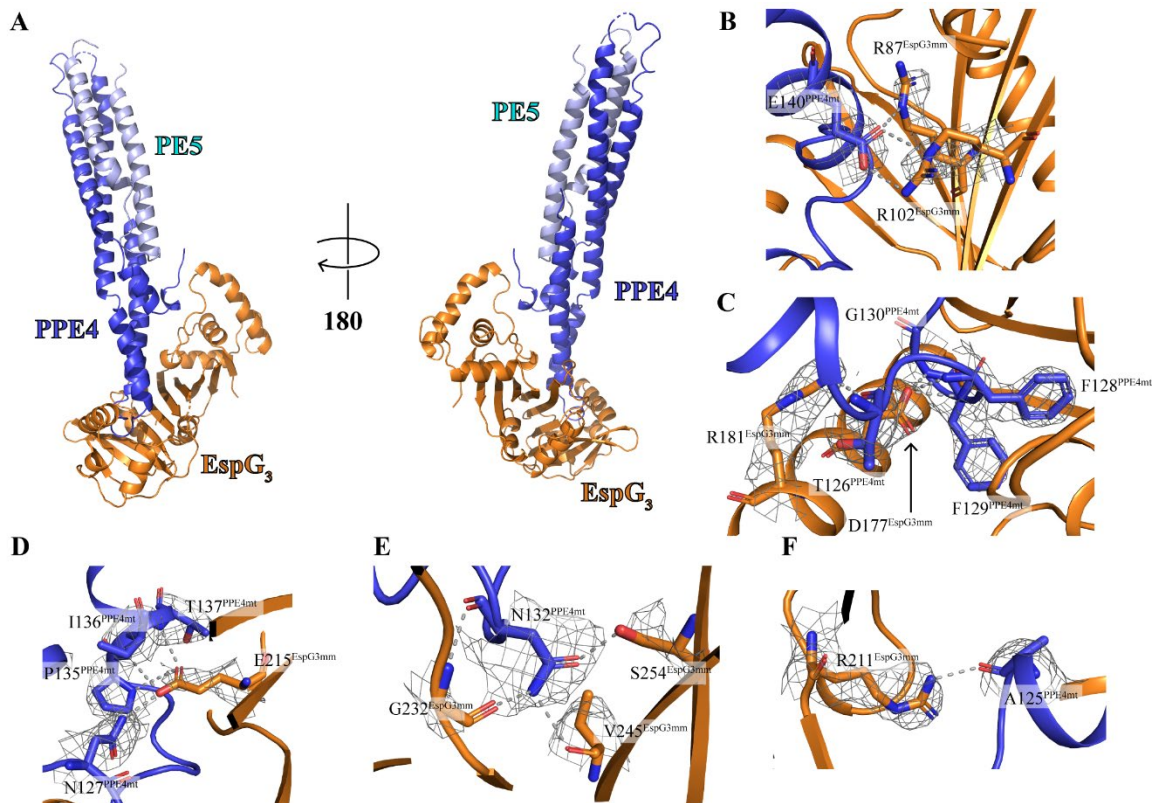
**Figure 4.1. Solution characterization of the PE5–PPE4–EspG<sub>3</sub> heterotrimer.**

*A*, schematic showing design and molecular masses for constructs used in this study. PE5 from *M. tuberculosis* (*Rv0285*) contains an N-terminal His<sub>6</sub> tag that is connected to the gene via a TEV protease cleavable linker. PPE4 (*Rv0286*) was truncated after its N-terminal PPE domain. Full-length copies of both *M. tuberculosis* (*Rv0289*) and *M. marinum* (*MMAR\_0548*) EspG<sub>3</sub> were used. *B* and *C*, elution profile of PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mt</sub> (*B*) and PE5<sub>mt</sub>–PPE4<sub>mt</sub>–EspG<sub>3mmm</sub> (*C*), with the *right y axis* showing the MALS-measured molecular mass. The *insets* show an SDS-PAGE image of the major peak fraction. *res.*, residues.



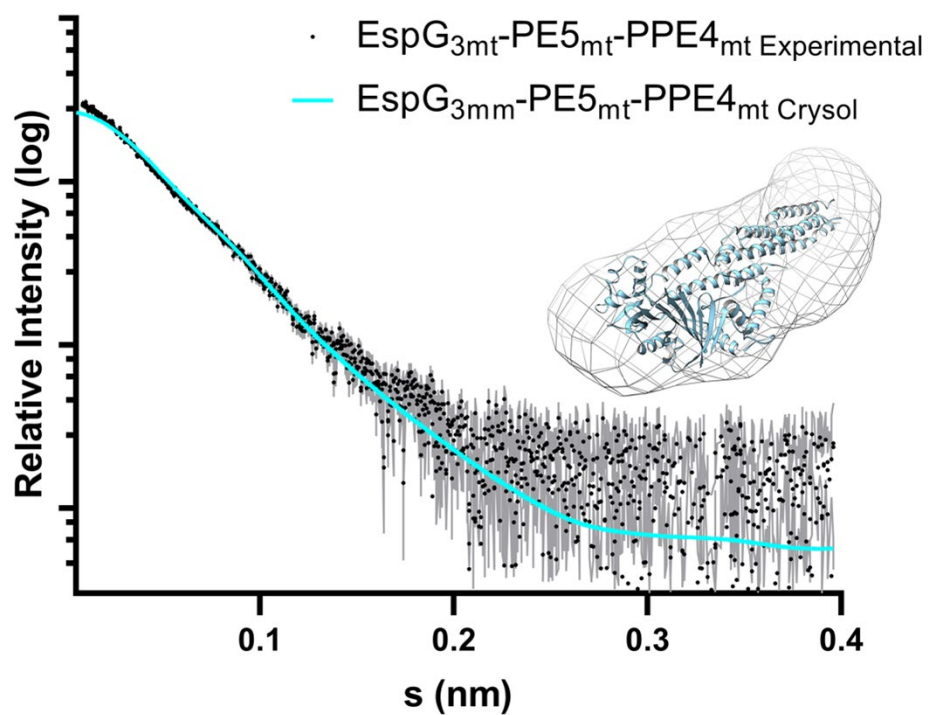
**Figure 4.2. PE5<sub>mt</sub>-PPE4<sub>mt</sub> dimer is bound by EspG<sub>3</sub> from various mycobacterial species.**

Copurification of PE5<sub>mt</sub>-PPE4<sub>mt</sub> with *a*, EspG<sub>3mt</sub>, *b*, EspG<sub>3mm</sub>, *c*, EspG<sub>3ms</sub>, *d*, EspG<sub>3mk</sub>, or *e*, EspG<sub>3mh</sub>. T is total lysate, I is insoluble lysate, S is soluble lysate, F is column flow through, W is column wash, and E is column elution.



**Figure 4.3. Crystal structure of the PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> heterotrimer and selected interactions in PPE4<sub>mt</sub>-EspG<sub>3mm</sub> interface.**

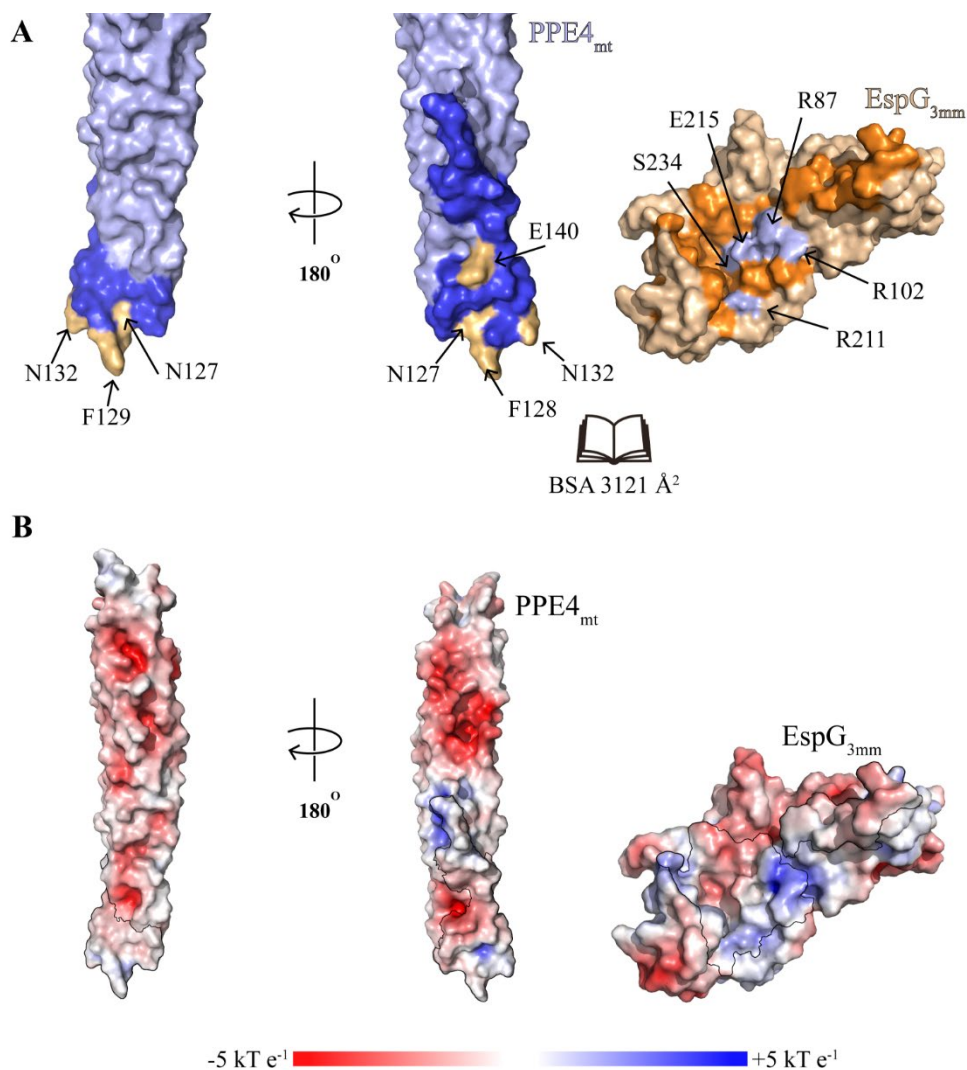
*A*, crystal structure of the PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> heterotrimer in a cartoon representation with two views related by a rotation of 180°. EspG<sub>3</sub> interacts exclusively with the tip PPE4, distal to PE5. *B-F*, interacting residues are shown with main chain and side chain in stick form with electron density map ( $2F_o - F_c$  shown at  $1.0 \sigma$ ) covering side chains and hydrogen bonding in gray dashed lines.



**Figure 4.4. Comparison of PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG3<sub>mm</sub> crystal structure and PE5<sub>ms</sub>-PPE4<sub>ms</sub>-EspG3<sub>ms</sub> SAXS data.**

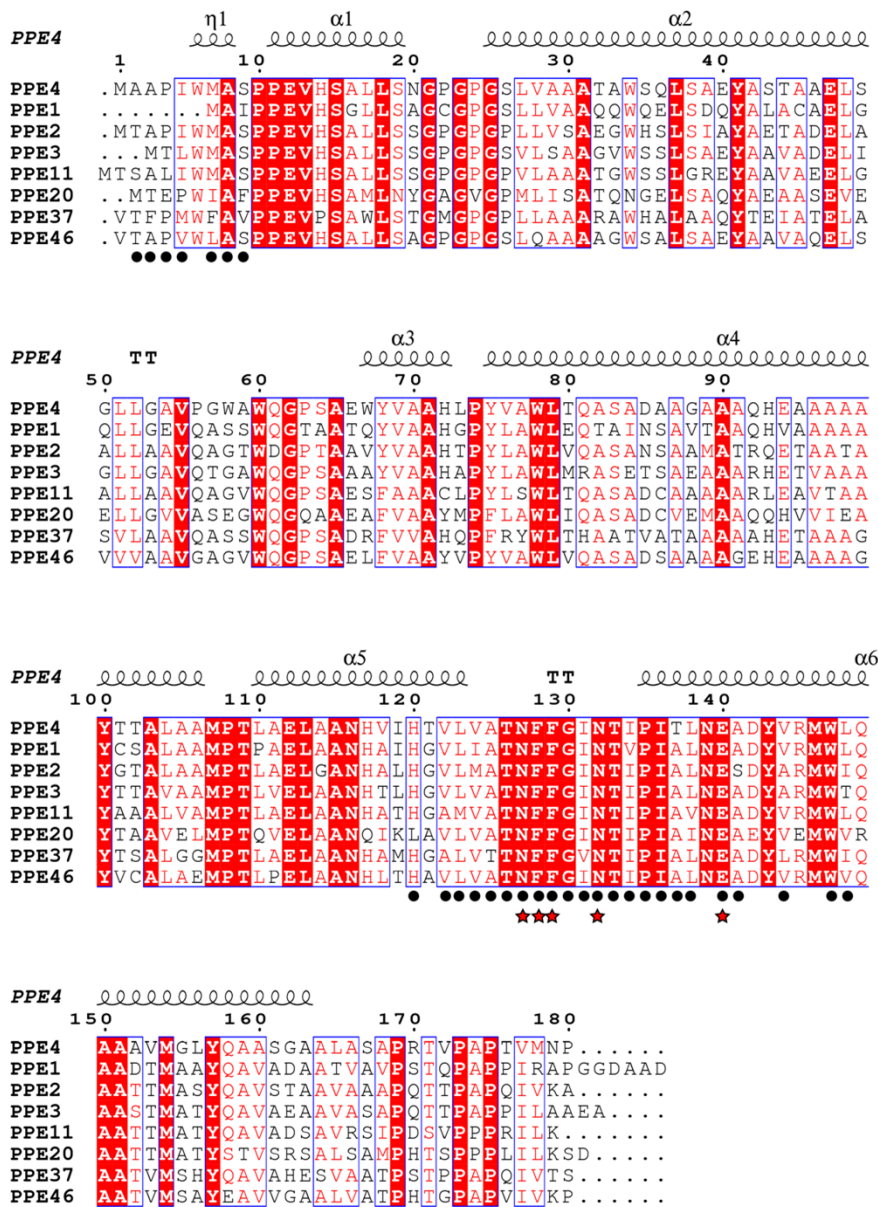
The SAXS data was originally collected in (202) and compared to the 6UUJ structure we obtained. The  $\chi^2$  between the crystal structure and SAXS data is 2.53. An insert shows the 6UUJ structure inside an envelope created by GABSOR (254).





**Figure 4.5. Interface between PPE4<sub>mt</sub> and EspG<sub>3mm</sub>.**

*A*, surface representation of PPE4<sub>mt</sub> and EspG<sub>3mm</sub> shown in an “open book” view. Interacting residues are colored in *blue* (PPE4<sub>mt</sub>) and *orange* (EspG<sub>3mm</sub>), with mutated residues highlighted in *light orange* (PPE4<sub>mt</sub>) and *light blue* (EspG<sub>3mm</sub>). Another view of PPE4<sub>mt</sub>, related by a 180° rotation, is also shown. *B*, the same orientations as *A* but with the surface colored according to surface potential as calculated by APBS (258). The interacting residues are highlighted with a *black outline*.



