

IDENTIFICATION OF A CHAPERONE FOR THE SECA2 PROTEIN EXPORT PATHWAY
OF *MYCOBACTERIUM TUBERCULOSIS*

Brittany Kristine Miller

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Microbiology and Immunology in the School of Medicine.

Chapel Hill
2018

Approved by:

Miriam Braunstein

Peggy Cotter

Jeff Dangel

Virginia Miller

Rob Nicholas

© 2018
Brittany Kristine Miller
ALL RIGHTS RESERVED

ABSTRACT

Brittany Kristine Miller: Identification of a chaperone for the SecA2 protein export pathway of
Mycobacterium tuberculosis
(Under the direction of Miriam Braunstein)

The bacterial pathogen *Mycobacterium tuberculosis* is responsible for the disease tuberculosis. To promote disease, *M. tuberculosis* exports proteins from the cytoplasm to the bacterial cell surface or out into the host environment. Exported proteins are in an ideal location to manipulate the host. All bacteria, including mycobacteria, utilize the Sec export system for the bulk of protein export. The Sec system is composed of an ATPase protein, SecA, and a membrane channel complex, SecYEG. Mycobacteria, along with some Gram-positive bacteria, contain a second, functionally distinct paralog of the SecA protein. In mycobacteria, the SecA responsible for housekeeping export is called SecA1 and is essential for bacterial survival, while the second SecA is called SecA2 and exports a smaller subset of proteins and is important for *M. tuberculosis* virulence. The mechanism of SecA2-dependent export is not well understood. Past data support a model where the mycobacterial SecA2 export pathway is integrated into the housekeeping Sec pathway, and SecA2 shares use of the same SecYEG channel as SecA1 to export its substrates. Like SecA1, SecA2 requires ATPase activity to function. In this dissertation, we take the approach of characterizing suppressors of a *secA2* mutant allele to better understand the mechanism of SecA2-dependent export. Intragenic suppressor mutations map to the surface of SecA2 and help identify functional regions of SecA2 that may promote interactions with SecYEG, SecA2 substrates or other partners of SecA2. Extragenic suppressor

mutations map to a new component of the SecA2 pathway that we named SatS. In *M. tuberculosis*, SatS is required for the export of a subset of SecA2 substrates and for pathogenesis. SatS functions as a protein export chaperone that protects and promotes export of its specific substrates. Structural studies of SatS reveal a new fold combined with hydrophobic grooves representing potential sites of substrate binding. Taken together, the findings presented in this dissertation advance our understanding of the mechanism of the SecA2 export pathway and expand our appreciation of the diversity among chaperones by identifying SatS as a new type of protein export chaperone.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS AND SYMBOLS	xi
CHAPTER 1 Introduction.....	1
Spread and survival of <i>M. tuberculosis</i>	1
Protein export in <i>M. tuberculosis</i>	2
Protein export by the conserved Sec (SecA1) pathway	3
Mechanism of protein export by the SecA1 pathway	3
The SecA2 Protein Export Pathway.....	5
The SecA2 Exportome	7
Mechanism of protein export by the SecA2 pathway	11
Features of SecA2-dependent substrates	14
Protein export chaperones	16
Summary	32
REFERENCES	38
CHAPTER 2 Structural Similarities and Differences between Two Functionally Distinct SecA Proteins, <i>Mycobacterium tuberculosis</i> SecA1 and SecA2	47
Introduction	48
Materials and Methods	52
Results	56
Crystal structure of <i>Mtb</i> SecA2.	56
Broad structural similarity between <i>Mtb</i> SecA1 and <i>Mtb</i> SecA2.....	57

Differences between the structures of <i>Mtb</i> SecA1 and <i>Mtb</i> SecA2.	57
Similarities between SecYEG binding regions of <i>M. tuberculosis</i> SecA1 and SecA2.	59
Mapping of suppressor mutations on the SecA2 structure.	60
Intragenic suppressors alter membrane localization of the dominant-negative SecA2.	64
Discussion	64
REFERENCES	84
CHAPTER 3 <i>Mycobacterium tuberculosis</i> SatS is a chaperone for the SecA2 protein export pathway	90
Introduction	90
Materials and Methods	92
Results	104
<i>satS</i> suppressors of <i>secA2 K129R</i>	104
SatS is required for export of the SecA2-dependent SapM phosphatase	105
Mce proteins exported by the SecA2 pathway require SatS for their export	107
SatS is required for <i>M. tuberculosis</i> growth in macrophages.....	108
SatS effects on cellular levels of its substrates	109
SatS and SapM interact	110
SatS functions prior to SecA2	112
SatS behaves as a chaperone to prevent SapM aggregation.....	112
SatS has a new fold and hydrophobic grooves that share similarity with the preprotein binding sites of the SecB chaperone	113
SatS has at least two separable roles in protein export.....	114
Discussion	116
Suppressor analysis led to the identification of SatS.....	116
SatS as a protein export chaperone.....	117

The SatS structure defines a new fold with hydrophobic grooves typical of substrate binding sites.....	119
SatS is required for growth of <i>M. tuberculosis</i> in macrophages	120
Conclusion.....	120
REFERENCES.....	143
CHAPTER 4 Discussion.....	147
SecA2 K129R: tool for studying the SecA2 pathway	149
The SecA2 crystal structure paired with intragenic suppressors provide insight into important domains of SecA2.....	151
Structural comparison of SecA2 and SecA1	151
Suppressors with mutations in SecA2 map to specific regions	153
Extragenic suppressors lead to the identification of a new player of the SecA2 export pathway	157
SatS has many similarities with protein export chaperones	158
SatS alignments and structure reveal two highly similar domains.....	160
Differences in electrostatic surface potential of SatS _N and SatS _C	161
Hydrophobic grooves in SatS _C resemble SecB substrate binding sites	162
Consideration of SatS being a SecB counterpart for the SecA2 pathway.....	165
SatS forms a multimer	165
SatS is the founding member of a new class of protein export chaperones	166
What are the functions of SatS?	167
A SecA2-targeting role for SatS.....	168
SatS substrates beyond SapM	170
SatS is required for growth of <i>M. tuberculosis</i> in macrophages	171
Conclusions	172
REFERENCES	188

LIST OF FIGURES

Figure 1.1. Models of SecA1 and SecA2 export in <i>M. tuberculosis</i> .	34
Figure 1.2. Solute binding proteins and Mce proteins are exported by the SecA2 pathway.	35
Figure 1.3. Multiple components of Mce transporters are reduced in the cell wall of the $\Delta secA2$ mutant.	36
Figure 2.1. Domain architecture of <i>MtbSecA2</i> .	72
Figure 2.2. Comparison of <i>MtbSecA2</i> with <i>MtbSecA1</i> .	73
Figure 2.3. Comparison of different orientations of the PPXD domain.	74
Figure 2.4. Illustration of how deletion of the HWD in SecA2 makes the signal peptide binding cleft more solvent-exposed.	75
Figure 2.5. Conformation of the two-helix finger (2HF).	77
Figure 2.6. Interface residues (red) of <i>MtbSecA2</i> (tan) that would contact SecYEG (cyan), based on superposition with <i>TmSecA</i> .	78
Figure 2.7. Intragenic suppressors suppress the azide sensitivity phenotype of <i>secA2 K129R</i> .	80
Figure 2.8. Suppressor mutations from <i>MsSecA2 K129R</i> mapped.	82
Figure 2.9. Subcellular localization of SecA2 is altered in intragenic suppressors of <i>secA2 K129R</i> .	83
Figure 3.1. $\Delta satS$ mutant suppresses SecA2 K129R phenotypes.	126
Figure 3.2. Characterization of the <i>M. smegmatis</i> $\Delta satS$ mutant.	127
Figure 3.3. <i>satS</i> and <i>sapM</i> are co-transcribed.	128
Figure 3.4. Characterization of the <i>M. tuberculosis</i> $\Delta satS$ mutant.	129
Figure 3.5. SatS is required for the export of SapM and Mce proteins.	131
Figure 3.6. SatS contributes to <i>M. tuberculosis</i> virulence.	132
Figure 3.7. SatS effects cellular levels of its substrates.	133
Figure 3.8. SatS does not affect <i>sapM</i> transcription or translation.	134
Figure 3.9. Epitope tags do not disrupt SapM or SatS functions.	135

Figure 3.10. SatS and SapM interact	136
Figure 3.11. SatS functions prior to SecA2	137
Figure 3.12. SatS and SatS _C prevent aggregation of SapM <i>in vitro</i>	138
Figure 3.13. SatS _C secondary structure and the predicted secondary structure of SatS _N	140
Figure 3.14. SatS has a new fold and hydrophobic grooves that share similarity with the preprotein binding sites of the SecB chaperone.....	141
Figure 3.15. SatS has at least two separable roles in protein export.....	142
Figure 4.1. Adding back <i>satS</i> to the extragenic suppressor 7S	175
Figure 4.2. SecA2 orthologs with a truncated or missing HWD	176
Figure 4.3. Intragenic suppressor mutations of <i>M. tuberculosis</i> SecA2 K115R mapped onto <i>M. tuberculosis</i> SecA2 structure using Pymol.	177
Figure 4.4. SatS _N structure is predicted to be remarkably similar to SatS _C structure.....	179
Figure 4.5. SatS _N and SatS _C domains are separated by a variable linker.....	180
Figure 4.6. SatS _N domain can restore stability of SapM in a $\Delta satS$ mutant.	181
Figure 4.7. The N domain and C domain of SatS _{Mtb} display different surface characteristics.....	183
Figure 4.8. Hydrophobic grooves in SatS with candidate residues for mutagenesis to test for roles in substrate binding.	184
Figure 4.9. SatS forms an orientation-specific multimer in a bacterial two-hybrid assay.	185
Figure 4.10. SatS G134D affects a SecA2-specific role of SatS.	186
Figure 4.11. SatS and Mce4E interact in <i>M. smegmatis</i>	187

LIST OF TABLES

Table 1.1. Structural and functional comparisons between known protein export chaperones.....	37
Table 2.1. Data collection and refinement statistics for SAD (SeMet) structure of SecA2.....	71
Table 2.2. Conservation of the 2HF among SecA homologs	76
Table 2.3. Residues of <i>Mtb</i> SecA2 predicted to be in contact with SecY based on structural superposition with <i>Tm</i> SecA in complex with SecYEG	79
Table 2.4. Suppressor mutations observed in <i>M. smegmatis</i> SecA2 K129R.....	81
Table 3.1. Primers used in this study.	121
Table 3.2. Plasmids used in this study.	123
Table 3.3. <i>M. tuberculosis</i> and <i>M. smegmatis</i> strains used in this study.	125
Table 3.4. SatS _C X-ray Structure Validation Details.	139
Table 4.1. All suppressor mutations identified using <i>M. smegmatis</i> SecA2 K129R or <i>M. tuberculosis</i> SecA2 K115R	174
Table 4.2. Protein export chaperone features shared by SatS	178

LIST OF ABBREVIATIONS AND SYMBOLS

°C	degrees Celsius
μL	microliter
2D	Two-dimensional
AA	amino acid
ABC	ATP-binding cassette
ADS	Albumin Dextrose Saline
Ag	antigen
ANOVA	analysis of variance
ATP	adenosine triphosphate
BCG	bacillus Calmette-Guerin attenuated vaccine strain
B-ME	beta-mercaptoethanol
Bp	nucleotide base pairs
BSL-3	Biosafety level 3
<i>C.</i>	<i>Clostridium</i> or <i>Corynebacteria</i>
C-	carboxy terminus of a protein
cDNA	complementary DNA
CFP	culture filtrate protein
cfu	colony forming units
Cl	chlorine
Cs	cesium
CW	cell wall
D	aspartic acid

dATP	deoxyadenosine triphosphate
dH ₂ O	distilled deionized water
DNA	deoxyribonucleic acid
Dpi	days post infection
E	glutamic acid
<i>E.</i>	<i>Escherichia</i>
Ecc	ESX conserved component
Esp	ESX specific protein
ESX	ESAT secretion system
G	glycine
g	grams
HA	Influenza hemagglutinin epitope
His	Hexahistidine tag
HRP	horseradish peroxidase
Hyg	hygromycin
K	Lysine
Kan	kanamycin
kb	kilobase
kDa	kiloDalton
LB	Luria Bertani media
<i>M.</i>	<i>Mycobacterium</i>
Mce	mammalian cell entry
MDR	multi-drug resistant

MEM	membrane
Mg	magnesium
mg	milligram
min	minute
mL	milliliter
<i>msmeg</i>	<i>Mycobacterium smegmatis</i> gene designation
Msmeg	<i>Mycobacterium smegmatis</i> protein designation
N-	amino-terminus of a protein
Na	sodium
OD	optical density
ORF	open reading frame
p	plasmid
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PE	proline glutamic acid repeats containing protein
PPE	proline proline glutamic acid repeats containing protein
Pro	proline
qRT-PCR	quantitative reverse-transcriptase PCR
R	arginine
<i>rv</i>	<i>Mycobacterium tuberculosis</i> gene designation
Rv	<i>Mycobacterium tuberculosis</i> protein designation
RNA	ribonucleic acid

Rv	<i>M. tuberculosis</i> strain H37Rv virulent
SDS	sodium dodecyl sulfate
Sec	secretory pathway
SOL	soluble
SP	signal peptide
SPI	<i>Salmonella</i> pathogenicity island
SRP	signal recognition particle
T3SC	Type III secretion chaperone
T3SS	Type III secretion system
Tat	Twin-arginine translocation
TM	transmembrane domain
Tnseq	Transposon deep sequencing
TraSH	Transposon site hybridization
Tw	polyoxyethylene sorbitan monoleate
Tween	polyoxyethylene sorbitan monoleate
Ty	Tyloxapol
WCL	whole cell lysate
WT	wild type
x	variable amino acid
XDR	extensively drug resistant
Y	tyrosine
Δ	deletion

CHAPTER 1¹

Introduction

In 2016 alone, there were over 10 million new cases of tuberculosis (TB) and 1.8 million deaths, making *Mycobacterium tuberculosis*, the bacterium that causes TB, the leading cause of death from a single infectious disease worldwide (1). By current estimates, up to one-third of the world's population is infected with *M. tuberculosis*. Despite over 70 years since the first anti-mycobacterial drug was introduced, we are still struggling with effective treatments for TB. Further, the World Health Organization reported that in 2016 over 490,000 people worldwide developed multidrug-resistant TB (MDR-TB), which is resistant to standard treatment with the first-line antibiotics isoniazid and rifampicin (1). Treatment is still possible with second-line antibiotics; however, these treatment plans are more expensive, have severe side effects, take longer to complete than the standard course of treatment for TB, and only have a 54% success rate (1). A detailed molecular understanding of the mechanisms of *M. tuberculosis* pathogenesis could lead to novel therapeutics for combating *M. tuberculosis*.

Spread and survival of *M. tuberculosis*

M. tuberculosis is spread via aerosols created when a person with an active case of TB coughs or sneezes. When aerosols are inhaled by an uninfected person, *M. tuberculosis* can reach the alveolar spaces of the lung, where the bacterium is phagocytosed by alveolar macrophages.

¹ Adapted for this dissertation with permission from: Miller BK, Zulauf KE, Braunstein M. 2017. The Sec Pathways and Exportomes of *Mycobacterium tuberculosis*. Microbiology Spectrum 5(2) doi: 10.1128/microbiolspec.TBTB2-0013-2016.

Typically, phagocytosed bacteria are delivered to membrane-bound compartments called phagosomes. The host then mounts multiple anti-microbial defenses against the bacteria. In contrast, *M. tuberculosis* is able to block these host responses in order to create a hospitable niche for replicating in the phagosome. Consequently, the mechanisms by which *M. tuberculosis* survives inside macrophages are critical for virulence. In order to promote survival in the macrophage, *M. tuberculosis* exports a variety of effector proteins outside of the bacterial cell to the host-pathogen interface.

Protein export in *M. tuberculosis*

Approximately 20% of bacterial proteins have functions outside the bacterial cytoplasm (2). All bacteria possess protein export pathways that transport proteins made in the cytoplasm beyond the cytoplasmic membrane. These exported proteins may remain in the bacterial cell envelope or be further secreted to the extracellular environment. Many exported proteins function in essential physiological processes. Additionally, in bacterial pathogens, many exported proteins have functions in virulence. Consequently, the pathways that export proteins are commonly essential and/or are important for pathogenesis. Across bacteria, including mycobacteria, there are conserved protein export pathways: the general Sec secretion (Sec) and the twin-arginine translocation (Tat) pathways. Both Sec and Tat pathways are essential to the viability of *Mycobacterium tuberculosis* and both also contribute to virulence (3-5). In addition to these conserved pathways, bacterial pathogens commonly have specialized protein export systems that are important for pathogenesis due to their role in exporting virulence factors. Mycobacteria also have specialized protein export systems: the SecA2 export pathway and five ESX (Type VII) pathways.

Protein export by the conserved Sec (SecA1) pathway

Due to the large bulk of protein export carried out by the Sec pathway and the importance of exported proteins to physiological processes and host interactions in bacterial pathogens, the Sec pathway is important to both bacterial homeostasis and virulence (6). Of the 4,024 open reading frames (ORFs) in the *M. tuberculosis* genome, 997 ORFs are predicted to be exported out of the cytoplasm due to the presence of an N-terminal signal sequence and/or transmembrane domain that direct them to be exported by the Sec pathway (7, 8). In support of these bioinformatics predictions, our lab recently identified 593 of these proteins as being exported during *M. tuberculosis* infection of mice, using an *in vivo* selection strategy to identify exported fusion proteins (9). While the specific pathway responsible for the export of most of these proteins has not been directly investigated, the majority of them are likely exported by the Sec pathway since they possess conserved Sec export signals and only a small subset are exported by the specialized SecA2 export pathway (see below). The essentiality of components of the Sec export machinery presents challenges to comprehensively identify members of the conserved Sec exportome of *M. tuberculosis*. Essential proteins are required for bacterial viability, and the genes that code for them cannot be deleted without providing another copy of the gene. However, a dedicated effort to conduct such an analysis is warranted, particularly for understanding the contribution of the conserved Sec export pathway to virulence.

Mechanism of protein export by the SecA1 pathway

The Sec pathway transports proteins from the cytoplasm across the cytoplasmic membrane. Sec exported proteins can then remain in the cell envelope or be fully secreted into the extracellular space. Sec exported proteins are transported in an unfolded state through the SecYEG membrane channel (Figure 1.1). SecY is an essential integral membrane protein, which

forms the central core of the channel (10). SecE is an essential integral membrane protein that is suggested to stabilize the open SecY conformation before and during translocation (11). SecG is not essential, but improves the efficiency of translocation (12). Additional membrane bound components of the Sec pathway also serve to improve export efficiency: SecD, SecF, and YajC (13). In addition to fully exporting proteins across the cytoplasmic membrane, the Sec pathway is also involved in the delivery and insertion of integral membrane proteins into the cytoplasmic membrane. Complete export across the membrane by the Sec pathway is a post-translational process; whereas, the role of the Sec pathway in integral membrane protein localization is a co-translational process that involves the conserved, essential, ribonucleoprotein signal recognition particle (SRP) (14). SRP is composed of the Ffh protein bound to SRP RNA (15). SRP recognizes the transmembrane domain on a nascent integral membrane protein and then delivers it as a ribosome-mRNA-nascent protein complex to the SRP receptor FtsY (16). In turn, FtsY delivers the integral membrane protein to SecYEG, where a lateral gate in SecY allows passage of transmembrane domains into the cytoplasmic membrane (17). Further discussion of Sec pathway mechanism below focuses on post-translational export of fully synthesized proteins across the cytoplasmic membrane.

In addition to the Sec proteins discussed above, Sec export requires the essential SecA ATPase that is peripherally associated with SecYEG. SecA is a multifunctional protein that binds to Sec exported proteins in the cytoplasm, targets them to SecYEG, and harnesses energy from repeated rounds of ATP binding and hydrolysis to drive stepwise export of proteins through the SecYEG channel (18). An unusual feature of mycobacteria is that two SecA paralogs (named SecA1 and SecA2) exist (19). These two SecAs have unique functions, with SecA1 being the

essential SecA that functions in the conserved Sec pathway of mycobacteria to transport the majority of exported proteins.

Proteins exported by the Sec pathway have specific characteristics. Sec exported proteins are synthesized as preproteins with N-terminal signal sequences. The signal sequence has a positively charged N-terminus, a hydrophobic central domain, and an uncharged polar C-terminus containing a cleavage site (20). Additionally, lipoproteins exported by the Sec pathway contain a lipobox motif at the C-terminal end of the signal sequence with an invariant cysteine that is the site of lipid attachment (21). The signal sequence and portions of the mature domain are recognized by SecA. After export, the signal sequence is removed by one of two signal peptidases (the type I signal peptidase LepB or the type II signal peptidase LspA) on the extracytoplasmic side of the membrane to release the protein, which then folds into its mature confirmation (22).

Another feature of Sec-exported preproteins is that they must be in an unfolded conformation in order to be exported through SecY (23). Cytosolic chaperones, such as SecB in Gram negative bacteria, can help preproteins maintain an unfolded and translocation-competent state compatible with transport through SecYEG (24). SecB also has a role in delivering preproteins to SecA (25). However, not all preproteins require a SecB chaperone and Gram positive bacteria lack a SecB ortholog. While there is a SecB-like protein (Rv1957) in *M. tuberculosis*, the available data indicate Rv1957 is a chaperone for the HigBA toxin-antitoxin system and not a chaperone for Sec export (26).

The SecA2 Protein Export Pathway

The first report of a second SecA paralog in bacteria was in mycobacteria (19). However, two SecA proteins also exist in other actinomycetes and a small set of Gram positive bacteria

including *Listeria*, *Staphylococcus*, and some *Streptococci*. In bacteria with two SecA proteins, SecA1 is the name given to the SecA with higher sequence similarity to the well-studied SecA protein of *E. coli* and SecA2 refers to the second paralog of SecA. In mycobacteria, the two SecAs are demonstrated to have unique functions. Even when SecA2 is overexpressed, *secA1* cannot be deleted, indicating that SecA2 cannot substitute for SecA1 (19). Similarly, overexpression of SecA1 does not rescue the phenotypes of a $\Delta secA2$ deletion mutant. As mentioned above, the mycobacterial SecA1 is essential and functions in the conserved Sec pathway. In contrast to SecA1, the mycobacterial SecA2 is not essential for growth *in vitro* and it exports a smaller subset of proteins. However, SecA2 contributes to the virulence of the pathogenic mycobacteria *M. tuberculosis* and *M. marinum* in macrophage and animal models of infection (27-31). In other bacterial pathogens with SecA2 proteins, such as *L. monocytogenes*, *Staphylococcus aureus*, *Streptococcus gordonii*, *Streptococcus parasanguinis*, SecA2 is also nonessential but has a role in virulence (32-35).

SecA2 systems can be divided into two groups: those that only possess a single copy of *secY* and those that also include an accessory copy of *secY*. Systems with an accessory SecY (SecY2) are referred to as SecA2/Y2 systems. The best studied SecA2/Y2 systems are those of *S. gordonii* and *S. parasanguinis*. In SecA2/Y2 systems, a single, large, glycosylated substrate is exported, presumably through the SecY2 channel. The genes encoding SecA2, SecY2, and the substrate are arranged in a conserved locus along with the glycosylation factors that modify the substrate (34). However, there is only a single copy of *secY* in mycobacteria. Furthermore, the mycobacterial SecA2 pathway exports multiple, diverse substrates and is referred to as a SecA2-only or multi-substrate SecA2 system (36, 37). In contrast to the SecA2/Y2 systems, studies suggest that the SecA2 proteins of the SecA2-only systems of mycobacteria, *L. monocytogenes*,

and *Clostridium difficile* likely work with the canonical SecYEG channel for the export of their specific sets of SecA2-dependent proteins (38-40). We will first discuss some examples of proteins exported by the SecA2 pathway in mycobacteria, and then describe what is known about the mechanism of mycobacterial SecA2-dependent export.

The SecA2 Exportome

As indicated by the virulence defects of $\Delta secA2$ mutants of *M. tuberculosis* and *M. marinum*, the SecA2-dependent pathway exports proteins with roles in pathogenesis (27-29). Compared to SecA1, which is responsible for the majority of protein export in the bacterial cell, SecA2 appears to export only a small number of proteins. Early studies identified a small number of SecA2 exported proteins using comparative two-dimensional gel electrophoresis of cell wall or secreted proteins of *M. smegmatis* and *M. tuberculosis* (19, 41). More recent quantitative mass spectrometry analyses of cell wall and cell envelope fractions of $\Delta secA2$ mutants of *M. tuberculosis* and *M. marinum* identified additional SecA2 substrates (29, 42). While the mass spectrometry studies dramatically increase our knowledge of the SecA2 exportome, further studies are still required to validate the SecA2-dependency of many of the more recently identified proteins. Below, we highlight some of the validated examples of SecA2 substrates.

Solute binding proteins (SBPs)

The best studied SecA2 substrates, the Msmeg1704 and Msmeg1712 lipoproteins of *M. smegmatis*, represent one class of SecA2-dependent substrates, solute binding proteins (SBPs). SBPs are cell wall localized proteins that deliver solutes to permease components of ATP-binding cassette (ABC) transporters for import using energy from ATP hydrolysis (Figure 1.2).

Although *M. tuberculosis* does not have a direct homolog of Msmeg 1704 or Msmeg1712, quantitative mass spectrometry reveals numerous SBPs that are also SecA2 dependent in *M. marinum* and *M. tuberculosis* (29, 42). In *M. tuberculosis*, nearly all of the SBPs

identified in the cell wall fraction (13 out of 15) are present at lower levels in the $\Delta secA2$ mutant cell wall (42). All of the *M. tuberculosis* SecA2-dependent SBPs are lipoproteins with predicted signal sequences. Three SBPs with predicted Sec signal sequences were also identified as SecA2 dependent in *M. marinum* (29). The trend of SBPs being SecA2 dependent may also extend to other SecA2-only systems, as there is some evidence that SecA2 may also export SBPs in *Listeria* (32).

Although most SBPs in *M. tuberculosis* have not been directly studied and their substrates remain unknown, SBPs can import a wide range of solutes (43). Thus, the role of SecA2 in exporting SBPs could be important for nutrient acquisition and affect the ability of *M. tuberculosis* to thrive in the host.

Mce transporters

Another class of proteins exported by SecA2 is Mce transporter components. Mce transporters are importer complexes that are thought to function similarly to ABC transporters, in that they recognize an extracytoplasmic substrate (in this case a lipid) and import it into the cytoplasm using ATP hydrolysis (44). Mce transporters are composed of two YrbE membrane proteins forming a putative permease, five exported Mce proteins, one exported Mce lipoprotein and in some cases, one or more Mam transmembrane proteins (Figure 1.2). All of these components are exported proteins that either possess a transmembrane domain or a signal sequence. The five Mce proteins and the Mce lipoprotein are proposed to recognize the lipid substrate and deliver it to the permease and, therefore, are functionally similar to SBPs of ABC transporters (44). Supporting this speculated similarity, there are presumed lipid importing SBPs of Gram negative bacteria that possess Mce-like domains (45). *M. tuberculosis* has four Mce transporters. The best studied Mce transporter is Mce4, which imports cholesterol (46). Because cholesterol catabolism is critical to *M. tuberculosis* pathogenesis, Mce4 has an important role in

virulence (46). Furthermore, studies in mice suggest that Mce4 is required for *M. tuberculosis* persistence during chronic infection (46, 47). Mce1 is proposed to be a mycolic acid re-importer (48, 49). Mce1 is required for optimal growth in mice and in macrophages; however, there are also conflicting reports concerning these phenotypes (48, 50-53).

Multiple exported components of Mce1 and Mce4 transporters are identified by quantitative mass spectrometry as being SecA2 dependent in *M. tuberculosis*. Six components of Mce1 and six components of Mce4 are significantly reduced in the $\Delta secA2$ mutant cell wall (42). Although Mce2 and Mce3 transporter components were not detected in the *M. tuberculosis* mass spectrometry dataset, Mce3E, along with Mce4D, was identified as reduced in the $\Delta secA2$ mutant of *M. marinum* (Figure 1.3) (29, 42). This suggests SecA2 may also export Mce2 and Mce3 system components, in addition to Mce1 and Mce4. Levels of MceG, the presumed ATPase for Mce transporters, are also reduced in the $\Delta secA2$ mutant of *M. tuberculosis* (42). While MceG is predicted to be cytoplasmic, MceG levels are shown to depend on the presence of other Mce transporter components (52). Therefore, the reduction in MceG may be a consequence of reduced export of Mce transporter components in the $\Delta secA2$ mutant. Collectively, these data suggest a link between SecA2 and lipid transport in mycobacteria.

It is striking that SecA2 impacts multiple components of Mce transporters. This result presents two possibilities. First, SecA2 may export multiple components of Mce transporters. Second, SecA2 may only export one or a small number of proteins that make up the Mce transporter complex, but when that protein(s) is not exported (i.e. in a $\Delta secA2$ mutant) the entire Mce complex may fail to assemble and/or be destabilized. Because of the proposed similarity with SBPs, our identification of Mce proteins and Mce lipoproteins as SecA2-dependent is intriguing in terms of common features of proteins in the SecA2 exportome (Figure 1.2). As Mce

systems are important for virulence, the role of SecA2 in the export of Mce transporters may also contribute to the pathogenesis of *M. tuberculosis*. More specifically, because of the impact on the Mce4 transporter, SecA2 may be required for cholesterol catabolism during infection, which could translate to a role for SecA2 in persistence.

PknG

The eukaryotic-like serine-threonine kinase PknG is also dependent upon SecA2 for export (29, 42). Recent studies in *M. tuberculosis* and *M. marinum* identify reduced levels of PknG in the cell wall of the $\Delta secA2$ mutant compared to wild type or complemented strains (29, 42). PknG lacks a signal sequence, yet is found in both the mycobacterial cytoplasm and its exported location (54, 55). Intriguingly, *pknG* is transcribed in a proposed operon downstream of the gene for the SecA2-dependent SBP *glnH*. The connection, if any, between these two SecA2 substrates remains to be investigated.

PknG functions in both mycobacterial metabolism and pathogenesis and an *M. tuberculosis* $\Delta pknG$ mutant has an *in vitro* growth defect and is attenuated in mice (54). In addition to regulating glutamate metabolism and redox homeostasis in the mycobacterial cytoplasm, during *M. tuberculosis* infection of macrophages, exported PknG is detected in the host cell cytosol and is considered a virulence effector (54, 56, 57). A hallmark of macrophage infections with *M. tuberculosis* is that the bacterium arrests the normal process of phagosome maturation (58, 59). Consequently, *M. tuberculosis* avoids delivery to a mature, fully acidified phagolysosome (58, 59). Although the mechanistic details remain unclear, PknG plays a role in phagosome maturation arrest. A $\Delta pknG$ mutant of BCG localizes to mature phagosomes, and a strain of *M. smegmatis* expressing the BCG *pknG* prevents phagosome-lysosome fusion unlike wild type *M. smegmatis* (56). This specific role of PknG in phagosome maturation is intriguing

because studies of $\Delta secA2$ mutants of *M. tuberculosis* and *M. marinum* demonstrate a role of the SecA2 pathway in phagosome maturation arrest (28, 29).

SapM

The most recently validated SecA2-dependent substrate in *M. tuberculosis* is the secreted phosphatase SapM. Like PknG, SapM has a demonstrated role in arresting phagosome maturation and preventing delivery of *M. tuberculosis*-containing phagosomes to phagolysosomes (60-62). SapM has a Sec signal sequence and is fully secreted out of the bacterium *in vitro* and *in vivo* (9, 62). One role of SapM is to dephosphorylate Phosphatidylinositol-3 phosphate (PI3P) on the *M. tuberculosis*-containing phagosomal membrane, which limits recruitment of PI3P binding proteins, such as Early endosome antigen 1 (EEA1) that promote downstream phagosome maturation (60). Consequently, the $\Delta secA2$ mutant fails to prevent EEA1 recruitment to *M. tuberculosis*-containing phagosomes (62), which contributes to the $\Delta secA2$ mutant's inability to arrest phagosome maturation. A $\Delta sapM$ mutant of *M. tuberculosis* is deficient for growth in macrophages and is attenuated in a guinea pig model of infection (61).

Mechanism of protein export by the SecA2 pathway

Studies of the mechanism of SecA2-dependent export in mycobacteria have been conducted in *M. tuberculosis* and *M. marinum* as well as the non-pathogenic model organism *M. smegmatis*. Cross-species complementation experiments show that the *secA2* of *M. tuberculosis* and *M. smegmatis* can substitute for one another to complement $\Delta secA2$ deletion mutant phenotypes, indicating that SecA2 is functionally conserved between *M. smegmatis* and *M. tuberculosis* (63).

Although it is understood that *M. tuberculosis* SecA1 and SecA2 are functionally distinct, the differences that contribute to their individual roles remain unknown. *M. tuberculosis* SecA1 is more closely related to SecA proteins of other bacteria than it is to *M. tuberculosis* SecA2. When aligned to the well-studied SecA proteins of *E. coli* and *Bacillus subtilis*, *M. tuberculosis* SecA1 shares 48% and 49% sequence identity respectively. *M. tuberculosis* SecA2 shares 36% sequence identity to *E. coli* SecA and 37% sequence identity to *B. subtilis* SecA. Additionally, SecA2 shares only 38% amino acid sequence identity with SecA1 and is approximately 20 kDa smaller. Most of the annotated domains from *E. coli* SecA are conserved in SecA1 and SecA2, and both have demonstrated ATPase activity (64). The difference in size between SecA1 and SecA2 is predicted to come from the lack of a helical wing domain (HWD) in SecA2. The function of the HWD is unknown; however, the HWD undergoes conformational changes in response to substrate binding and may be involved in signal sequence recognition and binding (65, 66). One distinct difference between SecA1 and SecA2 is their subcellular localization. Like *E. coli* SecA, SecA1 is evenly distributed between the cell membrane and cytoplasm (63). Alternatively, SecA2 is predominantly cytoplasmic, reinforcing that the roles of SecA1 and SecA2 in protein export differ from one another.

ATP binding is necessary for SecA2 function in *M. tuberculosis* and *M. smegmatis* as shown with a SecA2 variant harboring an amino acid substitution in the ATP-binding site of *M. smegmatis* SecA2 (K129R) and *M. tuberculosis* SecA2 (K115R). K129 and K115 are corresponding residues in *M. smegmatis* SecA2 and *M. tuberculosis* SecA2, respectively. Both *M. tuberculosis* SecA2 K115R and *M. smegmatis* SecA2 K129R mutants cannot bind ATP, and are defective in protein export (63, 64). In *M. tuberculosis*, SecA2 K115R fails to carry out the SecA2 role in promoting growth in macrophages (64). In *M. smegmatis*, SecA2 K129R is

dominant negative and is associated with phenotypes that are more severe than those exhibited by a *secA2* null mutant. These results are indicative of SecA2 K129R being locked in a nonfunctional complex with SecA2-interacting proteins. Expression of *secA2 K129R* also results in reduced levels of the sole SecY of mycobacteria and increased SecY levels suppress the severe phenotypes of the *M. smegmatis secA2 K129R* mutant (38). This suggests that SecA2 K129R is locked in a non-productive interaction with SecYEG and that increased SecY levels overcome the effect of this dominant negative interaction. These data also suggest that wild type SecA2 uses the canonical SecY channel to export its own set of substrates (Figure 1.1). In the SecA2-only system of *L. monocytogenes*, suppressor mutations of a *secA2* mutant also map to *secY*, which is consistent with SecA2 working with SecY in other SecA2-only systems as well (39). In *M. smegmatis*, we identified additional suppressor mutations that alleviated the detrimental SecA2 K129R phenotypes without directly increasing SecY levels (38). The causes of suppression in these strains are investigated in subsequent chapters of this dissertation.

Along with SecY, SecA1 may also be important for SecA2-dependent export. Numerous reports demonstrate the ability of SecA proteins to dimerize, and a recent study demonstrates heterodimer formation (as well as homodimer formation) *in vitro* of recombinant *M. tuberculosis* SecA1 and SecA2 purified out of *E. coli* (67, 68). However, it remains to be demonstrated if SecA1-SecA2 heterodimers exist and are biologically relevant in mycobacteria. Further, the oligomeric state of SecA during protein export is controversial. SecA is either suggested to function as a monomer or as a dimer during translocation (67). However, another finding in support of a role of SecA1 in the SecA2 pathway is that export of SecA2-dependent substrates in *M. smegmatis* is compromised when SecA1 is depleted, although it is also possible that this result could reflect a role for SecA1 in assembling the SecYEG channel in the membrane (63).

As current data supports a model where SecA2 works with the conserved SecYEG channel, and possibly SecA1 as well, to export its specific subset of preproteins, an outstanding question remaining to be answered is how SecA2 recognizes and exports its unique set of substrates but not the larger population of SecA1 preproteins. Recent observations suggest that ADP binding induces conformational changes of *M. tuberculosis* SecA2 that are not observed in *M. tuberculosis* SecA1 or *E. coli* SecA (69). While the biological significance of these structural changes are unknown, it is hypothesized that the conformational changes result in the closure of a clamp in SecA2 that is thought to bind the mature domain of the preprotein (69). It is possible that closure of the clamp prevents recognition of SecA1 substrates by ADP-bound SecA2. In this scenario, SecA2 binding to an unknown factor or a SecA2-dependent substrate could promote ADP release and open the clamp to enable SecA2 activity. Additionally, the absence of the helical wing domain (HWD) in SecA2 leaves the signal sequence binding cleft more open and solvent-exposed (70). This structural difference could help SecA2 recognize its unique substrates and exclude SecA1 substrates. An improved model of how SecA1 and SecA2 cooperate for export through the SecYEG channel, and identification of additional proteins that work with SecA2, are important next steps in understanding the distinctive functions of the SecA2 protein export pathway of mycobacteria.

Features of SecA2-dependent substrates

As described in the previous section, studies in *M. tuberculosis*, *M. marinum*, and *M. smegmatis* have identified proteins that are exported in a SecA2-dependent manner (29, 41, 42). The list of SecA2-dependent proteins includes examples with N-terminal Sec signal sequences as well as proteins lacking recognizable signal sequences (e.g. PknG). Interestingly, the SecA2-only

systems of other pathogens, such as *L. monocytogenes*, are also associated with export of proteins with or without signal sequences (71).

The best studied SecA2-dependent substrates are the *M. smegmatis* SBP lipoproteins Msmeg1704 and Msmeg1712 (41). The lipidated nature of these proteins does not confer their SecA2-dependency for export; an amino acid substitution of the invariant cysteine in the lipoboxes of Msmeg1704 and Msmeg1712, which prevents lipid attachment, does not eliminate the requirement for SecA2 in export (72). Further, although Msmeg1704 and Msmeg1712 require signal sequences to be exported, their signal sequences do not contain any distinguishing features targeting them to SecA2 for export (72). If their signal sequences are swapped for the signal sequence of a SecA1-dependent substrate, Msmeg1704 and Msmeg1712 retain their SecA2-dependency. Thus, the mature domains of these proteins, not signal sequences, impart the requirement for SecA2 in their export.

One possible defining feature of the mature domains of SecA2-dependent substrates is a propensity to fold in the cytoplasm prior to export. In support of this idea, when the Sec signal sequence of Msmeg1704 is exchanged for a signal sequence that directs preproteins for export through the twin-arginine translocase (Tat) pathway, Msmeg1704 is exported by the Tat pathway (72). The Tat pathway differs from the Sec system in that it requires preproteins to be folded prior to export. Thus, the fact that the mature domain of a SecA2 substrate is compatible with export by the Tat pathway suggests that SecA2-dependent substrates can fold in the cytoplasm prior to export. Furthermore, it suggests that the role of SecA2 may be to facilitate export of these problematic substrates through the SecYEG channel, which requires proteins to be unfolded. It remains unclear how SecA2 may assist in the export of these substrates. Three nonexclusive possibilities for how SecA2 could be influencing export are that SecA2 serves a

chaperone-like role of keeping preproteins unfolded prior to export, that SecA2 is capable of working with chaperones that keep preproteins unfolded prior to export, or that SecA2 cooperates with SecA1 to provide additional energy to translocate substrates that have a tendency to fold through SecYEG.

SecA2-dependent features of the mature domain may also help explain the SecA2-dependence of *M. tuberculosis* proteins lacking signal sequences (e.g. SodA and PknG). SodA proteins lacking signal sequences are reported to be exported by the canonical Sec pathway in *Rhizobium leguminosarum* and by the SecA2-only pathway in *L. monocytogenes* (73, 74). Further studies are needed for these unconventional exported proteins lacking signal sequences, as it also remains possible that the effect of SecA2 on exported proteins like SodA is indirect. There may be unidentified signal sequence-containing proteins exported by SecA2, that are themselves components of a different specialized export machinery, responsible for secreting unconventional proteins such as SodA.

Protein export chaperones

SecA2 exports a variety of proteins with a wide range of functions, some of which contribute to *M. tuberculosis* virulence. Despite progress in our understanding of the SecA2 protein export pathway, it remains unclear what distinguishes its substrates from those of the canonical Sec (SecA1-dependent) pathway. Prior to the studies presented in this dissertation, the only known difference between the canonical Sec export pathway and the SecA2 pathway were the ATPases, SecA1 and SecA2. We hypothesized that there were additional proteins involved in SecA2-dependent protein export that distinguish it from the SecA1 pathway. As described in chapter 3, we used a suppressor screen of the SecA2 K129R allele to identify one such protein, which we then named SatS and characterized as a protein export chaperone involved in the

export of some SecA2-dependent proteins. SatS is the first protein export chaperone identified for any SecA2 system.

Nearly all export systems utilize cytoplasmic chaperones for at least some of their exported substrates. In general, protein export chaperones maintain their substrates in an export competent state and target them to the appropriate export system. A comparison of some of the best studied export chaperones reveals key similarities and differences in their function, substrate recognition, and structure.

Molecular chaperones

To guard against inappropriate protein interactions, all organisms produce molecular chaperones that promote proper protein folding and prevent protein misfolding and aggregation. The best characterized molecular chaperones in prokaryotes, DnaJ/K and GroEL, are primarily responsible for performing these tasks in the bacterial cytoplasm (75). As a subset of molecular chaperones, protein export chaperones also prevent inappropriate interactions of their substrates prior to their export from the cytoplasm. Inappropriate interactions, if permitted to occur, prohibit the substrate from being exported. Protein export chaperones accomplish this task through roles such as preventing substrate folding, unfolding substrates, and preventing substrate aggregation or degradation. As an additional function, protein export chaperones contribute to targeting their substrates to the appropriate export machinery through interactions with components of the export apparatus. This targeting function helps to distinguish protein export chaperones from other molecular chaperones.

SecB

The best studied protein export chaperone is SecB of the canonical Sec export pathway. While the essential components of the Sec pathway including SecY and SecA are found in all bacteria, the cytoplasmic SecB only exists in Gram-negative bacteria and is a non-essential component of the pathway since SecB is only required for the export of a subset of Sec-exported proteins. A few well-studied examples of SecB substrates include maltose binding protein (MBP), outer membrane protein A (OmpA), and maltoporin (LamB) (76). Proteins exported through the Sec export system must be kept unfolded prior to and during export. Through transient interactions with substrates, SecB prevents inappropriate interactions by keeping its substrates in an unfolded, export competent state. In the absence of SecB, these substrates accumulate in the cytoplasm as folded precursors (77). SecB also functions to target its substrates to the Sec translocon through SecB interactions with SecA.

SecB function

Ultimately, SecB functions to ensure that its substrates properly engage SecYEG, while in an unfolded state for export. To accomplish this goal, SecB first inhibits substrate intracellular folding and/or aggregation. In an *E. coli* $\Delta secB$ mutant, signal sequence-containing MBP (pre-MBP) folds in the cytoplasm into a protease resistant, native conformation, which is not exported (77). Further, urea-treated purified pre-MBP folds quickly into its native state upon removal of urea. However, in the presence of SecB, pre-MBP is maintained in an unfolded conformation bound to SecB (78). SecB cannot prevent folding of the mature (periplasmic) form of MBP, which lacks the amino terminal signal sequence (78) because mature MBP folds much faster than pre-MBP. There are also cases of SecB-dependent substrates that misfold in the cytoplasm and form aggregates in the absence of *secB*, which are also not exported (79).

In addition to preventing folding of its substrates, SecB performs an additional role as a targeting protein. SecB delivers substrates to the Sec-translocon for export through the cytoplasmic membrane through its interactions with SecA. SecB affinity for SecA is greatly increased if the SecA is bound to SecYEG, encouraging SecB:substrate complexes to be targeted to an active translocon instead of cytoplasmic SecA (80). Once the substrate has been passed onto SecA for export and released from SecB, SecB remains in the cytoplasm, where it can engage another substrate.