**Figure 4.6. Sequence alignment of *M. tuberculosis* ESX-3-specific PPE genes.**

Genomic sequences of *M. tuberculosis* PPE proteins were aligned with Clustal (256). The secondary structure from PPE4<sub>tm</sub> from the trimer model (6UUJ) is shown above each row of the alignment. Residues that are identical across all of the ESX-3-specific PPEs are highlighted in red. Residues interacting with EspG<sub>3mm</sub> in the crystal are denoted with black circles and the ones that were chosen for mutagenesis are denoted with a red star.

*M. marinum*

β1 → α1 η1

1 10 20 30 40

*M. marinum* .....MESM PNAVELTVENAWFIAEMV GAGTFPWVLAIT T PYSDEAQR S  
*M. tuberculosis* .....MDAT PNAVELTVDNWAFIAETI GAGTFPWVLAIT T PYSDA AQRG  
*M. smegmatis* .....MG PNAVELTDDAWCLADV LGAGSYFWVLAIT T PYS DHSQRS  
*M. kansasii* MVAESAEPSQ PNAVELTVDNWAFIAESV GAGSFPWVLAIT T P Y T D S A Q R S  
*M. hassiacum* .....MP ANAVELTAEQAWFAADAT GAGNFPWVLAIT T P Y T D H A B R A

*M. marinum*

α2 α3 β2

50 60 70 80 90

*M. marinum* AF FARQRDELTQTGLLS SDGVVN PAVAEWIKVVC FPERWLDLRYV GP GTG  
*M. tuberculosis* AFVDRQRDELT RMGLLS PQGVIN PAVADWIKVVC FPERWLDLRYV GP PASA  
*M. smegmatis* AFLAAQSAELTRMGV VNSAGAVD PRVAQWITTVCRATQWLDLRFV S...  
*M. kansasii* AFD RQKDEL TGLGLMSPDGT VNSAVADWIKVVC FPERWLDLRYI GPASA  
*M. hassiacum* AVEDRLTR ELT E LGVMR. D GAVD PRVRRWIATTCRPRWLELRFV R...  
 \*

*M. marinum*

β3 β4 β5

100 110 120 130 140

*M. marinum* N G G E D L L R G I V A Q S A G I M G K A G A H P S F N T V V A L R N A Q L V T F T A M D I D D P  
*M. tuberculosis* D G A C E L L R G I V A L R T G T . . . G K T S N K T G N G V V A L R N A Q L V T F T A M D I D D P  
*M. smegmatis* . G P G D L L R G M V A R R S E . . . . . E T V V A L R N A Q L V T F T A M D I G H Q  
*M. kansasii* E G S G E L L R G I V A R R V G . . . A G G K T S R K F N T V V A L R S A Q L V T F T A M D I D D P  
*M. hassiacum* . G S G Q M L R G I V A R R G D G S . . . . . A A G E T T V V A L R S G G L V T F S E L A V D H P  
 \*

*M. marinum*

η2 α4 β6 α5

150 160 170 180 190

*M. marinum* R A L V P V L G V G L S A R P P A R F E E F S M P M R V G A R A D E R L R S G E S L D E V L D Y L G  
*M. tuberculosis* R A L V P I L G V G L A H R P P A R F D E F S L P T R V G A R A D E R L R S G V P L G E V V D Y L G  
*M. smegmatis* H A L V P V L T A G L S G R K P A R F D D E A L P A A G A R A D E Q I R N G A P L A E V L E F L G  
*M. kansasii* R A L V P V L G V G L S Q R P P A R F E E F S M P M R V G A R A D E R L R S G A P L E E V L D Y L G  
*M. hassiacum* Q A L V P I L T A G L S G R A P A R F A E E S I P A R A G A R A D E Q L R N G A D L A E I L E F L G  
 \*

*M. marinum*

α6 β7 β8 β9

200 210 220 230 240

*M. marinum* I P V S A R P V V Q A V F S G P R S Y V E T V A G C N R D G E H T T T D V G H S I V D T T A G R V L  
*M. tuberculosis* I P A S A R P V V E S V F S G P R S Y V E T V A G C N R D G R H T T T E V G H S I V D T S A G R V L  
*M. smegmatis* V P S A R P L V E S V F D G R R T Y V E T V A G E H R D G H R V T T E V G V S I I D T P H G R I L  
*M. kansasii* I P T S A R P V V E A V F S G P R S Y V E T V A G C N R D G Q H T T T D V G H S I V D T T A G R V L  
*M. hassiacum* I P P S A R P V V E A A Y A P D R S Y V E V V A G D H R D G H R V S T E V G V S I V D T R Q G R V L  
 \*

*M. marinum*

β10 β11 α7

250 260 270 280 290

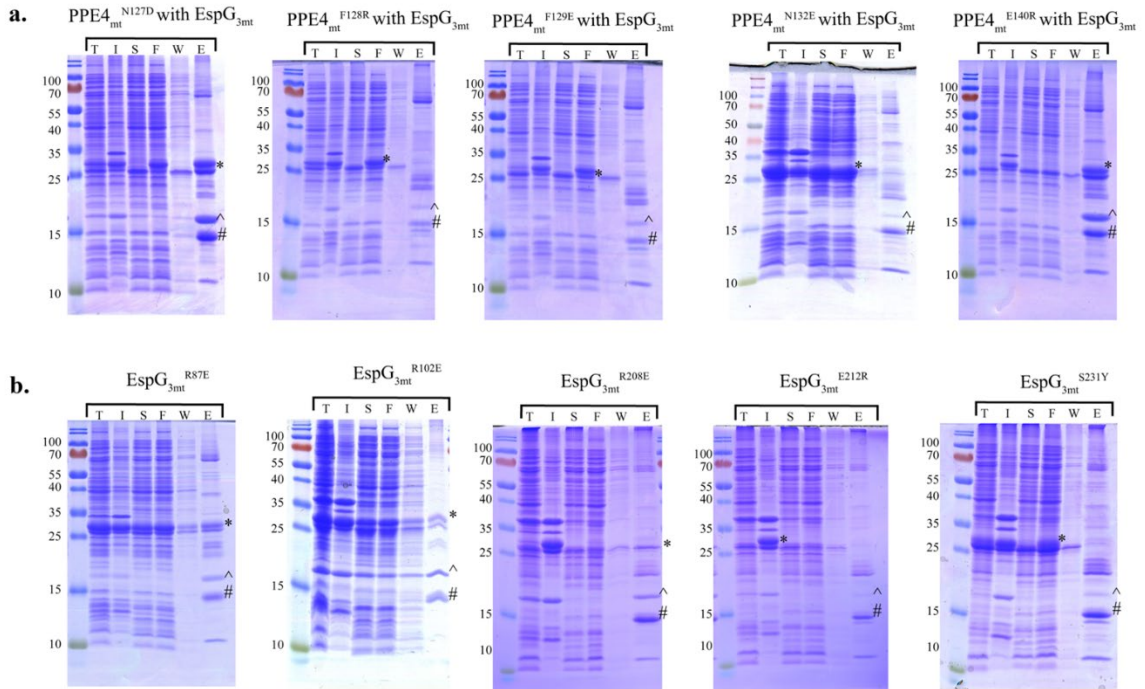
*M. marinum* V S P S R A F D G E W V S T F S A G T P F A T A V A I D Q L I A N L P D G Q W F P G Q R L S R D F S  
*M. tuberculosis* V S P S R A F D G E W V S T F S P G T P F A I A V A I Q T L T A C L P D G Q W F P Q R V S R D F S  
*M. smegmatis* V H P T K A F D G E W S T F T P G S A D A I A M A V E R L T A S L P S G S W F P D Q P L T R D F D  
*M. kansasii* V S P S R A F D G E W V S T F S P G T P F A T A V A I E Q L I A N L P E G Q W F P G Q R L S R D F S  
*M. hassiacum* V H P S R A Y D G E W V S T F T A G T P L A I A A A V E R L T A T L P D G P W F P D L H L T R D F D  
 \*

*M. marinum*

*M. marinum* SQPS.....  
*M. tuberculosis* TQSS.....  
*M. smegmatis* EDAATHREPVLQRRTQKA..  
*M. kansasii* GQSS.....  
*M. hassiacum* TDNRTEDQQWRQTLTHQARR

**Figure 4.7. Sequence alignment of selected EspG<sub>3</sub>'s shows interacting residues are conserved.**

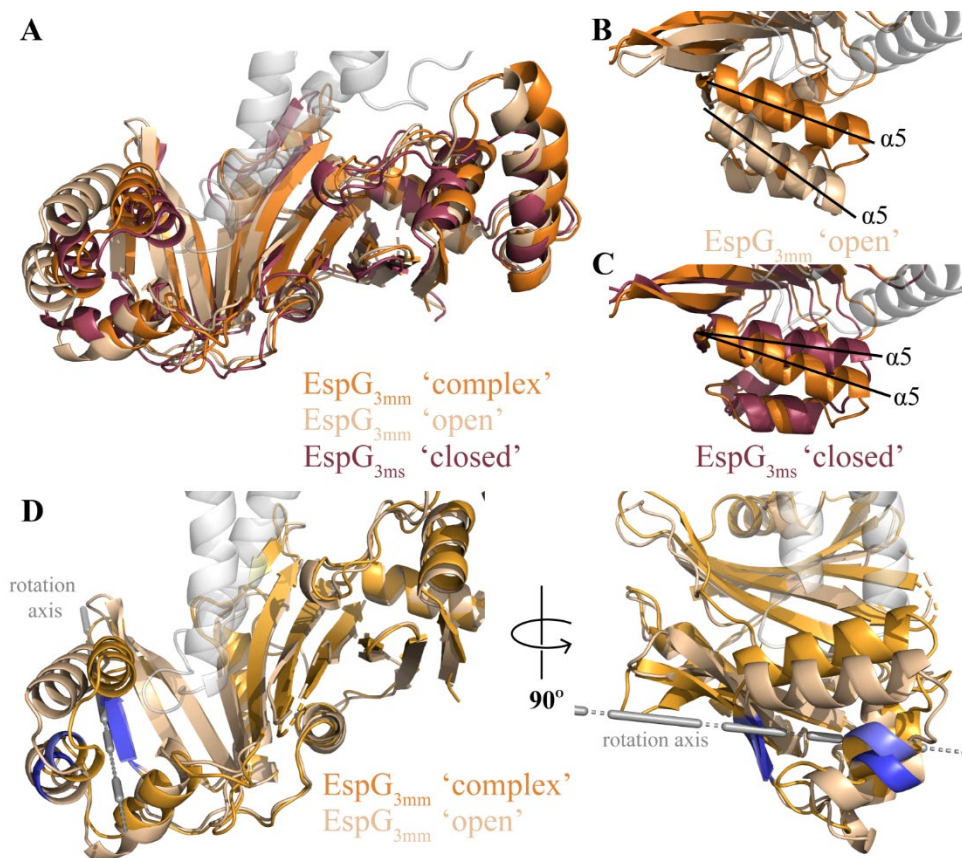
Genomic sequences of the five EspG<sub>3</sub>s used in this study were aligned using Clustal (256). The secondary structure from EspG<sub>3mm</sub> from our trimer model (6UUJ) is shown above each row of the alignment. Residues that are identical across the five species are highlighted in red. Residues interacting with PPE4<sub>mt</sub> in the crystal are denoted with black circles and the ones that were chosen for mutagenesis are denoted with a red star.



**Figure 4.8. Co-purification of selected PPE4<sub>mt</sub> and EspG<sub>3mt</sub> mutants with their wild-type partners.**

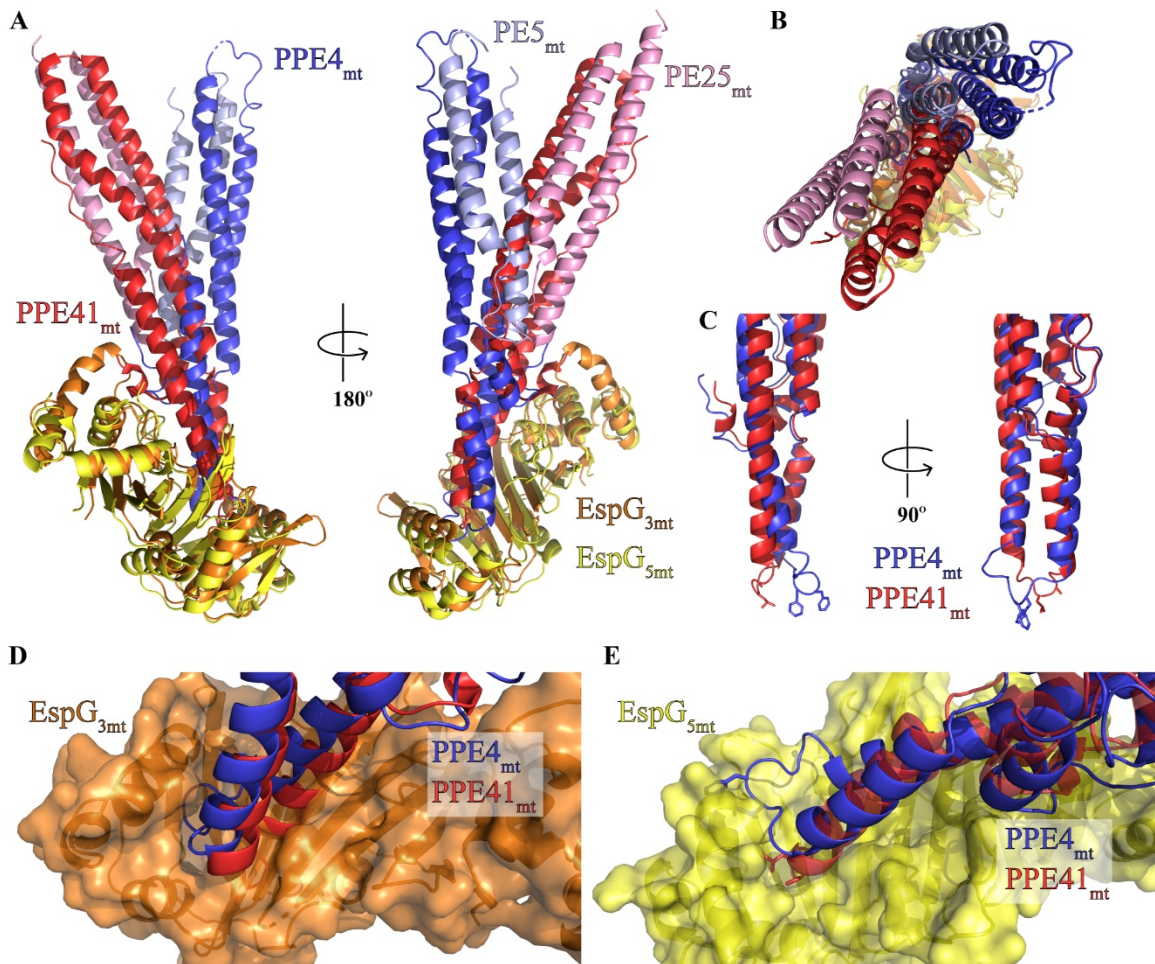
Gels from co-purification pulldowns of PPE4<sub>mt</sub> (a) and EspG<sub>3mt</sub> (b) mutations show which mutations disrupt the PPE4<sub>mt</sub>-EspG<sub>3mt</sub> interface (PPE4<sub>mt</sub><sup>F128R</sup>, PPE4<sub>mt</sub><sup>F129E</sup>, EspG<sub>3mt</sub><sup>E212R</sup>, EspG<sub>3mt</sub><sup>S231Y</sup>), and which do not. Results are summarized in Table 4.2. Each protein is denoted with a unique symbol in each gel; PPE5<sub>mt</sub> (^), PPE4<sub>mt</sub> (#), and EspG<sub>5mt</sub> (\*). T is total lysate, I is insoluble lysate, S is soluble lysate, F is column flow through, W is column wash, and E is column elution. The identity of PPE5<sub>mt</sub>, PPE4<sub>mt</sub>, and EspG<sub>5mt</sub> was confirmed by mass spectrometry analysis.





**Figure 4.9. EspG<sub>3</sub> exists in multiple structural forms.**

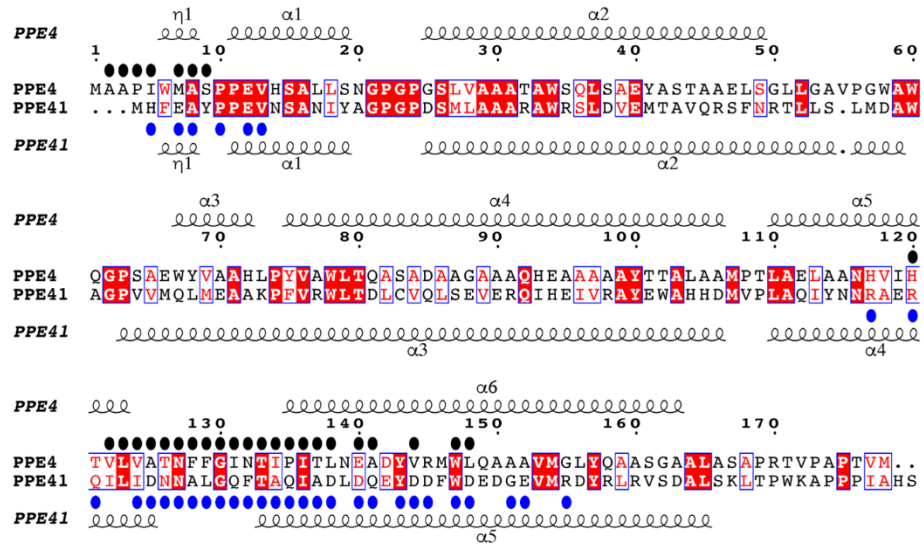
*A*, open (PDB code 5DLB, EspG<sub>3mm</sub>, *sand*) and closed (PDB code 4RCL, EspG<sub>3ms</sub>, *maroon*) conformations of EspG<sub>3</sub> were aligned to EspG<sub>3mm</sub> (PDB code 6UUJ, *orange*) as it is bound to PPE4<sub>mt</sub>. Overall the different conformations align well to the bound conformation of EspG<sub>3mm</sub> with RMSDs of 2.1 Å (open) and 1.9 Å (closed). *B* and *C*, closeups highlighting the different orientations of the α5 helices in the open (*B*) and closed (*C*) EspG<sub>3</sub> structures as compared with EspG<sub>3mm</sub> bound to PPE4<sub>mt</sub>. *D*, movement regions defined in EspG<sub>3</sub> as it moves from the open conformation to the bound conformation in two different views related by a 90° rotation. The rotation axis for the moving domain is shown in *gray*. Each conformation maintains the same coloring as in *A*, with the hinge between the moving and fixed domains colored *blue* (open) and *light blue* (bound).



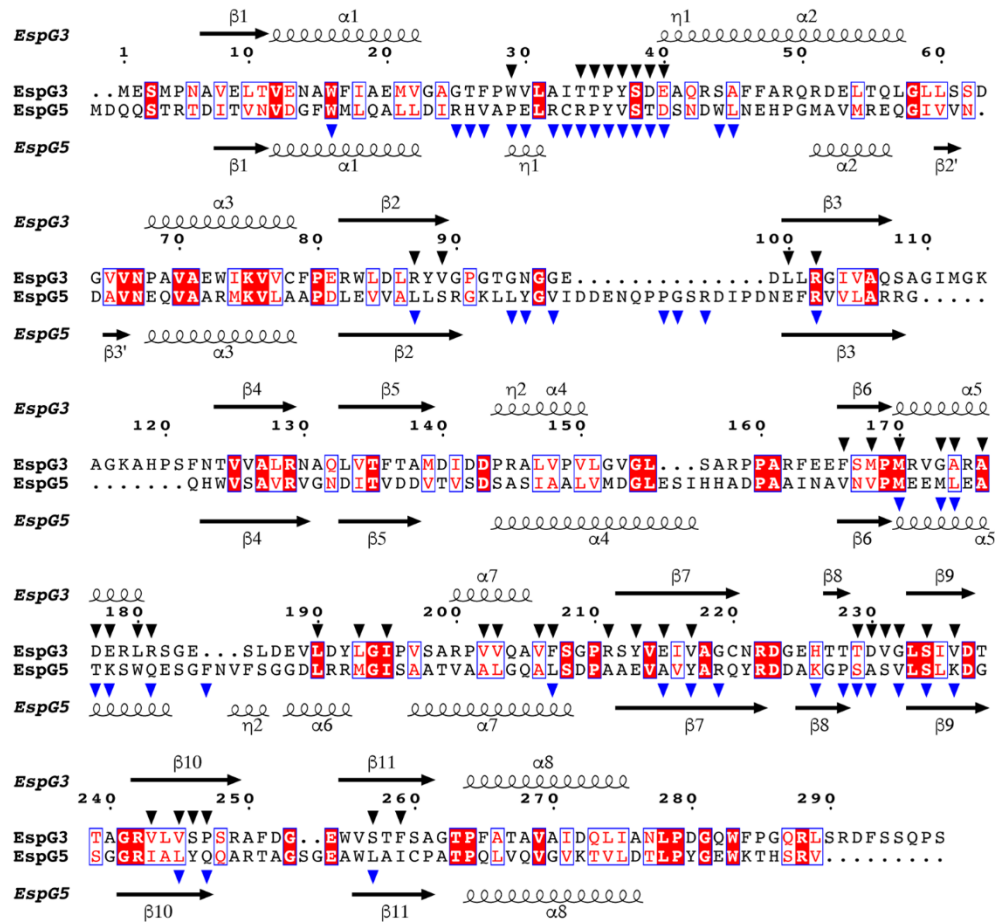
**Figure 4.10. PE5<sub>mt</sub>–PPE4<sub>mt</sub> interacts with EspG<sub>3mm</sub> chaperone in a unique mode compared with ESX-5 PE–PPE dimers.**

*A*, structural alignment (259) of the ESX-3 and ESX-5 heterotrimers via the EspG chaperones reveals a difference in the angle of interaction between the PE–PPE heterodimers with their respective chaperone. *B*, top view of alignment from *A*. *C*, superposition of PPE41<sub>mt</sub> and PPE4<sub>mt</sub> highlights difference in hh loop conformations between ESX-3 (PPE4<sub>mt</sub>) and ESX-5 (PPE41<sub>mt</sub>). *D* and *E*, superposition of PPE alignment from *C* in context of EspG<sub>3mm</sub> interaction (*D*) and EspG<sub>5mt</sub> interaction (*E*) shows the incompatibility of each PPE protein with noncognate chaperone binding.

A



B





**Figure 4.11. Comparison of interfaces in the ESX-3- and ESX-5-specific PPE-EspG complexes.**

*A*, Structure-based sequence alignment of PPE4 (6UUJ) and PPE41 (4KXR). Residues interacting with EspG<sub>3mm</sub> and EspG<sub>5mt</sub> chaperones are indicated with black and blue circles, respectively. *B*, Structure-based sequence alignment of EspG<sub>3mm</sub> and EspG<sub>5mt</sub>. Residues interacting with PPE4 and PPE41 are indicated with black and blue triangles, respectively.

## Chapter 5. Discussion.

### 5.1. Summary of this dissertation.

The work of this dissertation aimed to expand the knowledge of PE-PPE heterodimer interaction with EspG. Chapter 3 investigated the interaction of ESX-5-specific PE-PPE heterodimers and how EspG<sub>5</sub> accommodates the variety found within the PPE proteins it interacts with, particularly the variety found within the hh motif of ESX-5-specific PPE proteins. The results within Chapter 3, including the structure of PE25<sub>mt</sub>-PPE41<sup>A124L</sup><sub>mt</sub>-EspG<sub>5mm</sub> suggests that EspG<sub>5</sub> is dynamic and has structural plasticity to accommodate its repertoire of PPE proteins. Chapter 4 is focused on the ESX-3 system and aimed to uncover the mechanism of binding between EspG<sub>3</sub> and an ESX-3-specific PE-PPE heterodimer. The first ESX-3-specific PE-PPE-EspG<sub>3</sub> heterotrimer structure was described in Chapter 4, and along with the rest of the results contained in the chapter, identified a unique binding of EspG to PPE, that is different than what was previously observed with ESX-5-specific PE-PPE-EspG<sub>5</sub> heterotrimers. Our analysis in Chapter 4 also highlights a dynamic motion within EspG<sub>3</sub> that may be important for recognition and/or unloading of the PE-PPE heterodimer at the core ESX machinery. The work in this dissertation begins to give a more complete picture of PPE-EspG binding, yet still some questions exist in the PPE-EspG binding mechanism, along with how PE, PPE, and EspG proteins fit within the broader ESX systems.

### 5.2. Recognition of PPE by EspG and the ESX-1 heterotrimer mystery.

To date only three unique PE-PPE-EspG heterotrimers have been structurally characterized. The first was PE25-PPE41-EspG<sub>5</sub>, which was done by two groups in 2014 (204,205). Another ESX-5 heterotrimer, PE8-PPE15-EspG<sub>5</sub> was published in 2017 (206). The third is described in Chapter 4, was published in 2020 (247), and was the first to be from the ESX-3 system. Only ESX-1 is left without a characterized PE-PPE-EspG structure. It is still unclear what drives PPE recognition by the EspG chaperones. The hypothesis outlined in Chapter 4 is that shape complementarity between EspG and its cognate PPE proteins only allows for binding between the cognate PPE-EspG pairs.

The ESX-1-specific PE-PPE-EspG<sub>1</sub> heterotrimer has been an exceptionally challenging mystery to uncover. ESX-1 is hypothesized to be the most ancestral ESX system to secrete a PE-PPE heterodimer, and predates the expansion of *pe* and *ppe* genes

(124). This presents a challenge for structural characterization as the PE-PPE pair with experimental evidence as an ESX-1 substrate is PE35-PPE68\_1. PE35 belongs to *pe* sublineage 1 and contains only the conserved PE domain. Only one other sublineage 1 *pe* gene exists in *Mtb*, PE34, although its classification as a PE protein is only ‘probable’ and it is located in the genome alone without a putative PPE partner (123,273). PPE68\_1 belongs to *ppe* sublineage 1 and is the only sublineage 1 *ppe* gene in *Mtb* (123). The PE35-PPE68\_1 heterodimer has been purified in an *E. coli* system in the presence of EspG<sub>1</sub>, however it was done with full-length PPE68\_1, which contains a long extension in addition to the conserved PPE domain (123,205,273). Full-length PPE68\_1 may prove problematic to structural characterization due to the unknown nature of its C-terminal domain. Interestingly, all current PE-PPE-EspG heterotrimer structures contain just the conserved PE and PPE domains from the PE-PPE heterodimers (204-206,247). Although, with only three current PE-PPE-EspG heterotrimers published in the PDB, and with only one of those PE-PPE heterodimers published without EspG, PE25-PPE41 (144), there might be challenges related to the PE-PPE heterodimers themselves. EspG<sub>1</sub> also proved to be a difficult protein to structurally characterize. The first, and currently only, EspG<sub>1</sub> structure was published in 2019 (202). In the publication of the structure, the authors comment on the difficulty of obtaining crystals from the *Mtb* copy of the gene, along with the numerous homologs they screened (202). In fact, it was not until they added T4 lysozyme as an N-terminal fusion, did they obtain diffraction quality crystals (202).

Previous members of the Korotkov lab have attempted to obtain diffraction quality crystals of the PE35-PPE68\_1-EspG<sub>1</sub> heterotrimer from *Mtb*, which utilizes full-length copies of each protein without success. Future attempts to obtain diffraction quality crystals could utilize a truncated PPE68\_1 construct that only contains the conserved PPE domain. Also, the EspG<sub>1</sub> construct with the T4 lysozyme fusion utilized in (202) could also be used since the PPE-binding pocket of EspG<sub>1</sub> is not occluded in the structure. Alternatively, utilizing other EspG<sub>1</sub> homologs might prove a fruitful avenue as it was successful for the PE5<sub>Mt</sub>-PPE4<sub>Mt</sub>-EspG<sub>3Mm</sub> structure (247). The ability to utilize additional ESX-1-specific PE-PPE heterodimers is limited, as only the only option is PE35-PPE68\_1 from *Mtb*. Recently MMAR\_2894, a *pe* gene in *M. marinum*, was shown to be dependent on ESX-1 for secretion (274). MMAR\_2894 is located next to a *ppe* gene, MMAR\_2895, and this

could be its PPE partner although the genes are not located in an operon and are transcribed in opposite directions (273). MMAR\_2894 and MMAR\_2895 could easily be tested as a potential PE-PPE pair and if shown to interact, they then can be used for crystallographic studies.

The ability of EspG<sub>5</sub> to recognize its diverse group of cognate PPE proteins still requires more investigation. Chapter 3 began the work to understand EspG<sub>5</sub>'s ability to bind its cognate PPE proteins by making mutations in the hh motif of PPE41. However, this would be greatly enhanced by utilizing novel PE-PPE heterodimers and structurally characterizing them in complex with EspG<sub>5</sub>. This has been challenging because of the uncertainty in PE and PPE pairings for most genes. To date only PE25-PPE41 and PE8-PPE15 have been structurally characterized as complexes with EspG<sub>5</sub>. Even within the ESX-5 cluster it is unclear which *pe* genes partner with which *ppe* genes as there is an uneven number of each, two *pe* and three *ppe*. Some work has been done to computationally predict PE-PPE pairings, however ambiguity and multiple pairings still existed for many of the PE-PPE heterodimers studied (275). Only a small number of possible pairings outlined in the computational study have been experimentally verified as interacting, and individual PE or PPE proteins interacting with multiple partners has not yet been ruled out (206,275,276).

The exact molecular determinants of PPE-recognition by EspG are still unclear, but the work in this dissertation aided the formation of the hypothesis that shape complementarity plays an important role. A crucial missing piece is the structure of an ESX-1-specific PE-PPE-EspG heterotrimer.