SecB substrates

Of the hundreds of Sec substrates in *E. coli*, there are only 22 known SecB-dependent substrates (81). There is no specific sequence that distinguishes SecB substrates from the larger pool of Sec-exported proteins. Moreover, the genes encoding SecB substrates are distributed throughout the genome, rather than being clustered near *secB*. The commonality among SecB substrates is that they require interactions with SecB to avoid folding in the cytoplasm. The need for SecB in protein export is decreased if the substrate is mutated in order to fold sufficiently slowly (82); this result is consistent with the role of SecB being to keep substrates unfolded in the cytoplasm. The signal sequences of SecB substrates are indistinguishable from the signal sequences of SecB-independent substrates, and signal sequences are not directly recognized by SecB. However, the signal sequences of SecB substrates can play a role in SecB recognition by slowing the rate of folding of the mature domain of the preprotein to give SecB more opportunity to bind (83). Substrates must be unfolded to bind to SecB, and SecB will even bind to proteins that are not its natural ligands if they are unfolded (84). Unfolded substrates wrap around SecB to simultaneously occupy several low-specificity binding sites. This binding to SecB disrupts secondary structure in the substrate and prevents the substrate from folding.

SecB structure

SecB is a small (17.5 kDa) acidic (pI 4.1) protein that forms a homotetramer. Structural studies with and without bound substrates reveal that the tetramer is arranged as a dimer of dimers (78, 85). SecB monomers contain two long, continuous hydrophobic grooves that make up the primary binding sites for unfolded substrates. When bound to SecB, the mature domain of the substrate wraps around the SecB tetramer and ~9-12 amino acid long hydrophobic stretches of the substrate come in contact with the substrate-binding grooves on SecB (86, 87). There is no obvious consensus motif in the sequence of the substrates' SecB-binding stretches. Instead, peptide studies show SecB has a binding preference for sequences of 8 or 9 aminoacyl residues that are enriched in aromatic and basic amino acids (76). In SecB substrates, these binding-compatible sequences occur frequently, allowing the substrates to bind to SecB multiple times and prevent their folding.

Through defined mutagenesis, the hydrophobic grooves on SecB surface have proven to be critical for substrate binding (78). SecB likely binds its substrate with a stoichiometry of one substrate per SecB tetramer. It is interesting to note that in the course of its function, SecB displays two transient modes of binding. In its role preventing substrate folding, SecB binds to its substrates using multiple hydrophobic grooves on SecB, causing the unfolded substrates to wrap around the surface of SecB. Conversely, when SecB binds SecA in its targeting role, the interaction occurs at two specific sites on SecB. The hydrophobic grooves of SecB that interact with its substrates overlap with one of the SecA interaction sites, suggesting a mechanism for substrate transfer from SecB to SecA and from there to the SecYEG translocon (88, 89). The primary region on SecA that interacts with SecB is a small C-terminal domain (CTD). Of note,

mycobacteria SecA1 and SecA2 lack a CTD. A summary of the structural and functional details of SecB are found in Table 1.1.

Type III Secretion Chaperones (T3SC)

The Type III Secretion System (T3SS) is a mechanism used by many Gram-negative bacterial pathogens to facilitate host-pathogen interactions. The T3SS delivers bacterial effectors from their site of synthesis in the cytoplasm directly into the host cell where those effectors can manipulate host pathways including immune responses, cytoskeletal dynamics, and signal transduction pathways (90). T3SSs are needle-like secretion systems, composed of ~25 proteins that assemble together in a stepwise manner to form three components: (i) the extracellular needle filament, (ii) a basal body composed of inner and outer membrane-spanning rings, and (iii) an ATPase-containing sorting platform at the cytoplasmic face of the basal body (90). Once the T3SS is assembled, it remains ready for secretion until the bacterium senses contact with a host cell, triggering secretion of the effector substrates through the needle. The needle channel is too narrow to accommodate folded proteins; hence, the T3SS, like the Sec system, requires effectors to be unfolded for secretion. Secretion of the effector substrates, as well as export of the needle filament and basal body, requires specific cytoplasmic chaperones. The best studied T3SS chaperones (T3SCs) are those that function in exporting the T3SS effector substrates, referred to as Class I chaperones. Class I chaperones are further classified based on their specificity for just one cognate effector substrate (class Ia) or multiple effector substrates (class Ib). T3SCs form highly specific, transient interactions with their substrates to promote their secretion through the T3SS. T3SCs perform many roles, including stabilizing their substrates prior to secretion by preventing inappropriate interactions, targeting substrates to the appropriate T3SS, and orchestrating a hierarchy of secretion through the T3SS needle.

T3SC function

Ultimately, T3SCs function to ensure that their effector substrates properly engage the correct T3S apparatus while in a partially unfolded state for export. To accomplish this goal, T3SCs form transient interactions with their partially-unfolded effector substrates in the cytoplasm. In most cases, in the absence of their cognate T3SC, effector substrates are unstable and are rapidly degraded in the cytoplasm. Thus, one function of T3SCs is to prevent or protect their substrates from inappropriate interactions that would lead to their degradation. For example, in the absence of the enteropathogenic *E. coli* class 1b T3SC CesT, its primary substrate Tir is no longer secreted, and is present at much lower levels in the cytoplasm (91). There are also examples of aggregation-prone substrates such as the *Yersinia* effector YopO, which accumulates as insoluble aggregates in the cytoplasm in the absence of its cognate chaperone SycO (92). Binding of a T3SC to its effector masks an aggregation or degradation-prone region of the effector substrate and protects it in the cytoplasm. This binding also keeps the substrate partially unfolded and prepared for export.

T3SCs also function in targeting their effectors to the correct T3SS via interactions with the cytoplasmic T3SS ATPase, which then triggers chaperone-substrate disassembly and substrate unfolding to facilitate substrate entry into the T3S machinery. Because a single genome can encode multiple independent T3SSs, mechanisms must be in place to ensure that effectors are secreted through the appropriate system. Chaperone binding to the substrate ensures successful targeting of the substrate for secretion through the correct T3SS pathway. T3SS specificity is conferred through associations of chaperone-effector complexes with the ATPase on the cytoplasmic face of the T3S machinery (93, 94). T3SS ATPases form a hexameric ring-like structure that sit at the base of the T3S machinery peripherally associated with the inner

membrane (95). One function of the T3SS ATPase is to take the partially unfolded effector substrates delivered to them by T3SCs and unfold the remainder of the substrate prior to export through the T3S needle. In *S. enterica* Typhimurium, ATP hydrolysis activity of the InvC ATPase causes the SptP effector to disassociate from its chaperone SicP and fully unfold (96).

T3SCs may also orchestrate a hierarchal secretion among the multiple substrates being secreted through a single T3SS. The role of T3SCs in establishing a substrate hierarchy is still understudied; however, chaperone-bound substrates have a competitive advantage for secretion over chaperone-less substrates (97, 98). Further, multi-effector, class 1b, T3SCs may have mechanisms to establish a hierarchy of secretion among their multiple effectors. The multi-effector T3SC CesT binds some substrates with higher affinity than others. CesT interacts with at least 10 effectors, most of which require CesT for translocation into host cells (99). However, its primary substrate is the translocated intimin receptor (Tir). Tir secretion occurs prior to secretion of other effectors (100). Recently, two CesT-binding sites were identified in Tir as well as three other CesT-dependent, highly translocated effectors (101). The presence of two CesT-binding sites in these effector substrates may contribute to differential effector recognition and preference by CesT. Once the substrate has been passed onto the T3SS and released from the T3SC, the T3SC remains in the cytoplasm, where it can engage another substrate.

T3SC substrates

T3S effector substrates do not have Sec signal sequences, but contain non-cleavable, N-terminal, 15-20 residue long signal sequences (102). Adjacent to the signal sequence is often a Chaperone Binding Domain (CBD) that encompasses the next 50-100 residues of the substrate. While the N-terminal signal sequence is sufficient for protein secretion (102), the CBD directs

substrates to the correct T3SS through its interactions with the cognate chaperone (103, 104). When chaperone binding is prevented by deletion of their CBDs, the *Salmonella enterica* Typhimurium effectors SptP and SopE cannot bind their respective chaperones and are not targeted for secretion through their cognate T3SS (104). Instead, SptP and SopE are secreted through the non-cognate, flagellar export pathway. Not all T3SS effectors require a chaperone for export, and chaperone-less effectors do not have a CBD (98). When bound to its cognate T3SC the CBD of a substrate is unfolded. However, unlike SecB chaperone binding, the remainder of the T3SC substrate is usually folded and even retains enzymatic activity (98, 105).

A striking feature of T3S effector substrates and their cognate chaperones is the colocalization of their genes. Particularly for class 1a T3SCs, the gene encoding the T3SC is located adjacent to the gene of its target effector substrate(s), linking chaperone and substrate expression in the bacterial cytoplasm. In the case of CesT, a multi-substrate T3SC, the gene products for CesT and its primary substrate Tir are adjacent to one another, while the gene products for the rest of its substrates are elsewhere in the genome.

T3SC structure

T3SCs are all small (less than 20 kDa), acidic (pI 4-4.5) proteins. By and large, class Ia and Ib T3SC structures are remarkably similar despite their lack of amino acid sequence conservation (105). Tertiary structures of Class I T3SCs alone and in complex with cognate substrates reveal that T3SCs function as homodimers. T3SCs bind their substrates with a stoichiometry of one substrate per T3SC dimer. A few CBDs have been co-crystallized in complex with their cognate T3SCs, revealing that the CBD wraps around a chaperone dimer as a non-globular (unfolded) polypeptide, interacting with hydrophobic patches on the chaperone

surface. Like SecB, the T3SC structure does not depend on the presence of an effector; in the absence of a substrate, T3SCs maintain a dimeric state with similar structural folds and a surface dominated by electronegative charges with hydrophobic patches (105). These hydrophobic patches are required for T3SC:substrate interactions. Through defined mutagenesis, the hydrophobic patches on the T3SC surface have proven to be critical for substrate binding (106). In contrast, electronegative residues do not necessarily affect substrate binding, and they may be more important for recognizing a component of the T3SS machinery, such as the system ATPase. The highly conserved T3SC fold and substrate recognition pattern raise the question of how T3SCs are substrate-specific. It's possible that differences in the surface distribution of hydrophobic patches may be a determining factor in chaperone-substrate binding specificity. Although the crystal structure of a T3SC in complex with its ATPase has not been solved, T3SCs with mutations in their extreme C-terminus retain their ability to bind substrates, but are still defective for substrate secretion, suggesting that interactions with the ATPase may occur in the C terminus of the T3SC (95, 107). A summary of the structural and functional details of T3SCs are found in Table 1.1.

EspG

ESX or type VII secretion systems (T7SS) are specialized secretion systems present in mycobacteria and a small subset of Gram-positive bacteria including *L. monocytogenes*, *S. aureus*, and *Bacillus anthracis* (108). Pathogenic mycobacteria have up to five homologous T7SSs, called ESX-1 through ESX-5. Three of the five ESX systems in *M. tuberculosis* are important for virulence, with roles including immune evasion and iron acquisition (109). The known components of the ESX membrane channel likely only span the mycobacterial inner membrane, and the mechanism of ESX substrate transport across the mycobacterial outer

membrane remains unknown (110). ESX secretion systems are encoded in operons typically containing ~5 structural components of the secretion machinery called Ecc proteins, ESX substrates, and the protein export chaperone EspG; additional substrates are encoded outside of the ESX operons. The EspG proteins of different ESX systems are distinguished by a subscript number; for example, EspG₅ is encoded in the ESX-5 operon and works specifically with this system.

There are three families of substrates secreted by the five ESX systems in mycobacteria: Esx, PE/PPE and Esp proteins. All Esx proteins and many PE/PPE and Esp substrates are co-transcribed with a partner substrate. Several substrate pairs in each of these categories have been shown to form heterodimers in the cytoplasm and are predicted to be secreted in their folded, dimeric conformation as they are co-dependent on each other for secretion (111). While specific chaperones for the Esx and Esp substrates have not been identified, most PE/PPE pairs studied thus far require a cognate EspG chaperone to prevent inappropriate interactions (i.e. aggregation) and to target the substrate pair to the correct ESX machinery.

EspG function

EspG is a cytoplasmic protein that ultimately functions to prevent inappropriate interactions of its PE/PPE substrate heterodimer prior to export through the correct ESX machinery by forming transient interactions with the folded PE/PPE dimer. PE/PPE pairs are prone to aggregation in the absence of their EspG chaperone. In a $\Delta espG_I$ mutant of *M. tuberculosis*, PE35/PPE68 is found in insoluble aggregates that cannot be exported (112). Aggregation of PE/PPE dimers results in their degradation, and consequently PE/PPE dimers are also prone to degradation in the absence of their EspG chaperones. In a $\Delta espG_I$ mutant of *M.*

tuberculosis, there are reduced levels of the ESX-1 substrate PPE68 in the cytoplasm (113). In a $\Delta espG_5$ mutant of *M. marinum*, there are reduced levels of multiple PE/PPE substrates in the cytoplasm (114).

Like T3SCs, EspG is also predicted to be involved in determining system specificity of its substrates. Of the five ESX operons in the *M. tuberculosis* genome, ESX-1, ESX-2, ESX-3 and ESX-5 encode EspG, PE, and PPE paralogs. In addition, numerous other PE and PPE genes are scattered throughout the genome. Because each operon is thought to encode a complete, independent ESX system, it is generally assumed that PE, PPE, and EspG proteins from a given operon will interact specifically with each other and only associate with other components of that ESX system. Targeting and delivery of substrates to the proper ESX machinery may be orchestrated by interactions between EspG and ESX conserved component A (EccA). EccA is a cytoplasmic ATPase associated with membrane-bound ESX machinery that may be involved in dissociating the EspG-PPE interaction because (i) *eccA* mutants accumulate PPE-bound EspG (114), (ii) EccA interacts with both PPE proteins (115) and EspG (116) in yeast two-hybrid assays, and (iii) EccA proteins are only encoded in ESX gene clusters that also encode PE and PPE proteins. One model for the transfer of PE/PPE substrates from EspG to EccA is that transient binding of EccA to PE/PPE-EspG might allow EccA to dissociate the complex, passing the PE/PPE proteins off to the rest of the ESX system for secretion and recycling EspG to recruit additional PE/PPE proteins from the cytoplasm.

EspG substrates

Each EspG chaperone specifically interacts with the PE/PPE pair encoded in the same operon; for example, EspG₁ specifically interacts with PE35/PPE68 and EspG₅ interacts with

PE25/PPE41 (114). In both cases the PE/PPE pair is encoded in an operon with its cognate EspG encoding gene. Additional PE/PPE substrates depend on EspG for solubility and secretion but are encoded by genes located elsewhere in the genome (114). Of the four EspG chaperones in *M. tuberculosis*, EspG₅ has the largest substrate pool, and most PE/PPE proteins are secreted through the ESX-5 system (112, 114, 116). EspG does not interact with the other families of ESX secreted substrates, the Esx and Esp proteins (114). The Esx, PE/PPE, and Esp substrates all have conserved, uncleaved signal sequences (YxxxD/E) on one of the partner proteins, which are essential for secretion (114). For PE/PPE pairs, the secretion motif is always on the PE protein. The YxxxD/E signal sequence does not define system specificity, as the C-terminal 15 amino acids of an ESX-5 substrate containing this motif can be replaced by the homologous sequence of an ESX-1 substrate, and vice versa, without changing system specificity (114).

EspG does not bind to the YxxxD/E signal sequence. Rather, EspG specifically recognizes and binds to an EspG-binding domain on the PPE protein of the PE/PPE pair. This EspG-binding domain is a conserved, hydrophobic patch on the tip of the PPE protein. Mutation of the patch on PPE disrupts EspG-binding and reduces secretion of the PE/PPE pair (112). This patch also designates the specificity of the PPE proteins for its cognate EspG chaperone. For example, the ESX-1 secreted PE35/PPE68 pair can be rerouted to the ESX-5 system of *M. marinum* by replacing the EspG₁ binding domain of PPE68 with the equivalent EspG₅ binding domain of the ESX-5 substrate PPE18 (117). This domain replacement makes PE35/PPE68 independent of both EspG₁ and the ESX-1 membrane complex and instead dependent on EspG₅ and the ESX-5 complex for solubility and secretion.

EspG structure

EspG proteins are acidic (pI 4.5-6) proteins that are ~30 kDa in size. The crystal structures of several EspG proteins from *M. tuberculosis* (EspG₃ and EspG₅) and *M. smegmatis* (EspG₃) have been solved with and without bound substrates and reveal that despite low amino acid conservation (less than 25% pairwise sequence identity), the tertiary structures of the chaperones are highly similar and are not affected by substrate binding (112, 116). EspG proteins do not strongly resemble the solved structures of the smaller SecB or T3SCs (Table 1.1). Intriguingly, the overall EspG fold is approximately twofold symmetrical (112). The two halves, or subdomains, of EspG have the same topology and connectivity despite having low sequence similarity, suggesting that the EspG family of proteins may have evolved by an intergenic duplication event (112).

EspG proteins function as monomers, and bind substrates with a stoichiometry of one PE/PPE pair per EspG monomer (112). Unlike SecB and T3SCs, EspG binds to folded substrate pairs; the crystal structure of EspG₅ in complex with the PE25/PPE41 dimer reveals that the PE/PPE dimer is fully folded when bound to its chaperone (112, 116). EspG exclusively binds to the hydrophobic tip (EspG-binding domain) of the PPE protein, distal from the C-termini of both the PE and the PPE subunits (112). Because PE and PPE subunits often encode large C-terminal domains (of unknown function), binding of EspG far from the C-termini may prevent steric clashes with these additional domains. Additionally, binding of EspG at the EspG-binding domain does not interfere with the C-terminal YxxxD/E signal sequence, which is also located at the opposite end of the PE/PPE complex from EspG. The EspG-binding domain of the PPE protein interacts with a deep hydrophobic groove on EspG (112, 116). It is not known how EspG

interacts with the ATPase EccA. A summary of the structural and functional details of EspGs are found in Table 1.1.

Asp1, Asp2, and Asp3 of SecA2-SecY2 systems

As described briefly above, the SecA2-SecY2 accessory Sec system is conserved among a few streptococcal and staphylococcal Gram-positive bacteria, but not found in mycobacteria. The system is encoded within a single operon and mediates the transport of a single large serine-rich repeat, cell wall-anchored glycoprotein to the bacterial cell surface (34). In addition to genes encoding the serine-rich substrate, *secA2-secY2* operons contain genes that encode proteins mediating its export and glycosylation. In *Streptococcus gordonii* the substrate of the accessory Sec system is GspB (118). GspB export requires five highly conserved proteins: SecA2, SecY2, Asp1, Asp2, and Asp3. SecA2 and SecY2 are paralogs of canonical SecA and SecY.

Asp1, Asp2, and Asp3 do not have homologs beyond the accessory Sec system, and their functions in transport are largely unknown. One possibility is that Asp1-3 function as protein export chaperones for SecA2/SecY2 systems. In *S. gordonii*, the Asp1-3 proteins are present in the cytoplasm. Asp2 and Asp3 can bind the unstructured serine-rich-repeat domains of GspB suggesting that these two Asps may function to transport the substrate to the accessory Sec machinery by facilitating a specific interaction between GspB and SecA2 (119, 120). SecA2 serves as a docking site for Asp2, which in turn recruits Asp1 and Asp3 to the membrane (121). Although these Asps bind GspB directly, they do not meet the definition of a chaperone, since they are not required for GspB stability, prevention of folding, or anti-aggregation. Thus far, there are no examples of protein export chaperones that work with any SecA2 system.

Overall comparisons between protein export chaperones

Protein export chaperones are non-essential components of the export machinery, but are fully required for the proper export of their cognate substrates. In all cases, only subsets of the proteins exported by an export pathway require chaperones. All of the protein export chaperones highlighted in this chapter ensure that their specific substrates engage the appropriate export channel in a conformation compatible with export. In order to accomplish this goal, these protein export chaperones all form transient interactions with their substrates in the cytoplasm mediated by hydrophobic sites on the chaperone. For SecB and EspG, these sites are deep grooves in the chaperone. T3SC hydrophobic sites are more variable, and are generally referred to as hydrophobic patches. For their role in targeting to the export channel, all of these protein export chaperones target their substrate to an ATPase at the base of the channel. While the chaperones function in an energy-independent manner, ATP hydrolysis is required for disassembling the chaperone:substrate complex. However, this has not been demonstrated for EspG. Another similarity between these chaperones is that they are all acidic, although the contribution of acidic residues to chaperone function is unknown.

SecB, T3SCs and EspG do not resemble one another, either by primary amino acid sequence or by tertiary structure. Further, all of the protein export chaperones highlighted in this chapter have different oligomeric states; SecB forms a tetramer, T3SCs form dimers, and EspG functions as a monomer. Furthermore, although all of the chaperones interact with their substrates via hydrophobic sites, the effect of chaperone-binding for their substrates differs between all three chaperones. Sec substrates are fully unfolded when bound to SecB. T3SS substrates are only partially unfolded (at the CBD) and the remainder of the protein remains folded and enzymatically active. PE/PPE proteins are fully folded when bound to EspG. The

folding status of the chaperone-bound substrate is important for ensuring that it is compatible for export its with specific export pathway.

Bacteria often encode multiple ESX or T3S systems. Consequently, EspG and T3SC are required to target substrates to the correct machinery when the bacteria have paralogous export systems. It appears that both of these secretion systems have utilized similar strategies to ensure their substrate:chaperone pairs are appropriately targeted. In both cases, chaperones and substrates are typically encoded by genes adjacent to one another, linking expression of the chaperone and substrate. Additionally, the specificity of the T3SS substrate CBD and the ESX substrate EspG-binding domain ensures that substrates are only engaged by the correct chaperone. In the case of SecB, the substrate binding sites are less specific and can accommodate a larger variety of substrates, so long as they are unfolded.

Summary

It is clear that the mycobacterial SecA2 system is responsible for exporting a subset of proteins and it plays an important role in the virulence of *M. tuberculosis*. Nevertheless questions about the mechanistic steps of SecA2 export remain. A better understanding of the SecA2 protein export system has the potential to reveal new drug targets to combat infection with *M. tuberculosis*. The goal of my dissertation research was to fill in gaps in our understanding of the mechanism of SecA2 protein export by searching for additional proteins that are required for SecA2-dependent protein export. We previously established a suppressor screen carried out using the dominant-negative *M. smegmatis* *secA2 K129R* allele. In this dissertation, we describe two outcomes of this suppressor screen that improve our understanding of SecA2 protein export.

In chapter 2, we report the crystal structure of *M. tuberculosis* SecA2 and compare it with the previously solved SecA1 structure. This is the first SecA2 structure reported for any

organism. SecA1 and SecA2 exhibit a high level of structural similarity, and features of SecA1 that are important for the interaction between SecA1 and SecYEG are conserved in SecA2. We also identify some structural differences between SecA1 and SecA2, most notably the absence of the helical wing domain in SecA2. These structural differences may be important for recognizing and distinguishing SecA2-dependent substrates from the larger Sec substrate pool. Using this structure, we mapped intragenic suppressor mutations of *secA2 K129R* to the surface of SecA2. These suppressors allowed us to develop hypotheses for regions of SecA2 that may promote interactions with SecYEG, SecA2 substrates or other partners of SecA2.

In chapter 3, we identify several suppressors affecting the conserved hypothetical protein Msmeg_1684, which we renamed SatS for SecA2 (two) Suppressor. We provide evidence in both *M. smegmatis* and *M. tuberculosis* that SatS is required for the stability and export of some, but not all, SecA2-dependent substrates. Specifically, we show that SatS is required for SapM and Mce export. Our data led us to identify SatS as the founding member of a new class of protein export chaperones. Comparisons of the newly solved crystal structure of SatS with the well-studied protein export chaperone SecB reveal that, while SatS and SecB are not structurally similar, both proteins have similar hydrophobic grooves positioned on the surface which, for SecB, are crucial for substrate binding and unfolding.

Taken together, these studies greatly expand our model of SecA2-dependent protein export by structurally characterizing the mycobacterial SecA2 protein and chiefly by characterizing SatS, a protein export chaperone of the mycobacterial SecA2 system. By identifying SatS as the first protein export chaperone of a SecA2 pathway, our studies expand our appreciation of the diversity of chaperones in biological systems.

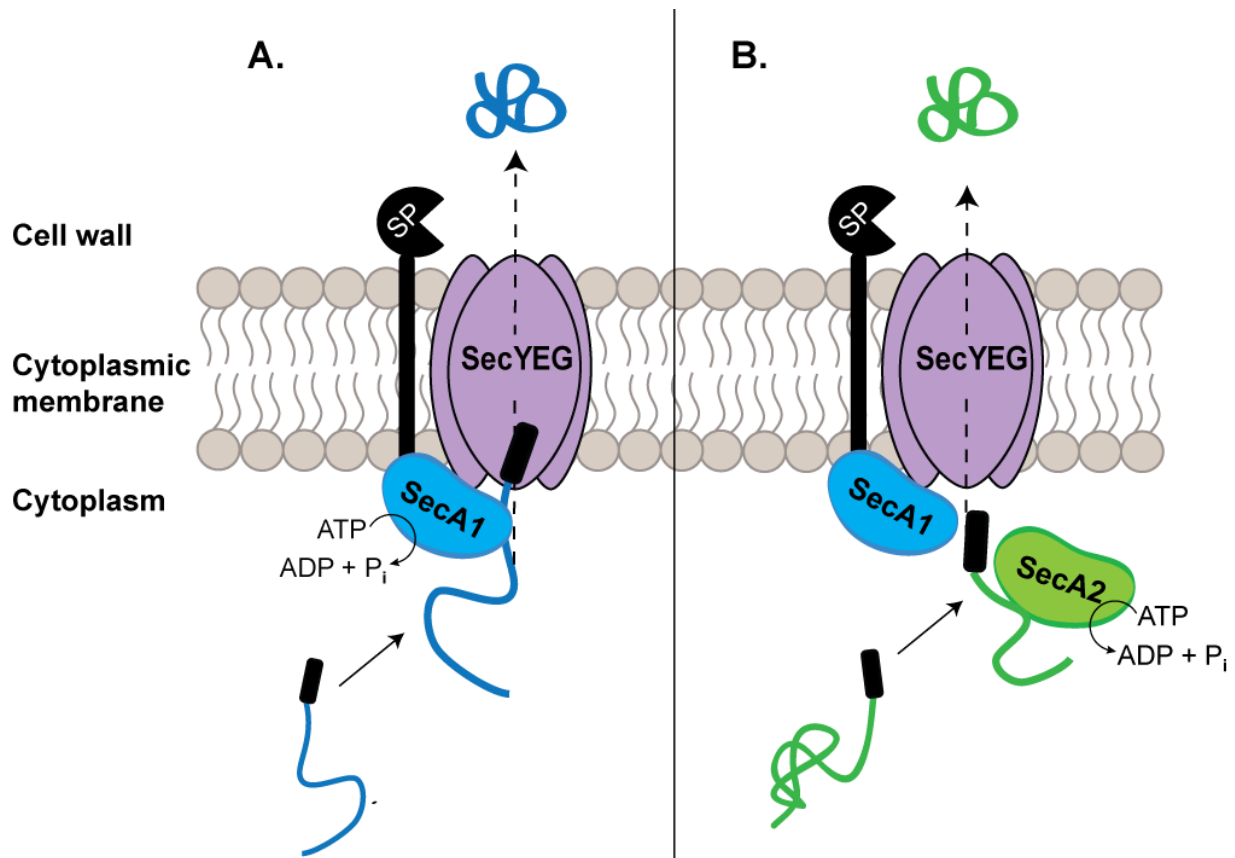


Figure 1.1. Models of SecA1 and SecA2 export in *M. tuberculosis*. A) SecA1 uses ATP hydrolysis to export preproteins through the SecYEG channel in an unfolded, export competent state. Sec signal sequences (black rectangle) target preproteins (blue ribbon) for export through SecYEG and are then cleaved by a signal peptidase (SP). B) SecA2 also uses the SecYEG channel and possibly SecA1 to export its own subset of preproteins (green ribbon). The signal sequence (black rectangle) is indistinguishable from canonical Sec signal sequences. Instead, the mature domain's propensity for cytoplasmic folding is predicted to confer specificity for SecA2.

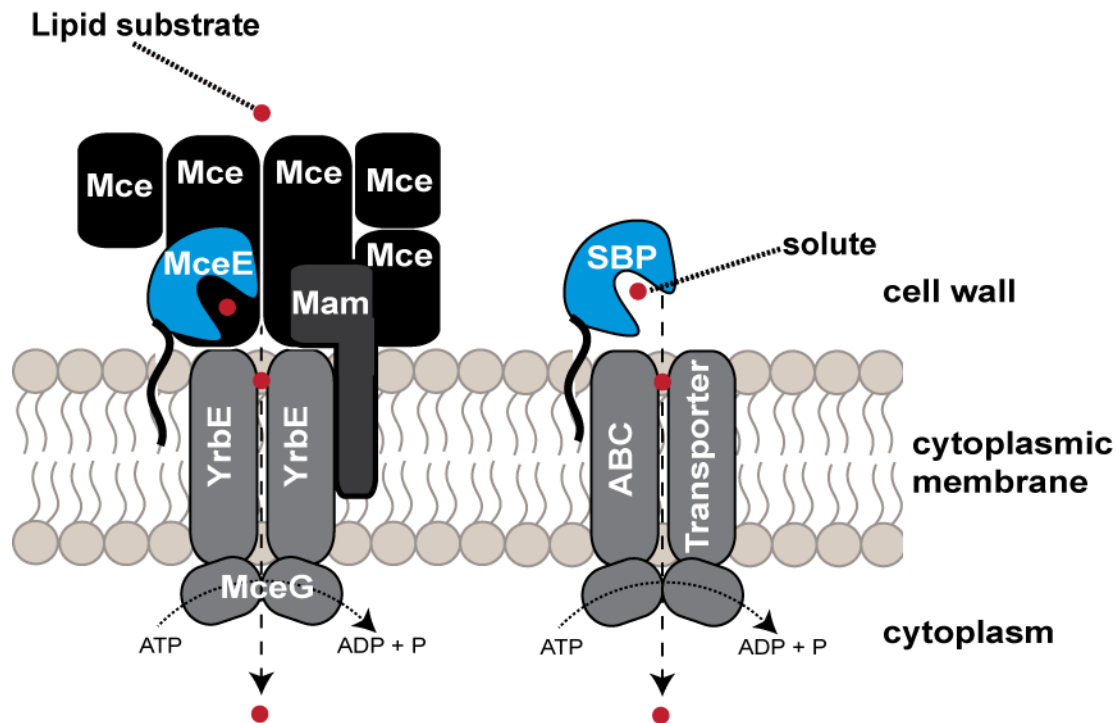


Figure 1.2. Solute binding proteins and Mce proteins are exported by the SecA2 pathway. Two classes of SecA2-dependent substrates are SBPs and Mce proteins. Both SBPs and Mce proteins are involved in solute acquisition. In the case of SBPs this involves import of a solute through an ABC transporter permease using energy provided by ATP hydrolysis. Mce transporters are thought to function in a similar manner as ABC transporters to import a lipid substrate through a YrbE permease in an ATP-dependent manner. Although the diagram of an Mce transporter is speculative the similarities between these two systems are compelling.

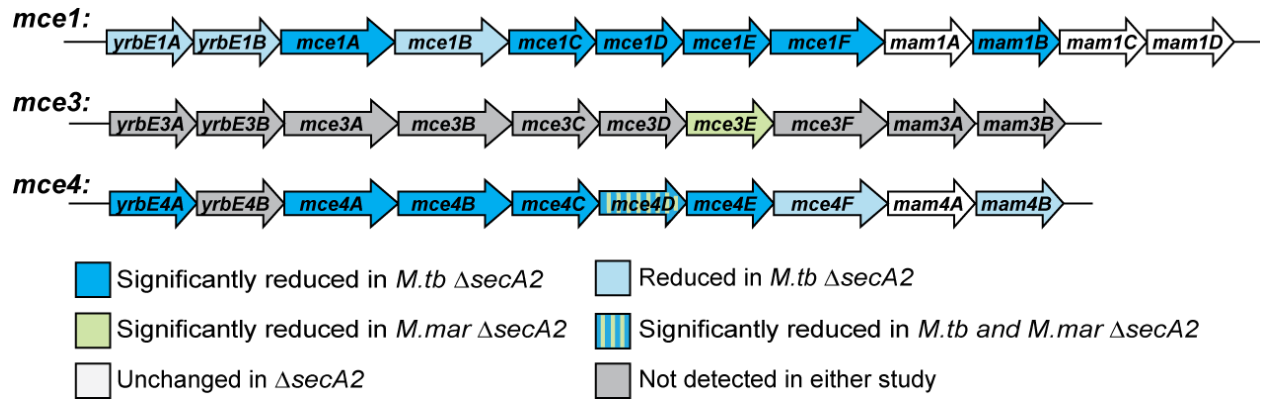
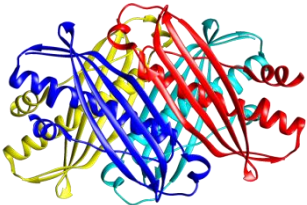
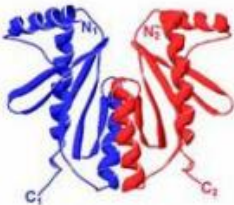
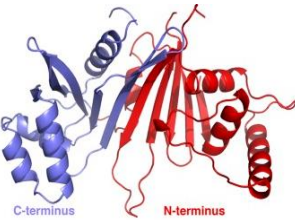


Figure 1.3. Multiple components of Mce transporters are reduced in the cell wall of the $\Delta secA2$ mutant. The *M. tuberculosis* genome contains four *mce* loci encoding putative lipid transporters. The genomic regions encoding Mce1, Mce3, and Mce4 transporters are shown with Open Reading Frames (ORF) colored for Mce proteins that are reduced in quantitative mass spectrometry studies of the *M. tuberculosis* and *M. marinum* $\Delta secA2$ mutant cell wall or cell envelope fractions (29, 42). In dark blue and/or green are *mce* genes for proteins that are significantly reduced ($p < 0.01$ for *M. tuberculosis* and $p < 0.05$ for *M. marinum*) in the $\Delta secA2$ mutant; in light blue are genes for *mce* proteins that are reduced in the *M. tuberculosis* $\Delta secA2$ mutant but did not reach statistical significance.

Table 1.1. Structural and functional comparisons between known protein export chaperones

Feature	SecB	T3SC (Class 1)	EspG
Location	Cytoplasmic	Cytoplasmic	Cytoplasmic
Size	17 kDa	15 - 20 kDa	30 kDa
Isoelectric point	Acidic (pI 4.1)	Acidic (pI ~ 4 - 4.5)	Acidic (pI ~4.5-6)
Oligomeric state	Tetramer	Dimer	Monomer
Interacting partners	Binds substrates and SecA ATPase	Binds substrates and T3SS ATPase	Binds substrates and EccA ATPase
Substrate pool	Not all Sec substrates use SecB; SecB has multiple substrates	Not all T3SS substrates use a given T3SC; some T3SCs have one substrate, some have multiple substrates	Not all ESX substrates use a given EspG; some EspGs have one substrate, some have multiple substrates
Gene location	No genomic context with substrates	Often in operon with substrate, additional substrates not in operon	In operon with substrate pair, additional substrates not in operon
Chaperone-bound substrate folding state	Whole substrate is unfolded when bound to SecB	Only CBD is unfolded when bound to T3SC; remainder of substrate is folded	Whole substrate pair is folded when bound to EspG
Chaperone/substrate interaction	Substrate binds in hydrophobic grooves, wraps around tetramer 1 substrate/SecB tetramer	Substrate CBD binds in hydrophobic patches, wraps around dimer 1 substrate/T3SC dimer	PPE substrate of PE/PPE pair binds in hydrophobic groove, 1 pair/EspG monomer
Misfolding phenotype of chaperone-null mutant	Without SecB, substrate is stable and accumulates in cytoplasm, some aggregate – no export	Without T3SC, substrate is unstable – no export	Without EspG, substrate aggregates in cytoplasm – no export
Additional functions	Performs multiple functions: prevent substrate folding and/or aggregation, target to Sec	Performs multiple functions: Stabilize substrates, target to correct T3SS, regulate hierarchy of substrates	Performs multiple functions: prevent substrate aggregation, target to correct ESX machinery
Structure	 SecB tetramer (PDB ID: 1QYN)	 SicP dimer (PDB ID: 1JYO)	 EspG ₃ monomer (PDB ID: 4W4I)

REFERENCES

1. Global tuberculosis report 2017. Geneva: World Health Organization, 2017 Contract No.: CC BY-NCSA 3.0 IGO.
2. Schneider G. 1999. How many potentially secreted proteins are contained in a bacterial genome? *Gene* 237(1):113-21.
3. Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48(1):77-84.
4. Saint-Joanis B, Demangel C, Jackson M, Brodin P, Marsollier L, Boshoff H, et al. 2006. Inactivation of Rv2525c, a substrate of the Twin Arginine Translocation (Tat) system of *Mycobacterium tuberculosis*, increases β -lactam susceptibility and virulence. *J Bacteriol* 188(18):6669-79.
5. McDonough JA, Hacker KE, Flores AR, Pavelka MS, Braunstein M. 2005. The Twin-Arginine Translocation Pathway of *Mycobacterium smegmatis* is functional and required for the export of mycobacterial β -lactamases. *J Bacteriol* 187(22):7667-79.
6. Vrontou E, Economou A. 2004. Structure and function of SecA, the preprotein translocase nanomotor. *Biochimica et Biophysica Acta - Molecular Cell Research* 1694(1):67-80.
7. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8:785-6.
8. Moller S, Croning MDR, Apweiler R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17(7):646-53.
9. Perkowski EF, Zulauf KE, Weerakoon D, Hayden JD, Ioerger TR, Oreper D, et al. 2017. The EXIT Strategy: an approach for identifying bacterial proteins exported during host infection. *mBio* 8(2).
10. Meyer TH, Ménétret J-F, Breitling R, Miller KR, Akey CW, Rapoport TA. 1999. The bacterial SecY/E translocation complex forms channel-like structures similar to those of the eukaryotic *sec61p* complex. *JMB* 285(4):1789-800.
11. Kihara A, Akiyama Y, Ito K. 1995. FtsH is required for proteolytic elimination of complexed forms of SecY, an essential protein translocase subunit. *PNAS* 92(10):4532-6.
12. Nishiyama K-i, Suzuki T, Tokuda H. 1996. Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. *Cell* 85(1):71-81.
13. Duong F, Wickner W. 1997. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *The EMBO Journal* 16(10):2756-68.
14. Lührink J, Sinning I. 2004. SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta - Molecular Cell Research* 1694(1–3):17-35.
15. Wild K, Rosendal KR, Sinning I. 2004. A structural step into the SRP cycle. *Mol Microbiol* 53(2):357-63.
16. Valent QA, De Gier J-WL, Heijne Gv, Kendall DA, Ten Hagen-Jongman CM, Oudega B, et al. 1997. Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Mol Microbiol* 25(1):53-64.

17. Egea PF, Stroud RM. 2010. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *PNAS* 107(40):17182-7.
18. Economou A, Wickner W. 1994. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* 78(5):835-43.
19. Braunstein M, Brown AM, Kurtz S, Jacobs WR. 2001. Two nonredundant SecA homologues function in mycobacteria. *J Bacteriol* 183(24):6979-90.
20. Heijne Gv. 1990. The signal peptide. *The Journal of Membrane Biology* 115(3):195-201.
21. Nakayama H, Kurokawa K, Lee BL. 2012. Lipoproteins in bacteria: structures and biosynthetic pathways. *FEBS Journal* 279(23):4247-68.
22. Paetzel M, Karla A, Strynadka NCJ, Dalbey RE. 2002. Signal peptidases. *Chemical Reviews* 102(12):4549-80.
23. Bassford PJ, Silhavy TJ, Beckwith JR. 1979. Use of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. *J Bacteriol* 139(1):19-31.
24. Sala A, Bordes P, Genevaux P. 2014. Multitasking SecB chaperones in bacteria. *Frontiers in Microbiology* 5:666.
25. Fisher AC, DeLisa MP. 2004. A little help from my friends: quality control of presecretory proteins in bacteria. *J Bacteriol* 186(22):7467-73.
26. Bordes P, Cirinesi A-M, Ummels R, Sala A, Sakr S, Bitter W, et al. 2011. SecB-like chaperone controls a toxin–antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *PNAS* 108(20):8438-43.
27. Braunstein M, Espinosa BJ, Chan J, Belisle JT, R. Jacobs W. 2003. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 48(2):453-64.
28. Sullivan JT, Young EF, McCann JR, Braunstein M. 2012. The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infection and Immunity* 80(3):996-1006.
29. van der Woude AD, Stoop EJM, Stiess M, Wang S, Ummels R, van Stempvoort G, et al. 2014. Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. *Cellular Microbiology* 16(2):280-95.
30. Kurtz S, McKinnon KP, Runge MS, Ting JPY, Braunstein M. 2006. The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. *Infection and Immunity* 74(12):6855-64.
31. Watkins BY, Joshi SA, Ball DA, Leggett H, Park S, Kim J, et al. 2012. *Mycobacterium marinum* SecA2 promotes stable granulomas and induces Tumor Necrosis Factor Alpha *in vivo*. *Infection and Immunity* 80(10):3512-20.
32. Lenz LL, Mohammadi S, Geissler A, Portnoy DA. 2003. SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *PNAS* 100(21):12432-7.
33. Siboo IR, Chambers HF, Sullam PM. 2005. Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. *Infection and Immunity* 73(4):2273-80.

34. Bensing BA, Sullam PM. 2002. An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. *Mol Microbiol* 44(4):1081-94.
35. Chen Q, Wu H, Fives-Taylor PM. 2004. Investigating the role of *secA2* in secretion and glycosylation of a fimbrial adhesin in *Streptococcus parasanguis* FW213. *Mol Microbiol* 53(3):843-56.
36. Rigel NW, Braunstein M. 2008. A new twist on an old pathway – accessory Sec systems. *Mol Microbiol* 69(2):291-302.
37. Bensing BA, Seepersaud R, Yen YT, Sullam PM. 2014. Selective transport by SecA2: an expanding family of customized motor proteins. *Biochim et biophys acta* 1843(8):1674-86.
38. Ligon LS, Rigel NW, Romanchuk A, Jones CD, Braunstein M. 2013. Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of mycobacteria. *J Bacteriol* 195(19):4456-65.
39. Durack J, Burke TP, Portnoy DA. 2015. A *prl* mutation in SecY suppresses secretion and virulence defects of *Listeria monocytogenes secA2* mutants. *J Bacteriol* 197(5):932-42.
40. Fagan RP, Fairweather NF. 2011. *Clostridium difficile* Has two parallel and essential Sec secretion systems. *The Journal of Biological Chemistry* 286(31):27483-93.
41. Gibbons HS, Wolschendorf F, Abshire M, Niederweis M, Braunstein M. 2007. Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* 189(14):5090-100.
42. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, et al. 2015. Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & Cellular Proteomics* 14(6):1501-16.
43. Braibant M, Gilot P, Content J. 2000. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiology Reviews* 24(4):449-67.
44. Casali N, Riley LW. 2007. A phylogenomic analysis of the Actinomycetales *mce* operons. *BMC Genomics* 8(1):60.
45. Malinverni JC, Silhavy TJ. 2009. An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. *PNAS* 106(19):8009-14.
46. Pandey AK, Sassetti CM. 2008. Mycobacterial persistence requires the utilization of host cholesterol. *PNAS* 105(11):4376-80.
47. Senaratne RH, Sidders B, Sequeira P, Saunders G, Dunphy K, Marjanovic O, et al. 2008. *Mycobacterium tuberculosis* strains disrupted in *mce3* and *mce4* operons are attenuated in mice. *Journal of Medical Microbiology* 57(2):164-70.
48. Forrellad MA, McNeil M, Santangelo MP, Blanco FC, García E, Klepp LI, et al. 2014. Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis* 94(2):170-7.

49. Cantrell SA, Leavell MD, Marjanovic O, Iavarone AT, Leary JA, Riley LW. 2013. Free mycolic acid accumulation in the cell wall of the *mce1* operon mutant strain of *Mycobacterium tuberculosis*. *Journal of Microbiology* 51(5):619-26.
50. Rengarajan J, Bloom BR, Rubin EJ. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *PNAS* 102(23):8327-32.
51. McCann JR, McDonough JA, Sullivan JT, Feltcher ME, Braunstein M. 2011. Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* 193(4):854-61.
52. Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sasseti CM. 2006. Characterization of mycobacterial virulence genes through genetic interaction mapping. *PNAS* 103(31):11760-5.
53. Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehrt S, et al. 2003. Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon. *PNAS* 100(26):15918-23.
54. Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, et al. 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(6):1691-702.
55. O'Hare HM, Durán R, Cerveñansky C, Bellinzoni M, Wehenkel AM, Pritsch O, et al. 2008. Regulation of glutamate metabolism by protein kinases in mycobacteria. *Mol Microbiol* 70(6):1408-23.
56. Walburger A, Koul A, Ferrari G, Nguyen L, Prescianotto-Baschong C, Huygen K, et al. 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* 304(5678):1800-4.
57. Wolff KA, de la Peña AH, Nguyen HT, Pham TH, Amzel LM, Gabelli SB, et al. 2015. A redox regulatory system critical for mycobacterial survival in macrophages and biofilm development. *PLOS Pathogens* 11(4):e1004839.
58. Bhat KH, Mukhopadhyay S. 2015. Macrophage takeover and the host–bacilli interplay during tuberculosis. *Future Microbiology* 10(5):853-72.
59. Armstrong JA, Hart PDA. 1971. Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. *The Journal of Experimental Medicine* 134(3):713-40.
60. Vergne I, Chua J, Deretic V. 2003. *Mycobacterium tuberculosis* phagosome maturation arrest: selective targeting of PI3P-dependent membrane trafficking. *Traffic* 4(9):600-6.
61. Puri RV, Reddy PV, Tyagi AK. 2013. Secreted acid phosphatase (SapM) of *Mycobacterium tuberculosis* is indispensable for arresting phagosomal maturation and growth of the pathogen in guinea pig tissues. *PLoS ONE* 8(7):e70514.
62. Zulauf KE, Sullivan JT, Braunstein M. 2018. The SecA2 pathway of *Mycobacterium tuberculosis* exports effectors that work in concert to arrest phagosome and autophagosome maturation. *PLOS Pathogens* 14(4):e1007011.

63. Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S, Braunstein M. 2009. The Accessory SecA2 system of mycobacteria requires ATP binding and the canonical SecA1. *The Journal of Biological Chemistry* 284(15):9927-36.
64. Hou JM, D'Lima NG, Rigel NW, Gibbons HS, McCann JR, Braunstein M, et al. 2008. ATPase activity of *Mycobacterium tuberculosis* SecA1 and SecA2 proteins and its importance for SecA2 function in macrophages. *J Bacteriol* 190(14):4880-7.
65. Bhanu MK, Zhao P, Kendall DA. 2013. Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility method. *J Bacteriol* 195(20):4709-15.
66. Auclair SM, Oliver DB, Mukerji I. 2013. Defining the solution-state dimer structure of *Escherichia coli* SecA using Förster resonance energy transfer. *Biochemistry* 52(14):2388-401.
67. Sardis MF, Economou A. 2010. SecA: a tale of two protomers. *Mol Microbiol* 76(5):1070-81.
68. Prabudiansyah I, Kusters I, Driessen AJM. 2015. *In vitro* interaction of the housekeeping SecA1 with the accessory SecA2 protein of *Mycobacterium tuberculosis*. *PLoS ONE* 10(6):e0128788.
69. D'Lima NG, Teschke CM. 2014. ADP-dependent conformational changes distinguish *Mycobacterium tuberculosis* SecA2 from SecA1. *The Journal of Biological Chemistry* 289(4):2307-17.
70. Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M, Sacchettini JC. 2015. Structural similarities and differences between two functionally distinct SecA proteins: the *Mycobacterium tuberculosis* SecA1 and SecA2. *J Bacteriol* 198(4):720-30.
71. Renier S, Chambon C, Viala D, Chagnot C, Hébraud M, Desvaux M. 2013. Exoproteomic analysis of the SecA2-dependent secretion in *Listeria monocytogenes* EGD-e. *Journal of Proteomics* 80:183-95.
72. Feltcher ME, Gibbons HS, Ligon LS, Braunstein M. 2013. Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *J Bacteriol* 195(4):672-81.
73. Krehenbrink M, Edwards A, Downie JA. 2011. The superoxide dismutase SodA is targeted to the periplasm in a SecA-dependent manner by a novel mechanism. *Mol Microbiol* 82(1):164-79.
74. Archambaud C, Nahori M-A, Pizarro-Cerda J, Cossart P, Dussurget O. 2006. Control of listeria superoxide dismutase by phosphorylation. *Journal of Biological Chemistry* 281(42):31812-22.
75. Ellis RJ. 1997. Molecular chaperones: Avoiding the crowd. *Current Biology* 7(9):R531-R3.
76. Knoblauch NTM, Rüdiger S, Schönfeld H-J, Driessen AJM, Schneider-Mergener J, Bukau B. 1999. Substrate specificity of the SecB chaperone. *Journal of Biological Chemistry* 274(48):34219-25.

77. Kumamoto CA, Gannon PM. 1988. Effects of *Escherichia coli* *secB* mutations on pre-maltose binding protein conformation and export kinetics. *Journal of Biological Chemistry* 263(23):11554-8.
78. Huang C, Rossi P, Saio T, Kalodimos CG. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature* 537:202.
79. Sakr S, Cirinesi A-M, Ullers RS, Schwager F, Georgopoulos C, Genevoux P. 2010. Lon protease quality control of presecretory proteins in *Escherichia coli* and its dependence on the SecB and DnaJ (Hsp40) chaperones. *Journal of Biological Chemistry* 285(30):23506-14.
80. Hartl F-U, Lecker S, Schiebel E, Hendrick JP, Wickner W. 1990. The binding cascade of SecB to SecA to SecY E mediates preprotein targeting to the *E. coli* plasma membrane. *Cell* 63(2):269-79.
81. Crane JM, Randall LL. 2017. The Sec system: protein export in *Escherichia coli*. *EcoSal Plus* 7(2):10.1128/ecosalplus.ESP-0002-2017.
82. Strobel SM, Cannon JG, Bassford PJ. 1997. Regions of maltose-binding protein that influence SecB-dependent and SecA-dependent export in *Escherichia coli*. *J Bacteriol* 175(21):6988-95.
83. Randall L, Topping T, Hardy S. 1990. No specific recognition of leader peptide by SecB, a chaperone involved in protein export. *Science* 248(4957):860-3.
84. Hardy S, Randall L. 1991. A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. *Science* 251(4992):439-43.
85. Xu Z, Knafels JD, Yoshino K. 2000. Crystal structure of the bacterial protein export chaperone SecB. *Nature Structural Biology* 7:1172.
86. Khisty VJ, Randall LL. 1995. Demonstration in vivo that interaction of maltose-binding protein with SecB is determined by a kinetic partitioning. *J Bacteriol* 177(11):3277-82.
87. Crane JM, Suo Y, Lilly AA, Mao C, Hubbell WL, Randall LL. 2006. Sites of interaction of a precursor polypeptide on the export chaperone SecB mapped by site-directed spin labeling. *Journal of Molecular Biology* 363(1):63-74.
88. Crane JM, Mao C, Lilly AA, Smith VF, Suo Y, Hubbell WL, et al. 2005. Mapping of the docking of SecA onto the chaperone SecB by site-directed spin labeling: insight into the mechanism of ligand transfer during protein export. *Journal of Molecular Biology* 353(2):295-307.
89. Gannon PM, Kumamoto CA. 1993. Mutations of the molecular chaperone protein SecB which alter the interaction between SecB and maltose-binding protein. *Journal of Biological Chemistry* 268(3):1590-5.
90. Deng W, Marshall NC, Rowland JL, McCoy JM, Worrall LJ, Santos AS, et al. 2017. Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* 15:323.
91. Thomas NA, Ma I, Prasad ME, Rafuse C. 2012. Expanded roles for multicargo and class 1B effector chaperones in type III secretion. *J Bacteriol* 194(15):3767-73.

92. Letzelter M, Sorg I, Mota LJ, Meyer S, Stalder J, Feldman M, et al. 2006. The discovery of SycO highlights a new function for type III secretion effector chaperones. *The EMBO Journal* 25(13):3223-33.
93. Gauthier A, Finlay BB. 2003. Translocated intimin receptor and its chaperone interact with ATPase of the type III secretion apparatus of enteropathogenic *Escherichia coli*. *J Bacteriol* 185(23):6747-55.
94. Thomas NA, Deng W, Puente JL, Frey EA, Yip CK, Strynadka NCJ, et al. 2005. CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol* 57(6):1762-79.
95. Zarivach R, Vuckovic M, Deng W, Finlay BB, Strynadka NCJ. 2007. Structural analysis of a prototypical ATPase from the type III secretion system. *Nature Structural & Molecular Biology* 14:131.
96. Akeda Y, Galán JE. 2005. Chaperone release and unfolding of substrates in type III secretion. *Nature* 437:911.
97. Boyd AP, Lambermont I, Cornelis GR. 2000. Competition between the Yops of *Yersinia enterocolitica* for delivery into eukaryotic cells: role of the SycE chaperone binding domain of YopE. *J Bacteriol* 182(17):4811-21.
98. Birtalan SC, Phillips RM, Ghosh P. 2002. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Molecular Cell* 9(5):971-80.
99. Thomas NA, Deng W, Puente JL, Frey EA, Yip CK, Strynadka NCJ, et al. 2005. CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol* 57(6):1762-79.
100. Mills E, Baruch K, Aviv G, Nitzan M, Rosenshine I. 2013. Dynamics of the type III secretion system activity of enteropathogenic *Escherichia coli*. *mBio* 4(4).
101. Little DJ, Coombes BK. 2014. Molecular basis for CesT recognition of type III secretion effectors in enteropathogenic *Escherichia coli*. *PLOS Pathogens* 14(8):e1007224.
102. Lloyd SA, Norman M, Rosqvist R, Wolf-Watz H. 2001. *Yersinia* YopE is targeted for type III secretion by N-terminal, not mRNA, signals. *Mol Microbiol* 39(2):520-32.
103. Cheng LW, Anderson DM, Schneewind O. 1997. Two independent type III secretion mechanisms for YopE in *Yersinia enterocolitica*. *Mol Microbiol* 24(4):757-65.
104. Lee SH, Galán JE. 2004. Salmonella type III secretion-associated chaperones confer secretion-pathway specificity. *Mol Microbiol* 51(2):483-95.
105. Luo Y, Bertero MG, Frey EA, Pfuetzner RA, Wenk MR, Creagh L, et al. 2001. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nature Structural Biology* 8:1031.
106. Knodler LA, Bertero M, Yip C, Strynadka NCJ, Steele-Mortimer O. 2006. Structure-based mutagenesis of SigE verifies the importance of hydrophobic and electrostatic residues in type III chaperone function. *Mol Microbiol* 62(4):928-40.

107. Thomas J, Stafford GP, Hughes C. 2004. Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. *PNAS* 101(11):3945-50.
108. Simeone R, Bottai D, Brosch R. 2009. ESX/type VII secretion systems and their role in host–pathogen interaction. *Current Opinion in Microbiology* 12(1):4-10.
109. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. *Nature Reviews Microbiology* 14:677.
110. Beckham KSH, Ciccarelli L, Bunduc CM, Mertens HDT, Ummels R, Lugmayr W, et al. 2017. Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nature Microbiology* 2:17047.
111. Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D. 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *PNAS* 103(21):8060-5.
112. Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, et al. 2014. Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25–PPE41 dimer. *Mol Microbiol* 94(2):367-82.
113. Bottai D, Majlessi L, Simeone R, Frigui W, Laurent C, Lenormand P, et al. 2011. ESAT-6 secretion-independent impact of ESX-1 genes *espF* and *espG1* on virulence of *Mycobacterium tuberculosis*. *The Journal of Infectious Diseases* 203(8):1155-64.
114. Daleke MH, van der Woude AD, Parret AHA, Ummels R, de Groot AM, Watson D, et al. 2012. Specific chaperones for the type VII protein secretion pathway. *Journal of Biological Chemistry* 287(38):31939-47.
115. Teutschbein J, Schumann G, Möllmann U, Grabley S, Cole ST, Munder T. 2009. A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. *Microbiological Research* 164(3):253-9.
116. Ekiert DC, Cox JS. 2014. Structure of a PE–PPE–EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. *PNAS* 111(41):14758-63.
117. Phan TH, Ummels R, Bitter W, Houben ENG. 2017. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Scientific Reports* 7:42704.
118. Takamatsu D, Bensing BA, Sullam PM. 2004. Genes in the accessory *sec* locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. *Mol Microbiol* 52(1):189-203.
119. Bensing BA, Yen YT, Seepersaud R, Sullam PM. 2012. A specific interaction between SecA2 and a region of the preprotein adjacent to the signal peptide occurs during transport via the accessory Sec system. *Journal of Biological Chemistry* 287(29):24438-47.
120. Yen YT, Seepersaud R, Bensing BA, Sullam PM. 2011. Asp2 and Asp3 interact directly with GspB, the export substrate of the *Streptococcus gordonii* accessory *sec* system. *J Bacteriol* 193(13):3165-74.

121. Yen YT, Cameron TA, Bensing BA, Seepersaud R, Zambryski PC, Sullam PM. 2013. Differential localization of the Streptococcal accessory Sec components and implications for substrate export. *J Bacteriol* 195(4):682-95.

CHAPTER 2²

Structural Similarities and Differences between Two Functionally Distinct SecA Proteins, *Mycobacterium tuberculosis* SecA1 and SecA2

While SecA is the ATPase component of the major bacterial secretory (Sec) system, mycobacteria and some Gram-positive pathogens have a second paralog, SecA2. In bacteria with two SecA paralogs, each SecA is functionally distinct, and they cannot compensate for one another. Compared to SecA1, SecA2 exports a distinct and smaller set of substrates, some of which have roles in virulence. In the mycobacterial system, some SecA2-dependent substrates lack a signal sequence, while others contain a single peptide but possess features in the mature protein that necessitate a role for SecA2 in their export. It is unclear how SecA2 functions in protein export, and one open question is whether SecA2 works with the canonical SecYEG channel to export proteins. In this study, we report the structure of *Mycobacterium tuberculosis* SecA2 (*Mtb*SecA2), which is the first structure of any SecA2 protein. A high level of structural similarity is observed between SecA2 and SecA1. The major structural difference is the absence of the helical wing domain, which is likely to play a role in how *Mtb*SecA2 recognizes its unique substrates. Importantly, structural features critical to the interaction between SecA1 and SecYEG are preserved in SecA2. Furthermore, suppressor mutations of a dominant-negative *secA2* mutant

² Adapted for this dissertation with permission from: Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M, Sacchettini JC. 2016. Structural similarities and differences between two functionally distinct SecA proteins, *Mycobacterium tuberculosis* SecA1 and SecA2. *Journal of Bacteriology* 198(4): 720-730.

map to the surface of SecA2 and help identify functional regions of SecA2 that may promote interactions with SecYEG or the translocating polypeptide substrate. These results support a model in which the mycobacterial SecA2 works with SecYEG.

Introduction

SecA is the ATPase component of the bacterial Sec secretion pathway (1). SecA recognizes proteins destined for export from the cytoplasm and provides energy to translocate them across the cytoplasmic membrane by way of the SecYEG translocase channel. The proteins exported by SecA are synthesized as preproteins with N-terminal signal sequences. Following translocation, the signal sequence is cleaved to release the mature protein species. Both the signal sequence and features of the mature protein are recognized by SecA (2). Some Gram-positive and acid-fast bacteria, including mycobacteria, have a SecA paralog referred to as SecA2. SecA1, the canonical SecA in these organisms, is essential for growth and responsible for the majority of protein export that occurs. In contrast, SecA2 is typically not essential and is required for the export of a more limited subset of proteins (3,4). Studies in mycobacteria show that even when overexpressed, the two SecA proteins are unable to compensate for each other (5). Thus, each SecA protein has distinct functions in protein export. In *Mycobacterium tuberculosis*, SecA2 (*Mtb*SecA2) is not essential for growth in culture, but it is essential for virulence *in vivo* (6,7). Furthermore, SecA2 is required for intracellular growth of *M. tuberculosis* in macrophages (8). The role of SecA2 in promoting growth in macrophages is attributed to a role in preventing phagosome maturation (9). In *Mycobacterium marinum*, export of protein kinase G (PknG) by the SecA2 pathway is suggested to at least be partially responsible for the SecA2 effect on phagosome maturation (10). In *M. tuberculosis*, the SecA2 pathway is additionally required to restrict apoptosis of infected macrophages. A possible explanation for the latter effect is the

SecA2-dependent secretion of superoxide dismutase, which may reduce reactive oxygen species (ROS)-mediated apoptosis (11, 12). An association between SecA2 and the secretion of virulence factors extends to other bacterial pathogens as well (13-16). There is also an intriguing association between the SecA2 pathway and the export of S-layer proteins by Gram-positive bacteria, such as *Bacillus anthracis* (17) and *Clostridium difficile* (18).

It is unclear how *Mtb*SecA2 carries out its unique function in protein export. In some organisms with two SecAs, there is a SecY paralog (SecY2), with which SecA2 likely interacts (19). In SecA2–SecY2 systems, SecY2 and several accessory Sec proteins (Asp) are thought to form an accessory protein translocation channel in the cytoplasmic membrane (4). Mycobacteria, however, are in a group of bacteria referred to as “SecA2-only” systems that lack a second SecY ortholog (3). Mycobacteria, as well as several Gram-positive species, including *Listeria monocytogenes* (13), *Corynebacterium glutamicum* (20), and *C. difficile* (18), are in the “SecA2-only” group. An important but unresolved question is whether SecA2 works with the canonical SecYEG channel to export proteins in these systems lacking a second SecY.

The mycobacterial proteins currently known to be exported by SecA2 include examples with typical Sec signal sequences, as well as proteins lacking signal sequences altogether (3). Superoxide dismutase (SodA) in *M. tuberculosis* and PknG in *M. tuberculosis* and *M. marinum* are examples of proteins lacking signal sequences that are exported in a SecA2-dependent manner (7, 10, 21). Of the signal sequence-containing proteins exported by the SecA2 systems of *Mycobacterium smegmatis* (22), *M. marinum* (10), and *M. tuberculosis* (21), the most thoroughly studied proteins are the *M. smegmatis* 1704 (*Ms*1704) and *Ms*1712 proteins (22). Studies of *Ms*1704 and *Ms*1712 demonstrate that they require their signal sequence for export, but it is a feature of the mature portions of these proteins that necessitates export via the SecA2-dependent

pathway (23). Interestingly, when fused to a signal sequence for the twin-arginine translocation (Tat) pathway, the mature domain of *Ms1704* is exported by the Tat pathway. This result suggests that the defining feature of SecA2 substrates may be a tendency to fold prior to export (23). This is because proteins that get translocated across the membrane by the Tat pathway must be folded in the cytoplasm prior to export (24). In contrast, preproteins exported by the canonical SecA must be unfolded (25), sometimes with the help of export chaperones (26, 27), due to the narrow diameter of the SecYEG central channel. Therefore, if SecA2 works with SecYEG, the role of SecA2 may be to facilitate the export of proteins that have a tendency to fold prior to export by either helping to maintain such proteins in an unfolded state or assisting in the recognition or export of such problematic substrates.

There is only 38% amino acid sequence identity between the *MtbSecA1* and *MtbSecA2* proteins. Yet, SecA2, like SecA1, has a DEAD box ATPase domain (28), and ATPase activity is required for SecA2 function (29). Furthermore, SecA2 variants lacking ATPase activity due to an amino acid substitution in the Walker box are dominant negative, and a *secA2* dominant-negative mutant exhibits *secA2* mutant phenotypes (a growth defect on rich agar and azide sensitivity) that are more severe than those exhibited by a $\Delta secA2$ null mutant (29). Extragenic suppressors of this dominant-negative *secA2* allele map to the *secY* promoter, and increased SecY levels suppress the *secA2* dominant-negative phenotype (30). These findings suggest that the SecA2 dominant-negative protein is locked in a nonproductive interaction with the essential SecYEG channel, which inhibits SecYEG function but can be overcome by increased SecY production. This is consistent with SecA2 working with SecYEG. In a recent study of the SecA2-only system of *L. monocytogenes*, suppressors of a *secA2* mutation also mapped to *secY* (74). Furthermore, the behavior of a dominant-negative SecA1 mutant in the *C. difficile* system is

consistent with the SecYEG translocase used by SecA1 also being used by SecA2 (18). Thus, it seems likely that in these SecA2-only systems, SecY is involved. However, a direct interaction between SecA2 and SecYEG has not been demonstrated in any system.

Previously, the crystal structure of the canonical SecA1 was solved in *M. tuberculosis* (31), as well as several other organisms, including *Escherichia coli* (32), *Bacillus subtilis* (33), *Thermotoga maritima* (34), and *Thermus thermophilus* (35). SecA structures contain five canonical domains, organized roughly in the shape of a barbell: a core helical scaffold domain (HSD), forming the “axis”; 2 nucleotide-binding domains (NBD1 and NBD2), which together form a DEAD box, RecA-like, or superfamily II helicase motor domain on one end of the barbell; and a helical-wing domain (HWD) and preprotein cross-linking domain (PPXD) on the other end of the barbell. In addition, a helix-loop-helix domain called IRA1 (for “intramolecular regulator of ATPase”) packs against the HSD, with helices aligned in parallel. The loop connecting the helices of IRA1 is known as the two-helix finger (2HF). The 2HF has been shown to insert into the SecYEG pore, and it is proposed to promote forward movement of the preprotein through the channel (34, 36), although the interaction between the 2HF and SecYEG could also serve an alternate role besides pushing the translocating protein through the channel (37). During preprotein translocation, SecA undergoes significant conformational changes, one of which involves the orientation of the PPXD domain. According to one model (38), the PPXD likely starts out oriented toward the HWD, forming a hydrophobic “cleft” for binding the signal sequence of the preprotein (39, 40), and then rotates toward NBD2 to form a “clamp” around the translocating polypeptide chain, which has been proposed to be initiated by docking with SecYEG (41).

In order to better understand the unique function of SecA2, we solved the crystal structure of *Mtb*SecA2, which is the first SecA2 structure to be determined in any organism. The structure reveals that the HWD domain is completely absent in *Mtb*SecA2. The HWD could play a role in interacting with protein substrates, as it forms part of a cleft with the PPXD that is implicated in peptide binding (40). Although the residues that directly bind the signal sequence (based on nuclear magnetic resonance [NMR] studies) are contributed by the PPXD and IRA1 domains (40), the HWD would likely be physically proximal to the untranslocated portion of protein substrates. Furthermore, residues in the HWD of *E. coli* SecA (along with the PPXD and HSD) have been shown to crosslink with synthetic signal sequences in cysteine substitution experiments (42). The lack of an HWD in SecA2 leads to a signal sequence binding cleft that is more highly solvent exposed than in SecA1, which we propose could account for recognition of specific SecA2-dependent substrates and prevent export of the larger number of SecA1-dependent preproteins. The structure also reveals conservation in *Mtb*SecA2 of features critical to the interaction between SecA and SecYEG proteins. Finally, by mapping intragenic suppressor mutations onto the SecA2 structure, we show that the mutated residues appear in surface-exposed regions and map to three functional domains that are likely involved in mediating interactions with other protein partners, such as SecYEG.

Materials and Methods

Protein expression and purification. The 778-residue open reading frame (ORF) of *Mtb*SecA2 was cloned into expression vector pNR14. Several genomic databases list *Mtb*SecA2 as having a total length of 808 amino acids (aa) (e.g., NCBI accession no. NP_216337). However, the start site in this annotation is likely to be incorrect, as the first 30 aa are not required for function and represent an N-terminal extension that is not observed in other SecA orthologs (28, 43).

Therefore, we designate the GTG codon corresponding to residue 31 in the NCBI annotation as the true start codon, yielding a total ORF length of 778 aa. The expression construct pNR14 produces a tag-less form of the protein (28). Selenomethionyl protein was produced by transforming the *E. coli* methionine auxotroph B834(DE3) (Novagen) with the pNR14 expression vector. A 6-liter culture was grown under standard conditions to mid-log phase. The cells were pelleted and used to inoculate 12 liters of M9 minimal medium supplemented with 50 mg/liter of L-selenomethionine (SeMet), 50 mg/liter of standard L-amino acids (excluding methionine), 100 nM vitamin B12, and trace elements (44). Expression was induced with 0.5 mM IPTG (isopropyl- β -D-thiogalactopyranoside) at 16°C for 12 h. Cells were then harvested and resuspended in lysis buffer containing 50 mM Tris (pH 8.0), 50 mM NaCl, 1 mM dithiothreitol (DTT), 10 mM MgCl₂, 20 μ g/ml DNase, and 1 x protease inhibitor cocktail V (EMD Biosciences). The cells were disrupted in a BeadBeater (Biospec) using 0.1-mm-diameter glass beads. Cellular debris was cleared from the lysate by spinning at 27,200 x g for 2 h. The supernatant was then filtered and loaded onto a Blue Sepharose column (GE Healthcare) that had been equilibrated in a mixture of 50 mM Tris (pH 8.0), 50 mM NaCl, and 1 mM DTT. Protein collected from the flowthrough was further purified by anion-exchange chromatography using a HiTrap Q high-performance (HP) column (GE Healthcare). The purified protein was dialyzed overnight against buffer containing 50 mM Tris (pH 8.0), 50 mM NaCl, and 1 mM DTT and was then concentrated to 10 mg/ml using a Centriprep centrifugal concentrator (Milipore) and flash frozen until further use.

Crystallization. Purified protein was crystallized in 20% polyethylene glycol 8000 (PEG 8000), 0.1 M Tris (pH 8.0), 0.2 M NaCl, 3% ethylene glycol, and 8 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS). Wells were set up using

sitting-drop vapor diffusion at 21°C, with drops consisting of 2 parts buffer and one part protein. Crystals grew to 100 µm within 3 to 4 days. Perfluoropolyether (Hampton Research) was used as a cryo-protectant. The protein crystallized in space group P21 with the unit cell parameters $a = 39$, $b = 165$, $c = 67$ Å, and $\beta = 97^\circ$. The corresponding unit cell volume can accommodate a single molecule in the asymmetric unit.

Crystal dehydration. A crystal dehydration method was developed that significantly improved the mosaic spread and diffraction power of the crystals (45). Both the well and drop solution were replaced with mother liquor that had a 3 to 5% increase in precipitant concentration. Crystals were left to dehydrate for a minimum of 48 h before making another incremental increase in the precipitant. Successfully dehydrated crystals had a reduced b unit cell parameter of up to 15 Å, with the largest difference resulting in a 10.5% decrease in the unit cell volume. The crystal that produced the best diffraction data and led to structure solution had only a 3-Å difference in the b unit cell parameter.

Data collection, structure determination, and refinement. The structure was solved by single-wavelength anomalous dispersion (SAD) using a selenomethionine (SeMet) derivative (46). Anomalous diffraction data were collected at beamline 23-ID of the GM/CA-CAT facilities of the Advanced Photon Source, Argonne National Laboratory. Crystals were partitioned using the 10-µm minibeam (47). This prevented global-scale radiation exposure and allowed for more data to be collected from a single crystal. The data were processed and reduced using the HKL2000 software package (48). The locations of 3 Se sites were found using SHELX C/D and were used as a starting point for locating additional sites in autoSHARP (49, 50). The resulting experimental phases extended to a 3.8-Å resolution and produced an electron density map in which approximately 60% of the backbone could be placed in NBD1, NBD2, and parts of the

HSD. Model building was performed in Coot (51). The phases from the partial model were then combined with the experimental phases using SigmaA and used as a starting point for progressive runs of density modification in DM (52, 53). This facilitated the placement of the backbone in the PPXD as well as in other parts of the model. Initially, sequence was assigned by the positions of the Se atoms and from the density of large side chains. Then a real-space cross-validation procedure called “ping-pong” cross-validation was used to complete the structure (54). Briefly, the model was split into two sets. Side chains that could be identified in the first set of residues were used during phase combination and density modification. The resulting map was used to place side chains for the second set of residues, and the process continued in alternation. Structure refinement was carried out in autoBuster (55).

Suppressor screen and reconstruction. Spontaneous suppressors of the *secA2 K129R* strain were isolated by plating onto Mueller-Hinton agar at 37°C, as described previously (30). The *secA2 K129R* strain has the chromosomal *secA2* gene deleted and carries a copy of the *secA2* gene encoding SecA2 K129R integrated at the chromosomal L5 *att* site. The *secA2 K129R* gene of the suppressors was PCR amplified and sequenced to identify intragenic suppressor mutations. To confirm that suppressor phenotypes were due to sequenced mutations in *secA2 K129R*, the intragenic suppressors were recreated in a fresh strain background. PCR-amplified *secA2 K129R* gene products from the intragenic suppressors were subcloned into pCR2.1 followed by cloning into pMV306. The resulting vectors were electroporated into the $\Delta secA2$ mutant of *M. smegmatis*, and transformants were tested for sensitivity to sodium azide and SecA2 localization.

Azide sensitivity assay. Cultures were plated for sensitivity to sodium azide as previously described (29). In brief, 200 μ l of saturated (optical density at 600 nm [OD₆₀₀] of 2.0) *M. smegmatis* culture was mixed with 3.5 ml of molten 7H9 top agar and then poured onto a 7H10

bottom agar plate lacking Tween. Sterile 6-mm-diameter filter discs were placed onto the surface of the cooled top agar. Ten microliters of 0.15 M sodium azide was then added to the disc. The plates were incubated for 2 days at 37°C, and the resulting zones of growth inhibition were measured. Each strain was tested in triplicate.

Subcellular fractionation and immunoblotting. To determine the subcellular localization of SecA2 in *M. smegmatis*, we fractionated bacterial whole-cell lysates as described previously (22, 29). Whole-cell lysates were generated by five passages through a French pressure cell. The lysates were separated into cell envelope (100,000 x g pellet) and soluble (100,000 x g supernatant) fractions. Protein derived from the same amount of starting cells for each fraction was analyzed by SDS-PAGE and immunoblots using an anti-SecA2 antibody at a 1:20,000 dilution (56). For quantification, secondary antibody conjugated to alkaline phosphatase was used and detected using the ECF reagent (GE Healthcare). Fluorescence was quantified using a phosphorimager and ImageQuant 5.2 (Molecular Dynamics).

Protein structure accession number. The structural coordinates of the *Mtb*SecA2 model used in this study have been deposited in the Protein Data Bank under identification no. 4UAQ.

Results

Crystal structure of *Mtb*SecA2. *Mtb*SecA2 (Rv1821) was crystallized in space group P2₁, and the structure was solved by single-wavelength anomalous dispersion (SAD) (57) to a resolution of 2.8 Å. The asymmetric unit of the crystal contains a single monomer, and there is no indication of a higher-order oligomer in the crystal lattice. A total of 705 out of 778 residues of the apoprotein were visible in the electron density and could be built. The crystallographic statistics are shown in Table 2.1.

Broad structural similarity between *Mtb*SecA1 and *Mtb*SecA2. The tertiary structure of SecA2 is very similar overall to those of *Mtb*SecA1 and other orthologs in the SecA family (Figure 2.1). SecA2 has a long 65-Å (45-amino-acid) helix scaffold domain (HSD), which interconnects four other domains, including two nucleotide-binding domains (NBD1 and NBD2), the IRA1 domain, and the PPXD domain. NBD1 and NBD2 pack together to form a DEAD box motor domain with an ATP-binding site between them. Catalytically important residues such as K115 and R545 are conserved, consistent with demonstrated ATPase activity of SecA2 (28). As in other SecA structures, the IRA1 domain consists of a pair of alpha-helices packed in parallel to the HSD (forming a 3-helix bundle) and connected by a 9-aa loop (known as the 2-helix finger [2HF]). SecA2 lacks the ~70-aa C-terminal domain (CTD) that is present in SecA1 orthologs. However, the short linker to this domain, called the C-terminal linker (CTL [residues 734 to 778]), is retained in the SecA2 sequence. The CTL is largely disordered in the crystal structure. However, as observed in previous SecA structures (33) part of the CTL of SecA2 (residues 749 to 759, shown in yellow in Figure 2.1) forms a third β -strand along the outside of the preprotein binding site. Note that this region is preceded by a disordered loop (residues 734 to 748), which appears as a discontinuity between IRA1 and CTL in the figure, and followed by only 19 residues at the C terminus, which are also disordered. During model building, sequence assignment in this strand was aided by the location of SeMet757 and the density of bulky side chains, which helped to rule out the possibility of a bound preprotein substrate.

Differences between the structures of *Mtb*SecA1 and *Mtb*SecA2. Despite the overall similarity between the structures of SecA1 and SecA2, there are several notable differences. One structural difference between SecA1 and SecA2 is found in the nucleotide-binding region. SecA2 lacks the VAR domain (58), which in other SecA orthologs consists of a pair of helices that

reach out from NBD2 and cover over the ATP-binding site (Figure 2.2). Consequently, the ATP binding site is more solvent exposed in SecA2. The VAR domain is present in some SecA orthologs, including *Mtb*SecA1 (21) and *E. coli* SecA (32), but it is absent in others, such as *B. subtilis* SecA (33) and *T. maritima* SecA (*Tm*SecA) (39). The functional significance of the absence of the VAR domain in SecA2 is unknown.

A second structural difference involves the orientation of the PPXD domain. As in other SecA structures, the PPXD domain consists of an α + β -fold that is attached to the NBD1 motor domain by a pair of anti-parallel β -strands that cross over the HSD. The PPXD of *M. tuberculosis* SecA2 occupies a distinct orientation compared to previous SecA structures, as illustrated in Figure 2.3. The PPXD in previous SecA structures has been observed in several different orientations, ranging from contact with the HWD (to form a “signal sequence binding-cleft closed” conformation, as observed in 1nl3) to contact with NBD2 (to form a “preprotein clamp closed” conformation, as observed in PDB code 3DIN) (38, 39). The PPXD in SecA2 occupies an intermediate position between these two extremes.

The most striking structural difference in SecA2 is that the HWD is missing (Figure 2.2) due to deletion of 70 aa that form a helical domain at the end of the HSD, as anticipated from the sequence alignment. In SecA2, the remaining 23 residues connect the HSD directly to IRA1, bypassing the helical wing domain. In other SecA structures, including *Mtb*SecA2, the body of the HWD forms a deep hydrophobic cleft with PPXD, which can be open or close against it (39), with the signal sequence binding site at the base (formed by residues from PPXD and IRA1) (39, 40). The absence of the HWD in SecA2 makes the cleft significantly more open and solvent exposed (Figure 2.4), which could help SecA2 recognize its unique substrates that are

distinguished by features of their mature domains – possibly a tendency to fold prior to export (23).

The functionally important two-helix finger (2HF), which is a 9-residue loop connecting two helices in the IRA1 domain that inserts into the SecYEG pore, is conserved in the *MtbSecA2* structure (residues 695 to 703) (Table 2.2). However, the 2HF loop in *MtbSecA2* adopts a different three-dimensional conformation compared to previous structures. In the *MtbSecA2* structure, the 2HF is observed to close down approximately 10 Å onto the HSD, like a jaw hinge (Figure 2.5), due to differences in how the ends of the helices unwind (even though the 2HF amino acid sequence itself is highly conserved, as shown in Table 2.2). This orientation contrasts with the conformation observed in most other SecA structure, in which the loop is more flipped out into solvent (Figure 2.5); however, the conformations of the 2HF loop are also quite variable among SecA crystal structures. Fluorescence studies also suggest that the 2HF loop is flexible and can adopt different conformations in solution (37).

Similarities between SecYEG binding regions of *M. tuberculosis* SecA1 and SecA2. The conservation of the overall structure of SecA2 is consistent with a model in which SecA2 works with SecYEG to translocate SecA2-dependent proteins across the membrane. Furthermore, the key regions of SecA2 that would interact with the SecYEG pore are conserved, including the 2HF. The helix-terminating proline in the 2HF is present in SecA2 (Pro703), as it is in all SecA homologs (Table 2.2). Tyr794 in *E. coli* SecA is another critical residue in the 2HF (36). Although it is substituted for by Leu698 in *MtbSecA2*, this tyrosine is substituted by large hydrophobic residues in 20% of SecA homologs (methionine in *MtbSecA1*). Furthermore, structural data from the *TmSecA*-SecYEG complex supports that hydrophobic substitutions, such

as leucine, can be accommodated at this position, as the side chain sits in a hydrophobic pocket in SecY (34).

Structural superposition of *Mtb*SecA2 onto *Tm*SecA in the *Tm*SecA-SecYEG complex (PDB code 3DIN [34]) (Figure 2.6) further indicates that SecA2 preserves many of the structural features of SecA implicated in binding to SecYEG. This includes amino acids in the *Mtb*SecA2 2HF and immediately adjacent regions of IRA1 that contact SecY in the *Tm*SecA-SecYEG complex (aa 687 to 715 in SecA2) (Table 2.3 and Figure 2.6). There are also regions of NBD2 and the HSD that are structurally conserved in the SecA2 structure and positioned for contact with SecY (Table 2.3 and Figure 2.6). These residues in NBD2, IRA1, and the HSD are clustered at the interface with SecY. In addition, although the PPXD of SecA2 is rotated away and does not appear to make direct contact with SecY in the superposition, if it were rotated into an orientation similar to that observed in *Tm*SecA in the complex, it would place additional SecA2 residues (listed in Table 2.3) in contact with SecY, as shown in Figure 2.6. It is notable that *Mtb*SecA2 D607 (in the HSD) corresponds to one of the residues in *E. coli* SecA (position 640) that can be cross-linked with SecY using photoactivatable unnatural amino acids (60).

Mapping of suppressor mutations on the SecA2 structure. Prior studies indicate that a SecA2 dominant-negative protein with an amino acid substitution in the ATP binding Walker box, making it unable to bind ATP, is locked in a nonfunctional complex, likely with SecYEG, at the membrane (29). In order to identify important residues in SecA2, we identified intragenic suppressor mutations that could overcome the *secA2* dominant-negative phenotypes (30) with the rationale being that such mutations might map to sites of protein interactions in SecA2 complexes. For convenience, these experiments were performed with the *M. smegmatis* ortholog of SecA2, which has 83% amino acid identity to *Mtb*SecA2 and is able to substitute for the

*Mtb*SecA2 in cross-species complementation experiments (29). An *M. smegmatis* strain expressing the dominant-negative *Ms*SecA2 K129R, which has an amino acid substitution in the Walker box (equivalent to K115 in *Mtb*SecA2) was used. All suppressors identified reversed the severe dominant-negative phenotypes caused by SecA2 K129R, as assessed by azide sensitivity assays (5) and colony size on rich agar (30) (Figure 2.7 and data not shown), but they still exhibited a phenotype similar to that of a $\Delta secA2$ null mutant.

Eight independent suppressors with mutations in the coding sequence of *secA2 K129R* were identified by sequencing, mapping to four different domains: NBD1, NBD2, PPXD, and IRA1 (Table 2.4). All eight suppressor mutants produced full-length SecA2 protein at normal levels, as confirmed by Western blot analysis. Each mutation was validated to be responsible for the suppression by retesting the phenotype of individual mutations when introduced into a fresh *secA2 K129R* mutant background (Figure 2.7).

When mapped to the SecA2 structure, all of the suppressor mutations were located on the surface of the protein (Figure 2.8). For simplicity, below we will refer to the suppressors using amino acid numbering that corresponds to *Mtb*SecA2 (Table 2.4). There were three categories of suppressors. The first set of suppressor mutations affected the same surface loop of NBD1. There were two suppressors derived from independent cultures with identical mutations in NBD1 and a third suppressor with a different mutational alteration that mapped to the same site in NBD1. These NBD1 suppressors involve a 4-residue loop, ¹⁶⁸STPD¹⁷², in *M. tuberculosis* connecting a β -strand and an α -helix; this loop was deleted in one mutant and duplicated in another. It is currently unknown what role these residues play, but it is striking that three out of eight suppressor mutations involved this surface-localized loop of the nucleotide binding domain, suggesting it is a functionally important point of contact for SecA2.

The second group of suppressors (three in total) clustered in the SecA “polypeptide clamp” region made up of PPXD and NBD2 domains. Two suppressor mutations mapped to the SecA2 PPXD domain: a nonsynonymous substitution, D316H, and an insertion of a second glutamate at E354. These amino acids are in separate loops in the PPXD domain, but they are proximal in the three-dimensional structure, approximately 7 Å apart (Figure 2.8a). The PPXD is positioned far from the NBD2 domain in the SecA2 structure. The distance between the closest residues of the two domains is 23 Å, representing a “clamp open” state. However, in the *Tm*SecA-SecYEG complex, the corresponding PPXD loops to which these suppressor mutations map come in contact with NBD2. Moreover, the *Tm*SecA residue corresponding to the D316H suppressor in the *Mtb*SecA2 PPXD is in direct contact with NBD2 in the *Tm*SecA-SecYEG complex (34) (Figure 2.8b). It should be noted that this *Tm*SecA complex with SecYEG represents an extreme conformation (induced by ADP and BeFx in the crystallization buffer) in which the preprotein channel is entirely collapsed (i.e., a loop of the PPXD actually inserts into the preprotein binding channel). In a structure of SecA bound to a preprotein substrate (PDB code 3JV2 [61]), the PPXD does not rotate quite as far toward NBD2 as in the SecA-SecYEG complex, but the residues corresponding to the suppressor mutations are still on the surface of the PPXD in a region that would be in position to interact with SecYEG or the lipid bilayer (similar to the red residues highlighted in Figure 2.6). Thus, these suppressor mutations could disrupt intramolecular interactions when PPXD rotates to form the “clamp” around the translocating polypeptide (41) or could lock it in the extreme closed state such that the preprotein channel is collapsed altogether. Strikingly, the NBD2 suppressor T449I in *M. tuberculosis* also maps to the SecA “preprotein clamp” region and is proximal (within 10 Å) to the two PPXD suppressor mutations when the clamp is closed (based on the analogous residues in the *Tm*SecA-

SecYEG docked structure [34]) (Figure 2.8b). Thus, these three suppressors in NBD2 and PPXD could conceivably cause a defect in clamp closure during translocation. In light of past studies suggesting that interactions between SecA2 K129R and SecYEG are responsible for the dominant-negative phenotype (30), these results suggest that a defect in clamp closure may dislodge or prevent SecA2 interactions with SecYEG by disrupting interactions with the polypeptide being translocated through the channel.

The final group of intragenic suppressors identified have deletions in IRA1. One suppressor has a deletion of residues 714 to 721 in *M. tuberculosis* and another suppressor has a very similar, yet distinct, deletion of residues 712 to 719 in *M. tuberculosis*. These deletions are in the middle of one of the α -helices, just downstream from the 2HF that forms part of the interface with SecYEG (Figures 2.5 and 2.7, and Table 2.3), and similar mutations in IRA1 have previously been shown to disrupt binding to SecYEG (62). Furthermore, one of the deleted residues in both of the IRA1 suppressors is Phe715, which is a conserved residue predicted to contact SecY (colored red in Figure 2.6) that is equivalent to the highly conserved Phe798 (in *TmSecA*). In the *TmSecA*-SecYEG structure, Phe798 (in *TmSecA*) forms an aromatic stacking interaction with Tyr418 in the C-terminal tail of *TmSecY* (34). This interaction appears to be crucial to docking as the equivalent tyrosine residue in *E. coli* SecY (Tyr429) is the location of a cold-sensitive mutation that prevented insertion of SecA into the membrane channel (63). These interacting residues are highly conserved in all Sec systems, including *MtbSecA2* (Phe715) and *MtbSecY* (Tyr436). The fact that this group of intragenic *secA2* suppressor mutants harbors deletions in a structurally conserved and critical SecY-interacting region of IRA1 (Figure 2.6 and Table 2.3) is consistent with their mode of suppression being avoidance of complex formation between SecA2 K129R and SecYEG.