### 5.3. Recognition of the PE-PPE-EspG heterotrimer by the core ESX machinery.

A conserved type VII secretion signal has been discovered (147). This signal is found in substrates from each ESX system, so it does not confer system specificity. What drives each substrate to its cognate secretion machinery remains unsolved. EccC may be the receptor for substrates at the core secretion machinery. Substrates bind EccC and activate its multimerization (138), and recently the linker 2 domain of EccC has been shown to mediate species-specificity of the ESX systems (277). If EccC is the receptor for substrates at the core machinery, it is unclear how it would recognize its cognate substrates.

I hypothesize that a system specificity code exists for ESX substrates and I will discuss this specificity signal in the context of the PE-PPE-EspG heterotrimer.

In terms of the PE-PPE-EspG heterotrimers three interesting options present themselves as the key to system specificity. The first is that the system specific signal is located within EspG. There is low sequence homology across the paralogs of EspG in *Mtb*, ranging from only 13% to 23% sequence identity (202), which leaves plenty of sequence space for unique, system specific codes to be encoded within each EspG. It is also simpler to encode the system specificity signal on EspG than it would be individual PE or PPE proteins. Evidence suggests that PE-PPE heterodimers require EspG for proper secretion (190,191,203,205). Therefore, EspG will always be present at the core machinery with its system specificity signal to ensure PE-PPE heterodimers are secreted by the correct ESX system. The linkage of PE-PPE heterodimers with EspG leads to the second option for system specificity, the signal is a part of the PPE-EspG interface. My hypothesis outlined in Chapter 4 is that each EspG is binding its cognate PPE proteins in unique modes due to the variety of the shape complementarity between EspG and PPE. This would present unique cross-faces to the core secretion machinery for each subset of PPE-EspG interactions. This hypothesis can be tested with more structures of PE-PPE-EspG heterotrimers and analyzing the combined PPE-EspG face. Currently only ESX-5 has multiple heterotrimers structurally characterized with PE25-PPE41-EspG<sub>5</sub> (204,205) and PE8-PPE15-EspG<sub>5</sub> (206).

The final option is linked with the uncoupling of the PE-PPE heterodimers from EspG. It is established the EspG is maintained cytosolically while PE-PPE heterodimers are secreted (203), with the low nanomolar affinity of PPE-EspG it is unlikely that this interaction is broken without an external energy input. EccA is the current leading candidate to provide this energy due to its ATPase activity and its ability to interact with both PPE and EspG (204,213-215). The *eccA* gene is also only encoded in ESX systems that also encode *pe*, *ppe*, and *espG* genes. EccC is another ATPase in the ESX core machinery, however it is hypothesized that it provides the energy for translocation across the plasma membrane, and it is unlikely that it would also provide the energy to uncouple PE-PPE heterodimers from EspG (127,136,137). The interaction platform between EccA and the PE-PPE-EspG heterotrimer is likely unique between each ESX system as each ESX

system encodes its own copy of *eccA* and knocking out *eccA* only effects specific ESX systems (103,121). I hypothesize that this action would be a two-step activation of EccA's ATPase activity. EccA would first bind the PE-PPE-EspG heterotrimer in a system-specific manner. The EccA-heterotrimer complex would then be recognized by the ESX core machinery, which would activate EccA's ATPase activity to uncouple the PE-PPE heterodimer from EspG and allow the PE-PPE heterodimer to be translocated through the ESX core machinery. The recognition of the EccA-heterotrimer complex by the ESX core machinery would be a key regulatory step to prevent the unproductive uncoupling of PE-PPE heterodimers in the cytosol. EccA is a soluble protein and is not a part of the stable core ESX machinery (125-128). Although EccA could be transiently interacting with the core machinery, similarly to how MycP interacts with the core machinery (129,133). If EccA does not interact with the core machinery it would be specifically recognizing cognate PE-PPE-EspG heterotrimers and the system specificity signal could be located in either manner described above: on EspG or as part of the PPE-EspG interface.

Recently it was postulated that the EspG-binding region of PPE proteins conferred system specificity. In so-called swapping experiments, the EspG-binding region of the ESX-5-specific PPE18 replaced the EspG-binding region of the ESX-1-specific PPE68\_1 (278). This "swapped" PPE68\_1 was co-expressed with its PE partner, PE35, and shown to be dependent on ESX-5 for proper secretion (278). It is unclear if this "swapped" PPE68\_1 is interacting with EspG<sub>5</sub>, and therefore it makes it difficult to piece out where exactly the system-specificity signal is located. If PPE68\_1 is interacting with EspG<sub>5</sub>, then EspG<sub>5</sub> could still be carrying the system specificity signal and "swapped" PPE68\_1 is simply re-routed to ESX-5 because of this EspG<sub>5</sub> interaction. A combined PPE-EspG interface signal is also plausible if "swapped" PPE68\_1 is interacting with EspG<sub>5</sub>. Since the "swapped" portion of PPE68\_1 contains the entirety of the PPE-EspG<sub>5</sub> interface from PPE18, whatever cross-face was present in the PPE18-EspG<sub>5</sub> interface would still be present in the "swapped" PPE68\_1-EspG<sub>5</sub> interface. Evidence of which specific EspG the "swapped" PPE68\_1 is bound to would clear up some of this uncertainty. It would also be beneficial to repeat these experiments in the opposite direction, or swap the EspG-binding region of PPE68\_1 into PPE18. I would expect this set of experiments to make the "swapped" PPE18 dependent on ESX-1 for proper secretion. The system specificity signal

would likely be carried in the same location for all ESX systems, and these experiments could also be extended to PPE proteins from ESX-3. These “swapping” experiments are made more interesting by the recent discovery that an ESX-1-specific Esx heterodimer could be re-routed to ESX-5, solely by manipulating the EspG-binding region of the ESX-1-specific PE35-PPE68\_1 (279). The ESX-1-specific Esx heterodimer’s secretion through ESX-1 was severely enhanced when the ESX-1-specific PE35-PPE68\_1 heterodimer was co-expressed, and evidence suggested that this was not transcriptional linkage between the Esx heterodimer and PE-PPE heterodimer (279). While the Esx heterodimer was able to be re-routed to ESX-5 by only modifying the EspG-binding region of PPE68\_1, the efficiency of re-routing for both Esx and PE-PPE heterodimers were enhanced when their C-terminal regions, which contain the YXXXD/E general ESX secretion motif, was also swapped for C-terminal regions of ESX-5-specific Esx and PE proteins, respectively (279). The reason for this enhanced efficacy with proper C-terminal regions could be because of known interactions with Esx proteins to the core secretion machinery through EccC (138,148). However, it still appears that for the PE-PPE heterodimers system specificity is largely controlled via the EspG-binding region as swapping only this region was sufficient to re-route the heterodimers in both studies (278,279). These “swapping” studies unfortunately do not indicate whether the system specificity signal just overlaps with the EspG-binding region, or if the system specificity signal is instead carried on EspG, or perhaps the third alternative that the system specificity signal is combination from both PPE and EspG. The work in Damen *et al.* (279) also highlights the previously known interconnected nature of ESX substrates and provides a solid foundation for the continued work to understand the specificity mechanisms of ESX substrates, including the PE-PPE heterodimers.

Understanding the determinants for system specificity can improve vaccine developments for TB. As outlined in chapter 1.1., *M. bovis* BCG is used as a vaccine strain in areas where TB is endemic to protect children from severe forms of *Mtb* infections (refs 11-13). Unfortunately, vaccination with *M. bovis* BCG does not protect against the major pulmonary form of TB (14). Recently work has been done to improve vaccination strategies. One such new strategy involves a heterologous boost of *M. bovis* BCG with a strain of *Mtb* that is deficient in secretion via ESX-5 (280). The heterologous boost strategy

provided better protection against clinically relevant *Mtb* strains in both murine and guinea pig models when compared to only *M. bovis* BCG vaccination (280). The mechanism for this enhanced protection was not determined, and the creation of the ESX-5 deficiency was accomplished by deletion of the entire ESX-5 genetic locus (280). The determination of functions for individual *pe* and *ppe* genes, in addition to the knowledge of the molecular mechanism of system specificity could allow for the development of enhanced booster vaccination strains that have specially selected secretion profiles.

While it is well established that PE and PPE proteins are secreted through specific ESX systems, it is still unclear what drives the system specificity of these substrates. Numerous potential locations still exist for the system specificity signal and more work is needed.

#### 5.4. Function of the conserved PE and PPE domains.

The conserved PE and PPE domains are well established to be the drivers of localization to the cell envelope (184-186,188,281,282). This makes sense for the PE and PPE proteins with expanded C-terminal domains, with their localization linked to their function, however there are PE and PPE proteins that contain just the conserved N-terminal domains. What is the function of these proteins? There are about 18 *pe* and 7 *ppe* genes that code for only the conserved N-terminals, and while some are likely pseudogenes, many are likely expressed. The best biochemically studied PE-PPE pair, PE25-PPE41, both only contain the conserved N-terminal domains. What function are these proteins serving? An interesting possibility is a role in pore formation.

The related EspB may provide insight into pore formation. EspB contains an N-terminal domain that is a fused PE domain and PPE domain in a single chain and shares a similar structure to PE25-PPE41, and forms a hexameric donut ring structure in solution with 45 Å in diameter pore (145,146,174). EspB also functions in membrane lysis, and its N-terminal domain's structure alludes to functions as a porin or transporter (166,174). There are some inconsistencies between the structure of the N-terminal domain of EspB and known porins and transporters, such as pore diameter and its negative surface charge (174). Yet, EspB still provides an interesting comparison to PE and PPE proteins, and perhaps some PE-PPE heterodimers could form similar ring structures, although no ring structures have been observed as of yet.



Recently, the EsxEF heterodimer was characterized as forming a heptameric oligomer structure (283). The EsxEF oligomer also contained a central pore with a diameter of 25 Å and this oligomer can embed itself in a membrane to form a membrane-spanning pore (283). EsxE and EsxF are both WXG-100 proteins and likely have similar structure as the previously characterized Esx heterodimers EsxAB, EsxGH, and EsxOP (284,285). Given the established role of EsxAB in cytolysis and potential for pore formation of EsxA (16,166,286) and the ring forming EspB oligomers (145,146,174), perhaps forming oligomeric rings is commonplace for ESX substrates.

There is potential for PE-PPE heterodimers consisting of just the conserved N-terminal domains to form similar oligomeric ring structures. However, to date none have been observed, although this could be because PE-PPE heterodimers are incredibly difficult to work with without their EspG chaperones, and EspG would likely be preventing any kind of PE-PPE oligomeric structure. An interesting function for these PE-PPE oligomeric structures is the formation of the mycomembrane transport channel for ESX substrates. How ESX substrates cross the mycomembrane remains unsolved. However, this is unlikely, as ESX-4 lacks any PE-PPE substrates encoded within its genetic locus and the only known ESX-3 PE-PPE substrates have functional roles in iron homeostasis (155,226,273). What function other than driving localization to the cell envelope, if any, remains to be uncovered for the conserved PE and PPE domains.

##### 5.5. Concluding remarks.

The work in this dissertation adds crucial information to the understanding of EspG recognition of PPE proteins. Chapter 3 provided a new structure of a mutated PE25-PPE41 heterodimer in complex with EspG<sub>5</sub>. This mutated heterodimer contained a novel hh motif, LL, and the results in Chapter 3 expanded the knowledge of EspG<sub>5</sub> interactions with its cognate PPE and provided evidence that EspG<sub>5</sub> has structural plasticity to bind its numerous cognate PPE proteins. Chapter 4 presented the first heterotrimer from the ESX-3 system, PE5-PPE4-EspG<sub>3</sub>. The results in Chapter 4 highlighted the unique binding mode between PPE4-EspG<sub>3</sub> and lead to the hypothesize that shape complementarity drives PPE recognition by cognate EspG. Chapter 4 also highlighted a dynamic domain located within EspG<sub>3</sub> that moves upon binding of PE-PPE heterodimers, which could allow for flexibility in bind a variety of PPE proteins and may also be critical for release of PE-PPE

heterodimers at the core secretion machinery. In total the work in this dissertation provides further understanding of the PE-PPE heterodimer interaction with the EspG chaperone.

There is still much work to be done on understanding the functions of individual PE and PPE proteins, and a more complete picture of their interactions with their chaperone EspG will aid in their studying. The importance of PE and PPE proteins in the virulence pathways of *Mtb* highlight the need for better understanding of these uniquely mycobacterial protein families. In the future this knowledge can be leveraged to develop better therapeutics. Perhaps by blocking PE-PPE heterodimer interactions in ESX-3 and thereby disrupting *Mtb* iron homeostasis. Additionally, being able to direct specific PE-PPE heterodimers to different ESX systems at will may be able to improve vaccine development by preventing secretion through damaging ESX systems, but still allowing specific immunogenic PE and PPE proteins to be secreted.