Intragenic suppressors alter membrane localization of the dominant-negative SecA2. In the wild type, *MsSecA2* is predominantly found in the soluble cytoplasm-containing fraction. In contrast, the localization of SecA2 K129R is almost exclusively in the membrane-containing cell envelope pellet (29) (Figure 2.9a). This is consistent with a model for SecA2 K129R being locked in a protein complex with SecYEG at the membrane. Since we predicted that some of the intragenic suppressors alleviate SecYEG interactions, we determined the membrane localization of SecA2 K129R in the intragenic suppressor mutant background. Strains were lysed and then fractionated into cell envelope (pellet) and soluble (cytoplasmic) fractions. Western blot analysis with anti-SecA2 antibodies on fractions was then carried out to localize the protein. In each of the representative intragenic suppressors analyzed, the distribution of SecA2 K129R shifted from the envelope, as seen in the starting *secA2 K129R* strain, to the soluble cytoplasmic fraction (Figure 2.9b). Suppressor mutations in the “clamp” (PPXD and NBD2) and IRA1 domains had the most dramatic effects, restoring partitioning of SecA2 between the cell envelope and cytoplasm to almost wild-type levels. These data support a model in which the intragenic suppressor mutations alleviate the dominant-negative phenotype by disrupting protein-protein interactions involving the SecYEG membrane complex and/or the translocating polypeptide.

Discussion

Over 30 years ago, SecA was identified as a critical component of the protein export system of bacteria (64). Since that time, there have been extensive genetic, molecular, biochemical, biophysical, and structural studies to understand SecA function. Of the two SecAs in *M. tuberculosis*, SecA1 is the counterpart of the well-studied canonical SecA, while SecA2 has a distinct function from SecA1 and nonoverlapping substrate specificity profile. The structure of *MtbSecA2* we report is the first structure of any SecA2 protein. The broad structural

similarity observed between the two solved *Mtb*SecA structures indicates that, even after decades of mechanistic studies, gaps in our understanding of SecA proteins remain.

The smaller size of SecA2 versus SecA1 and canonical SecA proteins appears to come from the absence of the HWD, the VAR domain and a C-terminal domain (CTD [although it still retains the CTL linker]) reducing the overall size of the protein product from 949 aa to 778 aa. The lack of an HWD is the most striking structural difference in SecA2. Without the HWD, the signal sequence recognition site of SecA2 is more solvent exposed and thus more accessible to protein substrates. This structural difference may help explain the ability of SecA2 to export substrates with distinctive features of their mature domain, possibly a propensity to fold prior to export (23). The “open” nature of the cleft created by the absence of a HWD could provide a broad surface against which folded proteins could possibly dock and unfold for translocation through the SecYEG transmembrane pore. Several pieces of experimental evidence support the possibility that the HWD could interact with preproteins. While a nuclear magnetic resonance (NMR) structure of SecA bound to a signal sequence did not identify any residues of the HWD that directly interact with the signal sequence (40), several residues of the HWD were found to form cysteine-based crosslinks with a synthetic signal sequence (42), which may result from transient states (i.e., alternative conformations of the HWD) sampled dynamically in solution. In fact, the HWD is observed to rotate by up to 15° between different crystal structures, depending on the oligomeric state (39). This suggests the HWD itself is mobile in solution, which is supported by fluorescence-based (fluorescence resonance energy transfer [FRET]) studies (65). Furthermore, the mobility of the HWD appears to be influenced by the presence of a preprotein (66). Thus, the absence of the HWD in SecA2 could potentially affect substrate recognition. The 70-residue deletion of the HWD observed in *Mtb*SecA2 is a general feature among

actinomycetes (including *Mycobacterium* and *Corynebacterium* species) . It should be noted that other Gram-positive SecA2 proteins also appear to have a truncated version of this domain (deletions of 13 to 18 residues for *Streptococcus gordonii* and *L. monocytogenes*, respectively). Until structures of these other SecA2 orthologs are solved, the potential consequences of these HWD truncations remain unknown. It is possible that a reduced HWD could open up the signal sequence binding cleft and/or increase the site of interaction with preproteins, as we propose for *MtbSecA2*. To achieve a complete picture of SecA2 function going forward, the consequences of a truncated or deleted HWD will need to be explored in both mycobacterial and Gram-positive SecA2 proteins.

The significance of the absence of the VAR domain in the SecA2 NBD region is less clear. The lack of the VAR domain leaves the nucleotide-binding site relatively solvent exposed. While other SecA2 orthologs also lack the VAR domain, one-third of bacterial SecA (1) proteins lack this domain as well (58). In *E. coli* SecA, the VAR domain has been shown to regulate ATPase activity and ADP release, as *secA Δvar* mutants display higher ATPase activity and faster ADP release rates (58). However, *MtbSecA2* (28) was recently reported to release ADP more slowly (not more quickly) than the VAR-containing *MtbSecA1* (67).

Mycobacterial SecA2 proteins, as well as SecA2s in many other organisms, lack the C-terminal domain (CTD). The CTD in SecA1 proteins consists of a tail of 70 aa that is disordered in all previous crystal structures (1). In most bacteria, the CTD of SecA contains a Zn²⁺ finger domain that binds to the protein export chaperone SecB (68). Mycobacteria are an exception, in that the CTD of SecA1 does not contain the conserved cysteines of a Zn²⁺ finger motif. However, this may not be too surprising because, like Gram-positive bacteria (69), no SecB ortholog with a function in protein export has so far been identified in mycobacteria. Thus, because of the lack of

Zn²⁺ finger motif in the CTD of SecA1 and lack of a SecB ortholog, the absence of a CTD in *Mtb*SecA2 seems unlikely to be a significant contributing factor to the unique function of SecA2.

In comparison to all prior SecA structures, the SecA2 structure also revealed new orientations of the PPXD and the 2HF loop. However, these differences probably reflect the conformational plasticity of these two structural elements. Given the mobility of the PPXD domain already established for canonical SecA proteins, it seems likely that the PPXD orientation observed in SecA2 represents a previously unobserved structural intermediate in the transition of the preprotein binding clamp from the open to closed position (38). The unique orientation of the 2HF loop observed in SecA2, which occurs at a key point of interaction with the translocation channel and varies considerably among SecA structures, is probably a consequence of the flexibility of this loop in solution.

Given that there is no corresponding SecY2 partner in the *M. tuberculosis* genome, an important mechanistic question to be answered is whether SecA2 works with the canonical SecYEG channel to export proteins. In prior studies, we described a dominant-negative *secA2* mutation that exhibits more severe phenotypes than a $\Delta secA2$ deletion mutant (29). Such phenotypes often result from a dominant-negative protein being locked in a nonproductive complex with its normal binding partners. Furthermore, we showed extragenic suppressors that overexpress SecY suppress the *secA2* K129R dominant-negative phenotype, which argues for an interaction between SecA2 K129R and SecY (30). Here, we identified intragenic suppressors of *secA2* K129R, and all of them mapped to the surface of the SecA2 structure. One group of suppressors mapped to the IRA1 domain of SecA2 in regions where similar mutations disrupt *E. coli* SecA binding to SecYEG (62). These IRA1 suppressors also restored cytoplasmic localization of SecA2 K129R. These results can be explained by the IRA1 suppressor mutations

preventing SecA2 K129R interactions at the membrane SecYEG channel, and they support the model for SecA2 working with SecY to promote export of its specific substrates. The suppressors that mapped to the “polypeptide clamp” region of SecA could similarly suppress the dominant-negative phenotype. However, in this case, the suppression would result from the inability of SecA2 to trap the translocating polypeptide in the center of the SecYEG channel, causing SecA2 to fail to engage SecYEG (without the substrate) or causing the ternary system (SecYEG-SecA2-preprotein) to dissociate.

The SecA structure reported here is of a monomer. In other studies, SecA proteins have been crystallized as monomers (54) or dimers (39), and the issue of the oligomeric state of SecA during protein translocation has remained controversial (37, 70–72). A recent study demonstrated the ability of recombinant *Mtb*SecA1 and *Mtb*SecA2 to physically interact in vitro (73). If SecA1-SecA2 heterodimers form, it is possible that interactions between SecA1 and SecY avoid the need of SecA2 to directly interact with SecY. However, it is currently unclear if SecA1-SecA2 dimers exist and/or are functional in mycobacteria. Furthermore, the dominant-negative SecA2 phenotypes and the intragenic suppressors reported here, combined with structural conservation of SecA-SecY contact sites in SecA2, argue for the ability of SecA2 and SecY to interact. Ultimately, to clarify the mechanistic details of SecA2-dependent protein export, it will be necessary to study the pathway with an in vitro reconstitution system, as was used to dissect the mechanistic details of the *E. coli* Sec pathway.

Since the SecYEG channel requires that proteins be unfolded for translocation (25), the possibility of SecA2 working with the SecYEG channel is intriguing, in light of experiments suggesting that SecA2 substrates are distinguished by a tendency to fold in the cytoplasm (23). The role of SecA2 could be to promote recognition of proteins that would normally be

overlooked by the canonical SecA1-SecYEG translocase or to help maintain proteins in an unfolded state prior to or during export. The regions of structural difference and suppressor mutations identified in this study represent exciting new directions for exploring the functional differences between SecA2 and SecA1 proteins.

Attributions

The experiments in this work performed by me are described in Figure 2.7, Table 2.4 and Figure 2.9. Other experiments were performed by the authors listed in the footnote on page 47.

Table 2.1. Data collection and refinement statistics for SAD (SeMet) structure of SecA2

Parameter	Value(s) for SecA2 ^a
Data collection	
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	39.60, 162.09, 67.31
α , β , γ (°)	90.00, 95.87, 90.00
Wavelength (Å)	0.97949
Resolution (Å)	35.64-2.8 (2.85-2.8)
R _{sym} or R _{merge} (%)	0.096 (0.171)
Avg <i>I</i> / σ (<i>I</i>) ^c	24.3 (1.4)
Completeness (%)	0.72 (0.168)
Redundancy	4.1 (1.9)
Refinement	
Resolution (Å)	2.8
No. of reflections	16,255
R _{work} /R _{free}	0.291 (0.212)
No. of atoms	
Protein	4,894
Water	67
<i>B</i> -factors	
Protein	85.2
Water	58.3
RMSD ^b	
Bond length (Å)	0.01
Bond angle (°)	1.26

^a Values in parentheses are for the highest-resolution shell.

^b RMSD, root mean square deviation.

^c *I*, intensity of a reflection.

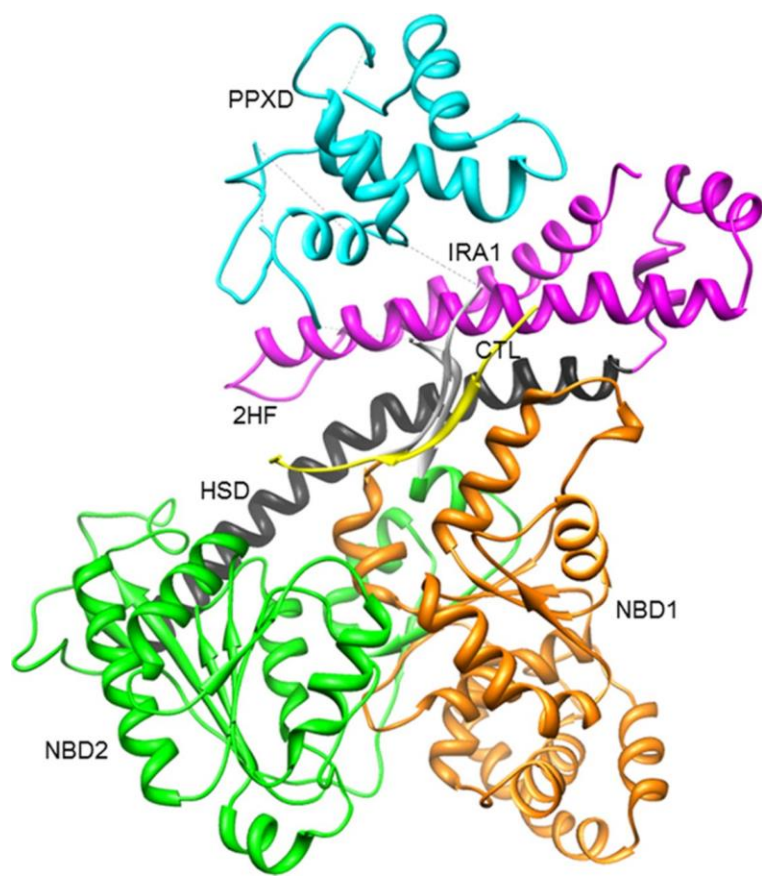


Figure 2.1. Domain architecture of *MtbSecA2*. Orange, NBD1; green, NBD2; cyan, PPXD; magenta, IRA1; black, HSD; yellow, C-terminal linker (CTL).

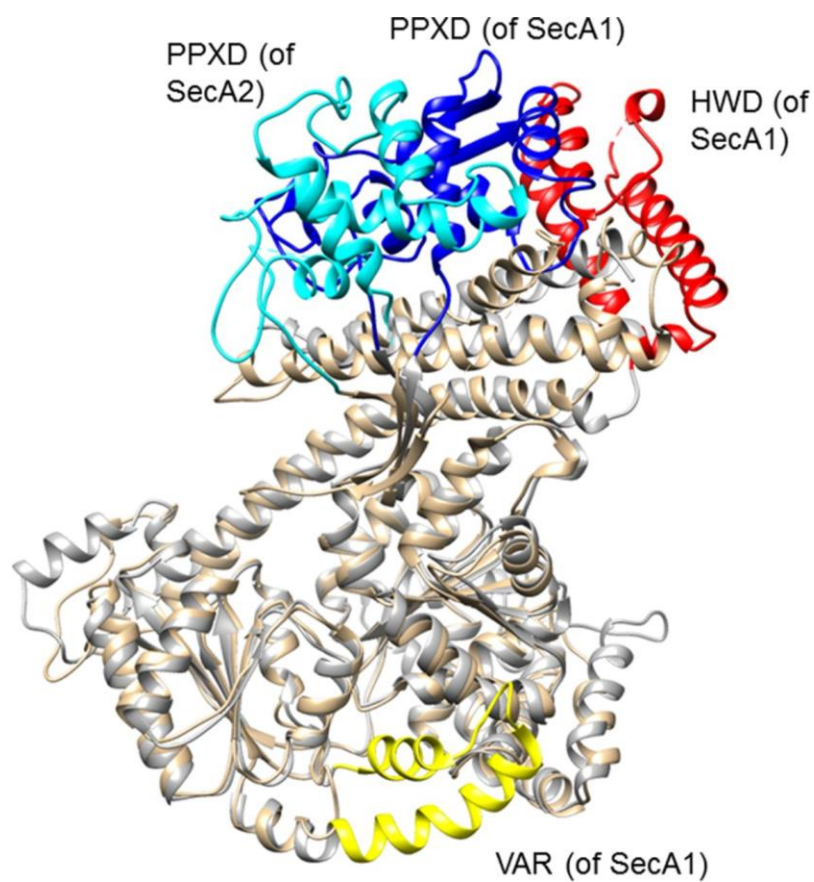


Figure 2.2. Comparison of *Mtb*SecA2 with *Mtb*SecA1. Relative to SecA1 (gray backbone) (PDB code 1NL3), SecA2 (tan backbone) is smaller, lacking the HWD (red) and the VAR domain (yellow). Also, the PPXD domain has undergone rotation (SecA1 [blue] → SecA2 [cyan]).

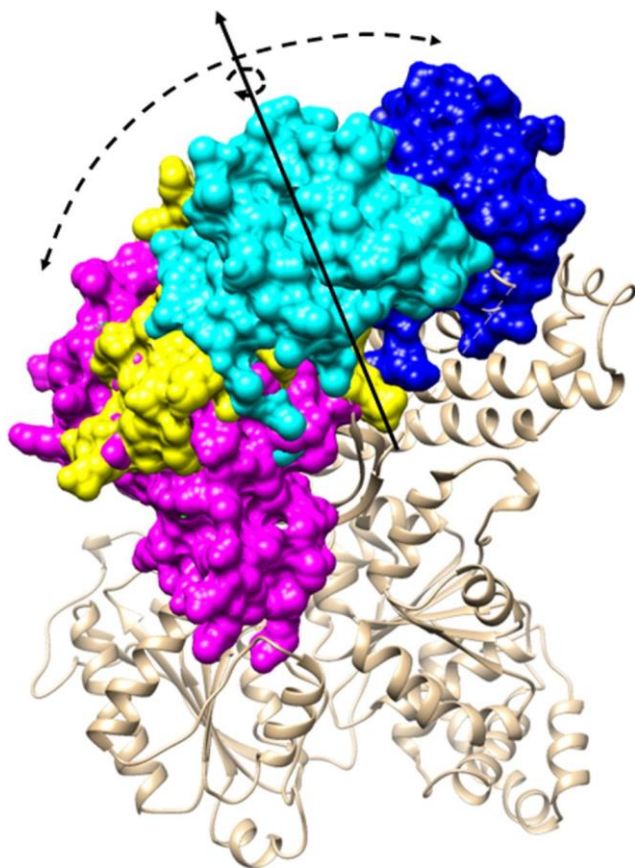


Figure 2.3. Comparison of different orientations of the PPXD domain. Shown are *Mtb*SecA2 (cyan), *Mtb*SecA1 (PDB code 1NL3) (dark blue), *B. subtilis* SecA (PDB code 1TF2) (yellow), and the *Tm*SecA-SecYEG complex (PDB code 3DIN) (red). All four PPXD domains are superposed onto the body of *Mtb*SecA2 (orange). At one extreme, in the *Mtb*SecA1 structure (right, dark blue), the PPXD is packed against the HWD (missing in SecA2), representing the signal sequence-recognition site closed conformation. At the other extreme (left, red), the PPXD from the *T. maritima* complex with SecYEG represents the “preprotein clamp closed” configuration, where contact is made with NBD2 (orange, lower left). The *Mtb*SecA2 PPXD occupies a unique intermediate position (cyan).

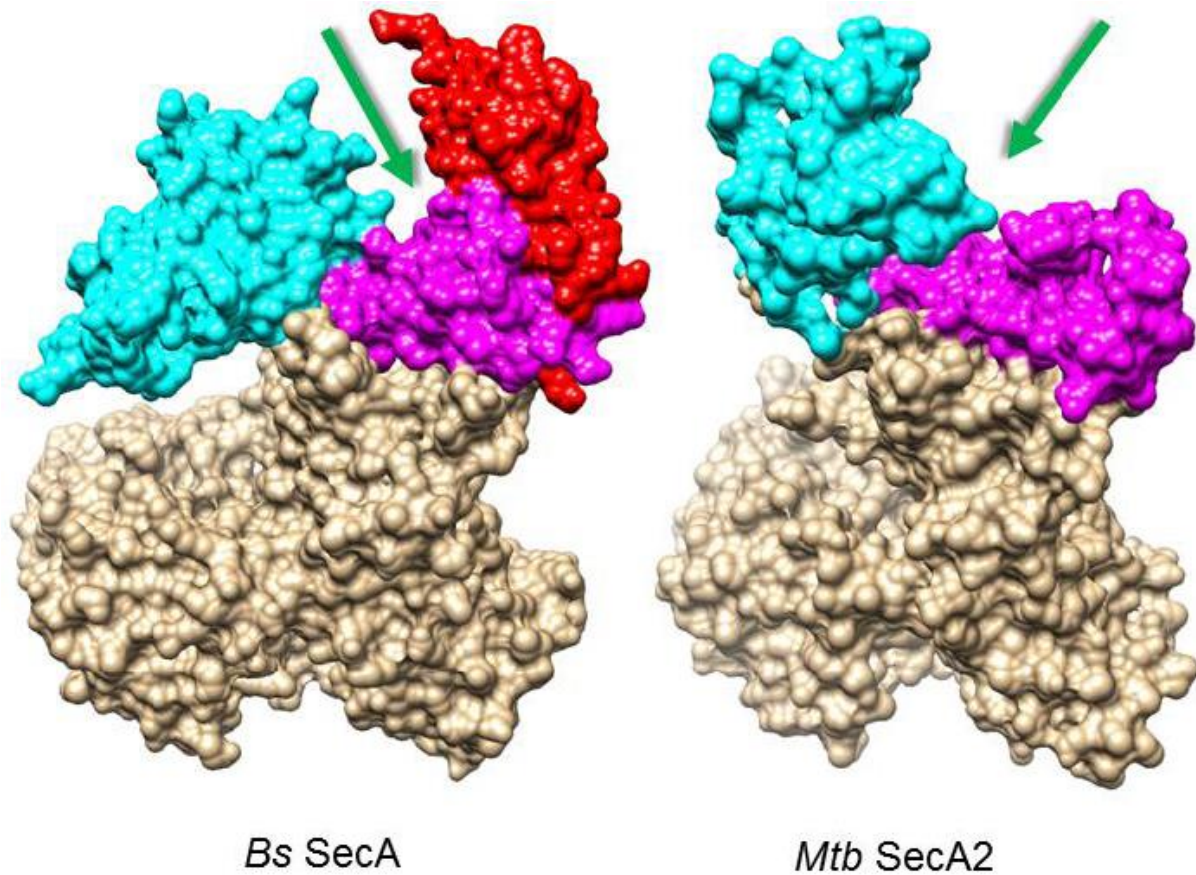


Figure 2.4. Illustration of how deletion of the HWD in SecA2 makes the signal peptide binding cleft more solvent-exposed. *B. subtilis* SecA (PDB ID: 1TF2) is shown on the left. NMR data suggests the signal sequence binds in the cleft formed between the PPXD (cyan), HWD (red), and IRA1 (magenta), indicated by the green arrow. *MtbSecA2* is shown on the right. The green arrow shows the relatively more open signal peptide binding site created by the absence of the HWD.

Table 2.2. Conservation of the 2HF among SecA homologs

Species	Protein	Sequence ^a
<i>E. coli</i>	SecA1	LRGY <u>A</u> QKDP
<i>T. maritima</i>	SecA1	LR <u>S</u> YGQKDP
<i>M. tuberculosis</i>	SecA1	LR <u>A</u> MAQRDP
<i>M. smegmatis</i>	SecA1	LR <u>A</u> MAQRDP
<i>M. tuberculosis</i>	SecA2	LRAL <u>G</u> RQNP
<i>M. avium</i>	SecA2	LRAL <u>G</u> RQNP
<i>M. smegmatis</i>	SecA2	LRAL <u>G</u> RQNP
<i>S. aureus</i>	SecA2	LR <u>S</u> YAQQNP
<i>L. monocytogenes</i>	SecA2	LRAY <u>G</u> QIDP
<i>S. gordonii</i>	SecA2	LRGY <u>A</u> QNNP
<i>C. difficile</i>	SecA2	LK <u>S</u> YAQKDP
<i>C. glutamicum</i>	SecA2	LRAI <u>A</u> RETP

^a Key residues are underlined.

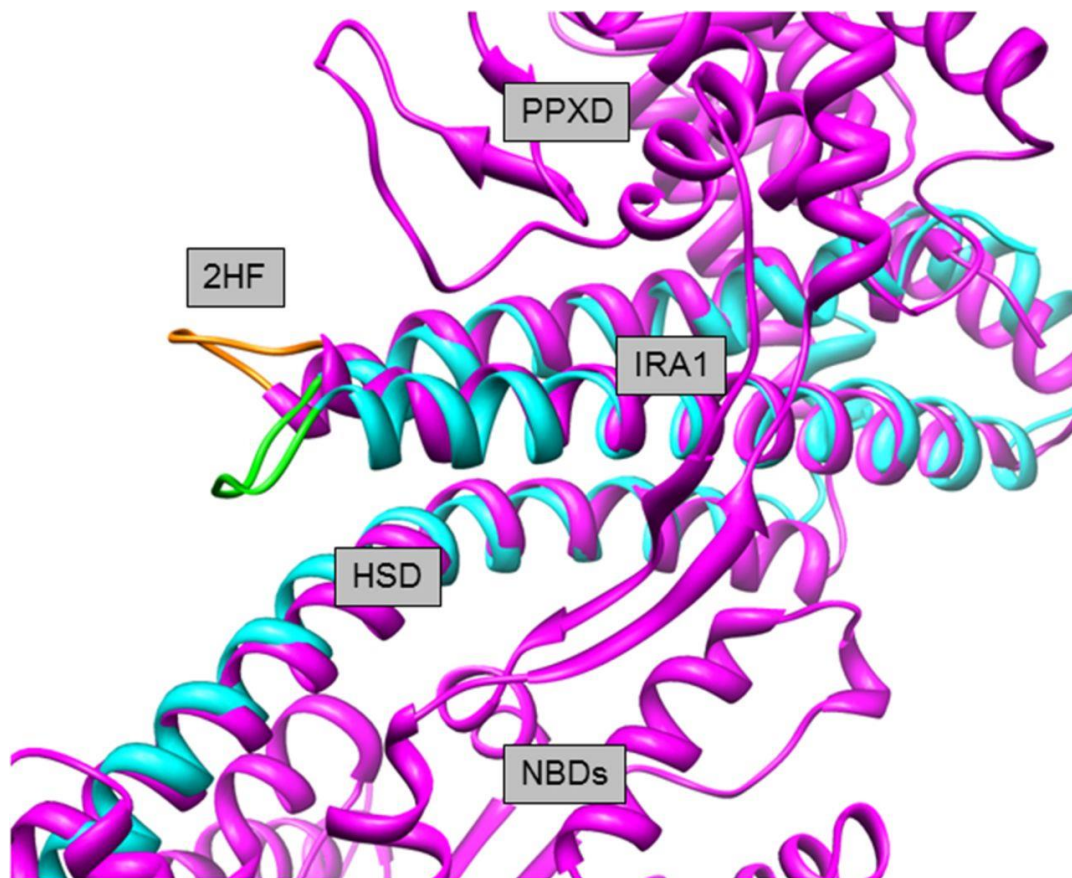


Figure 2.5. Conformation of the two-helix finger (2HF). The HSD of *MtbSecA2* (cyan) is shown superposed on the apo structure of *MtbSecA1* (purple). The loops connecting the two helices are shown in orange (SecA1) and green (SecA2).

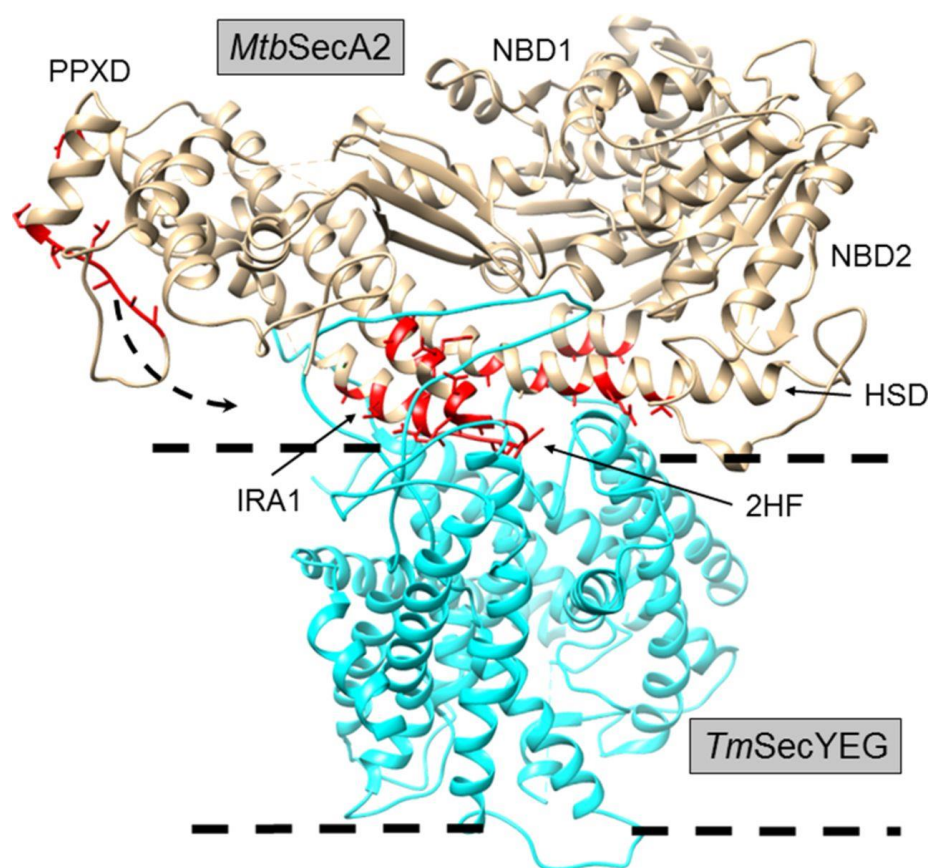


Figure 2.6. Interface residues (red) of *MtbSecA2* (tan) that would contact SecYEG (cyan), based on superposition with *TmSecA* (PDB code 3DIN). Note that the red residues highlighted in the PPXD correspond to residues of *TmSecA* that contact SecYEG as its PPXD is rotated into contact with SecYEG.

Table 2.3. Residues of *Mtb*SecA2 predicted to be in contact with SecY based on structural superposition with *Tm*SecA in complex with SecYEG (PDB code 3DIN)

Domain	<i>Mtb</i> SecA2 residues predicted to contact SecY
NBD1	None
NBD2	E392, R395, Q396
HSD	V600, R604, D607, A610, R614
IRA1	Most residues spanning 687-715 (includes residues of the 2HF and surrounding IRA1 helices)
PPXD	N270, H272, T274, E275, D289

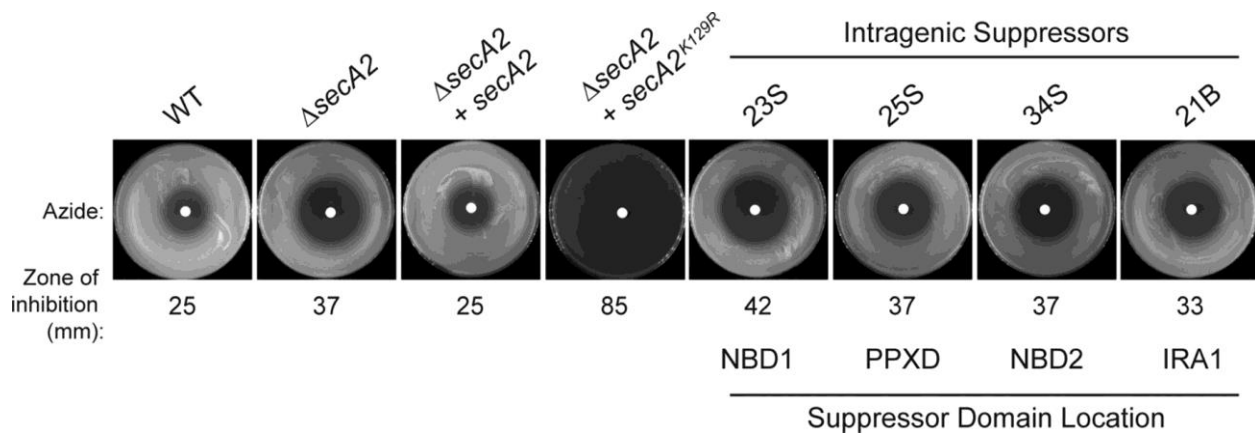


Figure 2.7. Intragenic suppressors suppress the azide sensitivity phenotype of *secA2 K129R*. Lawns of the indicated strains were plated and tested for sensitivity to 10 μ l of 0.15 M sodium azide (applied to a paper disk in the center of the plate) for 2 days at 37°C. Average inhibition was calculated by measuring the diameter of the zone of azide inhibition, and values are the means of three biological replicates. The $\Delta secA2$ mutant *M. smegmatis* strain was transformed with plasmids containing either *secA2*, *secA2 K129R*, or a reconstructed intragenic suppressor with the *secA2 K129R* mutation in combination with an intragenic suppressor mutation located in one of the following domains: NBD1, NBD2, PPXD, or IRA1.

Table 2.4. Suppressor mutations observed in *M. smegmatis* SecA2 K129R

Isolate(s) ^a	Effect on residue(s) in <i>Ms</i> SecA2	Domain	Corresponding residue(s) in <i>Mtb</i> SecA2
6S, 9S	Deletion of residues 182-185	NBD1	168-171 (STPD)
23S*	Duplication of residues 182-185	NBD1	168-171
2S	Asp326 → His	PPXD	D316
25S*	Glu insertion at residue 364	PPXD	E354
34S*	Thr459 → Ile	NBD2	T449
21B*	Deletion of residues 734-741	IRA1	714-721
38S	Deletion of residues 732-739	IRA1	712-719

^a An asterisk indicates the suppressor was subcloned and retested in *M. smegmatis*

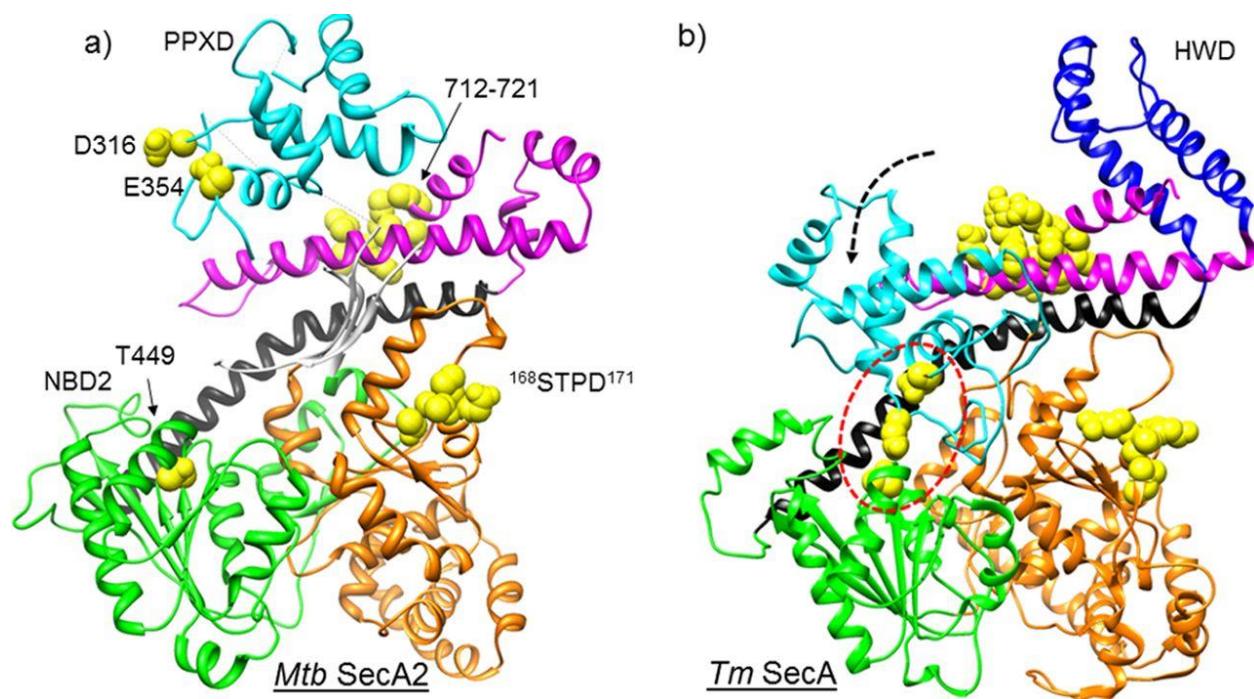


Figure 2.8. Suppressor mutations from *Ms*SecA2 K129R mapped onto (a) *Mtb*SecA2. Mutations are shown as yellow spheres. (b) Suppressor mutations mapped onto the *Tm*SecA complex with SecYEG (PDB code 3DIN). In the complex with SecYEG, SecA is in the “preprotein clamp closed” conformation, in which PPXD (cyan) is swung down (arrow) to make contact with NBD2 (green). In this conformation, the residues affected by the suppressor mutations in the PPXD and NBD2 domains can be seen to come into contact (circled).

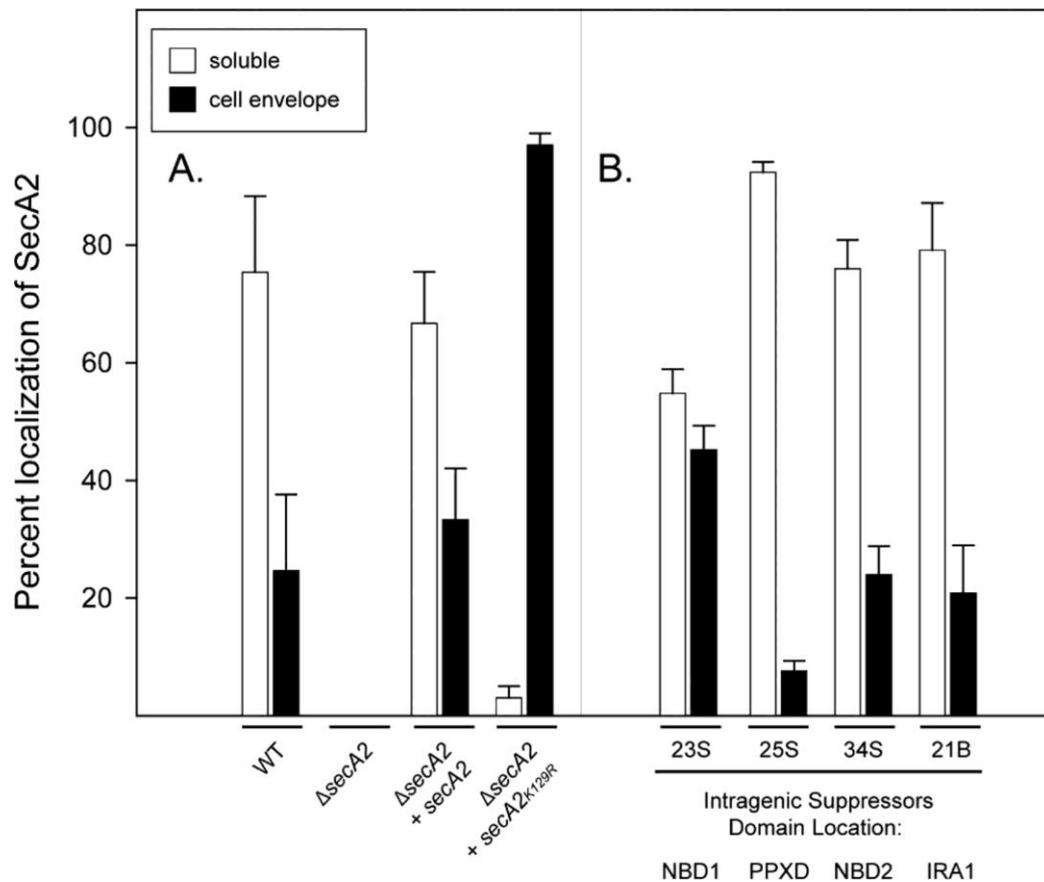


Figure 2.9. Subcellular localization of SecA2 is altered in intragenic suppressors of *secA2* *K129R*. Whole-cell lysates of the indicated strains were fractionated into a cytoplasm-containing soluble fraction and membrane-containing cell envelope fraction. Representative intragenic suppressors (PPXD, NBD2, IRA1, and NBD1) are indicated. Protein derived from an equal number of cells was analyzed by SDS-PAGE, and quantitative immunoblot analysis with anti-SecA2 antibodies was performed. The percentage of localization in a given fraction is plotted. Error bars represent the means from three independent replicates.

REFERENCES

1. Chatzi KE, Sardis MF, Economou A, Karamanou S. 2014. SecA mediated targeting and translocation of secretory proteins. *Biochim Biophys Acta* 1843:1466–1474.
2. Gouridis G, Karamanou S, Gelis I, Kalodimos CG, Economou A. 2009. Signal peptides are allosteric activators of the protein translocase. *Nature* 462:363–367.
3. Feltcher ME, Braunstein M. 2012. Emerging themes in SecA2-mediated protein export. *Nat Rev Microbiol* 10:779–789.
4. Bensing BA, Seepersaud R, Yen YT, Sullam PM. 2014. Selective transport by SecA2: an expanding family of customized motor proteins. *Biochim Biophys Acta* 1843:1674–1686.
5. Braunstein M, Brown AM, Kurtz S, Jacobs WR, Jr. 2001. Two nonredundant SecA homologues function in mycobacteria. *J Bacteriol* 183: 6979 – 6990.
6. Sasseti CM, Rubin EJ. 2003. Genetic requirements for mycobacterial survival during infection. *PNAS* 100:12989–12994.
7. Braunstein M, Espinosa BJ, Chan J, Belisle JT, Jacobs WR, Jr. 2003. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 48:453–464.
8. Kurtz S, McKinnon KP, Runge MS, Ting JP, 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.
9. Sullivan JT, Young EF, McCann JR, Braunstein M. 2012. The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infect Immun* 80:996–1006.
10. van der Woude AD, Stoop EJ, Stiess M, Wang S, Ummels R, van Stempvoort G, Piersma SR, Cascioferro A, Jimenez CR, Houben EN, Luirink J, Pieters J, van der Sar AM, 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.
11. Break TJ, Jun S, Indramohan M, Carr KD, Sieve AN, Dory L, Berg RE. 2012. Extracellular superoxide dismutase inhibits innate immune responses and clearance of an intracellular bacterial infection. *J Immunol* 188:3342–3350.
12. Hinchey J, Lee S, Jeon BY, Basaraba RJ, Venkataswamy MM, Chen B, Chan J, Braunstein M, Orme IM, Derrick SC, Morris SL, Jacobs WR, Jr, Porcelli SA. 2007. Enhanced priming of adaptive immunity by a proapoptotic mutant of *Mycobacterium tuberculosis*. *J Clin Invest* 117:2279–2288.
13. Lenz LL, Mohammadi S, Geissler A, Portnoy DA. 2003. SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *PNAS* 100:12432–12437.

14. Siboo IR, Chambers HF, Sullam PM. 2005. Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. *Infect Immun* 73:2273–2280.
15. Wu H, Mintz KP, Ladha M, Fives-Taylor PM. 1998. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. *Mol Microbiol* 28:487–500.
16. Xiong YQ, Bensing BA, Bayer AS, Chambers HF, Sullam PM. 2008. Role of the serine-rich surface glycoprotein GspB of *Streptococcus gordonii* in the pathogenesis of infective endocarditis. *Microb Pathog* 45:297–301.
17. Nguyen-Mau SM, Oh SY, Kern VJ, Missiakas DM, Schneewind O. 2012. Secretion genes as determinants of *Bacillus anthracis* chain length. *J Bacteriol* 194:3841–3850.
18. Fagan RP, Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J Biol Chem* 286:27483–27493.
19. Takamatsu D, Bensing BA, Sullam PM. 2004. Four proteins encoded in the *gspB-secY2A2* operon of *Streptococcus gordonii* mediate the intracellular glycosylation of the platelet-binding protein GspB. *J Bacteriol* 186: 7100–7111.
20. Caspers M, Freudl R. 2008. *Corynebacterium glutamicum* possesses two secA homologous genes that are essential for viability. *Arch Microbiol* 189:605–610.
21. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, Chen X, Braunstein M. 2015. Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Mol Cell Proteomics* 14:1501–1516.
22. Gibbons HS, Wolschendorf F, Abshire M, Niederweis M, Braunstein M. 2007. Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* 189:5090–5100.
23. Feltcher ME, Gibbons HS, Ligon LS, Braunstein M. 2013. Protein export by the mycobacterial SecA2 system is determined by the pre-protein mature domain. *J Bacteriol* 195:672–681.
24. DeLisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *PNAS* 100:6115–6120.
25. Nouwen N, Berrelkamp G, Driessen AJ. 2007. Bacterial Sec-translocase unfolds and translocates a class of folded protein domains. *J Mol Biol* 372:422–433.
26. Weiss JB, Ray PH, Bassford PJ, Jr. 1988. Purified SecB protein of *Escherichia coli* retards folding and promotes membrane translocation of the maltose-binding protein in vitro. *PNAS* 85:8978–8982.
27. Kusters I, Driessen AJ. 2011. SecA, a remarkable nanomachine. *Cell Mol Life Sci* 68:2053–2066.

28. Hou JM, D’Lima NG, Rigel NW, Gibbons HS, McCann JR, Braunstein M, Teschke CM. 2008. ATPase activity of *Mycobacterium tuberculosis* SecA1 and SecA2 proteins and its importance for SecA2 function in macrophages. *J Bacteriol* 190:4880 – 4887.
29. Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S, Braunstein M. 2009. The accessory SecA2 system of mycobacteria requires ATP binding and the canonical SecA1. *J Biol Chem* 284:9927–9936.
30. Ligon LS, Rigel NW, Romanchuk A, Jones CD, Braunstein M. 2013. Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of mycobacteria. *J Bacteriol* 195:4456 – 4465.
31. Sharma V, Arockiasamy A, Ronning DR, Savva CG, Holzenburg A, Braunstein M, Jacobs WR, Jr, Sacchettini JC. 2003. Crystal structure of *Mycobacterium tuberculosis* SecA, a preprotein translocating ATPase. *PNAS* 100:2243–2248.
32. Papanikolau Y, Papadovasilaki M, Ravelli RB, McCarthy AA, Cusack S, Economou A, Petratos K. 2007. Structure of dimeric SecA, the *Escherichia coli* preprotein translocase motor. *J Mol Biol* 366:1545–1557.
33. Hunt JF, Weinkauff S, Henry L, Fak JJ, McNicholas P, Oliver DB, Deisenhofer J. 2002. Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* 297:2018 –2026.
34. Zimmer J, Nam Y, Rapoport TA. 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936 –943.
35. Vassilyev DG, Mori H, Vassilyeva MN, Tsukazaki T, Kimura Y, Tahirov TH, Ito K. 2006. Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. *J Mol Biol* 364:248 –258.
36. Erlandson KJ, Miller SB, Nam Y, Osborne AR, Zimmer J, Rapoport TA. 2008. A role for the two-helix finger of the SecA ATPase in protein translocation. *Nature* 455:984 –987.
37. Whitehouse S, Gold VA, Robson A, Allen WJ, Sessions RB, Collinson I. 2012. Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. *J Cell Biol* 199:919 –929.
38. Chen Y, Bauer BW, Rapoport TA, Gumbart JC. 2015. Conformational changes of the clamp of the protein translocation ATPase SecA. *J Mol Biol* 427:2348 –2359.
39. Osborne AR, Clemons WM, Jr, Rapoport TA. 2004. A large conformational change of the translocation ATPase SecA. *PNAS* 101:10937–10942.
40. Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG. 2007. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131:756 –769.

41. Bauer BW, Rapoport TA. 2009. Mapping polypeptide interactions of the SecA ATPase during translocation. *PNAS* 106:20800 – 20805.
42. Bhanu MK, Zhao P, Kendall DA. 2013. Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility method. *J Bacteriol* 195:4709 – 4715.
43. DeJesus MA, Sacchettini JC, Ioerger TR. 2013. Reannotation of translational start sites in the genome of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 93:18 –25.
44. Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234.
45. Heras B, Martin JL. 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* 61:1173–1180.
46. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP. 1992. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 258:987–991.
47. Fischetti RF, Xu S, Yoder DW, Becker M, Nagarajan V, Sanishvili R, Hilgart MC, Stepanov S, Makarov O, Smith JL. 2009. Mini-beam collimator enables microcrystallography experiments on standard beamlines. *J Synchrotron Radiat* 16:217–225.
48. Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Macromol Crystallogr A* 276:307–326.
49. Sheldrick GM. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66:479 – 485.
50. Vonrhein C, Blanc E, Roversi P, Bricogne G. 2007. Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
51. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486 –501.
52. Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr A* 42:140 –149.
53. Cowtan KD, Main P. 1996. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D Biol Crystallogr* 52:43– 48.
54. Hunt JF, Deisenhofer J. 2003. Ping-pong cross-validation in real space: a method for increasing the phasing power of a partial model without risk of model bias. *Acta Crystallogr D Biol Crystallogr* 59:214 –224.
55. Bricogne GBE, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi PSA, Smart OS, Vonrhein C, Womack TO. 2011. BUSTER version 2.8.0. Global Phasing Ltd., Cambridge, United Kingdom.

56. Guo XV, Monteleone M, Klotzsche M, Kamionka A, Hillen W, Braunstein M, Ehrt S, Schnappinger D. 2007. Silencing *Mycobacterium smegmatis* by using tetracycline repressors. *J Bacteriol* 189:4614 – 4623.
57. Dauter Z, Dauter M, Dodson E. 2002. Jolly SAD. *Acta Crystallogr D Biol Crystallogr* 58:494 –506.
58. Das S, Grady LM, Michtav J, Zhou Y, Cohan FM, Hingorani MM, Oliver DB. 2012. The variable subdomain of *Escherichia coli* SecA functions to regulate SecA ATPase activity and ADP release. *J Bacteriol* 194: 2205–2213.
59. Gold VA, Whitehouse S, Robson A, Collinson I. 2013. The dynamic action of SecA during the initiation of protein translocation. *Biochem J* 449:695–705.
60. Das S, Oliver DB. 2011. Mapping of the SecA·SecY and SecA·SecE interfaces by site-directed in vivo photocross-linking. *J Biol Chem* 286:12371– 12380.
61. Zimmer J, Rapoport TA. 2009. Conformational flexibility and peptide interaction of the translocation ATPase SecA. *J Mol Biol* 394:606 – 612.
62. Vrontou E, Karamanou S, Baud C, Sianidis G, Economou A. 2004. Global co-ordination of protein translocation by the SecA IRA1 switch. *J Biol Chem* 279:22490 –22497.
63. Duong F, Wickner W. 1999. The PrlA and PrlG phenotypes are caused by a loosened association among the translocase SecYEG subunits. *EMBO J* 18:3263–3270.
64. Oliver DB, Beckwith J. 1982. Regulation of a membrane component required for protein secretion in *Escherichia coli*. *Cell* 30:311–319.
65. Ding H, Hunt JF, Mukerji I, Oliver D. 2003. *Bacillus subtilis* SecA ATPase exists as an antiparallel dimer in solution. *Biochemistry* 42:8729 – 8738.
66. Auclair SM, Oliver DB, Mukerji I. 2013. Defining the solution state dimer structure of *Escherichia coli* SecA using Forster resonance energy transfer. *Biochemistry* 52:2388 –2401.
67. D’Lima NG, Teschke CM. 2013. ADP-dependent conformational changes distinguish *Mycobacterium tuberculosis* SecA2 from SecA1. *J Biol Chem* 289:2307–2317.
68. Zhou J, Xu Z. 2003. Structural determinants of SecB recognition by SecA in bacterial protein translocation. *Nat Struct Biol* 10:942–947.
69. Schneewind O, Missiakas DM. 2012. Protein secretion and surface display in Gram-positive bacteria. *Philos Trans R Soc Lond B Biol Sci* 367: 1123–1139.
70. de Keyser J, van der Sluis EO, Spelbrink RE, Nijstad N, de Kruijff B, Nouwen N, van der Does C, Driessen AJ. 2005. Covalently dimerized SecA is functional in protein translocation. *J Biol Chem* 280:35255–35260.
71. Jilaveanu LB, Zito CR, Oliver D. 2005. Dimeric SecA is essential for protein translocation. *PNAS* 102:7511–7516.

72. Gouridis G, Karamanou S, Sardis MF, Scharer MA, Capitani G, Economou A. 2013. Quaternary dynamics of the SecA motor drive translocase catalysis. *Mol Cell* 52:655– 666.
73. Prabudiansyah I, Kusters I, Driessen AJ. 2015. In vitro interaction of the housekeeping SecA1 with the accessory SecA2 protein of *Mycobacterium tuberculosis*. *PLoS One* 10:e0128788.
74. Durack J, Burke TP, Portnoy DA. 2015. A *prl* mutation in SecY suppresses secretion and virulence defects of *Listeria monocytogenes secA2* mutants. *J Bacteriol* 197:932– 42.

CHAPTER 3¹

Mycobacterium tuberculosis SatS is a chaperone for the SecA2 protein export pathway

The SecA2 protein export system is critical for the virulence of *Mycobacterium tuberculosis*. However, the mechanism of this export pathway remains unclear. Through a screen for suppressors of a *secA2* mutant, we identified a new player in the mycobacterial SecA2 pathway that we named SatS for SecA2 (two) Suppressor. In *M. tuberculosis*, SatS is required for the export of a subset of SecA2 substrates and for pathogenesis. We further identify a role for SatS as a protein export chaperone. SatS exhibits multiple properties of a chaperone, including the ability to bind to and protect substrates from aggregation. Our structural studies of SatS reveal a distinct combination of a new fold and hydrophobic grooves resembling preprotein-binding sites of the SecB chaperone. These results are significant in better defining a molecular pathway for *M. tuberculosis* pathogenesis and in expanding our appreciation of the diversity among chaperones and protein export systems.

Introduction

With 1.7 million deaths from tuberculosis in 2016, *Mycobacterium tuberculosis* continues to have a significant impact on world health (1). For *M. tuberculosis* to cause disease, the bacillus must export effector proteins to the host-pathogen interface. These effectors enable *M. tuberculosis* to grow in macrophages and avoid clearance by the host immune response (2). At

¹ Contributing authors: Ryan Hughes and James Sacchettini (Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX); Lauren Ligon, Nathan Rigel, Seidu Malik, Brandon R. Anjuwon-Foster, and Miriam Braunstein (Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC).

least some of these effectors are exported by *M. tuberculosis* via the specialized SecA2 export pathway (3).

The mechanism of SecA2 export remains poorly understood. SecA2 is a paralog of the SecA ATPase of the general Sec protein export pathway. The general Sec pathway transports preproteins with N-terminal signal sequences across the inner membrane through a channel composed of SecY, SecE and SecG proteins (4). Preproteins must be in an unfolded state to travel through the SecYEG channel and, in Gram-negative bacteria, the SecB chaperone binds a subset of preproteins to maintain them in an unfolded translocation competent state. Following export across the membrane, the signal sequence is cleaved and the mature protein is released (5). While all bacteria possess an essential Sec pathway that carries out the majority of protein export, only mycobacteria and a subset of Gram-positive bacteria possess specialized Sec export systems that are defined by a second SecA (6, 7). In these organisms, SecA1 is the name given to the canonical SecA and the specialized SecA is named SecA2. For the mycobacterial SecA2 system, the housekeeping SecYEG channel, and possibly SecA1, as well, are also involved (8, 9). However, SecA1 and SecA2 are functionally distinct, as shown by their inability to compensate for the loss of one another (10, 11), and it remains unclear how SecA2 functions to export its relatively small and specific subset of proteins.

Here, we carried out a suppressor screen using a dominant negative *secA2 K129R* mutant of *Mycobacterium smegmatis*, a fast-growing model mycobacteria, as a means to identify new components of the mycobacterial SecA2 pathway. The K129R substitution is in the ATP binding site of SecA2, and past studies lead to a model where SecA2 K129R is defective for SecA2-dependent export but still able to interact with its normal binding partners that include SecYEG (8, 11). As a result, SecA2 K129R disrupts SecYEG channels at the membrane, which hinders

both general Sec and specialized SecA2 export as evidenced by more severe phenotypes of *secA2 K129R* than a $\Delta secA2$ null mutation (8). A large collection of *secA2 K129R* suppressor mutations mapped to *msmeg_1684*, a gene of unknown function that we renamed *satS* for SecA2(two) Suppressor. SatS is also present in *M. tuberculosis* and, remarkably, the *M. tuberculosis satS* gene is in an operon with the gene encoding SapM, which is a secreted phosphatase exported by the SecA2 pathway (12).

Here, we demonstrate that SatS, which we prove is required for *M. tuberculosis* growth in macrophages, functions in the export of SapM and an additional subset of the proteins exported by the SecA2 pathway. We further reveal properties of SatS that indicate a function as a protein export chaperone that protects its substrates from inappropriate interactions in the cytoplasm and additionally assists in their export. Finally, we determine the structure of the C-domain of SatS (SatS_C), which reveals a new fold lacking similarities to any solved chaperone structures, yet contains surface hydrophobic grooves resembling those of the SecB chaperone. The identification of SatS expands our understanding of SecA2 export in mycobacteria and provides another example of the diversity of molecular chaperones across biological systems.

Materials and Methods

Plasmids, bacterial strains and culture conditions.

For plasmid construction, PCR products were first ligated into TOPO cloning vectors (Invitrogen, Carlsbad, CA), digested with restriction enzymes described in Table 3.1, and ligated into their final vectors described in Table 3.2. In all cases, newly constructed plasmids were verified by sequencing and diagnostic digests. In the case of SatS G134D plasmids, *satS_{Msm}* was amplified by PCR from the 3S suppressor and *satS_{Mtb} G134D* was designed using site directed

mutagenesis (SDM) on the *satS_{Mtb}* complementation plasmid pBM81. Amino acid G134 was confirmed to be highly conserved in mycobacterial SatS homologs using ConSurf (13).

M. tuberculosis and *M. smegmatis* strains are described in Table 3.3. *M. tuberculosis* was grown at 37°C in Middlebrook 7H9/7H11 supplemented with 1x albumin dextrose saline (ADS), 0.5% glycerol and 0.025% Tween 80. *M. smegmatis* was grown at 37°C or 30°C in Middlebrook 7H9/7H10 or Mueller-Hinton medium. Media were supplemented with 0.5% glycerol plus 0.2% glucose (7H9/7H10 medium only) and 0.05% Tween 80 (all media). For all mycobacteria, the antibiotics kanamycin (20 µg/mL) and hygromycin B (50 µg/mL) were added as needed. *E. coli* strains were grown at 37°C in Miller LB broth or on Miller LB agar. The antibiotics kanamycin (40 µg/mL) and hygromycin B (150 µg/mL) were added as needed.

M. smegmatis growth was monitored using resazurin. At an OD_{600 nm} of 1, cells were diluted to 10⁵ c.f.u. ml⁻¹ in the same medium and 100 µl were added to 96-well plates. After 24 h of growth at 37 °C, resazurin (12.5 µg ml⁻¹ final concentration; Sigma-Aldrich, St. Louis, MO) was added and fluorescence with excitation at 530 nm and emission at 590 nm was monitored over time. *M. tuberculosis* growth was monitored by measuring the optical density (OD₆₀₀) of liquid broth cultures over time.

Suppressor collection and sequencing

The suppressor screen was performed as described previously (8). Suppressors of the *secA2 K129R* allele were isolated by plating independently grown cultures of the *secA2 K129R* strain onto Mueller-Hinton agar at 37°C. Genomic DNA from six suppressors was submitted for whole genome sequencing at the High-Throughput Sequencing Facility at the University of North Carolina at Chapel Hill. Sequencing was performed using Illumina GA II technology.

Reads were aligned to the *M. smegmatis* mc²155 reference genome (NCBI RefSeq accession number NC_008596.1) using SOAP (14).

***M. smegmatis* mutant construction**

The *M. smegmatis* unmarked $\Delta secA2/\Delta satS$ double mutant was created by two-step allelic exchange using plasmid pLL50 in the $\Delta secA2$ mutant strain NR116, resulting in strain BAF1. Briefly, the suicide plasmid pLL50, containing a hygromycin-resistance selectable marker, a *sacB* counter-selectable marker, and flanking regions for *satS* was transformed into *M. smegmatis*. Transformants were selected by plating on media containing hygromycin B. Hygromycin-resistant transformants were grown to saturation, diluted 1:100 in media lacking hygromycin B, and then grown overnight at 37°C. Bacteria in which a second recombination event occurred were selected by plating on 7H10 supplemented with 0.2% glucose and 4.5% sucrose. BAF1 was assessed for the desired chromosomal deletion by PCR and Southern blot.

The *M. smegmatis* unmarked $\Delta satS$ single mutant was created by adding back *secA2* into the BAF1 strain by two-step allelic exchange using plasmid pBM11, resulting in strain BM10. BM10 was assessed for the desired chromosomal insertion by PCR and Southern blot. Additionally, immunoblots of SecA2 were performed to ensure SecA2 levels were fully restored.

***M. tuberculosis* mutant construction**

The *satS* deletion mutant was created in H37Rv using the specialized transducing phage system as previously described (10). Briefly, cosmid pSM42 was created by subcloning *satS* upstream and downstream flanks into pYUB854 surrounding the hygromycin cassette. Cosmid pSM42 was ligated into phAE159 to generate recombinant phasmid pSM45. The recombinant phasmid, pSM45 was packaged into phage head using a λ *in vitro* packaging extract kit

(Gigapack III XL, Agilent, Santa Clara, CA) and was transduced into *E. coli*. Phasmid DNA was electroporated into *M. smegmatis* mc²155 to make phage (pSM60). Transduced phage was plaque purified and amplified for high titer phage lysate. H37Rv was transduced with high phage lysate as previously described (10). Transductants were grown at 37°C on Middlebrook 7H10 plates containing hygromycin for 4 weeks. To confirm the *satS* deletion in transductants, PCR and Southern blotting were used.

Azide sensitivity assays

Cultures were plated for azide sensitivity as previously described, by mixing 200 µL of a saturated culture with 7H9 top agar and pouring over 7H10 agar plates lacking tween in three technical replicates (8). The diameter of the zone of inhibition was measured after two days and reported as a percentage of the entire plate diameter, yielding percent azide inhibition.

Subcellular fractionation and Immunoblotting

Subcellular fractionations of *M. smegmatis* and irradiated *M. tuberculosis* were isolated as previously described (15, 16). Briefly, cells suspended in 1X PBS containing protease inhibitors were lysed by passage through a French pressure cell. Unlysed cells were removed by centrifugation at 3,000 x *g* for 30 min to generate clarified whole cell lysates (WCLs). The WCLs were either spun at 100,000 x *g* for 2 hr to collect the cell envelope fraction containing both the cell wall and membrane (ENV) or at 27,000 x *g* for 30 min to pellet the cell wall fraction only (CW). The supernatant following CW isolation was spun at 100,000 x *g* for 2 hr to separate the membrane fraction (MEM) and collect the soluble cytoplasm-containing fraction (SOL). Protein concentrations were determined by bicinchoninic acid assay (Pierce, ThermoFisher, Waltham, MA).

Samples containing equal protein were separated by SDS-PAGE and transferred to nitrocellulose membranes. After blocking, proteins were detected using the following antibodies: α SatS, α SapM (provided by Vojo Deretic, University of New Mexico), α Mce1A (17), α Mce1E (17), α Mce1D (15), α 19kD (provided by Douglas Young, Imperial College, United Kingdom), α PhoS1 (IT23, NIH Biodefense and Emerging Infections Research Resources Repository, NIAID), α PknG (provided by Yossef Av-Gay, University of British Columbia, Canada), α SecA2 (11), α SigA (provided by Murty Madiraju, The University of Texas Health Science Center Tyler), α MspA (16) and α SecY (8). FLAG-tagged proteins were detected with mouse α FLAG clone M2 antibody (Sigma-Aldrich) or rabbit α FLAG antibody (Sigma-Aldrich). HA-tagged proteins were detected with mouse α HA clone HA-7 (Sigma-Aldrich). α His (Abgent, San Diego, CA) was used to detect the mycobacterial GroEL1 which has a string of endogenous histidines. α Mouse and α Rabbit IgG conjugated horseradish peroxidase secondary antibodies (Bio-Rad, Hercules, CA) were used and signal was detected using Western Lightning Plus-ECL chemiluminescent detection reagent (Perkin-Elmer, Waltham, MA).

Culture filtrate protein preparation

M. tuberculosis culture filtrates were collected as described previously (12). Briefly, 200 mL cultures were grown in Sauton media at 37°C for 24 hours. The supernatants were double filtered through a 0.2 μ m filter. Supernatant proteins were concentrated 200 fold using 3,000 MW cut off centrifuge filters (Amicon) by centrifugation at 3,000 rpm at 4°C. For immunoblotting, protein was precipitated overnight at 4°C with 10% trichloroethanoic acid (TCA). Protein pellets were washed with acetone, resuspended in 250 μ L of 1 x SDS-PAGE buffer.

For *M. smegmatis* culture filtrate collection, samples were obtained as previously described (16). In brief, 10 mL cultures were grown without Tween 80 to an OD_{600 nm} of 0.4 to 0.7. Supernatant was separated from cells first by centrifugation at 3,000 x g and then filtration through a 0.2- μ m-pore-size filter. Protein from 2 mL of supernatant was TCA precipitated as described above and then resuspended in 50 μ L of 1 \times SDS-PAGE buffer.

Phosphatase activity assay

SapM activity was assayed as described previously (12, 18). In a 96 well plate, 3 μ g of CFP protein was diluted with water and added to 10X buffer (1M Tris base pH 6.8) with 20 mM Sodium tartrate to reduce background phosphatase activity, and 50 mM p-nitrophenyl phosphate (pNPP) for a total volume of 200 μ L (New England Biolabs, Ipswich, MA). Tartrate is an inhibitor of some phosphatases, but SapM activity is unaffected by tartrate (18). Despite the addition of tartrate, the phosphatase assay used is not specific for SapM. The residual activity in the Δ secA2 and Δ satS mutants can be attributed to SatS-independent phosphatases. The plate was incubated at 37°C in a plate reader, and the absorbance at 405 nm was measured every three minutes for four hours. Over the linear portion of the kinetic assay, we calculated the rate of pNPP conversion by calculating the slope of the line generated by plotting Abs_{405 nm} as a function of time. These slopes were then normalized to the WT rate of change, which we set to 100%.

Whole cell phosphatase activity assay

To perform the whole cell phosphatase activity assay, *M. smegmatis* strains expressing SapM (+/- ss) or an empty vector were grown in 7H9 medium to an OD₆₀₀ of 1, pelleted, and washed once in 7H9 medium. Cells were diluted to 6.25 x 10⁵ CFU/mL in 7H9 medium and 160 μ L was added in triplicate to a 96-well plate. Plates were incubated at 37°C for 24 hours. After 24 hours, 20 μ L of 10X buffer (1M Tris base pH 6.8) with 20 mM Sodium tartrate and 50 mM p-

nitrophenyl phosphate (pNPP) were added to the wells for a total volume of 200 μ L. The plate was incubated at 37°C in a plate reader, and the absorbance at 405 nm was measured every three minutes for four hours. We calculated the rate of pNPP conversion as described above.

Macrophage infection

To assess *M. tuberculosis* survival in macrophages, 2 x 10⁵ BMDMs from C57BL/6 mice were seeded 1 day prior to infection with *M. tuberculosis* (H37Rv, Δ *secA2*, Δ *satS*, or Δ *satS*+*psatS*) at an MOI of 1 as previously described (3, 12). At 4 hours post infection, macrophages were washed four times and at the indicated time points were lysed with 0.1% Triton X-100. Serial dilutions of the lysates were plated on 7H11 agar plates and CFUs were counted three weeks later.

Reverse transcriptase-PCR

To assess the operon nature of *sapM* and *satS*, RNA was extracted from mid-log phase cultures of *M. tuberculosis* H37Rv (see *qRT-PCR* for methods). Reverse transcription reaction was carried out using iScript cDNA Synthesis Kit (Bio-Rad) and random primers. PCR amplification of the intergenic regions on cDNA were performed using specific primers on *sapM* and *satS* (Supplemental File 1). Controls included primers for the housekeeping gene *sigA*, PCR amplification from genomic DNA, and PCR amplification from RNA lacking reverse transcriptase.

Quantitative Real-Time PCR

Triplicate *M. tuberculosis* cultures were grown in modified 7H9 medium to an OD₆₀₀ of 1 and RNA was isolated as previously described using a chloroform-methanol and Trizol (Invitrogen) extraction (15, 17). RNA samples were treated with DNase (Promega, Madison, WI)

and then column purified (Zymo RNA clean and concentrator Kit, Irvine, CA). Following RNA isolation, cDNA was synthesized with random primers using the iScript cDNA Synthesis Kit (Bio-Rad). Real-time PCR was completed using 25 ng of cDNA template in triplicate technical replicates using the SensiMix SYBR and fluorescein kit (Bioline, Toronto, Canada). Transcripts were normalized to the housekeeping gene *sigA*. Primer sequences are provided in Supplemental File 1.

LacZ (β -galactosidase) activity assays

LacZ activity assays in *M. tuberculosis* were performed using a modified protocol previously described for *M. smegmatis* (8). Strains were grown in 7AGT to mid-log phase and 800 μ L was pelleted. Pellets were resuspended in 800 μ L Z buffer (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, 50 mM β -mercaptoethanol), then lysed with 35 μ L chloroform and 1 μ L of 0.1% SDS by vortexing for 30 s followed by sonication. 640 μ g of *o*-nitrophenyl- β -D-galactopyranoside was added to each reaction and mixtures were incubated for 24 minutes at room temperature. Reactions were terminated by addition of 400 μ L of 1 M Na₂CO₃. Debris was removed by centrifugation at 3,000 rpm for 10 minutes, and the OD_{420 nm} was read from the supernatant. LacZ activity (Miller units) was calculated by the following formula: $(1000 \times \text{OD}_{420 \text{ nm}}) / ([\text{reaction time in minutes}] \times [\text{culture volume used in the reaction, in mL}] \times \text{OD}_{600 \text{ nm}})$.

SatS antiserum production

To generate polyclonal antisera against SatS, purified SatS_{Mtb} was produced in *E. coli* and injected into two rabbits using Titermax adjuvant (ThermoFisher). The serum from both rabbits was tested against wild-type *M. tuberculosis* and *M. smegmatis* and the Δ *satS* mutants for specificity. The serum from rabbit PA6753 recognizes SatS_{Mtb} but does not recognize SatS_{Msm}.

and is only used for SatS_{Mtb}. The sera from rabbit PA6754 has a non-specific band at the same size as SatS_{Mtb}, but recognizes SatS_{Msm} and is only used for SatS_{Msm}.

Co-immunoprecipitation

For *in vivo* co-immunoprecipitation, *M. smegmatis* cells were transformed with SatS_{Mtb} (+/- HA) tag and SapM-FLAG (+/- signal sequence). Transformed cells were grown in 50 mL of 7H9 medium to an OD_{600 nm} of 0.5. Cells were pelleted and resuspended in 2.5 mL 1X PBS buffer containing a protease inhibitor cocktail. Cells were lysed by passage through a French pressure cell. Unlysed cells were removed by centrifugation at 3,000 x g for 30 min to generate clarified whole cell lysates (WCLs). 200 µL of lysate was diluted in 1 mL of 1X PBS + protease inhibitors, added to 25 µL anti-HA agarose (Sigma-Aldrich), and mixed end to end at 4°C for 4 hrs, followed by four washes with 1X PBS. The immunoprecipitated SatS-HA along with co-immunoprecipitated proteins were eluted in 25 µL of 1X SDS-PAGE buffer, run on 15% SDS-PAGE gels for 4.5 hours, transferred onto nitrocellulose membranes, and immunoblotted.

Cloning, expression, and purification of SapM inclusion bodies (IBs)

The *sapM* full length gene was PCR amplified from genomic DNA of H37Rv using Phusion high-fidelity DNA polymerase and the primers sapM *E. coli* F and sapM *E. coli* R. The resulting PCR product treated with T4 polymerase and mixed with linear, T4 treated, pMSCG-28 vector, and transformed into chemically competent BL21 (DE3) cells as previously described (19).

SapM containing a C-His₆-tag and tobacco etch virus (TEV) protease cleavage site were grown in Luria-Bertani (LB) broth containing 100 µg/mL carbenicillin at 37°C to an OD of 0.8 (A_{600 nm}). SapM expression was induced with the addition of 0.5 mM isopropyl-β-D-

thiogalactoside (IPTG) and cells were grown for an additional 5 hours at 37°C. Cells were harvested and resuspended in lysis buffer (40 mM HEPES pH 7.4, 300 mM NaCl, 10 mM imidazole). Cells were broken using a high-pressure homogenizer in the presence of protease inhibitor cocktail (EMD Millipore, Burlington, MA) and centrifuged at 30,000 x g.

In order to clarify SapM IBs, a modified protocol was adopted (20). The supernatant obtained after cell lysis was decanted and the pellet resuspended in 40 mM HEPES pH 7.4, 2% Triton X-100, 5 mM EDTA. The suspension was then homogenized using sonication for 3 cycles for 30 seconds each, centrifuged at 30,000 x g for 15 minutes, and the supernatant decanted. This was repeated 3 times to remove cell wall, membrane material, and lipid/membrane associated proteins. In the final step, detergent was omitted, and SapM purity of greater than 95% was confirmed via SDS-PAGE.

Protein aggregation assay

Inclusion bodies of SapM pre protein with a 6X C-terminal His tag were denatured in 8 M urea, 40 mM HEPES pH 7.4, 100 mM NaCl, 1 mM EDTA to a final concentration of 150 µM. Denatured SapM (1 µL) was rapidly diluted into buffer (150 µL) containing 40 mM HEPES pH 7.4, 100 mM NaCl, and 1 mM EDTA. Protein aggregation was monitored in the absence or presence of SatS at 25°C by measuring light scattering in a time dependent manner using a Cary Eclipse Varian with excitation and emission at 350 nm.

Cloning, expression, purification, and crystallization of SatS and SatS_C

The *satS* and *satS_C* genes were PCR amplified from genomic DNA of H37Rv using Phusion high-fidelity DNA polymerase and the primers *satS E. coli* F, *satS_C E. coli* F, and *satS/satS_C E. coli* R. The resulting PCR products were digested with NdeI and HindIII, ligated

into Nde/HindIII digested Pet28b vector, and transformed into chemically competent BL21 (DE3) cells.

SatS and SatS_C with a N-His₆-tag and TEV protease cleavage site were grown separately in LB broth containing 50 µg/mL kanamycin at 37°C to an OD of 0.8 ($A_{600\text{ nm}}$). SatS and SatS_C expression was induced with the addition of 0.5 mM IPTG and cells were grown for an additional 5 hours at 37°C. Cells were harvested and resuspended in lysis buffer (40 mM HEPES pH 7.4, 300 mM NaCl, 10 mM imidazole). Cells were broken using a high-pressure homogenizer in the presence of protease inhibitor cocktail (EMD Millipore) and centrifuged at 30,000 x *g*. SatS and SatS_C were purified using a cOmplete His-tag purification resin (Roche, Basel, Switzerland), followed by removal of the tag using TEV protease at 25°C and further purified by size exclusion chromatography using a Superdex 200 26/60 (GE Healthcare, Chicago, IL). Protein purity was greater than 95% as determined by SDS-PAGE. Protein concentration was measured spectrophotometrically at 595 nm using Bradford reagent.

Crystals of SatS (20 mg/mL) were produced after screening 768 individual conditions using sitting drop vapor diffusion method at 16°C with a 50 µL well solution and a drop consisting of 1.2 µL of 0.6 µL protein and 0.6 µL of well solution. A single diffraction quality crystal appeared within 6 months in 3.5 M ammonium sulfate and 0.1 M sodium acetate trihydrate pH 4.6. SatS was indexed into space group P2₁2₁2₁ with the unit cell parameters *a* = 50, *b* = 51, *c* = 76. The unit cell was composed of a single molecule in the asymmetric unit.

Crystals of SatS_C (12mg/mL) were produced after screening 384 individual conditions using sitting drop vapor diffusion method at 16°C with a 50 µL well solution and a drop consisting of 1.2 µL of 0.6 µL protein and 0.6 µL of well solution. Initial crystal hits were optimized using

hanging drop vapor diffusion method at 16°C with a 1 mL well solution and a 4.0 µL drop consisting of random ratios of protein to well solution. The highest quality crystals appeared overnight in 3.5 M ammonium citrate pH 6.4 and continued to mature for an additional 2 weeks. SatS_C indexed into space group P2₁2₁2₁ with the unit cell parameters a = 50, b = 51, c = 76. The unit cell was composed of a single molecule in the asymmetric unit.

Data collection and structure determination of SatS_C

X-ray diffraction data were collected from a single crystal at beamline 23-ID of the GM/CA-CAT facilities of the Advanced Photon Source, Argonne National Laboratory. The structure of SatS was solved by single-wavelength anomalous dispersion (SAD) using a Bromine (Br) derivative. The data was processed and reduced using the HKL3000 software package. A single Br site was identified using Phenix HySS, and Phenix AutoSol was used to produce the initial electron density map. Simultaneous rounds of model building and structure refinement were performed manually in Coot and Phenix Refine. Additional structures of SatS_C were solved by molecular replacement using the initial structure of SatS_C as a model in Phenix Phaser MR. Simultaneous rounds of model building and structure refinement were carried out in Coot and Phenix Refine.

Statistical Analyses

For comparisons between the groups for the determination of (i) phosphatase activity in the mycobacterial culture filtrates, (ii) whole cell phosphatase activity (iii) '*lacZ* reporter fusions and (iv) growth in macrophages, one-way analysis of variance (ANOVA) with the Tukey post test was employed. For the statistical analysis and generation of graphs, Prism 5 software (version 7; GraphPad Software Inc., CA) was used.

Results

satS suppressors of secA2 K129R

A *secA2 K129R* mutant of *M. smegmatis* exhibits more exacerbated phenotypes (*i.e.* azide sensitivity and poor growth on Mueller-Hinton agar) than a $\Delta secA2$ deletion mutant (8) (Figure 3.1A). Starting with cultures of $\Delta secA2$ expressing the *secA2 K129R* allele on an integrating plasmid (this strain is referred to as *secA2 K129R* from hereon), we collected spontaneous suppressor mutants that restored growth on Mueller-Hinton agar. Whole-genome sequencing of six extragenic suppressors revealed mutations in the same gene *msmeg_1684* (Figure 3.1B). Three additional suppressors with mutations in *msmeg_1684* were identified by directly sequencing the *msmeg_1684* gene and upstream sequence in our pool of suppressors (Figure 3.1B). Msmeg_1684 is a highly acidic protein (pI 3.83) of unknown function with conserved homologs in all mycobacterial species, as well as other actinomycetes (21). However, no homologous proteins exist outside of actinomycetes and Msmeg_1684 does not have any conserved domains to predict function. Henceforth, we refer to *msmeg_1684* as *satS* (*secA2* (two) suppressor).

Seven of the nine suppressor mutations in *satS* were expected to be loss-of-function mutations (*i.e.* frameshifts or truncations). To validate that loss of *satS* suppresses *secA2 K129R* phenotypes, we deleted *satS* in the *secA2 K129R* mutant background. For future experiments, we also constructed a $\Delta satS$ mutant in a *secA2*⁺ background. The $\Delta satS$ mutant has no *in vitro* growth defect compared to wild-type *M. smegmatis* mc²155 (Figure 3.2A). Deletion of *satS* suppressed the exacerbated phenotypes of *secA2 K129R*, and the suppression phenotype of $\Delta satS$ could be complemented, as demonstrated by restoration of the severe *secA2 K129R* phenotypes upon adding back a copy of *satS* from *M. smegmatis* (Figure 3.1C). Complementation was also

successful with the *M. tuberculosis* *satS* homolog *rv3311* indicating that SatS function is conserved in *M. smegmatis* and *M. tuberculosis* (Figure 3.1C).

Unlike wild-type SecA2, which is predominantly in the cytoplasm, the majority of SecA2 K129R is localized to the membrane-containing cell envelope fraction, consistent with SecA2 K129R being trapped in a non-functional complex with SecYEG (11). We assessed the ability of Δ *satS* to suppress the mislocalization of SecA2 K129R. Using immunoblot analysis of cell envelope (cell wall and membrane) and soluble (cytoplasm) fractions with SecA2 antibodies, we observed that the absence of SatS suppressed the aberrant localization of SecA2 K129R (*i.e.* in the Δ *satS/secA2 K129R* strain) such that SecA2 K129R was now primarily localized to the cytoplasm, similar to wild-type SecA2 (Figure 3.1D). We immunoblotted for the cell wall MspA porin and the cytoplasmic GroEL protein as fractionation controls (Figure 3.2B). SecA2 K129R is also associated with reduced levels of SecY, which is a presumed mechanism to eliminate jammed SecA2 K129R-SecYEG channels (8). When we immunoblotted fractions from the Δ *satS/secA2 K129R* strain with SecY antibodies, we observed that the absence of SatS suppressed the SecA2 K129R effect on SecY levels. Both the rescued localization of SecA2 K129R and SecY levels observed in the Δ *satS* mutant background could be complemented by introduction of *satS*_{*Msm*} (Figure 3.1D and 3.1E). These results indicate that SatS is required for SecA2 K129R retention at the membrane in non-productive complexes with SecYEG. By extension, these results suggest a role for SatS in the SecA2 export pathway.

SatS is required for export of the SecA2-dependent SapM phosphatase

In *M. tuberculosis*, the gene encoding SatS is immediately downstream of the gene encoding SapM. Reverse transcriptase (RT) PCR performed on RNA from wild-type *M. tuberculosis* strain H37Rv was used to demonstrate that *sapM* and *satS* are in an operon (Figure

3.3). This genomic arrangement is striking as SapM, a secreted phosphatase of *M. tuberculosis* is exported by the SecA2 pathway (12). While SapM does not have an ortholog in *M. smegmatis*, we identified 26 different mycobacterial species in which the *sapM-satS* gene arrangement is conserved (22).

We constructed a $\Delta satS$ mutant of *M. tuberculosis* H37Rv to test if SatS is required for SapM secretion. The $\Delta satS$ mutant of *M. tuberculosis* did not exhibit an *in vitro* growth defect (Figure 3.4A). We monitored SapM secretion into culture media by immunoblotting culture filtrates prepared from H37Rv, the $\Delta secA2$ mutant, the $\Delta satS$ mutant, and the $\Delta satS$ mutant complemented with *satS_{Mtb}* using SapM antibodies. As expected, a SapM secretion defect was observed in the $\Delta secA2$ mutant. Even more striking was the SapM secretion defect of the $\Delta satS$ mutant, which was reproducibly more severe than the $\Delta secA2$ mutant (Figure 3.5A). This phenotype could be complemented with a *satS_{Mtb}* plasmid (Figure 3.5A). As controls, we immunoblotted for the 19kDa lipoprotein, which is exported in a SecA2-independent manner and was not affected by the $\Delta satS$ mutant, and also for the cytoplasmic SigA protein to rule out cell lysis contaminating the culture filtrates. Since SapM is a phosphatase, we also quantified SapM secretion by measuring phosphatase activity in the culture filtrates, using p-Nitrophenyl Phosphate (PNPP) as a substrate. Consistent with the immunoblot data, there was significantly less phosphatase activity in the supernatant of a $\Delta secA2$ mutant compared to H37Rv, the $\Delta satS$ mutant exhibited an even more severe reduction in secreted phosphatase activity, and the $\Delta satS$ mutant phenotype could be complemented (Figure 3.5B). These results extend our identification of SatS as a SecA2 suppressor in *M. smegmatis* by revealing a role of SatS in the SecA2-dependent secretion of SapM by *M. tuberculosis*.

Even though *M. smegmatis* lacks a SapM orthologue, when we expressed *M. tuberculosis* *sapM* in *M. smegmatis*, SapM was also secreted in a SecA2 and SatS dependent manner (Figure 3.5C). Again, the SapM secretion defect of a $\Delta satS$ mutant was more severe than that of a $\Delta secA2$ mutant, and this phenotype could be complemented (Figure 3.5C). This result indicates functional conservation of SatS in *M. smegmatis* and *M. tuberculosis*, and it indicates that the more amenable *M. smegmatis* is a valid model for studying SatS function in SapM secretion.

To develop a higher throughput and more quantitative method for monitoring SatS and SecA2-dependent secretion, we established a whole cell assay for measuring secreted SapM phosphatase activity from *M. smegmatis* grown in 96 well plates. Importantly, this assay is specific for secreted SapM; it did not detect cytoplasmic SapM, as demonstrated by background levels of phosphatase activity of a *M. smegmatis* strain expressing non-exported cytoplasmic SapM lacking a signal sequence (Δss -SapM). In contrast, *M. smegmatis* expressing full length SapM preprotein, which is secreted, exhibited significantly greater activity (Figure 3.5D). When the $\Delta secA2$ mutant and the $\Delta satS$ mutant were tested in this whole cell phosphatase assay, the results confirmed the immunoblot data. Secreted phosphatase activity was reduced in both $\Delta satS$ and $\Delta secA2$ mutants, and the reduction was significantly more dramatic in the *satS* mutant (Figure 3.5D). The reduced activity of the $\Delta satS$ mutant could be complemented with either SatS_{Mtb} or SatS_{Msm} (Figure 3.5D).

Mce proteins exported by the SecA2 pathway require SatS for their export

We tested if SatS is required for the export of additional SecA2 substrates. Multiple protein components of Mce transporters, which import lipids, depend on SecA2 to be exported to the cell wall (17). Immunoblot analysis of *M. tuberculosis* samples with Mce1A and Mce1E antibodies revealed that the levels of Mce1A and 1E were reduced in cell wall of a $\Delta satS$ mutant

(Figure 3.5E), with the defect in the $\Delta satS$ mutant again being more severe than the $\Delta secA2$ mutant. Mce importers are conserved in *M. smegmatis* and similar results were obtained upon immunoblotting *M. smegmatis* cell wall fractions with a Mce1D antibody (Figure 3.5F). The same effect was seen when a Mce4A-HA protein was expressed in *M. smegmatis* and localized to the cell wall (Figure 3.5F). In contrast to these results, the level of the SecA2-independent 19 kDa lipoprotein in *M. tuberculosis* and MspA porin in *M. smegmatis* were unchanged in cell wall fractions of $\Delta satS$ mutants (Figure 3.5E and 3.5F).

We next tested whether SatS contributes to export of the SecA2-dependent protein kinase PknG and solute binding protein PhoS1 of *M. tuberculosis*, as well as the solute binding protein Msmeg1704 of *M. smegmatis* (16, 17, 23). Immunoblotting of cell wall fractions confirmed that PknG, PhoS1, and Ms1704 depend on SecA2 for export; however, export of these proteins was not impaired in a $\Delta satS$ mutant (Figure 3.5G and 3.5H). These data demonstrate a level of specificity in the exported proteins affected by SatS. SatS affects multiple, but not all, of the proteins exported by the SecA2 pathway.

SatS is required for *M. tuberculosis* growth in macrophages

The dramatic reductions in export of SapM and Mce proteins in the $\Delta satS$ mutant suggested that SatS is required for the pathogenesis of *M. tuberculosis*. SapM functions in limiting *M. tuberculosis* delivery to degradative lysosomes in macrophages while Mce proteins import lipids and thereby contribute to *M. tuberculosis* growth in macrophages and persistence in the host (24-26). To test a role for SatS in pathogenesis, we infected murine bone marrow-derived macrophages with *M. tuberculosis* H37Rv, the $\Delta secA2$ mutant, $\Delta satS$ mutant, or $\Delta satS$ mutant complemented with *satS*_{Mtb}, and compared intracellular growth over time by plating macrophage lysates for viable bacilli. Compared to H37Rv, the $\Delta satS$ mutant demonstrated a

significant defect in intracellular growth that was comparable to the previously demonstrated, attenuated phenotype of the $\Delta secA2$ mutant, and the $\Delta satS$ mutant phenotype could be complemented (Figure 3.6) (3, 27). Thus, like SecA2, SatS plays an important role in enabling *M. tuberculosis* growth in macrophages even though only a subset of SecA2 substrates are affected by SatS.

SatS effects on cellular levels of its substrates

In addition to the reduced levels of SapM and Mce protein in exported locations (culture filtrates and cell wall fractions), the total cellular (*i.e.* in whole cell lysate) and cytoplasmic levels of these proteins were dramatically reduced in the *M. tuberculosis* $\Delta satS$ mutant compared to H37Rv, the $\Delta secA2$ mutant and the complemented $\Delta satS$ strains (Figure 3.7A and 3.7B). Notably, the reduction observed in the $\Delta satS$ mutant differed from a modest intracellular reduction of SapM and Mce1 proteins in the $\Delta secA2$ mutant (Figure 3.7A and 3.7B). The effect of SatS on cellular SapM and Mce1 levels was specific, as the cellular levels of the 19 kDa lipoprotein were equivalent across strains (Figure 3.7A). The same results were obtained with SapM-expressing *M. smegmatis* strains (Figure 3.7C).

The reduced cellular and cytoplasmic levels of SapM and Mce proteins in the $\Delta satS$ mutant could suggest transcriptional or translational effects of SatS on these proteins or that SatS acts post-translationally to stabilize the proteins prior to their export. Using quantitative Real-Time PCR (qRT-PCR) (Figure 3.8A) and a translational *sapM*'-'*lacZ* fusion (Figure 3.8B), we ruled out the possibilities of SatS functioning in transcription or translation. There was no difference in the level of *sapM* transcript or translation in the *M. tuberculosis* $\Delta satS$ mutant compared to H37Rv. Thus, the effect of SatS on SapM levels in the cytoplasm is post-translational and the likely consequence of SatS protecting SapM prior to export. Notably,

protein export chaperones of Type III and Type VII secretion systems (T3SS and T7SS) exhibit this behavior of stabilizing their substrates in the cytoplasm and protecting them from degradation prior to assisting in their secretion (28, 29). Thus, we considered the possibility that SatS serves as a protein export chaperone for its specific substrates.

SatS and SapM interact

We next tested whether SatS interacts with SapM in mycobacteria, as required for a chaperone:substrate pair, using co-immunoprecipitation experiments. Immunoprecipitations were performed from *M. smegmatis* strains co-expressing C-terminally tagged SapM-FLAG and C-terminally tagged SatS-HA proteins. These epitope tags did not disrupt SapM or SatS functions (Figure 3.9A and 3.9B).