## References

1. (2020) *Global tuberculosis report 2020*, World Health Organization, Geneva
2. Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S. R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., and Gagneux, S. (2013) Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* **45**, 1176-1182
3. Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K., and Spigelman, M. (2008) Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* **3**, e3426
4. Cambier, C. J., Falkow, S., and Ramakrishnan, L. (2014) Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. *Cell* **159**, 1497-1509
5. Kaufmann, S. H. (2005) Robert Koch, the Nobel Prize, and the ongoing threat of tuberculosis. *N Engl J Med* **353**, 2423-2426
6. Nathavitharana, R. R., and Friedland, J. S. (2015) A tale of two global emergencies: tuberculosis control efforts can learn from the Ebola outbreak. *Eur Respir J* **46**, 293-296
7. (2017) *Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update*, World Health Organization, Geneva
8. Awofeso, N. (2008) Anti-tuberculosis medication side-effects constitute major factor for poor adherence to tuberculosis treatment. *Bull World Health Organ* **86**, B-D
9. (2019) *WHO consolidated guidelines on drug-resistant tuberculosis treatment.*, World Health Organization, Geneva
10. Yang, T. W., Park, H. O., Jang, H. N., Yang, J. H., Kim, S. H., Moon, S. H., Byun, J. H., Lee, C. E., Kim, J. W., and Kang, D. H. (2017) Side effects associated with the treatment of multidrug-resistant tuberculosis at a tuberculosis referral hospital in South Korea: A retrospective study. *Medicine (Baltimore)* **96**, e7482
11. Tran, V., Liu, J., and Behr, M. A. (2014) BCG Vaccines. *Microbiol Spectr* **2**, MGM2-0028-2013
12. Colditz, G. A., Berkey, C. S., Mosteller, F., Brewer, T. F., Wilson, M. E., Burdick, E., and Fineberg, H. V. (1995) The efficacy of bacillus Calmette-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* **96**, 29-35
13. Trunz, B. B., Fine, P., and Dye, C. (2006) Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* **367**, 1173-1180
14. Kaufmann, S. H., Evans, T. G., and Hanekom, W. A. (2015) Tuberculosis vaccines: time for a global strategy. *Sci Transl Med* **7**, 276fs278
15. 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*. *Mol Microbiol* **46**, 709-717

16. Hsu, T., Hingley-Wilson, S. M., Chen, B., Chen, M., Dai, A. Z., Morin, P. M., Marks, C. B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R. G., Derrick, S. C., Collins, F. M., Morris, S. L., King, C. H., and Jacobs, W. R., Jr. (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A* **100**, 12420-12425
17. Kaufmann, S. H. (2001) How can immunology contribute to the control of tuberculosis? *Nat Rev Immunol* **1**, 20-30
18. Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., Bushman, F. D., and Collman, R. G. (2011) Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* **184**, 957-963
19. Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., Cosma, C. L., and Ramakrishnan, L. (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* **505**, 218-222
20. Bates, J. H., Potts, W. E., and Lewis, M. (1965) Epidemiology of Primary Tuberculosis in an Industrial School. *N Engl J Med* **272**, 714-717
21. Houk, V. N. (1980) Spread of tuberculosis via recirculated air in a naval vessel: the Byrd study. *Ann N Y Acad Sci* **353**, 10-24
22. Wells, W. F., Ratcliffe, H. L., and Grumb, C. (1948) On the mechanics of droplet nuclei infection; quantitative experimental air-borne tuberculosis in rabbits. *Am J Hyg* **47**, 11-28
23. Hart, P. D., Armstrong, J. A., Brown, C. A., and Draper, P. (1972) Ultrastructural study of the behavior of macrophages toward parasitic mycobacteria. *Infect Immun* **5**, 803-807
24. Sturgill-Koszycki, S., Schlesinger, P. H., Chakraborty, P., Haddix, P. L., Collins, H. L., Fok, A. K., Allen, R. D., Gluck, S. L., Heuser, J., and Russell, D. G. (1994) Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* **263**, 678-681
25. Pethe, K., Swenson, D. L., Alonso, S., Anderson, J., Wang, C., and Russell, D. G. (2004) Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci U S A* **101**, 13642-13647
26. Clemens, D. L., and Horwitz, M. A. (1996) The Mycobacterium tuberculosis phagosome interacts with early endosomes and is accessible to exogenously administered transferrin. *J Exp Med* **184**, 1349-1355
27. Clemens, D. L., Lee, B. Y., and Horwitz, M. A. (2000) Deviant expression of Rab5 on phagosomes containing the intracellular pathogens Mycobacterium tuberculosis and Legionella pneumophila is associated with altered phagosomal fate. *Infect Immun* **68**, 2671-2684
28. Via, L. E., Deretic, D., Ulmer, R. J., Hibler, N. S., Huber, L. A., and Deretic, V. (1997) Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7. *J Biol Chem* **272**, 13326-13331

29. Sturgill-Koszycki, S., Schaible, U. E., and Russell, D. G. (1996) Mycobacterium-containing phagosomes are accessible to early endosomes and reflect a transitional state in normal phagosome biogenesis. *EMBO J* **15**, 6960-6968
30. Russell, D. G., Dant, J., and Sturgill-Koszycki, S. (1996) Mycobacterium avium- and Mycobacterium tuberculosis-containing vacuoles are dynamic, fusion-competent vesicles that are accessible to glycosphingolipids from the host cell plasmalemma. *Journal of immunology (Baltimore, Md. : 1950)* **156**, 4764-4773
31. Fratti, R. A., Chua, J., Vergne, I., and Deretic, V. (2003) Mycobacterium tuberculosis glycosylated phosphatidylinositol causes phagosome maturation arrest. *Proc Natl Acad Sci U S A* **100**, 5437-5442
32. Hmama, Z., Sendide, K., Talal, A., Garcia, R., Dobos, K., and Reiner, N. E. (2004) Quantitative analysis of phagolysosome fusion in intact cells: inhibition by mycobacterial lipoarabinomannan and rescue by an 1alpha,25-dihydroxyvitamin D3-phosphoinositide 3-kinase pathway. *J Cell Sci* **117**, 2131-2140
33. Vergne, I., Chua, J., and Deretic, V. (2003) Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca<sup>2+</sup>/calmodulin-PI3K hVPS34 cascade. *J Exp Med* **198**, 653-659
34. Indrigo, J., Hunter, R. L., and Actor, J. K. (2003) Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. *Microbiology (Reading)* **149**, 2049-2059
35. Goren, M. B., D'Arcy Hart, P., Young, M. R., and Armstrong, J. A. (1976) Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **73**, 2510-2514
36. Vergne, I., Chua, J., Lee, H. H., Lucas, M., Belisle, J., and Deretic, V. (2005) Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **102**, 4033-4038
37. Cowley, S., Ko, M., Pick, N., Chow, R., Downing, K. J., Gordhan, B. G., Betts, J. C., Mizrahi, V., Smith, D. A., Stokes, R. W., and Av-Gay, Y. (2004) The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Mol Microbiol* **52**, 1691-1702
38. Via, L. E., Fratti, R. A., McFalone, M., Pagan-Ramos, E., Deretic, D., and Deretic, V. (1998) Effects of cytokines on mycobacterial phagosome maturation. *J Cell Sci* **111 ( Pt 7)**, 897-905
39. Schaible, U. E., Sturgill-Koszycki, S., Schlesinger, P. H., and Russell, D. G. (1998) Cytokine activation leads to acidification and increases maturation of Mycobacterium avium-containing phagosomes in murine macrophages. *J Immunol* **160**, 1290-1296
40. Gomes, M. S., Paul, S., Moreira, A. L., Appelberg, R., Rabinovitch, M., and Kaplan, G. (1999) Survival of Mycobacterium avium and Mycobacterium tuberculosis in acidified vacuoles of murine macrophages. *Infect Immun* **67**, 3199-3206
41. 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

42. Flynn, J. L., and Chan, J. (2001) Immunology of tuberculosis. *Annu Rev Immunol* **19**, 93-129
43. Adams, D. O. (1976) The granulomatous inflammatory response. A review. *Am J Pathol* **84**, 164-192
44. Spector, W. G. (1969) The granulomatous inflammatory exudate. *Int Rev Exp Pathol* **8**, 1-55
45. Williams, G. T., and Williams, W. J. (1983) Granulomatous inflammation--a review. *J Clin Pathol* **36**, 723-733
46. Ramakrishnan, L. (2012) Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol* **12**, 352-366
47. Sakula, A. (1982) Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax* **37**, 246-251
48. Sia, J. K., and Rengarajan, J. (2019) Immunology of Mycobacterium tuberculosis Infections. *Microbiol Spectr* **7**
49. Dannenberg, A. M., Jr. (1993) Immunopathogenesis of pulmonary tuberculosis. *Hosp Pract (Off Ed)* **28**, 51-58
50. Volkman, H. E., Clay, H., Beery, D., Chang, J. C., Sherman, D. R., and Ramakrishnan, L. (2004) Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. *PLoS Biol* **2**, e367
51. Volkman, H. E., Pozos, T. C., Zheng, J., Davis, J. M., Rawls, J. F., and Ramakrishnan, L. (2010) Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. *Science* **327**, 466-469
52. Davis, J. M., Clay, H., Lewis, J. L., Ghori, N., Herbomel, P., and Ramakrishnan, L. (2002) Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. *Immunity* **17**, 693-702
53. Connolly, L. E., Edelstein, P. H., and Ramakrishnan, L. (2007) Why is long-term therapy required to cure tuberculosis? *PLoS Med* **4**, e120
54. Davis, J. M., and Ramakrishnan, L. (2009) The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* **136**, 37-49
55. Choi, H. H., Shin, D. M., Kang, G., Kim, K. H., Park, J. B., Hur, G. M., Lee, H. M., Lim, Y. J., Park, J. K., Jo, E. K., and Song, C. H. (2010) Endoplasmic reticulum stress response is involved in Mycobacterium tuberculosis protein ESAT-6-mediated apoptosis. *FEBS Lett* **584**, 2445-2454
56. Derrick, S. C., and Morris, S. L. (2007) The ESAT6 protein of Mycobacterium tuberculosis induces apoptosis of macrophages by activating caspase expression. *Cell Microbiol* **9**, 1547-1555
57. Keane, J., Balcewicz-Sablinska, M. K., Remold, H. G., Chupp, G. L., Meek, B. B., Fenton, M. J., and Kornfeld, H. (1997) Infection by Mycobacterium tuberculosis promotes human alveolar macrophage apoptosis. *Infect Immun* **65**, 298-304
58. Mishra, B. B., Moura-Alves, P., Sonawane, A., Hacoheh, N., Griffiths, G., Moita, L. F., and Anes, E. (2010) Mycobacterium tuberculosis protein ESAT-6 is a potent activator of the NLRP3/ASC inflammasome. *Cell Microbiol* **12**, 1046-1063
59. Swaim, L. E., Connolly, L. E., Volkman, H. E., Humbert, O., Born, D. E., and Ramakrishnan, L. (2006) Mycobacterium marinum infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect Immun* **74**, 6108-6117

60. Via, L. E., Lin, P. L., Ray, S. M., Carrillo, J., Allen, S. S., Eum, S. Y., Taylor, K., Klein, E., Manjunatha, U., Gonzales, J., Lee, E. G., Park, S. K., Raleigh, J. A., Cho, S. N., McMurray, D. N., Flynn, J. L., and Barry, C. E., 3rd. (2008) Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* **76**, 2333-2340
61. Clay, H., Davis, J. M., Beery, D., Huttenlocher, A., Lyons, S. E., and Ramakrishnan, L. (2007) Dichotomous role of the macrophage in early *Mycobacterium marinum* infection of the zebrafish. *Cell Host Microbe* **2**, 29-39
62. Clay, H., Volkman, H. E., and Ramakrishnan, L. (2008) Tumor necrosis factor signaling mediates resistance to mycobacteria by inhibiting bacterial growth and macrophage death. *Immunity* **29**, 283-294
63. Bottai, D., Stinear, T. P., Supply, P., and Brosch, R. (2014) Mycobacterial Pathogenomics and Evolution. *Microbiol Spectr* **2**, MGM2-0025-2013
64. Barry, C. E., 3rd, Lee, R. E., Mdluli, K., Sampson, A. E., Schroeder, B. G., Slayden, R. A., and Yuan, Y. (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* **37**, 143-179
65. Daffé, M., and Draper, P. (1997) The Envelope Layers of Mycobacteria with Reference to their Pathogenicity. *Advances in Microbial Physiology* **39**, 131-203
66. Niederweis, M. (2003) Mycobacterial porins--new channel proteins in unique outer membranes. *Mol Microbiol* **49**, 1167-1177
67. Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffe, M. (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**, 5672-5680
68. Hoffmann, C., Leis, A., Niederweis, M., Pitzko, J. M., and Engelhardt, H. (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* **105**, 3963-3967
69. Green, E. R., and Mecsas, J. (2016) Bacterial Secretion Systems: An Overview. *Microbiol Spectr* **4**
70. 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
71. Vrontou, E., and Economou, A. (2004) Structure and function of SecA, the preprotein translocase nanomotor. *Biochim Biophys Acta* **1694**, 67-80
72. Miller, B. K., Zulauf, K. E., and Braunstein, M. (2017) The Sec Pathways and Exportomes of *Mycobacterium tuberculosis*. *Microbiol Spectr* **5**
73. Meyer, T. H., Menetret, J. F., Breitling, R., Miller, K. R., Akey, C. W., and Rapoport, T. A. (1999) The bacterial SecY/E translocation complex forms channel-like structures similar to those of the eukaryotic Sec61p complex. *J Mol Biol* **285**, 1789-1800
74. Kihara, A., Akiyama, Y., and Ito, K. (1995) FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* **92**, 4532-4536
75. Nishiyama, K., Suzuki, T., and Tokuda, H. (1996) Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. *Cell* **85**, 71-81

76. Bassford, P. J., Jr., Silhavy, T. J., and Beckwith, J. R. (1979) Use of gene fusion to study secretion of maltose-binding protein into Escherichia coli periplasm. *J Bacteriol* **139**, 19-31
77. von Heijne, G. (1990) The signal peptide. *J Membr Biol* **115**, 195-201
78. Luirink, J., and Sinning, I. (2004) SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta* **1694**, 17-35
79. Valent, Q. A., de Gier, J. W., von Heijne, G., Kendall, D. A., ten Hagen-Jongman, C. M., Oudega, B., and Luirink, J. (1997) Nascent membrane and presecretory proteins synthesized in Escherichia coli associate with signal recognition particle and trigger factor. *Mol Microbiol* **25**, 53-64
80. Egea, P. F., and Stroud, R. M. (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci USA* **107**, 17182-17187
81. Economou, A., and Wickner, W. (1994) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**, 835-843
82. Braunstein, M., Brown, A. M., Kurtz, S., and Jacobs, W. R., Jr. (2001) Two nonredundant SecA homologues function in mycobacteria. *J Bacteriol* **183**, 6979-6990
83. Ligon, L. S., Rigel, N. W., Romanchuk, A., Jones, C. D., and Braunstein, M. (2013) Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of Mycobacteria. *J Bacteriol* **195**, 4456-4465
84. Durack, J., Burke, T. P., and Portnoy, D. A. (2015) A prl mutation in SecY suppresses secretion and virulence defects of Listeria monocytogenes secA2 mutants. *J Bacteriol* **197**, 932-942
85. Fagan, R. P., and Fairweather, N. F. (2011) Clostridium difficile has two parallel and essential Sec secretion systems. *J Biol Chem* **286**, 27483-27493
86. Braunstein, M., Espinosa, B. J., Chan, J., Belisle, J. T., and Jacobs, W. R., Jr. (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis. *Mol Microbiol* **48**, 453-464
87. Kurtz, S., McKinnon, K. P., Runge, M. S., Ting, J. P., and Braunstein, M. (2006) The SecA2 secretion factor of Mycobacterium tuberculosis promotes growth in macrophages and inhibits the host immune response. *Infect Immun* **74**, 6855-6864
88. Sullivan, J. T., Young, E. F., McCann, J. R., and Braunstein, M. (2012) The Mycobacterium tuberculosis SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infect Immun* **80**, 996-1006
89. van der Woude, A. D., Stoop, E. J., Stiens, M., Wang, S., Ummels, R., van Stempvoort, G., Piersma, S. R., Cascioferro, A., Jimenez, C. R., Houben, E. N., Luirink, J., Pieters, J., van der Sar, A. M., and Bitter, W. (2014) Analysis of SecA2-dependent substrates in Mycobacterium marinum identifies protein kinase G (PknG) as a virulence effector. *Cell Microbiol* **16**, 280-295
90. Watkins, B. Y., Joshi, S. A., Ball, D. A., Leggett, H., Park, S., Kim, J., Austin, C. D., Paler-Martinez, A., Xu, M., Downing, K. H., and Brown, E. J. (2012) Mycobacterium marinum SecA2 promotes stable granulomas and induces tumor necrosis factor alpha in vivo. *Infect Immun* **80**, 3512-3520