Reasoning that it may be easier to detect a SatS-SapM interaction when SapM export was compromised, we first performed co-immunoprecipitations in a $\Delta secA2$ mutant background. For these experiments we used a $\Delta secA2/\Delta satS$ double mutant expressing SatS_{Mtb} +/- HA tag and SapM-FLAG and immunoprecipitated from whole cell lysates using anti-HA agarose. The resulting immunoprecipitates were analyzed by immunoblotting with FLAG antibodies to detect SapM and SatS antibodies to detect SatS. SapM-FLAG was detected in the immunoprecipitates of the samples from the strain expressing SatS_{Mtb}-HA (Figure 3.10A) indicating SatS and SapM interact. As a control, SapM-FLAG was not recovered when the anti-HA immunoprecipitation was performed from a strain expressing untagged SatS_{Mtb}. Using high percentage (15%) SDS-PAGE, we are able to detect two SapM-FLAG species: a ~31 kDa product corresponding to full length preproteins and a ~29 kDa product corresponding to the cleaved exported product. We confirmed the assignment of the smaller species as mature, cleaved SapM by immunoblotting lysate from a strain expressing Δss -SapM-FLAG (Figure 3.10A). It is striking that while the

smaller exported species is more abundant in the input lysate, the full length preprotein SapM was the species that co-immunoprecipitated with SatS (Figure 3.10A). This is consistent with SatS interacting with SapM preprotein in the cytoplasm, prior to its export and signal sequence cleavage. We investigated whether the signal sequence of SapM is required for the SatS-SapM interaction by immunoprecipitating from a strain co-expressing SatS-HA and Δ ss-SapM-FLAG. SatS-HA and Δ ss-SapM-FLAG co-immunoprecipitated indicating that the signal sequence of SapM is not required for the SatS-SapM interaction (Figure 3.10A).

We were also able to co-immunoprecipitate SapM-FLAG with SatS-HA from cell lysates of a *M. smegmatis* Δ satS strain expressing the same constructs (*i.e.* a *secA2* wild-type background), although there was reproducibly less SapM-FLAG recovered when export was not inhibited (Figure 3.10B). Once again, the SapM preprotein species preferentially co-immunoprecipitated. To address the specificity of SatS interacting with SapM, we also immunoblotted SatS immunoprecipitates with antibody to MspA, which is a cell wall porin that is exported in a SecA2 and SatS-independent manner (16, 30) (Figure 3.5F). MspA did not co-immunoprecipitate with SatS (Figure 3.10B).

The interaction between the preprotein species of SapM and SatS implies that SatS is a cytoplasmic protein. Using an antibody raised against SatS, we confirmed that in both *M. tuberculosis* and *M. smegmatis* SatS is cytoplasmic (Figure 3.4B and 3.2B). Interestingly, SatS in *M. tuberculosis* and *M. smegmatis* migrated on SDS-PAGE at ~65 kDa rather than at its predicted molecular weight of 46 kDa. SatS purified from *Escherichia coli* also ran at 65 kDa (data not shown).

SatS functions prior to SecA2

If SatS functions as a chaperone for preproteins exported by the SecA2 pathway, we predicted its role should come before the role of SecA2 in exporting SapM across the membrane. To test this order of events, we constructed a *M. smegmatis* $\Delta secA2/\Delta satS$ double mutant expressing SapM-FLAG and compared the cellular and secreted levels of SapM of the double mutant to the corresponding single $\Delta secA2$ or $\Delta satS$ mutants. If SatS acts prior to SecA2 in exporting SapM, the $\Delta satS$ mutation should be epistatic to the $\Delta secA2$ mutation, which proved to be the case. The $\Delta secA2/\Delta satS$ double mutant exhibited the equivalent dramatic reduction in cellular and secreted levels of SapM as exhibited by the $\Delta satS$ mutant (Figure 3.11A and 3.11B). Further, there was no additive effect evident on the SapM secretion defect in the $\Delta secA2/\Delta satS$ double mutant compared to the $\Delta satS$ mutant (Figure 3.11B).

SatS behaves as a chaperone to prevent SapM aggregation

The data so far is consistent with the hypothesis that SatS functions as a chaperone for a subset of SecA2 dependent substrates. The hallmark of a chaperone is that it binds to unfolded regions of proteins to prevent inappropriate interactions, such as aggregation (31). To obtain more direct evidence for chaperone activity of SatS, we purified SatS_{Mtb} and SapM-His from *E. coli* and tested the ability of SatS to prevent aggregation *in vitro* of SapM-His preprotein (containing the signal sequence). SapM-His was solubilized from inclusion bodies using 8 M urea, rapidly diluted into refolding buffer (150 fold), and its aggregation was followed by change in light scattering at 350 nm. In the absence of SatS, dilution of denatured SapM-His rapidly led to the formation of light scattering aggregates (Figure 3.12). However, inclusion of SatS in the dilution buffer prevented SapM-His aggregation in a dose-dependent manner (Figure 3.12). As controls, BSA and lysozyme did not reduce SapM-His aggregation (Figure 3.12). A SatS:SapM-

His molar ratio of 2.5:1 was sufficient to completely ablate SapM-His aggregation and even a 0.5:1 ratio was sufficient to reduce aggregation by 33%. The data from this *in vitro* anti-aggregation assay provides strong support for SatS acting as a chaperone and for a direct interaction between SatS and SapM preprotein.

SatS has a new fold and hydrophobic grooves that share similarity with the preprotein binding sites of the SecB chaperone

Although the amino acid sequence of SatS bears no similarity to any known chaperones, the data above supports a role for SatS as a protein export chaperone. To gain further insight into SatS function, we collected diffraction quality crystals and determined the crystal structure of SatS to 2.3 Å (Table 3.4). Upon inspection, the electron density map only corresponded to the last 185 amino acids (L237-E420) C-domain of the SatS sequence (SatS_C). The molecular weight of the SatS_C crystal was ~25 kD as determined by SDS-PAGE, indicating that the protein underwent *in situ* proteolysis in the crystallization buffer. Further investigation revealed striking similarity between the experimentally derived SatS_C secondary structure and the predicted secondary structure of the first ~180 amino acids of the N-domain of SatS (SatS_N) (Figure 3.13A and 3.13B). The SatS_C and SatS_N domains are also similar in size with 41% sequence similarity at the amino acid level (Figure 3.13A and 3.13B). This raises the possibility that SatS is composed of two similar domains with an intervening ~60 amino acids that is predicted to be a flexible, disordered linker (Figure 3.13A and 3.13B). Subsequently, constructs expressing SatS_C or SatS_N in *E. coli* were used to purify the individual domains to homogeneity for crystallization trials. Currently, only SatS_C has yielded diffraction quality crystals diffracting to 1.4 Å resolution (Table 3.4).

SatS_C displays α/β secondary structure composed of a mostly parallel, four stranded β -sheet core, flanked by seven α -helices (Figure 3.14A). The structure revealed a new fold sharing no similarities with any previously solved protein structure in the PDB based on DELTA-BLAST and VAST similarity searches (32, 33). Although the overall polypeptide fold is not similar to known proteins, the surface of SatS_C has pronounced electronegative charge potential that is comparable to many export chaperones, including SecB (34). Furthermore, the SatS_C structure features two surface localized hydrophobic grooves, mapped by the Kyte-Doolittle hydrophobicity scale (35) (Figure 3.14B), that bear similarity to the hydrophobic grooves on SecB (Figure 3.14C and 3.14D) that serve as primary and secondary client binding sites to regions of unfolded preproteins (36). The proximity of the smaller hydrophobic groove in SatS_C (Site 2) to the larger groove in SatS_C (Site 1) as well as their amino acid composition (aromatic and bulky side chains) resemble the arrangement and composition of the client binding sites of SecB. Further, the larger of the two hydrophobic grooves in SatS_C (Site 1) spans nearly the entire length of the domain and is comparable in total size to the ~60 Å long, main binding site in SecB. Because of the similarities between SecB and SatS_C, we speculated that SatS_C may be sufficient to perform SatS chaperone functions. In fact, when we tested SatS_C for chaperone activity in the *in vitro* anti-aggregation assay, SatS_C alone was sufficient to ablate SapM-HIS preprotein aggregation comparable to full length SatS (Figure 3.12).

SatS has at least two separable roles in protein export

The majority of *satS* suppressor mutations severely truncate SatS and are expected to behave like *satS* null mutations (Figure 3.1B). However, one mutation (3S) that caused a single amino acid substitution (G134D) produced wild-type levels of full length SatS protein when compared to wild-type SatS expressed from the same vector backbone (Figure 3.15A). Using this

expression plasmid, we tested the importance of the G134 residue, which maps to the SatS_N domain and is ubiquitous in SatS homologs in mycobacteria.

We first tested if *satS G134D* could complement the SapM-FLAG secretion defect of the *M. smegmatis* $\Delta satS$ mutant by immunoblotting culture filtrates. Since SatS G134D behaved like the $\Delta satS$ null mutant in suppressing *secA2 K129R*, we predicted that SatS G134D would fail to complement the SapM secretion defect of the $\Delta satS$ mutant. Along these lines, SatS G134D exhibited a SapM secretion defect (Figure 3.15A); however, the SapM secretion defect was comparable to the level of secretion in the $\Delta secA2$ mutant. Using the whole cell secreted phosphatase activity assay, expression of either SatS_{Msm}G134D or SatS_{Mtb}G134D in the $\Delta satS$ mutant resulted in a partial defect in secreted SapM, on the level exhibited by the $\Delta secA2$ mutant (Figure 3.15B). These results reveal *satS G134D* to have a secretion defect, although the defect is comparable to that of a $\Delta secA2$ mutant not a $\Delta satS$ mutant.

We also evaluated the effect of SatS G134D on cellular SapM levels. To our surprise, *satS G134D* did not behave like a $\Delta satS$ mutant. Rather, it fully complemented the dramatic $\Delta satS$ reduction in SapM levels seen in whole cell lysates (Figure 3.15A). Furthermore, we were able to co-immunoprecipitate SatS G134D-HA and SapM-FLAG preprotein (Figure 3.15C) indicating that SatS G134D retains the ability to interact with SapM-FLAG. Our discovery that SatS G134D still binds SapM preprotein and maintains cellular levels of SapM, yet SatS G134D exhibits a defect in SapM secretion equivalent to that of a $\Delta secA2$ mutant, indicates that SatS has more than one role in SapM secretion by the SecA2 pathway. Moreover, these multiple functions of SatS in export can be uncoupled.

Discussion

As with all bacterial pathogens, the protein export pathways of *M. tuberculosis* are critical to virulence. Here, we identified SatS, a previously uncharacterized protein of unknown function, as a new protein export factor with a role in intracellular growth of *M. tuberculosis*. We further discovered multiple properties of SatS that indicate a function as a protein export chaperone. As the amino acid sequence of SatS bears no similarity to chaperones and the structure of the SatS_C domain reveals a new fold, SatS appears to represent a new type of protein export chaperone.

Suppressor analysis led to the identification of SatS

Suppressor analysis is a classic approach for identifying genes in pathways, and it was used extensively in early studies of the general Sec pathway in *E. coli* (37, 38). Here, we carried out a suppressor screen using *secA2 K129R*, which encodes a variant of SecA2 that is unable to hydrolyze ATP (11). Past studies lead to a model where SecA2 K129R is locked in a nonfunctional complex with SecY while attempting to export its substrates (8). As a result, SecA2 K129R is trapped at the membrane and SecY proteins are degraded (8). Our discovery that loss-of-function *satS* mutations suppress *secA2 K129R* phenotypes suggests that SatS is required for the detrimental interaction of SecA2 K129R with SecYEG to occur. In fact, deletion of *satS* significantly reversed SecA2 K129R retention at the membrane and the associated SecY degradation, which is consistent with avoidance of the interaction. By extension, these results support a role for SatS in enabling wild-type SecA2 to interact with the SecYEG channel.

One possibility for how SatS promotes SecA2 interactions with the SecYEG channel is that in order for SecA2 to be delivered to or engage the SecYEG channel it must first be bound to a substrate in a translocation competent state and that SatS functions as a protein export

chaperone that facilitates this SecA2-substrate interaction. We favor this role for SatS as it would not only explain why phenotypes of *secA2 K129R* depend on the presence of SatS but it is also consistent with our identification of an interaction between SatS and SapM and the chaperone activities of SatS. An alternate possibility is that SatS is a core component of a SecA2-specific export apparatus with a function mediating the interaction between SecA2 and SecYEG. However, if SatS were to function this way, we would expect all SecA2-dependent substrates would require SatS for export, which was not the case.

SatS as a protein export chaperone

Molecular chaperones are defined by their ability to transiently bind unfolded regions of proteins and, thereby, protect them from inappropriate interactions, such as aggregation, incorrect/premature folding or degradation (31). Chaperones are a common component of protein export systems, with SecB of the general Sec pathway in Gram-negative bacteria and Type III Secretion System Chaperones (T3SCs) being the best characterized examples. In mycobacteria, EspG proteins of Type VII Secretion Systems are the only protein export chaperones identified so far (39, 40). As a subset of molecular chaperones, protein export chaperones have additional functions in export, such as targeting substrates to export machinery. Although there is a notable lack of amino acid and structural similarity between different types of protein export chaperones, commonalities exist. Protein export chaperones are all highly acidic (pI<5.0) proteins that transiently interact with their substrates in the cytoplasm and remain in the cytoplasm when the substrate is exported (28, 40, 41). Additionally, a hallmark of a protein export chaperone is that its role is limited to a subset of the proteins exported by a given system (39, 42, 43). Finally, in some cases, the genes encoding the chaperone and substrate are co-expressed and in an operon (44).

SatS has many features of a protein export chaperone. SatS is a highly acidic (pI 3.83), cytoplasmic protein with a role promoting export of a subset of the proteins exported by the SecA2 pathway. Further, the *satS* and *sapM* genes are co-transcribed in an operon and we obtained evidence of a SatS:SapM interaction occurring in mycobacteria. SatS preferentially interacted with the full-length preprotein of SapM indicating that the interaction occurs in the cytoplasm prior to SapM export. However, like other protein export chaperones (T3SCs, SecB, and EspG5) (40, 45, 46) where binding occurs in regions of the mature domain of the substrate, the signal sequence was not required for the SatS-SapM interaction. Finally, the *in vitro* anti-aggregation effect of SatS on SapM preprotein provided the most direct proof of a SatS:SapM interaction and a chaperone function for SatS.

Along with the above chaperone features, in the absence of SatS, the level of SapM in the cytoplasm was dramatically reduced. This effect of SatS on intracellular SapM levels is post-translational and reminiscent of effects of T3SCs and EspG chaperones protecting their cognate substrates from degradation prior to export (29, 43). In comparison to the dramatic reduction in intracellular SapM in the $\Delta satS$ mutant, the $\Delta secA2$ mutant exhibited only a modest effect on intracellular SapM levels. This difference explains why the $\Delta satS$ mutant secretion defect was much more dramatic than that of the $\Delta secA2$ mutant.

Because of the dramatically reduced levels of SapM in the whole cell lysate of the $\Delta satS$ mutant, it was not immediately clear if the role of SatS in SapM secretion was solely to maintain intracellular levels of SapM preprotein or if SatS had additional roles. By evaluating the *satS* *G134D* mutant, we revealed the existence of at least one additional role for SatS in promoting SapM secretion. In the *satS* *G134D* mutant, intracellular SapM was maintained at wild-type levels; yet, there remained a SapM secretion defect. Future studies should address this second

function, which could be a role for SatS in targeting substrates to the SecA2 pathway and/or in maintaining SapM in an unfolded state for protein translocation across SecYEG.

The SatS structure defines a new fold with hydrophobic grooves typical of substrate binding sites

Although we set out to solve the structure of SatS in its entirety, we were only able to obtain structural information for the C-terminal half of the protein (SatS_C), which arose during crystallization. However, the primary sequence and secondary structure similarity between the N-terminal and C-terminal halves of SatS raise the interesting possibility of SatS being composed of tandem SatS_C-like domains. Investigation of the SatS_C structure revealed a large network of negatively charged amino acids surrounding two surface exposed hydrophobic grooves, which are similar in arrangement, shape and size to the hydrophobic client binding sites of a SecB monomer (45). In the solution structure of SecB in complex with a preprotein, the unfolded preprotein wraps around the SecB tetramer through interactions with the hydrophobic client binding sites. This binding architecture helps explain the means by which SecB maintains Sec preproteins in an unfolded state, as is required for their transport through the SecYEG channel (5). The similarity in hydrophobic grooves in SatS and SecB is intriguing since SatS works with the SecA2 pathway, which also uses the SecYEG channel. Moreover, these similarities suggest that the hydrophobic grooves in SatS may serve as similar substrate binding sites. In fact, in the anti-aggregation assay the SatS_C domain was sufficient for preventing SapM aggregation, indicating that SatS_C is capable of directly interacting with SapM. Mycobacteria lack a canonical SecB protein export chaperone, although in *M. tuberculosis* there is a SecB-like protein that functions as a chaperone for a toxin-antitoxin system (47). Thus, even though SecB and SatS are not evolutionarily conserved, it is interesting to speculate a SecB-like function for SatS in the mycobacterial SecA2 pathway. It may be that SatS is a specific adaptation of the

mycobacterial/Actinomycetales SecA2 systems, since SecA2 systems outside of the order Actinomycetales do not have SatS orthologs

SatS is required for growth of *M. tuberculosis* in macrophages

Prior TraSH/Tnseq analyses using pooled libraries of transposon mutants predicted SatS to be required during murine and macrophage infections (48, 49); however, this prediction had never been validated. Here, using a $\Delta satS$ mutant and a complemented strain, we directly demonstrated a role for SatS in *M. tuberculosis* growth in macrophages. This data argues for an important role of SatS and its specific substrates in pathogenesis. Given that only a subset of SecA2 substrates are affected by SatS, future studies should include investigation of SatS substrates and their contribution to pathogenesis. Since our approach for identifying SatS substrates was not exhaustive, there may also exist SatS-dependent proteins that remain to be identified.

Conclusion

By way of a genetic screen in *M. smegmatis*, we identified a new protein SatS with roles in protein export in *M. tuberculosis*. This work not only expands our understanding of the specialized SecA2 protein export pathway of mycobacteria but it provides important functional information for a previously uncharacterized *M. tuberculosis* protein that contributes to pathogenesis. Further, by assigning a chaperone function to SatS, our studies expand our appreciation of the diversity of chaperones in biological systems. Although chaperones have common functions, substantial structural diversity exists among these proteins, which is further highlighted by the new fold revealed in the structure of SatS_C.

Table 3.1. Primers used in this study.

Name	Sequence (5' – 3')	Description
<i>satS_{Mtb}</i> US flank F	GCGGTACCGCCGTGGGTCAACT TCAGTAAC	Contains engineered KpnI site, used to amplify US flank for pSM42
<i>satS_{Mtb}</i> US flank R	GCGTCTAGAGGTGCTGATGATC TCGTTCGATG	Contains engineered XbaI site, used to amplify US flank for pSM42
<i>satS_{Mtb}</i> DS flank F	GCGAAGCTTATGATCGACCGAT CTTCCTG	Contains engineered HindIII site, used to amplify DS flank for pSM42
<i>satS_{Mtb}</i> DS flank R	GCGACTAGTCGGGCTGTTTTCTA CGTTGT	Contains engineered SpeI site, used to amplify DS flank for pSM42
<i>satS_{Msm}</i> US flank F	AACATATGCGCAACTGGGTGTG CCGTATCACTG	Contains engineered NdeI site, used to amplify US flank for pLL50
<i>satS_{Msm}</i> US flank R	AAGCTAGCAGCAGCCATGCGGC ACAGCCTAAC	Contains engineered NheI site, used to amplify US flank for pLL50
<i>satS_{Msm}</i> DS flank F	ATGCTAGCTCCCGGCTCCGTCAG GAGTAGCG	Contains engineered NheI site, used to amplify DS flank for pLL50
<i>satS_{Msm}</i> DS flank R	AACATATGAGCCACCCGGCGAA ATTGAAGCCAC	Contains engineered NdeI site, used to amplify DS flank for pLL50
1684-F-Native	AGTTAATTAACGTGTGCTCGAC GGCCTGGTTGCC	Contains engineered PacI site, used to construct <i>satS_{Msm}</i> plasmids pBM4, pBM22, and pBM23
1684-R	AATGGCCACTACTCCTGACGGA GCCGGGACTCCAC	Contains engineered BalI site, used to construct <i>satS_{Msm}</i> plasmids pBM4 and pBM22
1684-R-HA	ATTGGCCATCAGGCGTAGTCCG GCACGTCGTACGGGTACTCCTG ACGGAGCCGGGACTCCAC	Contains engineered HA tag and BamHI site, used to construct <i>satS_{Msm}</i> -HA plasmid pBM23
Rv3311-F	AATGGCCACTGACCTCGTACCC ATCCGCTTGAG	Contains engineered MscI site, used to construct <i>satS_{Mtb}</i> plasmids pBM13, pBM60, and pBM80
Rv3311-R	AATGGCCACTAGCCTTCGCCGG CTGAC	Contains engineered MscI site, used to construct <i>satS_{Mtb}</i> plasmid pBM80
Rv3311-HA-R	TTAAGCTTCGCGCCTGAGCCGC GACTCC	Contains engineered HindIII site, used to construct <i>satS_{Mtb}</i> plasmids pBM13 and pBM60
Rv3311-G134D-F	GCCCAGGATGGGATTGTCTGTTG AAGAACTTCGA	Used for site directed mutagenesis on pBM80 to generate pBM87
Rv3311-G134D-R	TCGAAGTTCTTCAACGACAATCC CATCCTGGGC	Used for site directed mutagenesis on pBM80 to generate pBM87
SapM-F	TGGCCAACCGCGGAATCCAGGC TCTC	Contains engineered MscI site, used to construct <i>sapM</i> plasmids pJTS130, and pBM56
Dss-SapM-F	TGGCCAAGACCTTCGCGCACGT GG	Contains engineered MscI site, used to construct <i>sapM</i> plasmids pJTS132 and pBM61

SapM-R	AAGCTTCCATGCGGCACAGAAT AGCGAC	Contains engineered HindIII site, used to construct <i>sapM</i> plasmids pJTS130 and pJTS132
SapM-FLAG-R	ACTAAGCTTTCACTTGTCGTCGT CGTCCTTGTAAGTCCGTATACGAG CCGCCGTCGCCCCAAATATCG	Contains engineered linker-FLAG and HindIII site, used to construct <i>sapM</i> plasmids pBM56 and pBM61
SapM-Promoter F	ACTGGTACCTTCACGCAGCGTG GTCAGTC	Used with EcoRI site in TOPO to construct <i>sapM-lacZ</i> reporter pBM94
PsapM-LacZ R	AGGATCCATTCCGCGGAGCATG CCGGGAG	Contains engineered BamHI site to construct <i>sapM-lacZ</i> reporter pBM94
SapM-3311 gap F	GACGGGTTATGCGACCAATG	Amplify the region between <i>sapM</i> and <i>satS</i> for RT-PCR
SapM-3311 gap R	CTCAAGCGGATGGGTACGAG	Amplify the region between <i>sapM</i> and <i>satS</i> for RT-PCR
Mce4A hsp60 F	AAGATATCCGAACGGAAACGCC AAACG	Contains engineered EcoRV site, used to construct <i>mce4A_{Msmeg}</i> -HA plasmids pBM44
Mce4A hsp60 R	TAAGCTTCGTCCCTTTCCGCGAA C	Contains engineered HindIII site, used to construct <i>mce4A_{Msmeg}</i> -HA plasmids pBM44
sigA RT F	AAGCGAACAGCGGCGAAGTC	qRT-PCR primer for <i>sigA</i>
sigA RT R	TTCGGGATGGTGCTGGTCGTAG	qRT-PCR primer for <i>sigA</i>
sapM RT F	ATCGTTGCTGGCCTCATGG	qRT-PCR primer for <i>sapM</i>
sapM RT R	AGGGAGCCGACTTGTTACC	qRT-PCR primer for <i>sapM</i>
sapM <i>E. coli</i> F	GTCTCTCCCATGCTCCGCGGAAT CCAG	Used to express <i>sapM</i> in the <i>E. coli</i> pMSCG-28 vector
sapM <i>E. coli</i> R	GGTTCTCCCCAGCGTCGCCCCAA ATATCGGTTATTGG	Used to express <i>sapM</i> in the <i>E. coli</i> pMSCG-28 vector
satS <i>E. coli</i> F	TTTTTTCATATGGTTGCTGACCT CGTACCCATC	Contains engineered NdeI site, used to express <i>satS</i> in <i>E. coli</i> Pet28b vector
satS _C <i>E. coli</i> F	TTTTTTCATATGCGGGACTTCTG GTTGCAG	Contains engineered NdeI site, used to express <i>satS_C</i> in <i>E. coli</i> Pet28b vector
satS/satS _C <i>E. coli</i> R	TTTTTTAAGCTTCTATTCGCGCC TGAGCC	Contains engineered HindIII site, used to express <i>satS/satS_C</i> in <i>E. coli</i> Pet28b vector

Table 3.2. Plasmids used in this study.

Plasmid	Antibiotic Resistance	Description	Source
pMV261.kan	Kan	Multicopy mycobacterial vector with <i>hsp60</i> promoter	(50)
pMV361.kan	Kan	Single-copy mycobacterial vector with <i>hsp60</i> promoter, integrates in mycobacteriophage L5 <i>attB</i> site	(50)
pMV306.kan	Kan	Single-copy, promoterless mycobacterial vector, integrates in mycobacteriophage L5 <i>attB</i> site	(50)
pJSC77	Kan	Multicopy mycobacterial vector, HA tag cloned into pMV261	(51)
pLL2	Hyg	single-copy mycobacterial shuttle vector, integrates in mycobacteriophage Tweety <i>attB</i> site	(8)
pYA810	Kan	Integrating <i>M. smegmatis</i> <i>secA2</i> complementation plasmid in pMV361.kan	(52)
pNR25	Kan	Integrating <i>M. smegmatis</i> <i>secA2 K129R</i> in pMV361.kan	(11)
pLL50	Hyg	Suicide vector pMP62 containing flanking regions to delete <i>satS_{Msm}</i>	This work
pBM11	Hyg	Suicide vector pMP62 containing <i>secA2_{Msm}</i> and flanking regions to reintroduce <i>secA2</i> to the BAF1 strain	This work
pBM4	Hyg	<i>satS_{Msm}</i> under native promoter in pLL2	This work
pBM80	Hyg	<i>satS_{Mtb}</i> under <i>hsp60</i> promoter in pLL2	This work
pSM42	Hyg	<i>satS_{Mtb}</i> upstream and downstream flanks inserted into pYUB854	This work
pSM45	Hyg	Phasmid for knocking out <i>satS_{Mtb}</i>	This work
pSM60	Hyg	Phage for knocking out <i>satS_{Mtb}</i>	This work
pBM13	Kan	<i>satS_{Mtb}</i> under <i>hsp60</i> promoter in pMV306.kan	This work
pJTS130	Kan	<i>sapM</i> under <i>hsp60</i> promoter in pMV261.kan	(12)
pJTS132	Kan	Δss - <i>sapM</i> under <i>hsp60</i> promoter in pMV261.kan	This work
pYUB76	Kan	Multicopy mycobacterial shuttle vector with promoterless <i>lacZ</i> gene	(53)
pBM94	Kan	<i>p_{sapM}</i> - <i>sapM</i> '- ' <i>lacZ</i> in pYUB76	This work
pBM56	Kan	<i>sapM</i> under <i>hsp60</i> promoter in pMV261.kan containing a C-terminal linker and FLAG tag	This work

pBM60	Hyg	<i>satS_{Mtb}</i> under <i>hsp60</i> promoter in pLL2 containing a C-terminal HA tag	This work
pBM61	Kan	$\Delta ss-sapM$ under <i>hsp60</i> promoter in pMV261.kan containing a C-terminal linker and FLAG tag	This work
pBM22	Hyg	<i>satS_{Msm}</i> under native promoter in pLL2 amplified from suppressor 3S to contain the G134D point mutation	This work
pBM23	Hyg	<i>satS_{Msm}</i> under native promoter in pLL2 amplified from suppressor 3S to contain the G134D point mutation and containing a C-terminal HA tag	This work
pBM87	Hyg	<i>satS_{Mtb}</i> under <i>hsp60</i> promoter in pLL2 with point mutation G134D generated by site directed mutagenesis	This work
pBM44	Kan	<i>Mce4A_{Msm}</i> under <i>hsp60</i> promoter in pJSC77 containing a C-terminal HA tag	This work
pHSG58	Kan	Multi-copy Ms1704-HA expression vector under <i>hsp60</i> promoter	(52)
pRH1	Carb	<i>sapM</i> in the <i>E. coli</i> pMSCG-28 vector containing a C terminal His tag	This work
pRH2	Kan	<i>satS</i> in <i>E. coli</i> Pet28b vector	This work
pRH3	Kan	<i>satS_C</i> in <i>E. coli</i> Pet28b vector	This work

Table 3.3. *M. tuberculosis* and *M. smegmatis* strains used in this study.

<i>M. tuberculosis</i> Strain	Description	Source
H37Rv	<i>M. tuberculosis</i> wild-type (WT) + Ev (pMV306.kan)	(12)
MBTB443	$\Delta secA2$ + Ev (pMV306.kan)	(12)
MBTB512	$\Delta satS$ + Ev (pMV306.kan)	This work
MBTB513	$\Delta satS$ + <i>psatS</i> (pBM13)	This work
<i>M. smegmatis</i> Strain		
mc ² 155	<i>M. smegmatis</i> wild-type (WT)	(54)
NR116	$\Delta secA2$	(11)
BAF1	$\Delta secA2/\Delta satS$	This work
BM10	$\Delta satS$	This work

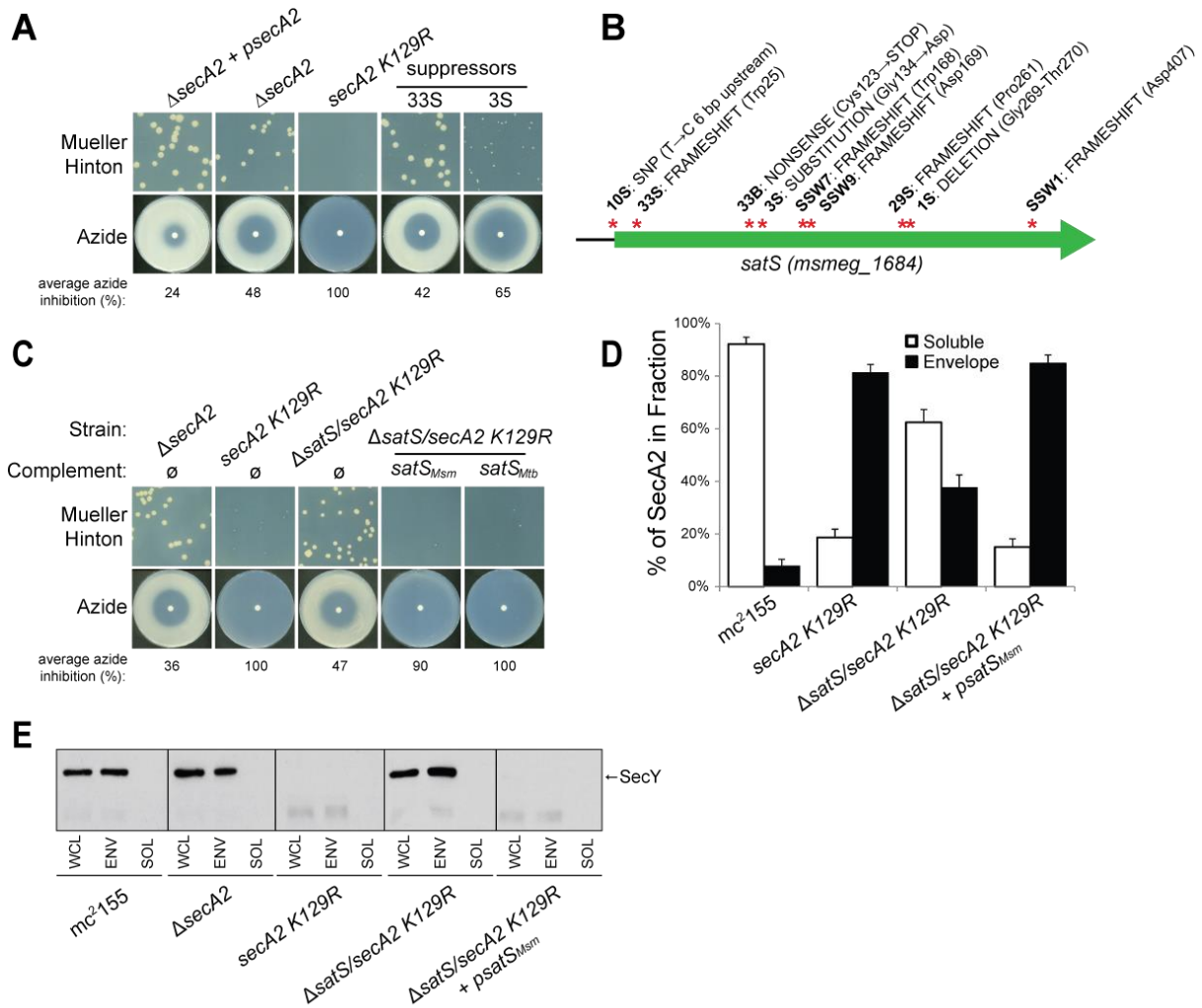


Figure 3.1. $\Delta satS$ mutant suppresses SecA2 K129R phenotypes (A) Mueller-Hinton growth phenotypes and azide sensitivity of *M. smegmatis* $\Delta secA2$ mutant expressing wild-type $secA2$ ($\Delta secA2 + psecA2$), an empty vector, or $secA2 K129R$ ($secA2 K129R$), and two representative suppressors. (B) Nine suppressor mutations affected $satS$ (*msmeg_1684*). Red stars indicated approximate locations of suppressor mutations. For panels C, D, and E, wild-type *M. smegmatis* mc^2155 , $\Delta secA2$, $secA2 K129R$, $\Delta satS/secA2 K129R$, and $\Delta satS/secA2 K129R$ complemented with $satS_{Msm}$ or $satS_{Mtb}$ were used. (C) Mueller-Hinton growth phenotypes and azide sensitivity of the strains described above. (D) Subcellular fractions were separated by SDS-PAGE, and SecA2 protein was detected by Immunoblot. Densitometry was used to quantify SecA2 levels in the soluble and envelope fractions (ImageJ). Percent localization to a given fraction for SecA2 is reported as the percentage of the total (soluble + envelope). Error bars indicate the standard deviation of the mean of three independent replicates for each strain. (E) Whole cell lysates (WCL) and subcellular envelope (ENV) and soluble (SOL) fractions were separated by SDS-PAGE and SecY protein was detected by Immunoblot. All results shown are representative of at least three independent experiments.

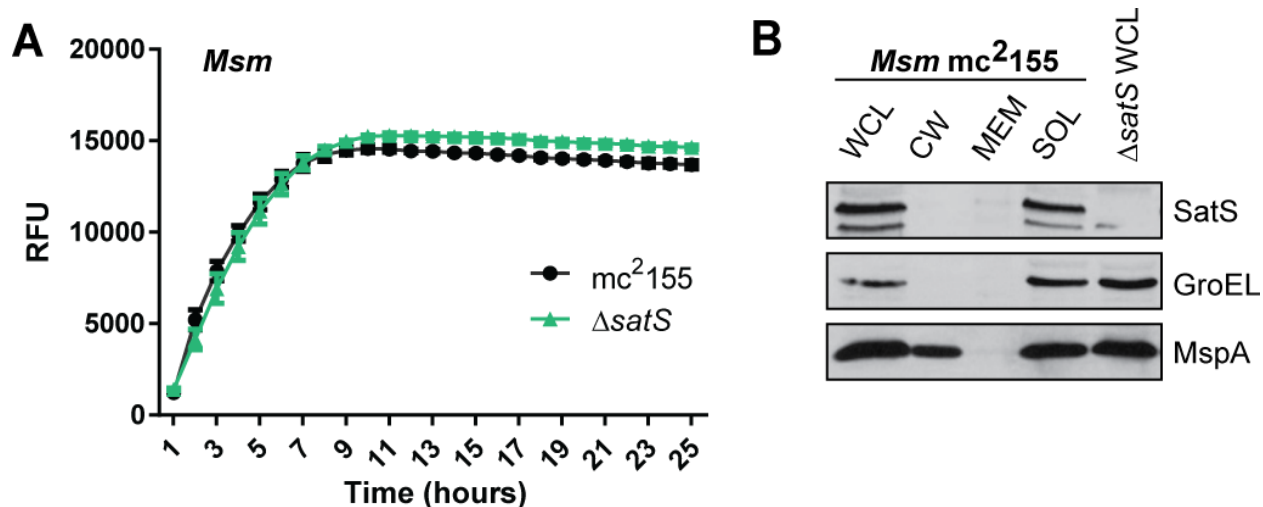


Figure 3.2. Characterization of the *M. smegmatis* $\Delta satS$ mutant. (A) Liquid cultures of *M. smegmatis* mc²155 and $\Delta satS$ were grown in 96-well plates with a starting cell count of 1×10^4 cells/well. After 24 h, resazurin was added and relative fluorescence units (RFU) were measured over the next 20 h. Polyclonal antibodies against SatS were generated in rabbits and used to localize SatS in *M. smegmatis* mc²155 by Immunoblot, using the $\Delta satS$ mutant as a control for specificity. GroEL and MspA were used to validate the purity of the fractions.

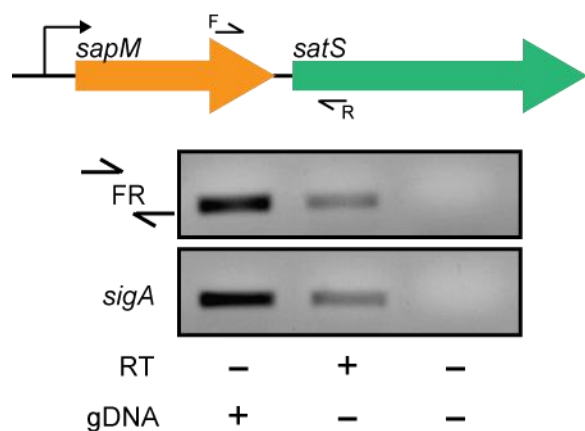


Figure 3.3. *satS* and *sapM* are co-transcribed. Visualization of the operon structure of *sapM* and *satS* and the location of the primers used. Primer set FR flanking the intergenic region between *sapM* and *satS* and a control primer set in the housekeeping gene *sigA* were used to amplify *M. tuberculosis* H37Rv RNA in the presence or absence of RT.

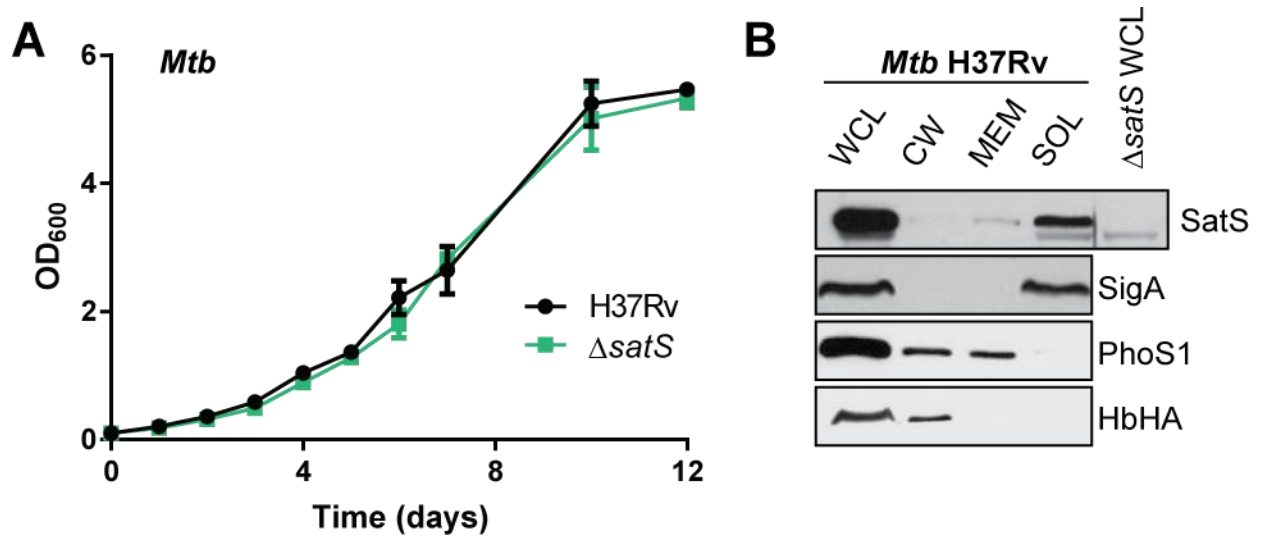


Figure 3.4. Characterization of the *M. tuberculosis* $\Delta satS$ mutant. (A) The $\Delta satS$ mutant of *M. tuberculosis* and the parental H37Rv strain were grown in liquid 7H9 medium with ADS supplementation. Growth was monitored by measuring optical density (OD₆₀₀) over time. Polyclonal antibodies against SatS were generated in rabbits and used to localize SatS in *M. tuberculosis* H37Rv by Immunoblot, using the $\Delta satS$ mutant as a control for specificity. SigA, PhoS1 and HbHA were used to validate the purity of the fractions.

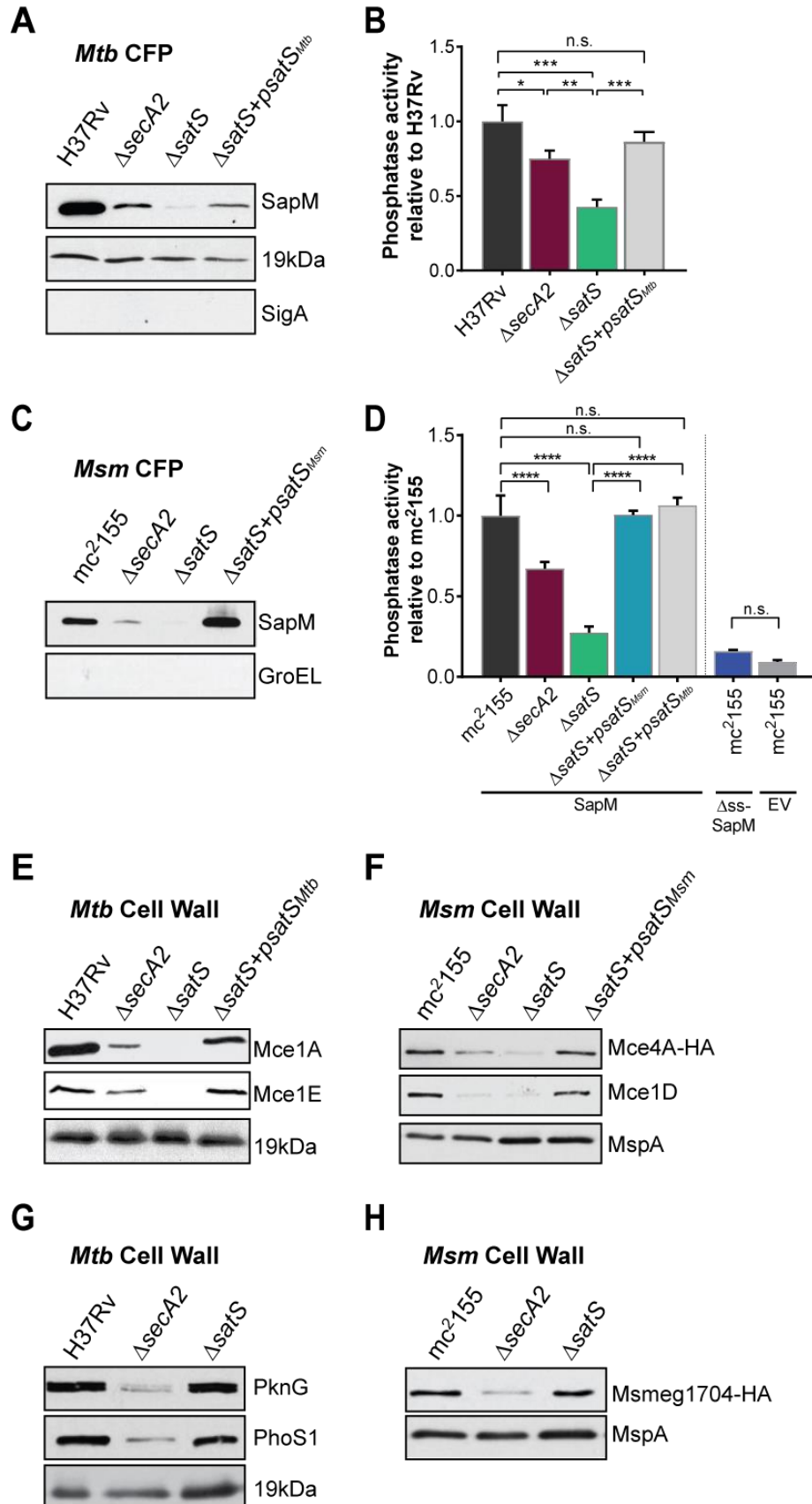


Figure 3.5. SatS is required for the export of SapM and Mce proteins (A) Equal protein from culture supernatants from *M. tuberculosis* H37Rv, $\Delta secA2$, $\Delta satS$ and the complemented strain ($\Delta satS+psatS$) were immunoblotted for SapM, 19kDa, or the cytoplasmic protein SigA as a control for lysis. (B) Phosphatase activity in triplicate culture supernatant samples was examined by quantifying cleavage of pNPP. Rates of pNPP cleavage were normalized to H37Rv. (C) Equal protein from culture supernatants from *M. smegmatis* mc²155, $\Delta secA2$, $\Delta satS$ and the complemented strain ($\Delta satS+psatS_{msm}$) were examined for levels of SapM by Immunoblot, using the cytoplasmic protein GroEL as a control for lysis. (D) Whole cell phosphatase activity assay in *M. smegmatis*. All strains are expressing SapM, SapM lacking its signal sequence (Δss -SapM), or an empty vector as indicated. Triplicate wells containing 2×10^5 cells/well were grown in a 96 well plate for 24 hours at 37°C before measuring phosphatase activity by quantifying cleavage of pNPP. Rates were normalized to mc²155 + SapM. (E) Equalized cell wall fractions of *M. tuberculosis* H37Rv, $\Delta secA2$, $\Delta satS$ and complemented ($\Delta satS+psatS_{Mtb}$) strains were analyzed by immunoblot using Mce1A, Mce1E, and 19KDa antibodies to monitor differences in protein levels. (F) Equalized *M. smegmatis* mc²155, $\Delta secA2$, $\Delta satS$, and $\Delta satS+psatS_{Msm}$ cell wall fractions were analyzed by immunoblot using HA (Mce4A-HA), Mce1D, and MspA antibodies. (G) Equalized cell wall fractions of *M. tuberculosis* H37Rv, $\Delta secA2$, and $\Delta satS$ strains were analyzed by immunoblot using PknG, PhoS1, and 19kDa antibodies. (H) Equalized *M. smegmatis* mc²155, $\Delta secA2$, and $\Delta satS$, cell wall fractions were analyzed by immunoblot using HA (Msmeg1704-HA) and MspA antibodies. All data is representative of at least three independent experiments and all error bars represent standard deviation of the mean of three independent replicates for each strain. n.s. – no significant difference; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$ by ANOVA and Tukey's post hoc test.

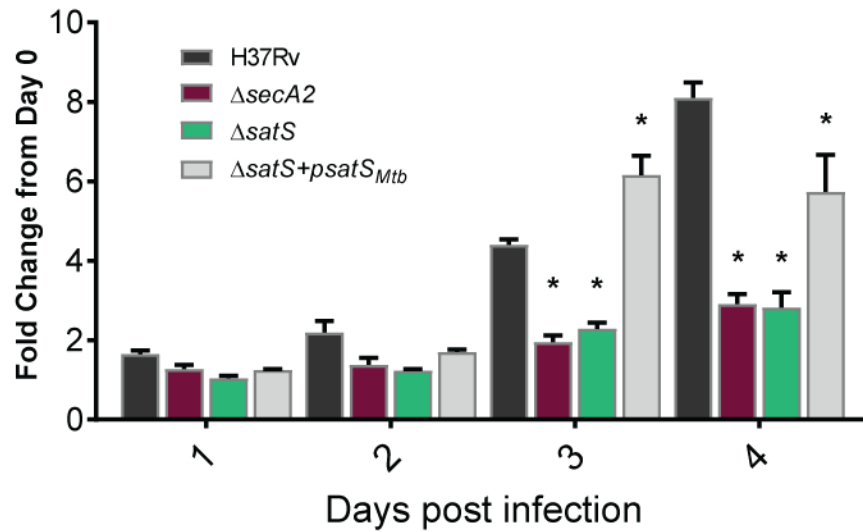


Figure 3.6. SatS contributes to *M. tuberculosis* virulence. Nonactivated BMDM were infected at an MOI of one with *M. tuberculosis* H37Rv, $\Delta secA2$, $\Delta satS$, or $\Delta satS+psatS_{Mtb}$, and CFU burden was monitored over the course of a 4 day infection. The fold change in CFU over the course of the 4 day macrophage infection for each strain was calculated; the points represent means of triplicate wells, and the error bars represent standard deviations (SD). *, $p < 0.01$; when compared to H37Rv by ANOVA and Tukey's post hoc test. Shown is a representative experiment of four independent experiments.

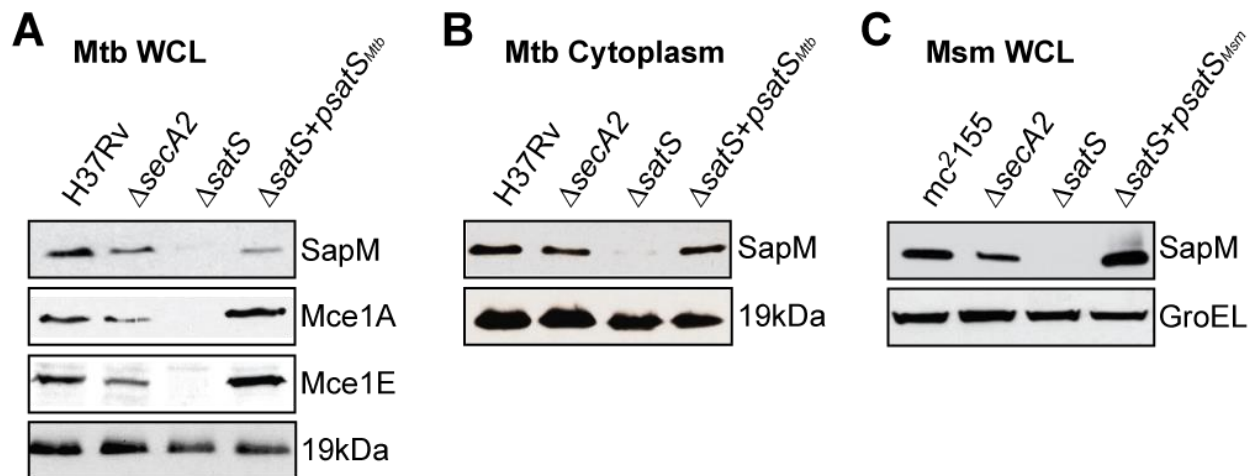


Figure 3.7. SatS effects cellular levels of its substrates (A) Equal protein levels from whole cell lysates or (B) the soluble, cytoplasmic fraction of *M. tuberculosis* H37Rv, $\Delta secA2$, $\Delta satS$ and the complemented strain ($\Delta satS+psatS_{Mtb}$) were immunoblotted for SapM, Mce1A (Figure 4A only), Mce1E (Figure 4A only) and 19kDa. (C) Equal protein levels whole cell lysates of *M. smegmatis* mc²155, $\Delta secA2$, $\Delta satS$ and the complemented strain ($\Delta satS+psatS_{Msm}$) were immunoblotted for SapM and GroEL. All data is representative of at least three independent experiments.

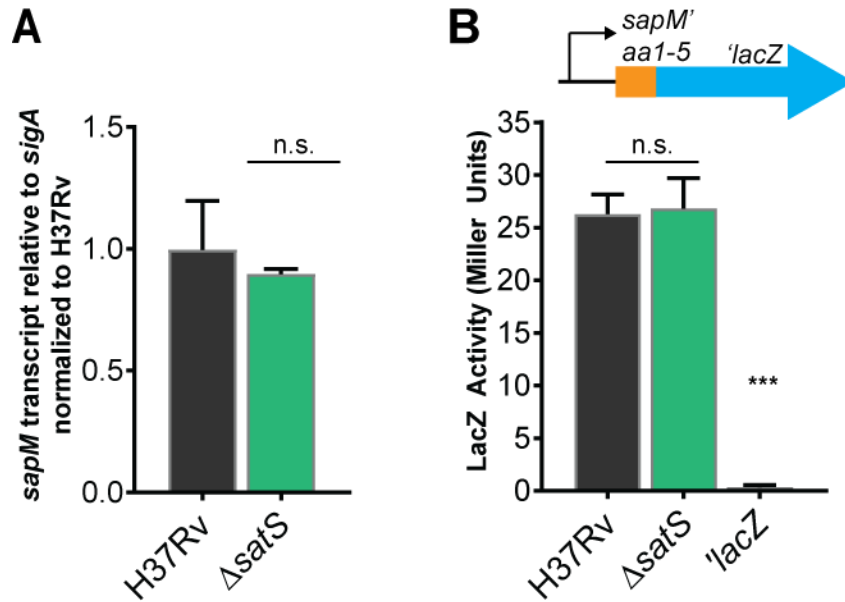


Figure 3.8. SatS does not affect *sapM* transcription or translation (A) RNA was isolated from *M. tuberculosis* H37Rv and Δ*satS*. *sapM* transcript levels were measured by quantitative RT-PCR and normalized to the level of *sigA* transcript. Data shown is for the mean of three biological replicates. (B) Expression of the promoterless *lacZ* gene in pYUB76 when fused to the 170-bp region upstream of *sapM*, containing the putative promoter and first 15 bp of *sapM*, was measured in liquid medium by measuring hydrolysis of *o*-nitrophenyl-β-D-galactoside. The results are from a representative experiment performed in triplicate. In all experiments, the error bars represent standard deviations. n.s. – no significant difference; ***, $p < 0.001$ when compared to H37Rv or Δ*satS* by ANOVA.

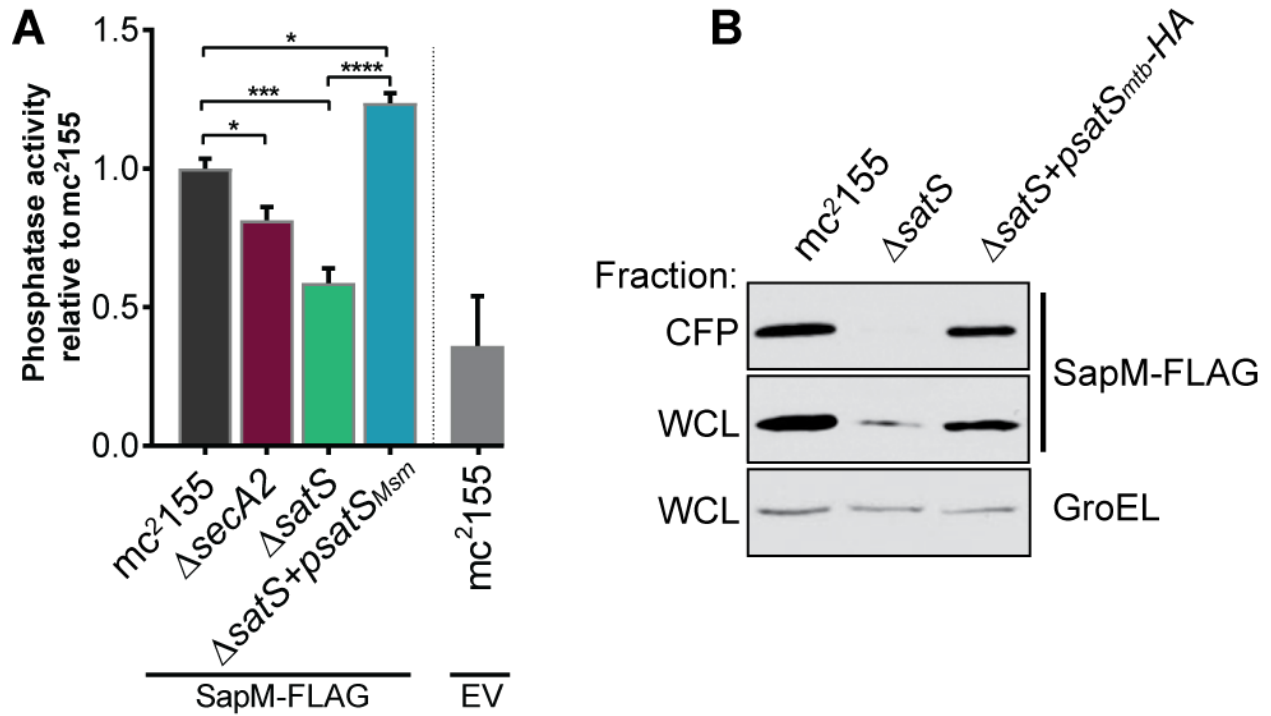


Figure 3.9. Epitope tags do not disrupt SapM or SatS functions (A) Whole cell phosphatase activity assay of *M. smegmatis* strains expressing SapM-FLAG or an empty vector. The error bars represent standard deviation from the mean of three independent replicates. *, $p < 0.05$; ***, $p < 0.001$; ****, $p < 0.0001$ by ANOVA. (B) Equal protein levels from whole cell lysates and culture supernatants of mc²155, ΔsatS, and the HA-tagged complemented strain (ΔsatS+psatS-HA) were examined for levels of SapM-FLAG and GroEL by Immunoblot.

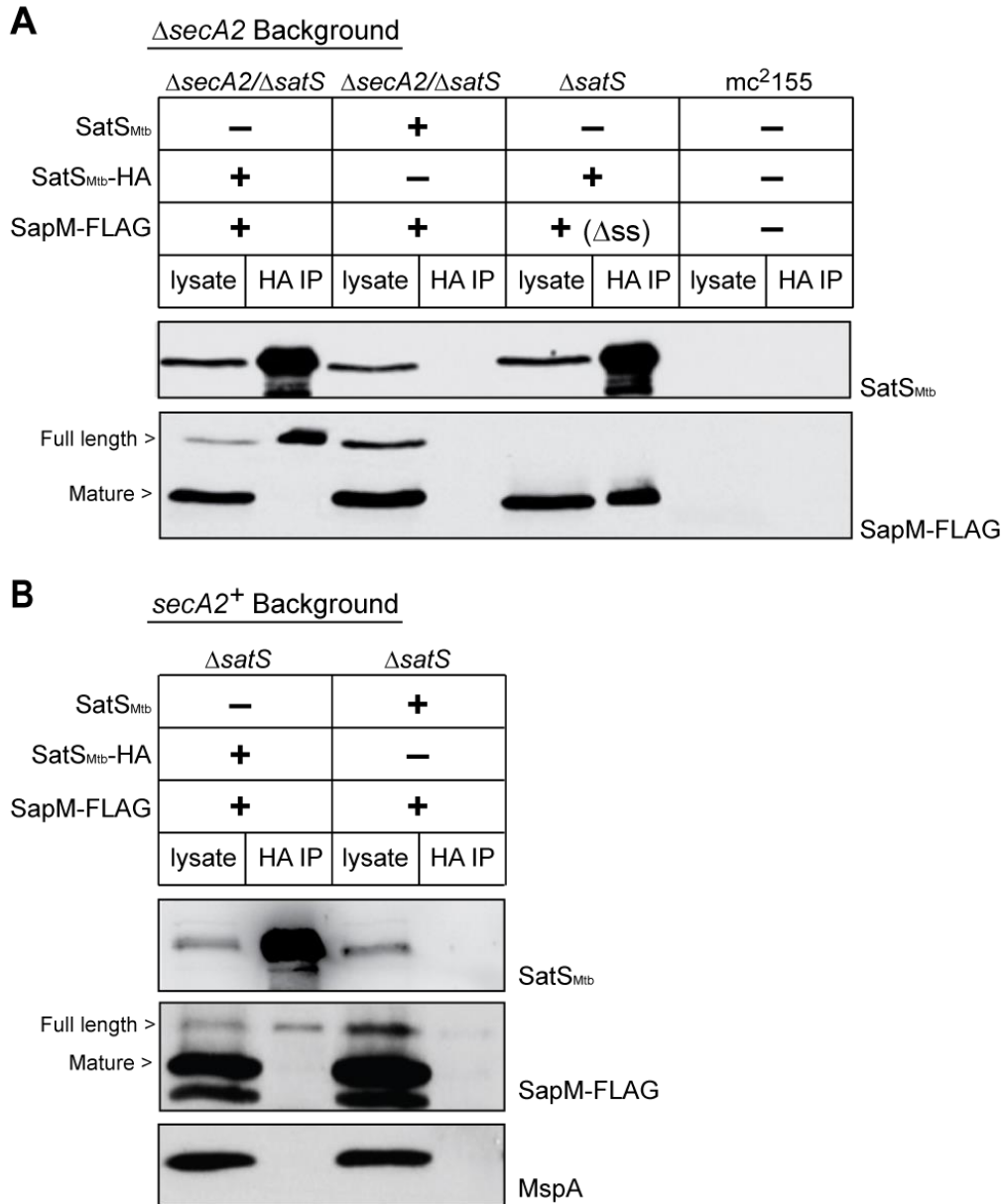


Figure 3.10. SatS and SapM interact (A) Lysate from *M. smegmatis* $\Delta secA2/\Delta satS$ expressing SapM-FLAG and either SatS_{Mtb}-HA or SatS_{Mtb} without a tag, $\Delta satS$ expressing Δss -SapM-FLAG and SatS_{Mtb}-HA, or mc²155 with two empty vectors (as shown above the blot) were used for co-immunoprecipitation using anti-HA conjugated beads. Lysates (left) and immunoprecipitations (right) for each strain were probed with SatS antibody and FLAG antibody for SapM. Two different sizes of SapM-FLAG corresponding to the full-length (signal sequence-containing) and mature (cleaved signal sequence) species were detected. (B) Lysate from *M. smegmatis* $\Delta satS$ expressing SapM-FLAG and either SatS_{Mtb}-HA or SatS_{Mtb} without a tag were used for co-immunoprecipitation using anti-HA conjugated beads. Lysates (left) and co-immunoprecipitations (right) for each strain were probed with SatS antibody, FLAG antibody, and MspA antibody. All data is representative of at least three independent experiments.

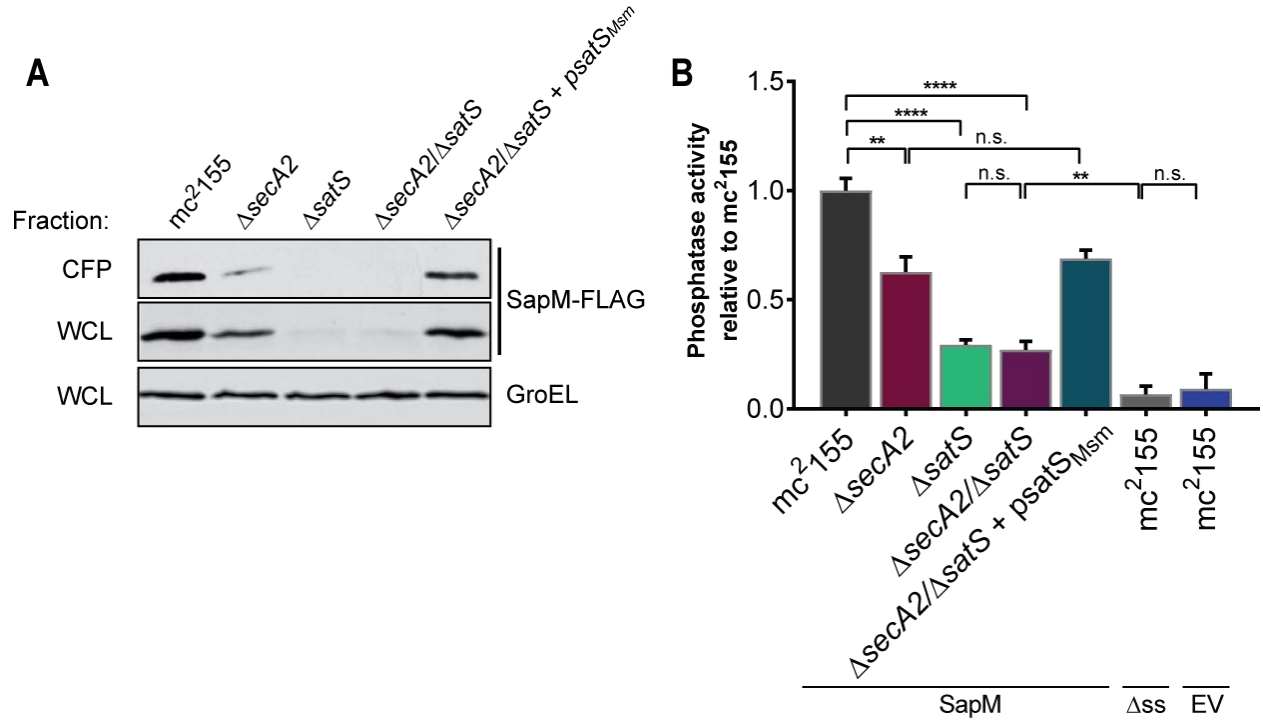


Figure 3.11. SatS functions prior to SecA2 (A) Equal protein from culture supernatants (CFP) and whole cell lysates (WCL) from *M. smegmatis* mc²155, ΔsecA2, ΔsatS, the ΔsecA2ΔsatS double mutant, and ΔsecA2ΔsatS expressing wild-type SatS (ΔsatS+psatS_{Msm}) were examined for levels of SapM-FLAG, SatS, and GroEL by Immunoblot. (B) Whole cell phosphatase activity assay using the above *M. smegmatis* strains. All strains are expressing SapM, SapM lacking its signal sequence (Δss-SapM), or an empty vector as indicated. Triplicate wells were grown in a 96 well plate for 24 hours at 37°C before measuring phosphatase activity by quantifying cleavage of pNPP. Rates were normalized as described above. All data is representative of at least three independent experiments and all error bars represent standard deviation of the mean of three independent replicates for each strain. n.s. – no significant difference; **, p<0.01; ****, p < 0.0001 by ANOVA and Tukey's post hoc test.

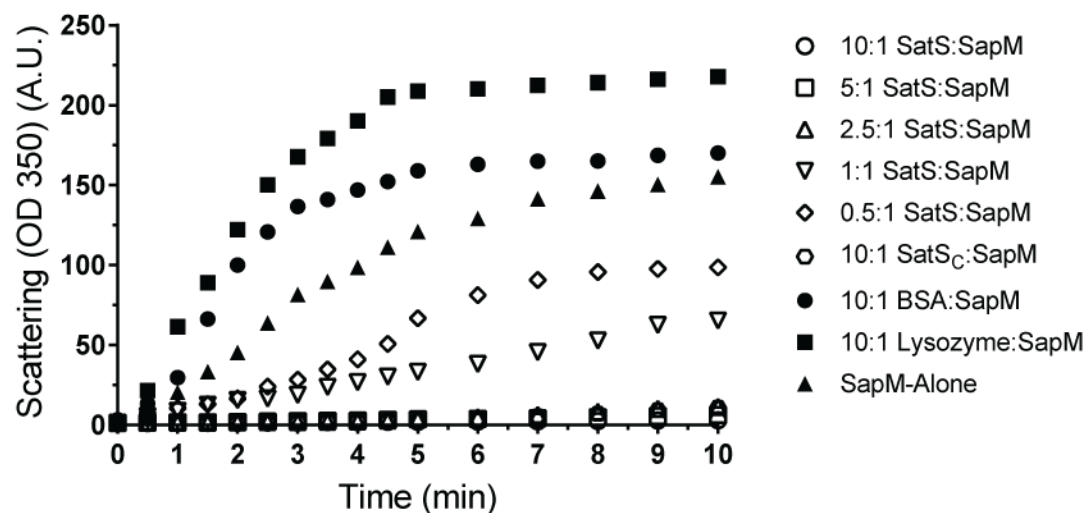


Figure 3.12. SatS and SatSc prevent aggregation of SapM *in vitro*. Denatured SapM-His was diluted 150 fold to a final concentration of 1 μ M in 40 mM HEPES, 100 mM NaCl, pH 7.4. SapM-His aggregation was monitored by light scattered (350 nm) at 25°C in the presence or absence of SatS/SatSc or, as controls, lysozyme or BSA. A molar ratio of 2.5:1 of SatS:SapM-His could prevent SapM-His aggregation and aggregation was significantly reduced using a molar ratio of 0.5:1.

Table 3.4. SatS_C X-ray Structure Validation Details.

	SatS_C(Br)	SatS_C(Apo)
PDB ID	6DRQ	6DNM
Data collection		
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	48.8, 50.6, 76.1	48.9, 50.6, 76.5
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 90	90, 90, 90
Resolution (Å)	2.3	1.4
<i>R</i> _{merge}	0.050 (0.105)	0.081 (0.254)
<i>I</i> / <i>σ</i> (<i>I</i>)	27.4	14.4
Completeness (%)	99.2 (100)	92.6 (70.4)
Redundancy	7.0 (7.3)	5.9 (1.4)
Refinement		
Resolution (Å)	42.129-2.3	21.19-1.4
No. reflections	8721	35457
<i>R</i> _{work} / <i>R</i> _{free}	0.2095/0.2973	0.193/0.231
No. atoms		
Protein	1463	1467
Ligand/ion	3	
Water	13	110
B-factors		
Protein	40	19.9
Ligand/ion	44	
R.m.s. deviations		
Bond lengths (Å)	0.007	0.005
Bond angles(°)	0.867	0.736

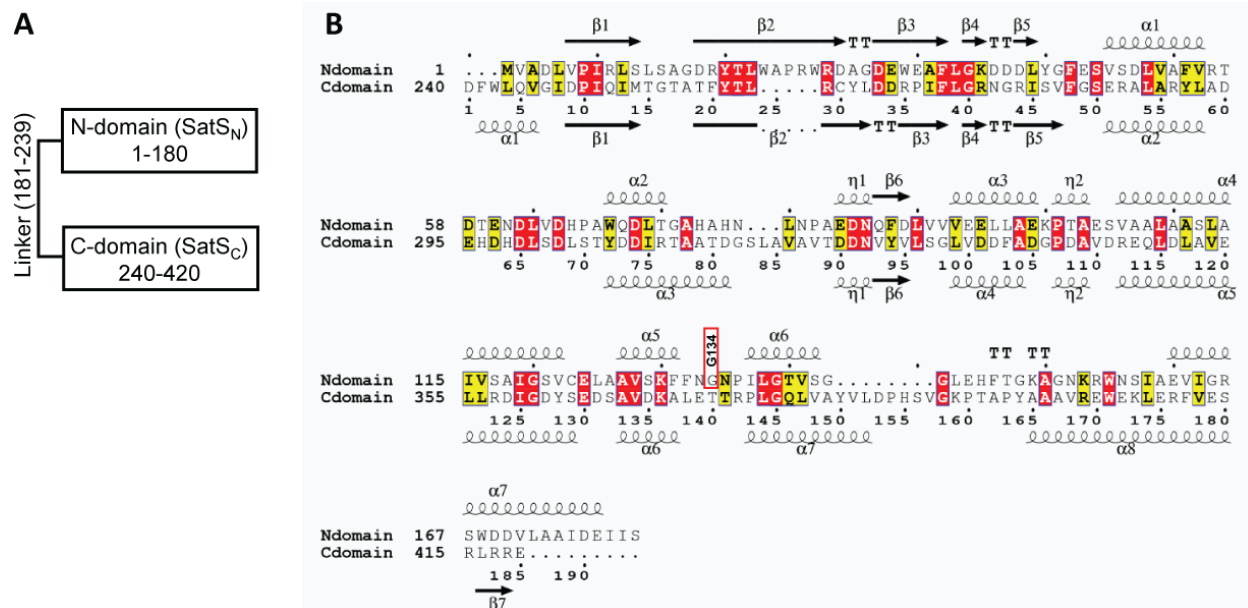


Figure 3.13. SatS_C secondary structure and the predicted secondary structure of SatS_N (A) SatS is composed of two domains of similar secondary structure joined by a potential linker of 60 amino acids with little predicted secondary structure. The disordered linker region was predicted using PrDOS (56). (B) Multiple sequence alignment of N-domain (SatS_N) with C-domain (SatS_C) without the linker region. Secondary structure elements for each domain are depicted above and below their corresponding sequence. SatS_N has 22% sequence identity (highlighted in red) and 41% sequence similarity to SatS_C (highlighted in yellow and red); TT - tight turn.

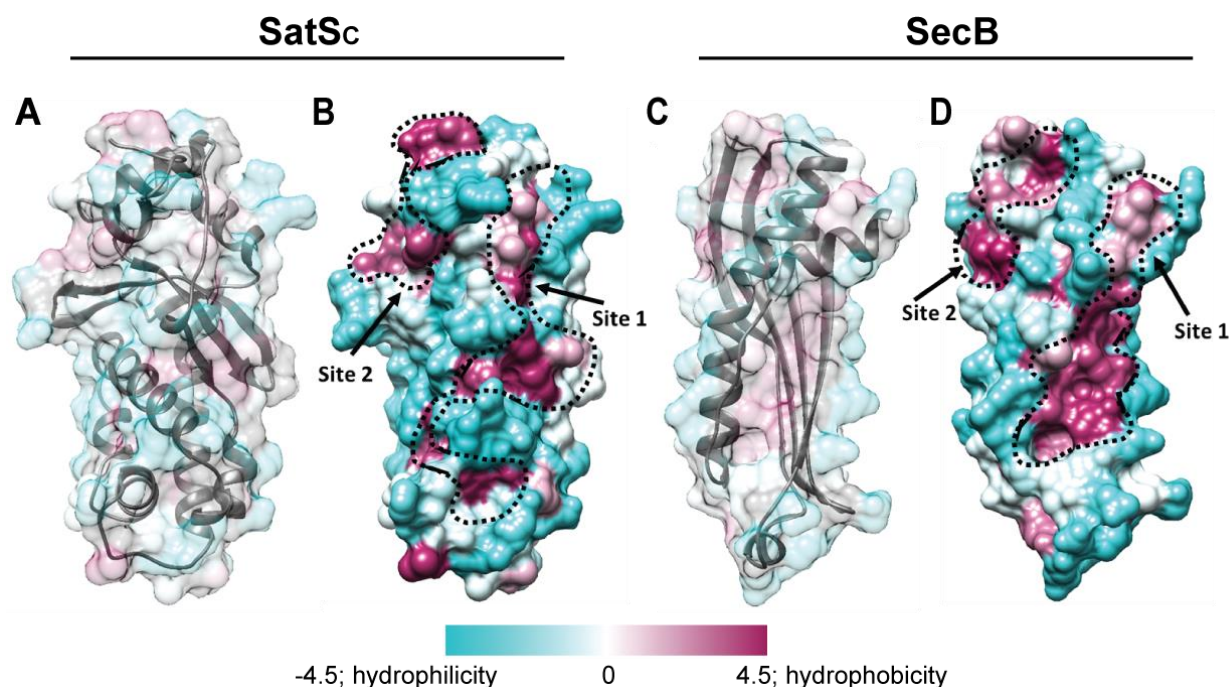


Figure 3.14. SatS has a new fold and hydrophobic grooves that share similarity with the preprotein binding sites of the SecB chaperone. (A) The overall secondary structure of SatS_C. (B) The hydrophilicity of SatS_C is a colored gradient from cyan (hydrophilic) to maroon (hydrophobic). SatS_C exposes ~2,900 Å² of hydrophobic surface. The predicted primary and secondary polypeptide binding site(s) are delineated. (C) The overall secondary structure of SecB monomer (PDB ID:1QYN) (1). (D) The hydrophilicity of SecB is a colored gradient from cyan (hydrophilic) to maroon (hydrophobic). The primary and secondary client binding site(s) are delineated. Each SecB monomer exposes ~1,900 Å² of hydrophobic surface for client protein interactions (45). Molecular graphics and analyses were performed with the UCSF Chimera package (55).

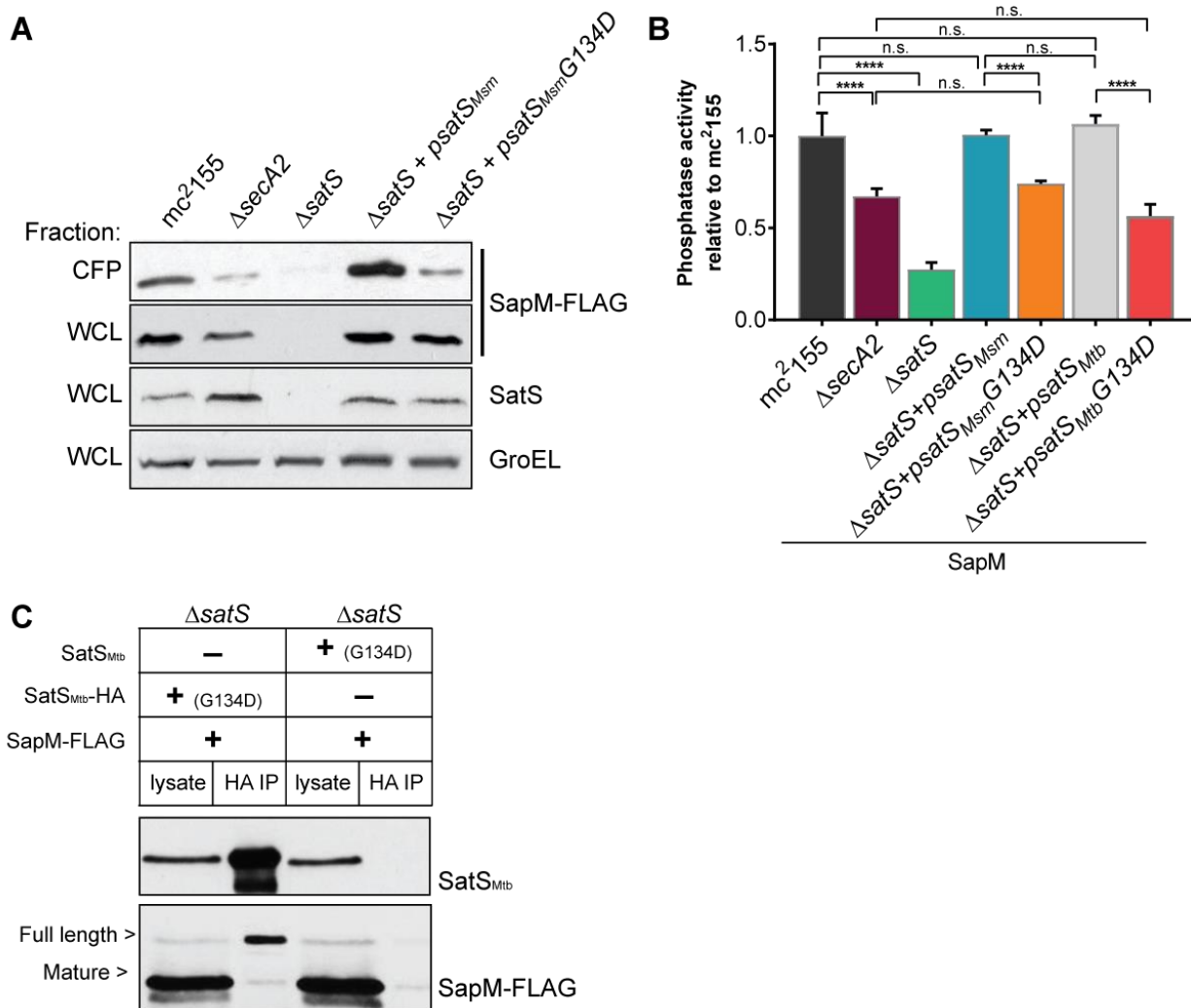


Figure 3.15. SatS has at least two separable roles in protein export (A) Equal protein from culture supernatants (CFP) and whole cell lysates (WCL) from *M. smegmatis* mc²155, ΔsecA2, ΔsatS and the ΔsatS mutant expressing either wild-type SatS (ΔsatS+psatS_{Msm}) or SatS G134D (ΔsatS+psatS_{Msm}G134D) were examined for levels of SapM-FLAG, SatS, and GroEL by Immunoblot. (B) Whole cell phosphatase activity assay of *M. smegmatis* strains expressing SapM. Triplicate wells were grown in a 96 well plate for 24 hours at 37°C before measuring phosphatase activity by quantifying cleavage of pNPP. Rates were normalized as described above. (C) Lysate from *M. smegmatis* ΔsatS expressing SapM-FLAG and either SatS_{Mtb}G134D-HA or SatS_{Mtb}G134D without a tag were used for co-immunoprecipitation using anti-HA conjugated beads. Lysates (left) and co-immunoprecipitations (right) for each strain were probed with SatS antibody and FLAG antibody. All data is representative of at least three independent experiments and all error bars represent standard deviation of the mean of three independent replicates for each strain. n.s. – no significant difference; ****, *p* < 0.0001 by ANOVA and Tukey's post hoc test.

REFERENCES

1. Global tuberculosis report 2017. Geneva: World Health Organization, 2017 Contract No.: CC BY-NCSA 3.0 IGO.
2. Awuh JA, Flo TH. 2017. Molecular basis of mycobacterial survival in macrophages. *Cellular and Molecular Life Sciences* 74(9):1625-48.
3. Sullivan JT, Young EF, McCann JR, Braunstein M. 2012. The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infection and Immunity* 80(3):996-1006.
4. Brundage L, Hendrick JP, Schiebel E, Driessen AJM, Wickner W. 1990. The purified *E. coli* integral membrane protein SecYE is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 62(4):649-57.
5. Tsirigotaki A, De Geyter J, Šoštaric' N, Economou A, Karamanou S. 2016. Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology* 15:21.
6. Bensing BA, Seepersaud R, Yen YT, Sullam PM. 2014. Selective transport by SecA2: an expanding family of customized motor proteins. *Biochimica et biophysica acta* 1843(8):1674-86.
7. Bensing BA, Yen YT, Seepersaud R, Sullam PM. 2012. A specific interaction between SecA2 and a region of the preprotein adjacent to the signal peptide occurs during transport via the accessory Sec system. *Journal of Biological Chemistry* 287(29):24438-47.
8. Ligon LS, Rigel NW, Romanchuk A, Jones CD, Braunstein M. 2013. Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of Mycobacteria. *J Bacteriol* 195(19):4456-65.
9. Prabudiansyah I, Kusters I, Driessen AJM. 2015. *In vitro* interaction of the housekeeping SecA1 with the accessory SecA2 protein of *Mycobacterium tuberculosis*. *PLoS ONE* 10(6):e0128788.
10. Braunstein M, Brown AM, Kurtz S, Jacobs WR. 2001. Two nonredundant SecA homologues function in Mycobacteria. *J Bacteriol* 183(24):6979-90.
11. Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S, Braunstein M. 2009. The accessory SecA2 system of Mycobacteria requires ATP binding and the canonical SecA1. *The Journal of Biological Chemistry* 284(15):9927-36.
12. Zulauf KE, Sullivan JT, Braunstein M. 2018. The SecA2 pathway of *Mycobacterium tuberculosis* exports effectors that work in concert to arrest phagosome and autophagosome maturation. *PLoS Pathogens* 14(4):e1007011.
13. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research* 44(W1):W344-W50.
14. Vergne I, Chua J, Deretic V. 2003. *Mycobacterium tuberculosis* phagosome maturation arrest: selective targeting of PI3P-dependent membrane trafficking. *Traffic* 4(9):600-6.

15. Perkowski EF, Miller BK, McCann JR, Sullivan JT, Malik S, Allen IC, Godfrey V, Hayden JD, Braunstein M. 2016. An orphaned Mce-associated membrane protein of *Mycobacterium tuberculosis* is a virulence factor that stabilizes Mce transporters. *Mol Microbiol* 100(1):90-107.
16. Feltcher ME, Gibbons HS, Ligon LS, Braunstein M. 2013. Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *J Bacteriol* 195(4):672-81.
17. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, Chen X, Braunstein M. 2015. Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & Cellular Proteomics* 14(6):1501-16.
18. Saleh MT, Belisle JT. 2000. Secretion of an acid phosphatase (SapM) by *Mycobacterium tuberculosis* that is similar to eukaryotic acid phosphatases. *J Bacteriol* 182(23):6850-3.
19. Eschenfeldt WH, Maltseva N, Stols L, Donnelly MI, Gu M, Nocek B, Tan K, Kim Y, Joachimiak A. 2010. Cleavable C-terminal His-tag vectors for structure determination. *Journal of structural and functional genomics* 11(1):31-9.
20. Palmer I, Wingfield PT. 2004. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. *Current protocols in protein science* Chapter: Unit-6.3.
21. Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, Glaser P, Cole ST, Brosch R. 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* 150(2):483-96.
22. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Research* 45(D1):D535-D42.
23. van der Woude AD, Stoop EJM, Stiess M, Wang S, Ummels R, van Stempvoort G, Piersma SR, Cascioferro A, Jiménez CR, Houben ENG, Luirink J, Pieters J, van der Sar AM, Bitter W. 2014. Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. *Cellular Microbiology* 16(2):280-95.
24. Yen YT, Cameron TA, Bensing BA, Seepersaud R, Zambryski PC, Sullam PM. 2013. Differential localization of the streptococcal accessory Sec components and implications for substrate export. *J Bacteriol* 195(4):682-95.
25. Puri RV, Reddy PV, Tyagi AK. 2013. Secreted acid phosphatase (SapM) of *Mycobacterium tuberculosis* is indispensable for arresting phagosomal maturation and growth of the pathogen in guinea pig tissues. *PLoS ONE* 8(7):e70514.
26. Perkowski EF, Zulauf KE, Weerakoon D, Hayden JD, Ioerger TR, Oreper D, Gomez SM, Sacchettini JC, Braunstein M. 2017. The EXIT Strategy: an approach for identifying bacterial proteins exported during host infection. *mBio* 8(2).

27. Kurtz S, McKinnon KP, Runge MS, Ting JPY, Braunstein M. 2006. The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. *Infection and Immunity* 74(12):6855-64.
28. Thomas NA, Ma I, Prasad ME, Rafuse C. 2012. Expanded roles for multicargo and class 1B effector chaperones in type III secretion. *J Bacteriol* 194(15):3767-73.
29. Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, Wilmanns M, Bitter W, Parret AHA, Houben ENG, Korotkov KV. 2014. Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG(5) in complex with PE25–PPE41 dimer. *Mol Microbiol* 94(2):367-82.
30. Wolschendorf F, Mahfoud M, Niederweis M. 2007. Porins are required for uptake of phosphates by *Mycobacterium smegmatis*. *J Bacteriol* 189(6):2435-42.
31. Ellis RJ. 1997. Molecular chaperones: Avoiding the crowd. *Current Biology* 7(9):R531-R3.
32. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biology Direct* 7:12-.
33. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. 2014. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Research* 42(Database issue):D297-D303.
34. Teutschbein J, Schumann G, Möllmann U, Grabley S, Cole ST, Munder T. 2009. A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. *Microbiological Research* 164(3):253-9.
35. Phan TH, Ummels R, Bitter W, Houben ENG. 2017. Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. *Scientific Reports* 7:42704.
36. Xu Z, Knafels JD, Yoshino K. 2000. Crystal structure of the bacterial protein export chaperone SecB. *Nature Structural Biology* 7:1172.
37. Bieker-Brady K, Silhavy TJ. 1992. Suppressor analysis suggests a multistep, cyclic mechanism for protein secretion in *Escherichia coli*. *The EMBO Journal* 11(9):3165-74.
38. Flower AM, Osborne RS, Silhavy TJ. 1995. The allele-specific synthetic lethality of *prlA-prlG* double mutants predicts interactive domains of SecY and SecE. *The EMBO Journal* 14(5):884-93.
39. Daleke MH, van der Woude AD, Parret AHA, Ummels R, de Groot AM, Watson D, Piersma SR, Jiménez CR, Luirink J, Bitter W, Houben ENG. 2012. Specific chaperones for the type VII protein secretion pathway. *Journal of Biological Chemistry* 287(38):31939-47.
40. Ekiert DC, Cox JS. 2014. Structure of a PE–PPE–EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. *PNAS* 111(41):14758-63.
41. Randall L, Hardy SJs. 2002. SecB, one small chaperone in the complex milieu of the cell. *Cellular and Molecular Life Sciences* 59(10): 1617-23.
42. Collier DN, Strobel SM, Bassford PJ. 1990. SecB-independent export of *Escherichia coli* ribose-binding protein (RBP): some comparisons with export of maltose-binding protein (MBP) and studies with RBP-MBP hybrid proteins. *J Bacteriol* 172(12):6875-84.