91. van Winden, V. J. C., Houben, E. N. G., and Braunstein, M. (2019) Protein Export into and across the Atypical Diderm Cell Envelope of Mycobacteria. *Microbiol Spectr* **7**
92. McDonough, J. A., Hacker, K. E., Flores, A. R., Pavelka, M. S., Jr., and Braunstein, M. (2005) The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial beta-lactamases. *J Bacteriol* **187**, 7667-7679
93. Posey, J. E., Shinnick, T. M., and Quinn, F. D. (2006) Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis. *J Bacteriol* **188**, 1332-1340
94. Saint-Joanis, B., Demangel, C., Jackson, M., Brodin, P., Marsollier, L., Boshoff, H., and Cole, S. T. (2006) Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of Mycobacterium tuberculosis, increases beta-lactam susceptibility and virulence. *J Bacteriol* **188**, 6669-6679
95. Behrendt, J., Standar, K., Lindenstrauss, U., and Bruser, T. (2004) Topological studies on the twin-arginine translocase component TatC. *FEMS Microbiol Lett* **234**, 303-308
96. Alami, M., Lüke, I., Deitermann, S., Eisner, G., Koch, H.-G., Brunner, J., and Müller, M. (2003) Differential Interactions between a Twin-Arginine Signal Peptide and Its Translocase in Escherichia coli. *Molecular Cell* **12**, 937-946
97. Yahr, T. L., and Wickner, W. T. (2001) Functional reconstitution of bacterial Tat translocation in vitro. *EMBO J* **20**, 2472-2479
98. Bageshwar, U. K., and Musser, S. M. (2007) Two electrical potential-dependent steps are required for transport by the Escherichia coli Tat machinery. *J Cell Biol* **179**, 87-99
99. Gohlke, U., Pullan, L., McDevitt, C. A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H. R., and Berks, B. C. (2005) The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc Natl Acad Sci U S A* **102**, 10482-10486
100. Bruser, T., and Sanders, C. (2003) An alternative model of the twin arginine translocation system. *Microbiol Res* **158**, 7-17
101. Berks, B. C. (1996) A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* **22**, 393-404
102. Palmer, T., and Berks, B. C. (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* **10**, 483-496
103. 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., Gey van Pittius, N. C., Pym, A. S., Rubin, E. J., Sherman, D. R., Cole, S. T., and Brosch, R. (2009) Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog* **5**, e1000507
104. Newton-Foot, M., Warren, R. M., Sampson, S. L., van Helden, P. D., and Gey van Pittius, N. C. (2016) The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. *BMC Evol Biol* **16**, 62
105. Houben, E. N., Korotkov, K. V., and Bitter, W. (2014) Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta* **1843**, 1707-1716
106. Ates, L. S., Houben, E. N. G., and Bitter, W. (2016) Type VII Secretion: A Highly Versatile Secretion System. *Microbiol Spectr* **4**

107. Pallen, M. J. (2002) The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system? *Trends Microbiol* **10**, 209-212
108. Garufi, G., Butler, E., and Missiakas, D. (2008) ESAT-6-like protein secretion in *Bacillus anthracis*. *J Bacteriol* **190**, 7004-7011
109. Burts, M. L., DeDent, A. C., and Missiakas, D. M. (2008) EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. *Mol Microbiol* **69**, 736-746
110. Burts, M. L., Williams, W. A., DeBord, K., and Missiakas, D. M. (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc Natl Acad Sci U S A* **102**, 1169-1174
111. Baptista, C., Barreto, H. C., and Sao-Jose, C. (2013) High levels of DegU-P activate an Esat-6-like secretion system in *Bacillus subtilis*. *PLoS One* **8**, e67840
112. Anderson, M., Chen, Y. H., Butler, E. K., and Missiakas, D. M. (2011) EsaD, a secretion factor for the Ess pathway in *Staphylococcus aureus*. *J Bacteriol* **193**, 1583-1589
113. Huppert, L. A., Ramsdell, T. L., Chase, M. R., Sarracino, D. A., Fortune, S. M., and Burton, B. M. (2014) The ESX system in *Bacillus subtilis* mediates protein secretion. *PLoS One* **9**, e96267
114. Laencina, L., Dubois, V., Le Moigne, V., Viljoen, A., Majlessi, L., Pritchard, J., Bernut, A., Piel, L., Roux, A. L., Gaillard, J. L., Lombard, B., Loew, D., Rubin, E. J., Brosch, R., Kremer, L., Herrmann, J. L., and Girard-Misguich, F. (2018) Identification of genes required for *Mycobacterium abscessus* growth in vivo with a prominent role of the ESX-4 locus. *Proc Natl Acad Sci U S A* **115**, E1002-E1011
115. Clark, R. R., Judd, J., Lasek-Nesselquist, E., Montgomery, S. A., Hoffmann, J. G., Derbyshire, K. M., and Gray, T. A. (2018) Direct cell-cell contact activates SigM to express the ESX-4 secretion system in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A* **115**, E6595-E6603
116. Gray, T. A., Clark, R. R., Boucher, N., Lapierre, P., Smith, C., and Derbyshire, K. M. (2016) Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria. *Science* **354**, 347-350
117. Izquierdo Lafuente, B., Ummels, R., Kuijl, C., Bitter, W., and Speer, A. (2021) *Mycobacterium tuberculosis* Toxin CpnT Is an ESX-5 Substrate and Requires Three Type VII Secretion Systems for Intracellular Secretion. *mBio* **12**
118. Sasseti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84
119. Sasseti, C. M., and Rubin, E. J. (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**, 12989-12994
120. 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. *Proc Natl Acad Sci U S A* **106**, 18792-18797
121. Bottai, D., Di Luca, M., Majlessi, L., Frigui, W., Simeone, R., Sayes, F., Bitter, W., Brennan, M. J., Leclerc, C., Batoni, G., Campa, M., Brosch, R., and Esin, S. (2012) Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE

- protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol* **83**, 1195-1209
122. Weerdenburg, E. M., Abdallah, A. M., Mitra, S., de Punder, K., van der Wel, N. N., Bird, S., Appelmelk, B. J., Bitter, W., and van der Sar, A. M. (2012) ESX-5-deficient *Mycobacterium marinum* is hypervirulent in adult zebrafish. *Cell Microbiol* **14**, 728-739
  123. Ates, L. S. (2020) New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Mol Microbiol* **113**, 4-21
  124. 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 Evol Biol* **6**, 95
  125. Houben, E. N., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S. R., Jimenez, C. R., Ottenhoff, T. H., Luirink, J., and Bitter, W. (2012) Composition of the type VII secretion system membrane complex. *Mol Microbiol* **86**, 472-484
  126. Beckham, K. S., Ciccarelli, L., Bunduc, C. M., Mertens, H. D., Ummels, R., Lugmayr, W., Mayr, J., Rettel, M., Savitski, M. M., Svergun, D. I., Bitter, W., Wilmanns, M., Marlovits, T. C., Parret, A. H., and Houben, E. N. (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nat Microbiol* **2**, 17047
  127. Famelis, N., Rivera-Calzada, A., Degliesposti, G., Wingender, M., Mietrach, N., Skehel, J. M., Fernandez-Leiro, R., Bottcher, B., Schlosser, A., Llorca, O., and Geibel, S. (2019) Architecture of the mycobacterial type VII secretion system. *Nature* **576**, 321-325
  128. Poweleit, N., Czudnochowski, N., Nakagawa, R., Trinidad, D. D., Murphy, K. C., Sassetti, C. M., and Rosenberg, O. S. (2019) The structure of the endogenous ESX-3 secretion system. *Elife* **8**
  129. van Winden, V. J., Ummels, R., Piersma, S. R., Jimenez, C. R., Korotkov, K. V., Bitter, W., and Houben, E. N. (2016) Mycosins Are Required for the Stabilization of the ESX-1 and ESX-5 Type VII Secretion Membrane Complexes. *mBio* **7**
  130. Beckham, K. S. H., Ritter, C., Chojnowski, G., Mullapudi, E., Rettel, M., Savitski, M. M., Mortensen, S. A., Kosinski, J., and Wilmanns, M. (2020) Structure of the mycobacterial ESX-5 Type VII Secretion System hexameric pore complex. *bioRxiv*, 2020.2011.2017.387225
  131. Bunduc, C. M., Fahrenkamp, D., Wald, J., Ummels, R., Bitter, W., Houben, E. N. G., and Marlovits, T. C. (2021) Structure and dynamics of a mycobacterial type VII secretion system. *Nature* **593**, 445-448
  132. Di Luca, M., Bottai, D., Batoni, G., Orgeur, M., Aulicino, A., Counoupas, C., Campa, M., Brosch, R., and Esin, S. (2012) The ESX-5 associated *eccB*-*EccC* locus is essential for *Mycobacterium tuberculosis* viability. *PLoS One* **7**, e52059
  133. van Winden, V. J. C., Bunduc, C. M., Ummels, R., Bitter, W., and Houben, E. N. G. (2020) A Chimeric *EccB*-*MycP* Fusion Protein is Functional and a Stable Component of the ESX-5 Type VII Secretion System Membrane Complex. *J Mol Biol* **432**, 1265-1278
  134. Wagner, J. M., Chan, S., Evans, T. J., Kahng, S., Kim, J., Arbing, M. A., Eisenberg, D., and Korotkov, K. V. (2016) Structures of *EccB1* and *EccD1* from the core

- complex of the mycobacterial ESX-1 type VII secretion system. *BMC Struct Biol* **16**, 5
135. Zhang, X. L., Li, D. F., Fleming, J., Wang, L. W., Zhou, Y., Wang, D. C., Zhang, X. E., and Bi, L. J. (2015) Core component EccB1 of the Mycobacterium tuberculosis type VII secretion system is a periplasmic ATPase. *FASEB J* **29**, 4804-4814
  136. 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. *Proc Natl Acad Sci U S A* **100**, 13001-13006
  137. Zoltner, M., Ng, W. M., Money, J. J., Fyfe, P. K., Kneuper, H., Palmer, T., and Hunter, W. N. (2016) EssC: domain structures inform on the elusive translocation channel in the Type VII secretion system. *Biochem J* **473**, 1941-1952
  138. Rosenberg, O. S., Dovala, D., Li, X., Connolly, L., Bendebury, A., Finer-Moore, J., Holton, J., Cheng, Y., Stroud, R. M., and Cox, J. S. (2015) Substrates Control Multimerization and Activation of the Multi-Domain ATPase Motor of Type VII Secretion. *Cell* **161**, 501-512
  139. Wang, S., Zhou, K., Yang, X., Zhang, B., Zhao, Y., Xiao, Y., Yang, X., Yang, H., Guddat, L. W., Li, J., and Rao, Z. (2020) Structural insights into substrate recognition by the type VII secretion system. *Protein Cell* **11**, 124-137
  140. Yu, L., Fang, W., He, Y., Cai, W., Wei, W., and Tian, C. (2021) Secondary structure and transmembrane topology analysis of the N-terminal domain of the inner membrane protein EccE1 from *M. smegmatis* using site-directed spin labeling EPR. *Biochim Biophys Acta Biomembr* **1863**, 183515
  141. Soler-Arnedo, P., Sala, C., Zhang, M., Cole, S. T., and Piton, J. (2020) Polarly Localized EccE1 Is Required for ESX-1 Function and Stabilization of ESX-1 Membrane Proteins in Mycobacterium tuberculosis. *J Bacteriol* **202**
  142. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-230
  143. Renshaw, P. S., Lightbody, K. L., Veverka, V., Muskett, F. W., Kelly, G., Frenkiel, T. A., Gordon, S. V., Hewinson, R. G., Burke, B., Norman, J., Williamson, R. A., and Carr, M. D. (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* **24**, 2491-2498
  144. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 8060-8065
  145. Korotkova, N., Piton, J., Wagner, J. M., Boy-Rottger, S., Japaridze, A., Evans, T. J., Cole, S. T., Pojer, F., and Korotkov, K. V. (2015) Structure of EspB, a secreted substrate of the ESX-1 secretion system of Mycobacterium tuberculosis. *J Struct Biol* **191**, 236-244
  146. Solomonson, M., Setiawati, D., Makepeace, K. A. T., Lameignere, E., Petrotchenko, E. V., Conrady, D. G., Bergeron, J. R., Vuckovic, M., DiMaio, F., Borchers, C. H., Yip, C. K., and Strynadka, N. C. J. (2015) Structure of EspB from

- the ESX-1 type VII secretion system and insights into its export mechanism. *Structure* **23**, 571-583
147. Daleke, M. H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C. M., Luirink, J., and Bitter, W. (2012) General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A* **109**, 11342-11347
  148. Champion, P. A., Stanley, S. A., Champion, M. M., Brown, E. J., and Cox, J. S. (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* **313**, 1632-1636
  149. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544
  150. 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
  151. Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, 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. *Infect Immun* **76**, 5478-5487
  152. 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 Pathog* **8**, e1002507
  153. 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. *Infect Immun* **75**, 2668-2678
  154. Samten, B., Wang, X., and Barnes, P. F. (2009) *Mycobacterium tuberculosis* ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses. *Tuberculosis* **89**, S74-S76
  155. Tufariello, J. M., Chapman, J. R., Kerantzas, C. A., Wong, K. W., Vilcheze, C., Jones, C. M., Cole, L. E., Tinaztepe, E., Thompson, V., Fenyó, D., Niederweis, M., Ueberheide, B., Philips, J. A., and Jacobs, W. R., Jr. (2016) Separable roles for *Mycobacterium tuberculosis* ESX-3 effectors in iron acquisition and virulence. *Proc Natl Acad Sci U S A* **113**, E348-357
  156. 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
  157. 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 (Reading)* **144** ( Pt 11), 3195-3203

158. 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. *J Biol Chem* **277**, 21598-21603
159. 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: functional implications and comparisons with other M. tuberculosis Esx family complexes. *J Biol Chem* **286**, 29993-30002
160. Sundaramoorthy, R., Fyfe, P. K., and Hunter, W. N. (2008) Structure of Staphylococcus aureus EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. *J Mol Biol* **383**, 603-614
161. 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
162. 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. *Proc Natl Acad Sci U S A* **102**, 10676-10681
163. 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. *Proc Natl Acad Sci U S A* **108**, 5730-5735
164. Sani, M., Houben, E. N., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S. R., Jimenez, C. R., Daffe, M., Appelmelk, B. J., Bitter, W., van der Wel, N., and Peters, P. J. (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* **6**, e1000794
165. 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* **13**, 5151-5164
166. 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. *Mol Microbiol* **53**, 1677-1693
167. 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 Pathog* **3**, e105
168. Xu, J., Laine, O., Masciocchi, M., Manoranjan, J., Smith, J., Du, S. J., Edwards, N., Zhu, X., 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. *Mol Microbiol* **66**, 787-800
169. 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. *Mol Microbiol* **57**, 1653-1663

170. Chen, J. M., Boy-Rottger, S., Dhar, N., Sweeney, N., Buxton, R. S., Pojer, F., Rosenkrands, I., and Cole, S. T. (2012) EspD is critical for the virulence-mediating ESX-1 secretion system in Mycobacterium tuberculosis. *J Bacteriol* **194**, 884-893
171. Sala, C., Odermatt, N. T., Soler-Arnedo, P., Gulen, M. F., von Schultz, S., Benjak, A., and Cole, S. T. (2018) EspL is essential for virulence and stabilizes EspE, EspF and EspH levels in Mycobacterium tuberculosis. *PLoS Pathog* **14**, e1007491
172. Chirakos, A. E., Nicholson, K. R., Huffman, A., and Champion, P. A. (2020) Conserved ESX-1 Substrates EspE and EspF Are Virulence Factors That Regulate Gene Expression. *Infect Immun* **88**
173. Phan, T. H., van Leeuwen, L. M., Kuijl, C., Ummels, R., van Stempvoort, G., Rubio-Canalejas, A., Piersma, S. R., Jimenez, C. R., van der Sar, A. M., Houben, E. N. G., and Bitter, W. (2018) EspH is a hypervirulence factor for Mycobacterium marinum and essential for the secretion of the ESX-1 substrates EspE and EspF. *PLoS Pathog* **14**, e1007247
174. Piton, J., Pojer, F., Wakatsuki, S., Gati, C., and Cole, S. T. (2020) High resolution CryoEM structure of the ring-shaped virulence factor EspB from Mycobacterium tuberculosis. *J Struct Biol X* **4**, 100029
175. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrell, B. G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007-1011
176. 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
177. 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. *Mol Microbiol* **69**, 794-808
178. Chen, J. M., Zhang, M., Rybniker, J., Boy-Rottger, S., Dhar, N., Pojer, F., and Cole, S. T. (2013) Mycobacterium tuberculosis EspB binds phospholipids and mediates EsxA-independent virulence. *Mol Microbiol* **89**, 1154-1166
179. 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
180. Fishbein, S., van Wyk, N., Warren, R. M., and Sampson, S. L. (2015) Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. *Mol Microbiol* **96**, 901-916
181. Phan, T. H., and Houben, E. N. G. (2018) Bacterial secretion chaperones: the mycobacterial type VII case. *FEMS Microbiol Lett* **365**
182. Iantomasi, R., Sali, M., Cascioferro, A., Palucci, I., Zumbo, A., Soldini, S., Rocca, S., Greco, E., Maulucci, G., De Spirito, M., Fraziano, M., Fadda, G., Manganello,

- R., and Delogu, G. (2012) PE\_PGRS30 is required for the full virulence of *Mycobacterium tuberculosis*. *Cell Microbiol* **14**, 356-367
183. Mitra, A., Speer, A., Lin, K., Ehrt, S., and Niederweis, M. (2017) PPE Surface Proteins Are Required for Heme Utilization by *Mycobacterium tuberculosis*. *mBio* **8**, e01720-01716
184. 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. *Mol Microbiol* **66**, 1536-1547
185. Daleke, M. H., Cascioferro, A., de Punder, K., Ummels, R., Abdallah, A. M., van der Wel, N., Peters, P. J., Luirink, J., Manganelli, 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. *J Biol Chem* **286**, 19024-19034
186. Zumbo, A., Palucci, I., Cascioferro, A., Sali, M., Ventura, M., D'Alfonso, P., Iantomasi, R., Di Sante, G., Ria, F., Sanguinetti, M., Fadda, G., Manganelli, R., and Delogu, G. (2013) Functional dissection of protein domains involved in the immunomodulatory properties of PE\_PGRS33 of *Mycobacterium tuberculosis*. *Pathog Dis* **69**, 232-239
187. Chatrath, S., Gupta, V. K., and Garg, L. C. (2014) The PGRS domain is responsible for translocation of PE\_PGRS30 to cell poles while the PE and the C-terminal domains localize it to the cell wall. *FEBS Lett* **588**, 990-994
188. Dona, V., Ventura, M., Sali, M., Cascioferro, A., Provvedi, R., Palu, G., Delogu, G., and Manganelli, R. (2013) The PPE domain of PPE17 is responsible for its surface localization and can be used to express heterologous proteins on the mycobacterial surface. *PLoS One* **8**, e57517
189. Basu, S., Pathak, S. K., Banerjee, A., Pathak, S., Bhattacharyya, A., Yang, Z., Talarico, S., Kundu, M., and Basu, J. (2007) Execution of macrophage apoptosis by PE\_PGRS33 of *Mycobacterium tuberculosis* is mediated by Toll-like receptor 2-dependent release of tumor necrosis factor-alpha. *J Biol Chem* **282**, 1039-1050
190. Abdallah, A. M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R. J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., Luirink, J., and Bitter, W. (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol* **62**, 667-679
191. Abdallah, A. M., Verboom, T., Weerdenburg, E. M., Gey van Pittius, N. C., Mahasha, P. W., Jimenez, C., Parra, M., Cadieux, N., Brennan, M. J., Appelmelk, B. J., and Bitter, W. (2009) PPE and PE\_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol* **73**, 329-340
192. Akhter, Y., Ehebauer, M. T., Mukhopadhyay, S., and Hasnain, S. E. (2012) The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? *Biochimie* **94**, 110-116
193. McGuire, A. M., Weiner, B., Park, S. T., Wapinski, I., Raman, S., Dolganov, G., Peterson, M., Riley, R., Zucker, J., Abeel, T., White, J., Sisk, P., Stolte, C., Koehrsen, M., Yamamoto, R. T., Iacobelli-Martinez, M., Kidd, M. J., Maer, A. M., Schoolnik, G. K., Regev, A., and Galagan, J. (2012) Comparative analysis of



- Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. *BMC Genomics* **13**, 120
194. Brennan, M. J., and Delogu, G. (2002) The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends in microbiology* **10**, 246-249
  195. Vordermeier, H. M., Hewinson, R. G., Wilkinson, R. J., Wilkinson, K. A., Gideon, H. P., Young, D. B., and Sampson, S. L. (2012) Conserved immune recognition hierarchy of mycobacterial PE/PPE proteins during infection in natural hosts. *PLoS One* **7**, e40890
  196. Cohen, I., Parada, C., Acosta-Gio, E., and Espitia, C. (2014) The PGRS Domain from PE\_PGRS33 of Mycobacterium tuberculosis is Target of Humoral Immune Response in Mice and Humans. *Front Immunol* **5**, 236
  197. Palucci, I., Camassa, S., Cascioferro, A., Sali, M., Anosheh, S., Zumbo, A., Minerva, M., Iantomasi, R., De Maio, F., Di Sante, G., Ria, F., Sanguinetti, M., Palu, G., Brennan, M. J., Manganelli, R., and Delogu, G. (2016) PE\_PGRS33 Contributes to Mycobacterium tuberculosis Entry in Macrophages through Interaction with TLR2. *PLoS One* **11**, e0150800
  198. Copin, R., Coscolla, M., Seiffert, S. N., Bothamley, G., Sutherland, J., Mbayo, G., Gagneux, S., and Ernst, J. D. (2014) Sequence diversity in the pe\_pgrs genes of Mycobacterium tuberculosis is independent of human T cell recognition. *mBio* **5**, e00960-00913
  199. De Maio, F., Berisio, R., Manganelli, R., and Delogu, G. (2020) PE\_PGRS proteins of Mycobacterium tuberculosis: A specialized molecular task force at the forefront of host-pathogen interaction. *Virulence* **11**, 898-915
  200. Ates, L. S., Ummels, R., Commandeur, S., van de Weerd, R., Sparrius, M., Weerdenburg, E., Alber, M., Kalscheuer, R., Piersma, S. R., Abdallah, A. M., Abd El Ghany, M., Abdel-Haleem, A. M., Pain, A., Jimenez, C. R., Bitter, W., and Houben, E. N. (2015) Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. *PLoS Genet* **11**, e1005190
  201. Sayes, F., Sun, L., Di Luca, M., Simeone, R., Degaiffier, N., Fiette, L., Esin, S., Brosch, R., Bottai, D., Leclerc, C., and Majlessi, L. (2012) Strong immunogenicity and cross-reactivity of Mycobacterium tuberculosis ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. *Cell Host Microbe* **11**, 352-363
  202. Tuukkanen, A. T., Freire, D., Chan, S., Arbing, M. A., Reed, R. W., Evans, T. J., Zenkeviciute, G., Kim, J., Kahng, S., Sawaya, M. R., Chaton, C. T., Wilmanns, M., Eisenberg, D., Parret, A. H. A., and Korotkov, K. V. (2019) Structural Variability of EspG Chaperones from Mycobacterial ESX-1, ESX-3, and ESX-5 Type VII Secretion Systems. *J Mol Biol* **431**, 289-307
  203. 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., and Houben, E. N. (2012) Specific chaperones for the type VII protein secretion pathway. *J Biol Chem* **287**, 31939-31947
  204. 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. *Proc Natl Acad Sci U S A* **111**, 14758-14763

205. Korotkova, N., Freire, D., Phan, T. H., Ummels, R., Creekmore, C. C., Evans, T. J., Wilmanns, M., Bitter, W., Parret, A. H., Houben, E. N., and Korotkov, K. V. (2014) Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. *Mol Microbiol* **94**, 367-382
206. Chen, X., Cheng, H. F., Zhou, J., Chan, C. Y., Lau, K. F., Tsui, S. K., and Au, S. W. (2017) Structural basis of the PE-PPE protein interaction in Mycobacterium tuberculosis. *J Biol Chem* **292**, 16880-16890
207. Wagner, J. M., Evans, T. J., and Korotkov, K. V. (2014) Crystal structure of the N-terminal domain of EccA(1) ATPase from the ESX-1 secretion system of Mycobacterium tuberculosis. *Proteins* **82**, 159-163
208. Zeytuni, N., and Zarivach, R. (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20**, 397-405
209. Cervený, L., Strasková, A., Danková, V., Hartlova, A., Cecková, M., Staud, F., and Stulik, J. (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect Immun* **81**, 629-635
210. Zhang, Y., and Chan, D. C. (2007) Structural basis for recruitment of mitochondrial fission complexes by Fis1. *Proc Natl Acad Sci U S A* **104**, 18526-18530
211. Buttner, C. R., Sorg, I., Cornelis, G. R., Heinz, D. W., and Niemann, H. H. (2008) Structure of the Yersinia enterocolitica type III secretion translocator chaperone SycD. *J Mol Biol* **375**, 997-1012
212. Lapouge, K., Smith, S. J., Walker, P. A., Gamblin, S. J., Smerdon, S. J., and Rittinger, K. (2000) Structure of the TPR domain of p67phox in complex with Rac.GTP. *Mol Cell* **6**, 899-907
213. Teutschbein, J., Schumann, G., Mollmann, U., Grabley, S., Cole, S. T., and Munder, T. (2009) A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of Mycobacterium tuberculosis. *Microbiol Res* **164**, 253-259
214. Luthra, A., Mahmood, A., Arora, A., and Ramachandran, R. (2008) Characterization of Rv3868, an essential hypothetical protein of the ESX-1 secretion system in Mycobacterium tuberculosis. *J Biol Chem* **283**, 36532-36541
215. Gaur, A., Sharma, V. K., Shree, S., Rai, N., and Ramachandran, R. (2017) Characterization of EccA3, a CbbX family ATPase from the ESX-3 secretion pathway of M. tuberculosis. *Biochim Biophys Acta Proteins Proteom* **1865**, 715-724
216. Lewis, K. N., Liao, R., 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. *J Infect Dis* **187**, 117-123
217. Bottai, D., Groschel, M. I., and Brosch, R. (2017) Type VII Secretion Systems in Gram-Positive Bacteria. *Curr Top Microbiol Immunol* **404**, 235-265
218. Simeone, R., Bottai, D., Frigui, W., Majlessi, L., and Brosch, R. (2015) ESX/type VII secretion systems of mycobacteria: Insights into evolution, pathogenicity and protection. *Tuberculosis (Edinb)* **95 Suppl 1**, S150-154
219. Kupz, A., Zedler, U., Staber, M., Perdomo, C., Dorhoi, A., Brosch, R., and Kaufmann, S. H. (2016) ESAT-6-dependent cytosolic pattern recognition drives noncognate tuberculosis control in vivo. *J Clin Invest* **126**, 2109-2122

220. 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*. *Proc Natl Acad Sci U S A* **101**, 12598-12603
221. 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
222. Maciag, A., Dainese, E., Rodriguez, G. M., Milano, A., Provvedi, R., Pasca, M. R., Smith, I., Palu, G., Riccardi, G., and Manganelli, R. (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol* **189**, 730-740
223. Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K., and Smith, I. (2002) *ideR*, An essential gene in mycobacterium tuberculosis: role of *IdeR* in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70**, 3371-3381
224. Pandey, R., Russo, R., Ghanny, S., Huang, X., Helmann, J., and Rodriguez, G. M. (2015) *MntR*(Rv2788): a transcriptional regulator that controls manganese homeostasis in *Mycobacterium tuberculosis*. *Mol Microbiol* **98**, 1168-1183
225. Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003) Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A* **100**, 3584-3588
226. Siegrist, M. S., Steigedal, M., Ahmad, R., Mehra, A., Dragset, M. S., Schuster, B. M., Philips, J. A., Carr, S. A., and Rubin, E. J. (2014) Mycobacterial *Esx-3* requires multiple components for iron acquisition. *mBio* **5**, e01073-01014
227. Abdallah, A. M., Savage, N. D., van Zon, M., Wilson, L., Vandenbroucke-Grauls, C. M., van der Wel, N. N., Ottenhoff, T. H., and Bitter, W. (2008) The *ESX-5* secretion system of *Mycobacterium marinum* modulates the macrophage response. *J Immunol* **181**, 7166-7175
228. Bosserman, R. E., and Champion, P. A. (2017) *Esx* Systems and the Mycobacterial Cell Envelope: What's the Connection? *J Bacteriol* **199**, e00131-00117
229. Sayes, F., Pawlik, A., Frigui, W., Groschel, M. I., Crommelynck, S., Fayolle, C., Cia, F., Bancroft, G. J., Bottai, D., Leclerc, C., Brosch, R., and Majlessi, L. (2016) CD4+ T Cells Recognizing PE/PPE Antigens Directly or via Cross Reactivity Are Protective against Pulmonary *Mycobacterium tuberculosis* Infection. *PLoS Pathog* **12**, e1005770
230. Sayes, F., Blanc, C., Ates, L. S., Deboosere, N., Orgeur, M., Le Chevalier, F., Groschel, M. I., Frigui, W., Song, O. R., Lo-Man, R., Brossier, F., Sougakoff, W., Bottai, D., Brodin, P., Charneau, P., Brosch, R., and Majlessi, L. (2018) Multiplexed Quantitation of Intraphagocyte *Mycobacterium tuberculosis* Secreted Protein Effectors. *Cell Rep* **23**, 1072-1084
231. Ates, L. S., Sayes, F., Frigui, W., Ummels, R., Damen, M. P. M., Bottai, D., Behr, M. A., van Heijst, J. W. J., Bitter, W., Majlessi, L., and Brosch, R. (2018) RD5-mediated lack of PE\_PGRS and PPE-MPTR export in BCG vaccine strains results in strong reduction of antigenic repertoire but little impact on protection. *PLoS Pathog* **14**, e1007139
232. Wallace, R. J., Jr., Brown, B. A., and Griffith, D. E. (1998) Nosocomial outbreaks/pseudo-outbreaks caused by nontuberculous mycobacteria. *Annu Rev Microbiol* **52**, 453-490

233. Duarte, R. S., Lourenco, M. C., Fonseca Lde, S., Leao, S. C., Amorim Ede, L., Rocha, I. L., Coelho, F. S., Viana-Niero, C., Gomes, K. M., da Silva, M. G., Lorena, N. S., Pitombo, M. B., Ferreira, R. M., Garcia, M. H., de Oliveira, G. P., Lupi, O., Vilaca, B. R., Serradas, L. R., Chebabo, A., Marques, E. A., Teixeira, L. M., Dalcolmo, M., Senna, S. G., and Sampaio, J. L. (2009) Epidemic of postsurgical infections caused by *Mycobacterium massiliense*. *J Clin Microbiol* **47**, 2149-2155
234. Olivier, K. N., Weber, D. J., Wallace, R. J., Jr., Faiz, A. R., Lee, J. H., Zhang, Y., Brown-Elliott, B. A., Handler, A., Wilson, R. W., Schechter, M. S., Edwards, L. J., Chakraborti, S., Knowles, M. R., and Nontuberculous Mycobacteria in Cystic Fibrosis Study, G. (2003) Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis. *Am J Respir Crit Care Med* **167**, 828-834
235. Roux, A. L., Catherinot, E., Ripoll, F., Soismier, N., Macheras, E., Ravilly, S., Bellis, G., Vibet, M. A., Le Roux, E., Lemonnier, L., Gutierrez, C., Vincent, V., Fauroux, B., Rottman, M., Guillemot, D., Gaillard, J. L., and Jean-Louis Herrmann for the, O. M. A. G. (2009) Multicenter study of prevalence of nontuberculous mycobacteria in patients with cystic fibrosis in france. *J Clin Microbiol* **47**, 4124-4128
236. Griffith, D. E., Girard, W. M., and Wallace, R. J., Jr. (1993) Clinical features of pulmonary disease caused by rapidly growing mycobacteria. An analysis of 154 patients. *Am Rev Respir Dis* **147**, 1271-1278
237. Qvist, T., Taylor-Robinson, D., Waldmann, E., Olesen, H. V., Hansen, C. R., Mathiesen, I. H., Hoiby, N., Katzenstein, T. L., Smyth, R. L., Diggle, P. J., and Pressler, T. (2016) Comparing the harmful effects of nontuberculous mycobacteria and Gram negative bacteria on lung function in patients with cystic fibrosis. *J Cyst Fibros* **15**, 380-385
238. Jonsson, B. E., Gilljam, M., Lindblad, A., Ridell, M., Wold, A. E., and Welinder-Olsson, C. (2007) Molecular epidemiology of *Mycobacterium abscessus*, with focus on cystic fibrosis. *J Clin Microbiol* **45**, 1497-1504
239. Bryant, J. M., Grogono, D. M., Rodriguez-Rincon, D., Everall, I., Brown, K. P., Moreno, P., Verma, D., Hill, E., Drijkoningen, J., Gilligan, P., Esther, C. R., Noone, P. G., Giddings, O., Bell, S. C., Thomson, R., Wainwright, C. E., Coulter, C., Pandey, S., Wood, M. E., Stockwell, R. E., Ramsay, K. A., Sherrard, L. J., Kidd, T. J., Jabbour, N., Johnson, G. R., Knibbs, L. D., Morawska, L., Sly, P. D., Jones, A., Bilton, D., Laurenson, I., Ruddy, M., Bourke, S., Bowler, I. C., Chapman, S. J., Clayton, A., Cullen, M., Daniels, T., Dempsey, O., Denton, M., Desai, M., Drew, R. J., Edenborough, F., Evans, J., Folb, J., Humphrey, H., Isalska, B., Jensen-Fangel, S., Jonsson, B., Jones, A. M., Katzenstein, T. L., Lillebaek, T., MacGregor, G., Mayell, S., Millar, M., Modha, D., Nash, E. F., O'Brien, C., O'Brien, D., Ohri, C., Pao, C. S., Peckham, D., Perrin, F., Perry, A., Pressler, T., Prtak, L., Qvist, T., Robb, A., Rodgers, H., Schaffer, K., Shafi, N., van Ingen, J., Walshaw, M., Watson, D., West, N., Whitehouse, J., Haworth, C. S., Harris, S. R., Ordway, D., Parkhill, J., and Floto, R. A. (2016) Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. *Science* **354**, 751-757
240. Leao, S. C., Viana-Niero, C., Matsumoto, C. K., Lima, K. V., Lopes, M. L., Palaci, M., Hadad, D. J., Vinhas, S., Duarte, R. S., Lourenco, M. C., Kipnis, A., das Neves, Z. C., Gabardo, B. M., Ribeiro, M. O., Baethgen, L., de Assis, D. B., Madalosso,

- G., Chimara, E., and Dalcolmo, M. P. (2010) Epidemic of surgical-site infections by a single clone of rapidly growing mycobacteria in Brazil. *Future Microbiol* **5**, 971-980
241. Bryant, J. M., Grogono, D. M., Greaves, D., Foweraker, J., Roddick, I., Inns, T., Reacher, M., Haworth, C. S., Curran, M. D., Harris, S. R., Peacock, S. J., Parkhill, J., and Floto, R. A. (2013) Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *The Lancet* **381**, 1551-1560
242. Adekambi, T., Ben Salah, S., Khlif, M., Raoult, D., and Drancourt, M. (2006) Survival of environmental mycobacteria in *Acanthamoeba polyphaga*. *Appl Environ Microbiol* **72**, 5974-5981
243. Lamrabet, O., Merhej, V., Pontarotti, P., Raoult, D., and Drancourt, M. (2012) The genealogic tree of mycobacteria reveals a long-standing sympatric life into free-living protozoa. *PLoS One* **7**, e34754
244. Bakala N'Goma, J. C., Le Moigne, V., Soismier, N., Laencina, L., Le Chevalier, F., Roux, A. L., Poncin, I., Serveau-Avesque, C., Rottman, M., Gaillard, J. L., Etienne, G., Brosch, R., Herrmann, J. L., Canaan, S., and Girard-Misguich, F. (2015) *Mycobacterium abscessus* phospholipase C expression is induced during coculture within amoebae and enhances *M. abscessus* virulence in mice. *Infect Immun* **83**, 780-791
245. Dumas, E., Christina Boritsch, E., Vandenbogaert, M., Rodriguez de la Vega, R. C., Thiberge, J. M., Caro, V., Gaillard, J. L., Heym, B., Girard-Misguich, F., Brosch, R., and Sapriel, G. (2016) Mycobacterial Pan-Genome Analysis Suggests Important Role of Plasmids in the Radiation of Type VII Secretion Systems. *Genome Biol Evol* **8**, 387-402
246. Korotkov, K. V., Delarosa, J. R., and Hol, W. G. J. (2013) A dodecameric ring-like structure of the N0 domain of the type II secretin from enterotoxigenic *Escherichia coli*. *J Struct Biol* **183**, 354-362
247. Williamson, Z. A., Chaton, C. T., Ciocca, W. A., Korotkova, N., and Korotkov, K. V. (2020) PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems. *J Biol Chem* **295**, 12706-12715
248. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
249. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
250. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
251. 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., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
252. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M.,

- Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315
253. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRY SOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768-773
254. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953
255. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864
256. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641
257. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324
258. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128
259. Holm, L. (2019) Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326-5327
260. Rother, K., Hildebrand, P. W., Goede, A., Gruening, B., and Preissner, R. (2009) Voronia: analyzing packing in protein structures. *Nucleic Acids Res* **37**, D393-395
261. Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999) The packing density in proteins: standard radii and volumes. *J Mol Biol* **290**, 253-266
262. Lyskov, S., Chou, F. C., Conchuir, S. O., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J. J., and Das, R. (2013) Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS One* **8**, e63906
263. Smith, C. A., and Kortemme, T. (2010) Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J Mol Biol* **402**, 460-474
264. Smith, C. A., and Kortemme, T. (2011) Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design. *PLoS One* **6**, e20451
265. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797
266. Krissinel, E. (2012) Enhanced fold recognition using efficient short fragment clustering. *J Mol Biolchem* **1**, 76-85
267. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin,

- A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
268. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383
269. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030-1033
270. Bosch, J., Buscaglia, C. A., Krumm, B., Ingason, B. P., Lucas, R., Roach, C., Cardozo, T., Nussenzweig, V., and Hol, W. G. (2007) Aldolase provides an unusual binding site for thrombospondin-related anonymous protein in the invasion machinery of the malaria parasite. *Proc Natl Acad Sci U S A* **104**, 7015-7020
271. Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-950
272. Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144-154
273. Kapopoulou, A., Lew, J. M., and Cole, S. T. (2011) The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)* **91**, 8-13
274. Bosserman, R. E., Nicholson, K. R., Champion, M. M., and Champion, P. A. (2019) A New ESX-1 Substrate in *Mycobacterium marinum* That Is Required for Hemolysis but Not Host Cell Lysis. *J Bacteriol* **201**, e00760-00718
275. 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 Comput Biol* **4**, e1000174
276. Tiwari, B., Ramakrishnan, U. M., and Raghunand, T. R. (2015) The *Mycobacterium tuberculosis* protein pair PE9 (Rv1088)-PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through Toll-like receptor 4. *Cell Microbiol* **17**, 1653-1669
277. Bunduc, C. M., Ummels, R., Bitter, W., and Houben, E. N. G. (2020) Species-specific secretion of ESX-5 type VII substrates is determined by the linker 2 of EccC5. *Mol Microbiol* **114**, 66-76
278. Phan, T. H., Ummels, R., Bitter, W., and Houben, E. N. (2017) Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Sci Rep* **7**, 42704
279. Damen, M. P. M., Phan, T. H., Ummels, R., Rubio-Canalejas, A., Bitter, W., and Houben, E. N. G. (2020) Modification of a PE/PPE substrate pair reroutes an Esx substrate pair from the mycobacterial ESX-1 type VII secretion system to the ESX-5 system. *J Biol Chem* **295**, 5960-5969
280. Tiwari, S., Dutt, T. S., Chen, B., Chen, M., Kim, J., Dai, A. Z., Lukose, R., Shanley, C., Fox, A., Karger, B. R., Porcelli, S. A., Chan, J., Podell, B. K., Obregon-Henao, A., Orme, I. M., Jacobs, W. R., Jr., and Henao-Tamayo, M. (2020) BCG-Prime and boost with Esx-5 secretion system deletion mutant leads to better protection against clinical strains of *Mycobacterium tuberculosis*. *Vaccine* **38**, 7156-7165

281. Cascioferro, A., Daleke, M. H., Ventura, M., Dona, V., Delogu, G., Palu, G., Bitter, W., and Manganelli, R. (2011) Functional dissection of the PE domain responsible for translocation of PE\_PGRS33 across the mycobacterial cell wall. *PLoS One* **6**, e27713
282. Sali, M., Di Sante, G., Cascioferro, A., Zumbo, A., Nicolo, C., Dona, V., Rocca, S., Procoli, A., Morandi, M., Ria, F., Palu, G., Fadda, G., Manganelli, R., and Delogu, G. (2010) Surface expression of MPT64 as a fusion with the PE domain of PE\_PGRS33 enhances *Mycobacterium bovis* BCG protective activity against *Mycobacterium tuberculosis* in mice. *Infect Immun* **78**, 5202-5213
283. Tak, U., Dokland, T., and Niederweis, M. (2021) Pore-forming Esx proteins mediate toxin secretion by *Mycobacterium tuberculosis*. *Nat Commun* **12**, 394
284. Poulsen, C., Panjekar, 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**, e89313
285. Arbing, M. A., Chan, S., Harris, L., Kuo, E., Zhou, T. T., Ahn, C. J., Nguyen, L., He, Q., Lu, J., Menchavez, P. T., Shin, A., Holton, T., Sawaya, M. R., Cascio, D., and Eisenberg, D. (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
286. Conrad, W. H., Osman, M. M., Shanahan, J. K., Chu, F., Takaki, K. K., Cameron, J., Hopkinson-Woolley, D., Brosch, R., and Ramakrishnan, L. (2017) Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions. *Proc Natl Acad Sci U S A* **114**, 1371-1376



## Vita

### Zachary Alan Williamson

#### *Education*

- 2015-present University of Kentucky, College of Medicine  
Lexington, KY  
**PhD Candidate** in Molecular and Cellular Biochemistry
- 2011-2015 State University of New York College at Oneonta  
Oneonta, NY  
**BS** in Chemistry  
**BS** in Biology

#### *Research and Relevant Work Experience*

- 2015-present **Graduate Research Assistant.**  
Lab of Konstantin V. Korotkov, PhD  
Department of Molecular and Cellular Biochemistry  
University of Kentucky, College of Medicine
- 2013-2015 **Undergraduate Research Assistant.**  
Lab of Vicky Lentz, PhD  
Department of Biology  
State University of New York College at Oneonta

#### Publications

1. **Williamson, Z.A.**, Evans, T.J., Korotkov, K.V. "Structural plasticity of EspG<sub>5</sub> accommodates variety in ESX-5-specific PPE proteins." In preparation.
2. Smith, C.N., Wei, M., **Williamson, Z.A.**, Chernyavskaya, Y., Chow, K.M., Hersh, L.B., Korotkov, K.V., Blackburn, J.S. "Development and validation of nanobodies specific to the oncogenic phosphatase Protein Tyrosine Phosphatase 4A3 (PTP4A3 or PRL-3)." bioRxiv. 2020.10.311787 (2020). In preparation for J Biol Chem.
3. **Williamson, Z.A.**, Chaton, C.T., Ciocca, W.A., Korotkova, N., Korotkov, K.V. "PE5-PPE4-EspG<sub>3</sub> heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems." J Biol Chem. 295, 36 (2020).
4. Sikora, A.E., Mills, R.H., Weber, J.V., Hamza, A., Passow, B.W., Romaine, A., **Williamson, Z.A.**, Reed, R.W., Zielke, R.A., Korotkov, K.V. "Peptide Inhibitors Targeting the Neisseria gonorrhoeae Pivotal Anaerobic Respiration Factor AniA." Antimicrob Agents and Chemother. 61, 8 (2017)

#### *Awards*

- 2019 Jun **Poster Presentation Award, Second Place.**  
Department of Molecular & Cellular Biochemistry Retreat.  
University of Kentucky. Lexington, KY.
- 2017 Oct **Fellowship for X-Ray Methods in Structural Biology.**  
National Institute of General Medical Sciences. Cold Spring Harbor, NY.

2017 Apr                   **Student Travel Award.**  
University of Kentucky Graduate School. Lexington, KY.

*Leadership*

2020 May                   **Student Invited Speaker Host.**  
Department of Molecular & Cellular Biochemistry  
University of Kentucky. Lexington, KY.

2018-2020               **Vice President.**  
Biomedical Graduate Student Organization.  
University of Kentucky College of Medicine. Lexington, KY.

2016-2018               **Social Committee Chair.**  
Biomedical Graduate Student Organization.  
University of Kentucky College of Medicine. Lexington, KY.