43. Thomas NA, Deng W, Puente JL, Frey EA, Yip CK, Strynadka NCJ, Finlay BB. 2005. CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol* 57(6):1762-79.
44. Takamatsu D, Bensing BA, Sullam PM. 2004. Genes in the accessory *sec* locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. *Mol Microbiol* 52(1):189-203.
45. Huang C, Rossi P, Saio T, Kalodimos CG. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature* 537:202.
46. Stebbins CE, Galán JE. 2001. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* 414:77.
47. Bordes P, Cirinesi A-M, Ummels R, Sala A, Sakr S, Bitter W, Genevoux P. 2011. SecB-like chaperone controls a toxin–antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *PNAS* 108(20):8438-43.
48. Rengarajan J, Bloom BR, Rubin EJ. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *PNAS* 102(23):8327-32.
49. Zhang YJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, Trauner A, Wallis D, Galaviz S, Huttenhower C, Sacchettini JC, Behar SM, Rubin EJ. 2013. Tryptophan biosynthesis protects mycobacteria from CD4 T cell-mediated killing. *Cell* 155(6):1296-308.
50. Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, Bennett LT, Bansal GP, Young JF, Lee MH, Hatfull GF, Snapper SB, Barletta RG, Jacobs Jr WR, Bloom BR. 1991. New use of BCG for recombinant vaccines. *Nature* 351:456.
51. Glickman MS, Cox JS, Jacobs WR. 2000. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Molecular Cell* 5(4):717-27.
52. Gibbons HS, Wolschendorf F, Abshire M, Niederweis M, Braunstein M. 2007. Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* 189(14):5090-100.
53. Barletta RG, Kim DD, Snapper SB, Bloom BR, Jacobs WR. 1992. Identification of expression signals of the mycobacteriophages Bxb1, L1 and TM4 using the Escherichia-Mycobacterium shuttle plasmids pYUB75 and pYUB76 designed to create translational fusions to the lacZ gene. *Microbiology* 138(1):23-30.
54. Snapper SB, Melton RE, Mustafa S, Kieser T, Jr WRJ. 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Molecular Microbiology* 4(11):1911-9.
55. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8:785-6.
56. Ishida T, Kinoshita K. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research* 35(2):W460-W4.

CHAPTER 4

Discussion

Mycobacterium tuberculosis, the causative agent of the disease tuberculosis, is responsible for approximately 1.8 million deaths each year, the equivalent of a death every 20 seconds (1). Moreover, a staggering one third of the world's population is estimated to be infected with *M. tuberculosis*. While antibiotic treatment for tuberculosis exists, the drug regimen is lengthy, which is a challenge for full patient compliance. Furthermore, 600,000 people infected with *M. tuberculosis* will not respond to standard drug treatment due to the emergence of multi-drug resistant (MDR) and extensively drug resistant (XDR) tuberculosis. Drug resistant *M. tuberculosis* requires more expensive treatment that takes longer to complete, has more severe side effects, and is unsuccessful in 50% of cases (1). Combined with the lack of an efficacious vaccine, the challenges of effectively treating *M. tuberculosis* have created a global health emergency that necessitates a better understanding of the mechanisms of *M. tuberculosis* pathogenesis.

In order to promote disease, bacteria export proteins outside of the bacterial cell into the host environment where the proteins can interfere with host defense mechanisms that would otherwise stop the infection. Specialized protein export pathways are often used by bacterial pathogens to export virulence factors into the host. *M. tuberculosis* possesses several specialized protein export pathways, including the SecA2 pathway. The SecA2 protein export pathway is

defined by the presence of a paralog of the essential SecA1 ATPase, which is called SecA2. Unlike the SecA1 ATPase, which is responsible for the bulk of housekeeping export and is essential for bacterial viability, SecA2 is a non-essential specialized SecA ATPase that is required for the virulence of *M. tuberculosis* in both murine and macrophage models of infection (2, 3). The SecA2 pathway is required for the export of a relatively small subset of proteins, some of which contribute to *M. tuberculosis* pathogenesis (4, 5).

SecA1 and SecA2 have unique functions; the two proteins are not functionally redundant. Overexpression of SecA1 does not compensate for the absence of SecA2 and vice versa (6). The differences between SecA1 and SecA2 are likely due to different activities and/or different proteins with which they interact even though current data support a model where SecA2 works with the same SecY channel protein as SecA1 to export its substrates. The goal of my dissertation research was to learn more about the mechanism of SecA2-dependent protein export. This dissertation describes the characterization of spontaneous intragenic and extragenic suppressors of the exacerbated rich agar growth defect of the *M. smegmatis secA2 K129R* strain. As discussed in Chapter 2, the intragenic suppressors mapped to different subdomains of SecA2. These intragenic suppressors may disrupt SecA2 interactions with SecY, its substrates, or other proteins. As discussed in Chapter 3, we identified extragenic suppressors with loss-of-function mutations in *satS* that suppressed *secA2 K129R* phenotypes. We continued our investigation of SatS, a previously uncharacterized protein of unknown function, and identified it as a new protein export chaperone with a role in intracellular growth of *M. tuberculosis* in murine macrophages. As the amino acid sequence of SatS bears no similarity to chaperones and the structure of the SatS_C domain reveals a new fold, SatS appears to represent a new type of protein export chaperone.

SecA2 K129R: tool for studying the SecA2 pathway

Suppressor analysis is a classical genetic method that can be used to identify unknown genes involved in various bacterial pathways, and has been used extensively to study the Sec system in *E. coli* (7). Here, we described the analysis of both intragenic and extragenic suppressors of the *M. smegmatis* *secA2* K129R allele. Our original suppressor collection included sixty-two independent suppressor strains capable of improved growth on Mueller-Hinton agar relative to the parent strain *secA2* K129R (8). Of these sixty-two suppressors, forty suppressors lost expression of the *secA2* K129R allele and were not studied further, since the loss of SecA2 K129R expression explains the suppression in these strains. These suppressors likely had an additional mutation in *secA2* K129R or its promoter that alleviated *secA2* K129R phenotypes by preventing SecA2 K129R production. In addition to the suppressors of *M. smegmatis* *secA2* K129R, we isolated a separate group of seven suppressors from an *M. smegmatis* strain expressing *M. tuberculosis* *secA2* K115R (an equivalent mutation to *M. smegmatis* *secA2* K129R) (8). These two separate screens resulted in a total of twenty-seven suppressors that expressed full length SecA2 protein (Table 4.1). Ten of these suppressors have intragenic mutations within the *secA2*_{Msm} K129R or *secA2*_{Mtb} K115R allele (Table 4.1, Chapter 2). The remaining seventeen suppressors expressing SecA2 protein do not have mutations in the *secA2* K129R allele and are extragenic suppressors. Among the seventeen extragenic suppressors, two have mutations in the promoter of *secY* [Table 4.1 and (9)] and twelve have mutations either in the *satS* gene or upstream of *satS* in its untranslated region (Table 4.1 and Chapter 3).

One interesting group of suppressors is made up of three suppressors with large overlapping chromosomal deletions (Table 4.1). Suppressors 7S, 20B, and SSW2 have a genomic deletion spanning more than 43 kbp and ≥ 42 genes from *msmeg1684* (*satS*) to

msmeg1726. After identifying suppressors that mapped to *satS*, we realized that the loss of *satS* might explain the suppression associated with these large deletion mutants. Importantly, adding back *satS* to 7S (Δ *msmeg1678-msmeg1726*) was sufficient to restore the SecA2 K129R phenotypes in this strain (Figure 4.1), which is consistent with suppression phenotypes of these large deletion mutants being due to the loss of *satS*. Several genes at the right end of the 7S and 20B deletions are annotated to encode transposase enzymes. These genes could conceivably have played a role in initiating deletion of this large region of the *M. smegmatis* genome.

Currently, there remain four suppressors whose suppressing mutations remain to be identified. These strains do not carry mutations in *secA2 K129R*, the *secY* promoter, or *satS* including its promoter. Three of the unknown suppressors, strains 5S, 13B, and SSW10, express SecA2 K129R at very low levels relative to the parent *secA2 K129R* strain and one additional suppressor, strain 7B, does not express SatS despite having no mutations in *satS* or its promoter. The reduced level of SecA2 K129R or SatS in these strains is likely responsible for the suppression, but we do not know what mutations are causing the reduction. In contrast to these other four remaining suppressors, 39S is a strong suppressor that produces wild type levels of SecA2 K129R, SecY and SatS (data not shown). Furthermore, 39S has no mutations in *secA2 K129R*, the *secY* promoter, or *satS*. Given how useful *secA2 K129R* suppressors have been in identifying members of the SecA2 export pathway thus far, whole genome sequencing of 39S in the future is warranted as it may provide additional new information about the SecA2 pathway.

The SecA2 crystal structure paired with intragenic suppressors provide insight into important domains of SecA2

Structural comparison of SecA2 and SecA1

M. tuberculosis SecA1 and SecA2 are 50% similar to each other by amino acid sequence. Sequence alignments prior to our solved SecA2 structure revealed that the similarity is spread throughout much of the two proteins, even though SecA1 is 20 kDa larger than SecA2. Using these sequence alignments, we originally predicted that the smaller size of SecA2 came from the loss of the variable domain (VAR) and helical wing domain (HWD), both of which are present in SecA1 and canonical SecA proteins (8). Genomic databases annotate *M. tuberculosis* SecA2 as having a total length of 808 amino acids (AA) (e.g., NCBI accession no. NP_216337). However, the start site in this annotation is likely incorrect, as the first 30 AA are not required for function and represent an N-terminal extension that is not observed in other SecA orthologs (10, 11). Therefore, we designate the GTG codon corresponding to residue 31 in the NCBI annotation as the true start codon, yielding a total SecA2 ORF length of 778 AA.

The structure of SecA2 we report in Chapter 2 is the first, and so far only, structure of any SecA2 protein. The structure first confirmed that *M. tuberculosis* SecA2 contains key functional domains conserved in all SecA proteins. These domains are made up of the following amino acids in SecA2: nucleotide binding domain 1 (NBD1) 1-229; preprotein cross-linking domain (PPXD) 240-360; NBD2 369-562; intramolecular regulator of ATPase (IRA1) [includes the helical scaffold domain (HSD) and two helix finger (2HF)] 590-733 [HSD 591-634; 2HF 695-703]; and C-terminal linker (CTL) 734-778 (Thomas Ioerger, Texas A&M University, personal communication). In addition to its structural domains, SecA2 has two important surface features that are conserved in SecA proteins. The first is a “substrate clamp” that is formed from

PPXD, HSD, and NBD2 and is where SecA interacts with the mature domain of its substrates (12). Upon interaction with a substrate, the flexible PPXD domain rotates and wraps around the substrate, resulting in a “closed clamp” conformation (12). The second important surface feature of SecA proteins is a “signal peptide binding cleft” that is formed from PPXD, IRA1, and HWD which is the docking site for the signal peptide of its substrates (13).

The crystal structure also confirmed differences predicted by sequence alignment; the smaller size of SecA2 versus SecA1 and canonical SecA proteins comes from the absence of the HWD and the VAR domain. The lack of a HWD is the most striking structural difference in SecA2. Without the HWD, the signal peptide recognition cleft of SecA2 is larger and more solvent exposed in SecA2 (Figure 2.4). This could make the cleft more accessible to substrates and affect the substrates that SecA2 recognizes. However, the function of the HWD is not well understood in any SecA proteins. Thus, alternate or additional consequences of its absence in SecA2, beyond exposing a larger signal peptide cleft, may exist. Absence or truncation of the HWD is a general feature among SecA2 proteins in actinomycetes (including *Mycobacterium* and *Corynebacterium* species) (Figure 4.2). Interestingly, the SecA2 protein of the SecA2/SecY2 system in *Streptococcus gordonii* and the SecA2 protein of the SecA2-only system in *L. monocytogenes* also appear to have truncated versions of the HWD (deletions of 13 or 18 residues for *S. gordonii* or *L. monocytogenes*, respectively) based on sequence alignments. However, until structures of these other SecA2 orthologues are solved, the potential consequences of these HWD truncations remain unknown. To our knowledge, no mutational studies have been performed on the HWD of any SecA protein, and our understanding of its function comes primarily from crystal structures and cross linking experiments (12, 13). To

achieve a complete picture of SecA2 function, the consequences of a truncated or deleted HWD will need to be explored in SecA2, SecA1 and canonical SecA proteins.

The significance of the absence of the VAR domain in the SecA2 NBD region is also unclear. The lack of the VAR domain leaves the nucleotide-binding site relatively solvent exposed. While other SecA2 orthologues also lack the VAR domain, one-third of bacterial SecA proteins lack this domain as well (14). In *E. coli* SecA, the VAR domain regulates ATPase activity and ADP release, as *secA* Δ *var* mutants display higher ATPase activity and faster ADP release rates (14). However, *M. tuberculosis* SecA2 is reported to release ADP more slowly (not more quickly) than the VAR-containing *M. tuberculosis* SecA1 (15). Constructing a mycobacteria *secA1* Δ *var* mutant may reveal insights into the function of the VAR domain and the consequences of its absence in SecA2.

Suppressors with mutations in SecA2 map to specific regions

Intragenic suppressors of *secA2* K129R map to the surface of SecA2. The suppressors can be organized into three groups based on the location of the amino acid alterations: suppressors with mutations in the IRA1 domain, suppressors with mutations in PPXD or NBD2 domains, and suppressors with mutations in the NBD1 domain. We subcloned and retested a representative intragenic suppressor from each of these groups to confirm the intragenic mutation was the source of the suppression and to investigate how these suppressors' mutations affect the subcellular localization of SecA2 K129R. While wildtype SecA2 is primarily cytoplasmic, SecA2 K129R is retained at the membrane (6), presumably because it is locked in a complex with SecY. SecA2 K129R jamming the SecY channel is thought to be the source of its exacerbated phenotypes on sodium azide and Mueller-Hinton plates. The group of suppressors that mapped to IRA1 fully restored cytoplasmic localization of SecA2 K129R (Figure 2.9).

Because similar mutations in the IRA1 domain of *E. coli* SecA disrupt SecA binding to SecYEG (16), these results are consistent with IRA1 suppressor mutations preventing SecA2 K129R interactions at the membrane SecYEG channel, which supports the model for SecA2 working with SecY to promote export of its specific substrates.

The second group of suppressors mapped to the “substrate clamp” of SecA2. We identified suppressors with mutations on both sides of the clamp, one on the NBD2 domain and two on the PPXD domain. The clamp suppressors restored cytoplasmic localization of SecA2 K129R to levels comparable to wildtype SecA2, suggesting that these mutations prevent SecA2 K129R from jamming the channel (Figure 2.9). For this group of suppressors, the suppression may result from an inability of SecA2 to either interact with its substrate or to trap the substrate in the SecYEG channel. This model may also suggest that SecA2 K129R can only jam the channel if it is first bound to a substrate.

The last group of suppressors mapped to a surface loop near the top of the NBD1 domain, close to the IRA1 domain, in a region of NBD1 that has not been assigned a function. This group of suppressors either resulted in a deletion or duplication of the same loop (residues 168 to 171). Interestingly, the NBD1 suppressors only partially restore cytoplasmic localization of SecA2 K129R (Figure 2.9), despite their full suppression of the *secA2 K129R* phenotype on sodium azide plates. The inability of these suppressors to fully restore SecA2 K129R localization to the cytoplasm suggests that even with these secondary mutations, some SecA2 K129R is still able to jam in the SecYEG channel (although the occurrence is rare enough to alleviate sensitivity to sodium azide). It is noteworthy that when SecA2 K129R is suppressed by $\Delta satS$ null mutations, we also observe a partial (not complete) restoration of SecA2 K129R localization to the cytoplasm (Figure 3.1D). One possibility for the partial restoration of SecA2 K129R localization

is that some SecA2 K129R can still bind to substrates and engage the channel in a $\Delta satS$ null strain and that full suppression of SecA2 K129R is not required to alleviate sensitivity to sodium azide (Figure 3.1A). There are examples of SecA2 substrates that do not require SatS for export. In our model, when presented with a SatS-independent substrate such as Msmeg1704, SecA2 K129R can still initiate export and jam SecYEG, even in the absence of SatS. The similarities between the NBD1 domain suppressors and the *satS* null suppressors, that they both restore resistance to sodium azide to that of a $\Delta secA2$ strain while only partially restoring SecA2 K129R localization to the cytoplasm, raise the possibility that the NBD1 suppressors may map to a region of SecA2 that is important for interacting with SatS. However, in order to test this hypothesis, we will first need to establish that SatS and SecA2 interact.

In addition to the intragenic suppressors described in Chapter 2, we identified two additional intragenic suppressors that map to *secA2* in our suppressor screen of the *M. tuberculosis* SecA2 K115R allele. SSW11 has a 7 amino acid deletion from amino acid 651 to 658, and SSW8 has a single amino acid substitution from Val 528 to Gly (Table 4.1 and Figure 4.3). The SSW11 deletion maps to the surface of the IRA1 domain (Figure 4.3). Intriguingly, the residues deleted in SSW11 are adjacent to the region of the IRA1 domain that typically forms the HWD of SecA and are not conserved in SecA1 (Figure 4.2). SSW11 is interesting as it suggests that despite the lack of a HWD, SecA2 requires residues in this region in order to function properly and it identifies a fourth location on the SecA2 structure as affecting SecA2KR phenotypes. To begin characterizing the cause of suppression in SSW11, it will be important to determine the subcellular localization of SecA2 K115R in this suppressor to classify it either as a suppressor that completely restores cytoplasmic localization (like the PPXD, NBD1, and other class of IRA1 suppressors) or only partially restores cytoplasmic localization (like the NBD2

suppressors). If the SSW11 suppressor also partially restores SecA2 K115R localization, residues 651 to 658 will be another good candidate for a SatS-interacting site on SecA2, particularly because this site is unique to SecA2; in SecA1 the same region is occupied by the HWD. The SSW8 substitution is in the NBD2 domain with substitution from Val528 to Gly. Val528 is near the residue mutated in NBD2 suppressor 34S; however, Val528 is not surface exposed (Figure 4.3). Of note, of all the intragenic suppressors identified thus far SSW8 is the only one affecting residues that are not localized to the surface of SecA2.

In general, we predict that our intragenic suppressors map to sites on SecA2 that are important for its interactions with other proteins, such as SecY and its substrates. We also predict that SecA2 interacts with SatS, potentially as SatS hands off its substrates to SecA2. In this case, some of our suppressors could be in sites on SecA2 that are required for interacting with SatS. Thus far, we have not been able to establish a direct interaction between SecA2 and SatS by coimmunoprecipitation. In the absence of a physical interaction, it is difficult to test whether any of the suppressors affect SecA2:SatS interactions. An alternative approach would be to ask whether these intragenic mutations, in the context of the wild type K129, prevent SecA2 from exporting both SatS-dependent and SatS-independent substrates. We can use site-directed mutagenesis to restore the Walker Box of these intragenic suppressors, resulting in *secA2* mutant alleles with the ATP binding site restored. Three of these suppressor alleles have already been made (8). By expressing these mutant *secA2* alleles in a $\Delta secA2$ strain, we can investigate their ability to secrete SatS-dependent and independent substrates. We would expect that mutations in the presumed SecY binding site on the IRA1 domain would reduce export of all SecA2-dependent substrates. Alternatively, if the mutations disrupt SecA2 interactions with SatS, we would expect that only export of SecA2-dependent and SatS-dependent substrates, such as

SapM, would be compromised. In this case, substrates such as Msmeg1704 and Msmeg1712 would have wildtype levels of export, despite mutations in *secA2*. Evaluating suppressors for their effects on different substrates could provide mechanistic insights on the relationship between SecA2 and SatS without requiring us to first establish protein-protein interactions. Identifying regions of SecA2 required for working with SatS may guide efforts to demonstrate a direct interaction between SatS and SecA2.

Extragenic suppressors lead to the identification of a new player of the SecA2 export pathway

In Chapter 3, we identified ten suppressors of the *M. smegmatis* *secA2* K129R allele with unique mutations in the same gene, *msmeg_1684* which we renamed *satS*. We subsequently validated that loss of SatS function suppresses *secA2* K129R phenotypes. Because previously characterized suppressors restored *secY* expression and/or reduced SecA2 K129R levels (9), we first ruled out the possibility of these suppressors in *satS* affecting transcription or translation of *secY* or reducing SecA2 K129R levels. Loss-of-function *satS* suppressors restore SecY levels but do not affect *secY* transcription (9), and SecA2 K129R levels are equivalent in the presence or absence of SatS (data not shown). However, deletion of *satS* significantly reversed SecA2 K129R retention at the membrane and the associated SecY degradation. As described above, when SecA2 K129R is suppressed by $\Delta satS$ null mutations, we also observe a partial (not complete) restoration of SecA2 K129R localization to the cytoplasm (Figure 3.1D). These data suggest that in the absence of SatS, SecA2 K129R cannot engage in detrimental interactions with SecYEG. By extension, these results support a role for SatS in enabling wild-type SecA2 to interact with the SecYEG channel. We identified SecA2-dependent substrates that require SatS

for export, namely SapM and Mce proteins. We also identified SecA2-dependent proteins that do not require SatS for export, such as the solute binding protein Msmeg1704.

One possibility for how SatS promotes SecA2 interactions with the SecYEG channel is that in order for SecA2 to be delivered to or engage the SecYEG channel it must first be bound to a substrate in a translocation competent state and that SatS functions as a protein export chaperone that facilitates a SecA2-substrate interaction for some of the substrates. We favor this role for SatS as it would not only explain why phenotypes of *secA2 K129R* depend on the presence of SatS but it is also consistent with our identification of an interaction between SatS and SapM and the chaperone activities of SatS. An alternate possibility is that SatS is a core component of a SecA2-specific export apparatus with a function mediating the interaction between SecA2 and SecYEG. However, if SatS were to function this way, we would expect all SecA2-dependent substrates would require SatS for export, which was not the case. The existence of SatS-independent substrates raises the question of how loss-of-function mutations in *satS* are able to suppress SecA2 K129R, since some substrates can still engage SecA2 K129R in the absence of SatS. We reason that the loss of SatS results in enough SecA2 K129R being diverted from the SecYEG channel that the detrimental phenotypes are alleviated. The incomplete restoring of SecA2 K129R to the cytoplasm in a *satS* null suppressor is reminiscent of the intragenic suppressors in the SecA2 NBD1 domain described earlier.

SatS has many similarities with protein export chaperones

Our hypothesis that SatS functions as a protein export chaperone stems from our initial observation that in the absence of SatS, cytoplasmic levels of its substrate SapM are dramatically reduced. After ruling out the possibility of SatS affecting transcription or translation of *sapM*, we

hypothesized that SatS acts as a protein export chaperone with functions that include post-translational stabilization of SapM prior to its export from the cytoplasm.

In Chapter 1, we reviewed common features of well-studied protein export chaperones (Table 1.1). Nearly all of these features are also shared by SatS (Table 4.2). SatS is a highly acidic (pI 3.83), cytoplasmic protein that promotes the export of a subset, but not all, of the proteins exported by the SecA2 pathway. Further, the *satS* and *sapM* genes are co-transcribed in an operon and SatS interacts with SapM in mycobacteria. SatS preferentially interacts with the full-length preprotein of SapM indicating that the interaction occurs in the cytoplasm prior to SapM export. Further, like other protein export chaperones where binding occurs in regions of the mature domain of the substrate (17-19), the signal sequence of SapM is not required for SatS and SapM to interact. The SatS interaction with SapM is much easier to detect in a $\Delta secA2$ mutant versus a *secA2*^{WT} background. We reason that interactions between SatS and SapM are transient in the presence of a functional SecA2 system and that in the absence of SapM export (i.e. in a $\Delta secA2$ mutant) there is more SatS:SapM complex in the cytoplasm to detect. Of note, in studying the interaction between the ESX chaperone EspG₅ and its PE/PPE substrate pair in *M. marinum* a similar strategy of using an export defective mutant background was utilized (20).

The *in vitro* anti-aggregation effect of SatS on SapM preprotein provided additional proof of a SatS:SapM interaction and the most direct evidence of a chaperone function for SatS. Anti-aggregation activity is also reported for the SecB, EspG₅, and the Yersinia T3SC SycO protein export chaperones (21-23). Interestingly, SecB can prevent aggregation of non-exported, cytoplasmic proteins such as luciferase (21). In the future, it will be interesting to perform *in vitro* aggregation assays with SatS using non-exported proteins, such as luciferase or citrate synthase, to determine the specificity of the anti-aggregation effect of SatS.

Along with the above chaperone features, in the absence of SatS, the level of SapM in the cytoplasm was dramatically reduced. The effect of SatS on intracellular SapM levels is post-translational and reminiscent of effects of T3SCs and EspG chaperones protecting their cognate substrates from degradation prior to export.

SatS alignments and structure reveal two highly similar domains

Although we set out to solve the structure of SatS in its entirety, we were only able to obtain structural information for the C-terminal half of the protein (SatS_C), which arose during crystallization. The structure of SatS_C revealed a new fold sharing no similarities with any previously solved protein structure in the PDB based on DELTA-BLAST and VAST similarity searches (24, 25). Interestingly, inspection of the full length SatS sequence reveals 26% identity and 44% similarity in amino acid sequence between the N-terminal (SatS_N) and C-terminal halves of SatS. There is additionally striking similarity in the predicted secondary structure of SatS_N and SatS_C (Figure 3.14), which raises the interesting possibility of SatS being composed of tandem SatS_C-like domains. Comparison of the experimentally derived SatS_C structure (AA 242-420) with an *in silico* model of SatS_N (AA 1-178) revealed exceptional conservation between the two domains (RMSD of 0.234) (Figure 4.4). The close predicted structural similarity between SatS_N and SatS_C suggests that SatS proteins may have evolved by an intergenic duplication event, with the two duplicated domains connected by a variable linker. Among SatS homologs, the linker varies widely in both amino acid sequence (Figure 4.5A) and length (Figure 4.5B). The linker is predicted by IUPred2A to be disordered (Figure 4.5C) (26). It is worth noting that the mycobacterial protein export chaperone EspG also consists of two subdomains that are twofold symmetrical to one another. However, in the case of EspG the subdomains only

share 13% sequence identity and there does not appear to be a linker between the two subdomains (22).

If SatS is composed of two structurally similar domains, it raises the question of whether the individual SatS domain (SatS_N or SatS_C) have identical functions and/or are sufficient to carry out SatS function. In the *in vitro* anti-aggregation assay, the SatS_C domain was sufficient for preventing SapM aggregation, indicating that SatS_C is capable of directly interacting with SapM and is sufficient to carry out at least this function of SatS on its own. In the future, the *in vitro* anti-aggregation assay should be performed with purified SatS_N to ask whether it is also sufficient to prevent SapM aggregation. Future experiments should also address whether single domains of SatS can complement the $\Delta satS$ mutant phenotypes of mycobacteria (i.e. the reduced stability and secretion of SapM) and whether SatS_N and SatS_C are capable of interacting with SapM as assessed by coimmunoprecipitation. Our preliminary data reveal that expressing SatS_N is sufficient to restore SapM levels in the whole cell lysate (i.e. restore SapM stability), suggesting that SatS_N can also interact with SapM (Figure 4.6A). Understanding the functions of the two domains of SatS will be an important next step in our investigation of SatS as a new type of protein export chaperone.

Differences in electrostatic surface potential of SatS_N and SatS_C

In accordance with their acidic pI, both SatS_N (pI 4.29) and SatS_C (pI 4.40) have significant levels of negatively charged electrostatic surface area (Figure 4.7A and 4.7B). Interestingly, the distribution of negative and positive charges on the SatS_N model and SatS_C structure are not mirror images of one another. While both surfaces are overall dominated by negative charges, SatS_C has a large stretch of positive, basic residues going along the length of the domain (Figure 4.7A). Alternatively, SatS_N is predicted to have a positive, basic tip at the

bottom of the domain (Figure 4.7B). In general, protein export chaperones have an abundance of electronegative charges distributed broadly over their surfaces, but the function of their negatively charged areas is unknown. In one study of the T3SC SigE, substitutions of electronegative surface residues for alanines results in decreased substrate secretion but no decrease in substrate stability, suggesting that these mutations are not defective in substrate binding but rather in other chaperones functions, possibly in targeting (27). The observation that SatS_N and SatS_C have different surface charge distributions may indicate that the two domains perform different roles in protein export.

Hydrophobic grooves in SatS_C resemble SecB substrate binding sites

Investigation of the SatS_C structure revealed a large network of negatively charged amino acids surrounding two surface exposed hydrophobic grooves, which are similar in arrangement, shape and size to the hydrophobic client binding sites of a SecB monomer (Figure 3.13 and Figure 4.7C). In the solution structure of SecB in complex with a preprotein, the unfolded preprotein wraps around the SecB tetramer through interactions with the hydrophobic client binding sites. This binding architecture helps explain the means by which SecB maintains Sec preproteins in an unfolded state, as is required for their transport through the SecYEG channel (17). The similarity in hydrophobic grooves in SatS and SecB is intriguing since SatS works with the SecA2 pathway that, like SecB and SecA of Gram negative bacteria, uses the SecYEG channel to export proteins. Moreover, these similarities suggest that the hydrophobic grooves in SatS are substrate binding sites.

Whether SatS uses these hydrophobic grooves to interact with its substrates remains to be investigated. Structure-based mutagenesis of SatS will be an important next step in identifying regions (i.e. Site 1 and Site 2) and residues that are important for substrate interactions. For

example, amino acid substitutions in the substrate-binding grooves in SecB resulted in a substantial decrease in the affinity for unfolded proteins and a strong reduction in its antifolding activity (28). These decreases were observed after the cumulative substitution of three hydrophobic residues (V40/L42/L44) to alanines in the primary substrate binding groove of SecB. SatS_C has hydrophobic residues in Sites 1 and 2 that are good candidates for participating in substrate binding based on the criteria that they are (i) surface exposed, (ii) clustered with other hydrophobic residues, and (iii) located in a concave groove on the SatS surface (criteria established with Konstantin Korotkov, University of Kentucky, personal communication). Within Site 1, there are two distinguishable grooves which we annotated middle and low (Figure 4.8A). Within these grooves, we selected several residues, highlighted in Figure 4.8A, for future evaluation in substrate binding using SDM. In the middle of Site 1, residues Val323, Val328 and Leu352 form a continuous groove along the side of SatS_C. In the low part of Site 1, residues Val354, Val383 and Leu387 form a concave pocket. Site 2 is considerably smaller than Site 1, but still has several good candidates for SDM including Ile248, Leu267, Ile272, and Leu304 that also form a concave pocket. By substituting alanines for these residues, we can ask whether disruption of the hydrophobic grooves in SatS affects binding to SapM using our coimmunoprecipitation assay.

SatS_N also has two surface exposed grooves containing hydrophobic residues: Site A and Site B (Figure 4.7D). Although the model of SatS_N is structurally very similar to the solved SatS_C structure, the distribution of surface hydrophobic residues in our model SatS_N structure is not a mirror image of Site 1 and Site 2 of SatS_C. The most pronounced difference between the sites on SatS_N and SatS_C is their proximity to one another. Sites 1 and 2 on SatS_C are close together whereas sites A and B on SatS_N are on opposite sides of the domain (Figure 4.7C and

D). Within Site A, Leu5, Val6, Leu20, and Leu72 are good candidate amino acids for SDM to disrupt a deep hydrophobic groove (Figure 4.8B). In Site B, two adjacent areas are good targets for SDM and are annotated as the Site B_i and the Site B_{ii} in Figure 4.8B. Site B_i is made up of residues Leu10, Leu12, Leu90, Ile119 and Ile172. The hydrophobic residues in Site B_{ii} that would make good candidates for SDM are Leu125 and Ile178. Structure-based mutagenesis targeting the hydrophobic sites on SatS, as well as solving a structure for SatS bound to a substrate, such as SapM, will be important steps to determine whether SatS interacts with its substrates in a similar manner to other protein export chaperones. As was the case for SecB, it may be difficult to generate a crystal structure of SatS bound to unfolded SapM. However, the structure of unfolded PhoA bound to SecB was recently solved using nuclear magnetic resonance (NMR) and isotope labelling (17). Prior to the solved structures, the recognition motif of SecB was identified by screening cellulose-bound, short (13-mer) peptides derived from different substrates (21). If we cannot obtain a cocrystal of SatS and SapM, these would be viable alternative approaches to study how SatS and SapM interact.

While evaluating the surface of SatS_N and SatS_C for hydrophobic residues that met our criteria listed above, we observed that each domain had a single, surface exposed cysteine in one of the hydrophobic grooves (Figure 4.8 A and B). We compared the amino acid sequence of 150 SatS homologs throughout actinomycetes using ConSurf (29) and discovered that these two cysteine residues are universally conserved in all 150 SatS proteins examined. While cysteine residues have not specifically been demonstrated to be important for SecB, T3SCs or EspG substrate interactions, these two residues are also appealing candidates for SDM based on their high conservation and proximity to a hydrophobic groove.

Consideration of SatS being a SecB counterpart for the SecA2 pathway

The similar hydrophobic grooves on SatS_C and SecB are intriguing since SatS works with the SecA2 pathway, which uses the SecYEG channel as is also the case for SecB working with SecA in Gram negative bacteria. Thus, even though SecB and SatS are not evolutionarily conserved, it is interesting to speculate a SecB-like function for SatS in the mycobacterial SecA2 pathway. Mycobacteria lack a canonical SecB protein export chaperone. However, in *M. tuberculosis* there is a SecB-like protein Rv1957 that functions as a chaperone for the HigB-HigA stress responsive toxin-antitoxin system by preventing aggregation of the HigA antitoxin (30). Although Rv1957 shares only 13% sequence identity with *E. coli* SecB, Rv1957 is designated a SecB-like protein because overexpression of Rv1957 in *E. coli* can substitute for SecB export function and prevent aggregation of SecB substrates (30). However, it is important to note that there is no evidence that Rv1957, which is not conserved across mycobacteria, functions in mycobacterial protein export. Our attempts at complementing *E. coli* $\Delta secB$ mutant phenotypes by overexpressing SatS have been unsuccessful (data not shown). Considering that Rv1957 is able to suppress $\Delta secB$ mutant phenotypes, we find this negative result interesting in suggesting that SatS may not be a true SecB homolog for the mycobacterial SecA2 pathway. Rather, the novel fold and interesting functions of SatS suggest to us that SatS is the founding member of a new class of protein export chaperones.

SatS forms a multimer

SecB and T3SCs form multimers; SecB forms a tetramer and T3SCs form dimers. In their multimeric state, both of these chaperones interact with one substrate. EspG on the other hand does not oligomerize, and binds a PE-PPE substrate pair. Understanding the functional quaternary structure of SatS will be another important step in its characterization. We began

studying SatS multimerization using a bacterial two hybrid system (BACTH). BACTH is a commonly used approach to monitor protein-protein interactions in *E. coli*. In this system, *Bordetella pertussis* adenylate cyclase CyaA is cleaved into two domains, T25 and T18, which cannot interact on their own. When fused to two interacting proteins, T25 and T18 can be brought together to reconstitute CyaA activity. These domains were separately fused to SatS and expressed in *E. coli* lacking an endogenous CyaA. SatS:SatS interactions resulted in reconstituted CyaA, as assessed by CyaA activation of downstream gene expression including both an operon required for fermenting maltose as well as β -galactosidase (Figure 4.9A and B). Surprisingly, detecting a positive SatS:SatS interaction is limited to specific SatS-CyaA domain fusions. Specifically, we are able to detect a positive SatS-SatS interaction using BACTH when one CyaA domain (either T25 or T18) is fused onto the N-terminus of SatS and the other CyaA domain is fused onto the C-terminus of SatS (Figure 4.9C). If both CyaA domains are fused to the N-terminus or C-terminus of SatS proteins, we do not observe an interaction. Our interpretation of this result is that due to the size of SatS, the orientation of the CyaA domains is important to obtaining a positive interaction. SatS oligomerization may be head-to-tail such that the N terminus of one SatS monomer comes in contact with the C terminus of another SatS monomer. Unfortunately, BACTH cannot distinguish between different multimeric states. We will need to use different approaches, such as size exclusion chromatography, to determine if SatS forms a dimer or a larger homomeric complex.

SatS is the founding member of a new class of protein export chaperones

SatS shares many conceptual and functional characteristics with well-studied protein export chaperones: SecB, Type III secretion chaperones (T3SCs), and EspG (Table 4.2). These similarities initially led us to hypothesize a protein export chaperone function for SatS. However,

there are also key differences between SatS and other protein export chaperones. The structure of SatS_C is a new fold lacking similarities to any solved structure. Additionally, SatS is larger than these other export chaperones; the predicted molecular weight of SatS is 43 kDa. SecB, T3SCs and EspG are smaller (between 15 and 30 kDa in size). Furthermore, when run on an SDS-PAGE gel, SatS migrates at a predicted size of 67 kDa regardless of whether SatS is expressed in *M. tuberculosis*, *M. smegmatis* (Figure 4.6B), or *E. coli*. We have confirmed the identity of the aberrantly migrating band as SatS using polyclonal antibodies raised against SatS, mass spectrometry, and Edman sequencing. The altered migration may be due to the high percentage (20%) of acidic residues in the SatS sequence (resulting in a pI of 3.8), which can alter protein migration on SDS-PAGE (31). However, SecB (pI 4.1) migrates close to its predicted MW of 17 kDa [(32) and data not shown] and the T3SC SycP (pI 4.5) also runs at 16 kDa, its predicted MW (33). The difference in size and migration pattern of SatS on SDS-PAGE as compared to other protein export chaperones may be important in distinguishing how SatS carries out its functions in protein export.

What are the functions of SatS?

We first identified a role for SatS in stabilizing its substrates in the cytoplasm prior to export. Because of the dramatically reduced levels of SapM in the whole cell lysate of the $\Delta satS$ mutant, it was not immediately clear if the role of SatS in SapM secretion was solely to maintain intracellular levels of SapM preprotein or if SatS had additional roles. By evaluating the *satS* G134D mutant, we revealed the existence of at least one additional role for SatS in promoting SapM secretion. In the *satS* G134D mutant, intracellular SapM was maintained at wild-type levels; yet, there remained a SapM secretion defect. Furthermore, SatS G134D is still able to interact with SapM in the cytoplasm of *M. smegmatis*. Based on our modeled SatS_N structure, the

G134 residue is surface exposed and is not located in either of the predicted hydrophobic groove sites (Figure 4.4). Instead, G134 appears to be in the positively charged tip of SatS_N (Figure 4.7B). When overlapping the structure of SatS_C with the modeled structure of SatS_N, T375 occupies nearly the same space as G134 (Figure 4.4). Future experiments should also include site directed mutagenesis of T375 to ask whether mutations in the same area on the C domain of SatS produce similar phenotypes to the G134D allele.

SatS G134D can still bind to and stabilize SapM, but is deficient for SapM export. With other protein export chaperones, there are also examples of mutations that eliminate one chaperone function while retaining others. The E77K amino acid substitution in SecB blocks productive interactions with SecA without significantly disrupting its oligomeric state or SecB:substrate complex formation (34). In other words, SecB E77K can still bind to its substrates and keep them unfolded, but cannot target them to SecA. Likewise, the E142G amino acid substitution in the T3SC chaperone CesT blocks CesT interactions with the T3SS ATPase EscN without disrupting its ability to bind its substrate Tir (35). The CesT E142G mutant still stabilizes Tir, but is deficient for Tir export because it cannot target it to the T3SS. Future studies should address the additional function(s) of SatS, which we predict to be a role in targeting substrates to the SecA2 pathway. Alternatively (or in addition), SatS could have a separate role maintaining SapM in an unfolded state for protein translocation across SecYEG.

A SecA2-targeting role for SatS

To begin addressing the second role of SatS, beyond stabilizing substrates, that is disrupted in G134D we asked if the role affected by the G134D mutation is specific to the SecA2 pathway. In our whole cell phosphatase activity assay (described in Chapter 3) the $\Delta secA2/\Delta satS$ double mutant can be complemented by expressing wild type SatS to restore secreted SapM

phosphatase activity to that of a $\Delta secA2$ -only mutant. If we instead express SatS G134D in the $\Delta secA2/\Delta satS$ double mutant, we can ask whether SatS G134D restores secreted SapM phosphatase activity in this background. If the G134D mutation only affects export of SapM through the SecA2 pathway, we would expect the phosphatase activity of the $\Delta secA2/\Delta satS$ double mutant expressing SatS G134D to be equivalent to the $\Delta secA2$ -only mutant. Alternatively, if the G134D mutation affected export of SapM through multiple pathways (e.g. the SecA1 and SecA2 pathways), we would expect the phosphatase activity of the $\Delta secA2/\Delta satS$ double mutant expressing SatS G134D to be less than the $\Delta secA2$ -only mutant, which still allows a residual level of SapM export.

Expression of *satS G134D* in the $\Delta secA2/\Delta satS$ double mutant restored secreted SapM phosphatase activity to the level of a $\Delta secA2$ -only mutant (Figure 4.10). This result suggests that the function of SatS affected by the G134D mutation is specific to the SecA2 pathway. These preliminary data suggest that SatS may have a role in targeting its substrates to the SecA2 pathway and that the targeting function is disrupted in SatS G134D. This interpretation is consistent with the fact that the secretion defect in a $\Delta satS$ mutant complemented with SatS G134D is at the level of a $\Delta secA2$ mutant and not more severe (Figure 3.15).

If SatS has a role in targeting SapM to the SecA2 pathway, we expect that the affected role in SatS G134D occurs after stabilization of SapM. To validate this hypothesis, it is worthwhile to purify SatS G134D and test its ability to prevent SapM aggregation in our *in vitro* aggregation assay. If stabilizing SapM in the cytoplasm prior to export and preventing its aggregation are coupled, SatS G134D should behave like wildtype SatS in this assay. We expect that SatS G134D will prevent SapM aggregation, since we can still observe a SatS G134D:SapM

interaction. To study a direct role of SatS in targeting SapM to the SecA2 pathway, we will ultimately need to identify new interacting partners of SatS.

SatS substrates beyond SapM

Besides SapM, we found that SatS was required for the stability and export of multiple components of the SecA2-dependent Mce1 and Mce4 lipid transporters. The interpretation of the effects on multiple Mce proteins is complicated because the stability of multi-protein Mce transporter complexes could be affected by the removal of a single component (36).

Multiple Mce proteins from the same transporter were reduced in the whole cell lysate and cell wall of a $\Delta satS$ mutant. This could reflect a role for SatS in exporting numerous Mce transporter components. However, it is also possible that the role of SatS is in the export of only one Mce component, and that a defect in localizing that single Mce protein destabilizes the entire complex. To begin addressing which Mce proteins directly depend on SatS, we performed coimmunoprecipitations between SatS and different components of the *M. smegmatis* Mce4 transporter. When we pulled down Mce4E expressing an HA tag, it coimmunoprecipitated with SatS (lacking an epitope tag) (Figure 4.11). This result is significant for two reasons. First, it indicates that at least one component of the Mce4 transporter physically interacts with SatS. We interpret these results to suggest that Mce4E is a specific substrate of SatS. Secondly, these results demonstrate that we can immunoprecipitate a substrate and detect an interaction with SatS. We had previously only been able to detect SatS interactions in the the reverse coimmunoprecipitation (i.e. immunoprecipitate SatS and detect an interaction with the substrate SapM). Since these results do not indicate whether Mce4E is the only Mce4 substrate of SatS, future studies should include looking at SatS interactions with other Mce proteins.

Thus far, our approach to identifying SatS-dependent substrates has been biased towards our panel of validated SecA2-dependent substrates, plus two SecA2-independent loading controls, 19 kDa and MspA. From this panel, we have identified a group of SecA2-dependent substrates that also require SatS for export and a group that does not require SatS. While this approach was extremely useful in providing us with tools to study SatS function, it limits the scope of possible SatS-dependent substrates. In particular, we have not addressed the possibility that SatS could also be required for export of substrates through other pathways, specifically the SecA1 pathway. Because of the strong phenotype in a $\Delta satS$ mutant (i.e. a dramatic reduction of both pre-exported and exported levels of substrates), a good approach to address this question will be to do quantitative proteomics in the whole cell lysate, the cell wall, and the culture filtrate (fully secreted proteins) of a $\Delta satS$ mutant compared to wild type *M. tuberculosis*. The possibility of SatS-dependent, SecA2-independent substrates could introduce a role for SatS beyond the SecA2 pathway. There is precedent for Sec protein export chaperones working with more than one export pathway. In *Serratia marcesens*, SecB is required for the secretion of HasA through a Type I secretion system (T1SS) in addition to its role in Sec protein export (37). Learning more about the SatS exportome would be extremely useful not only for determining the role of SatS in *M. tuberculosis* but also in determining what features of an exported protein define it as SatS-dependent for export. Currently, we do not have an obvious connection between SapM and Mce proteins to account for them being SatS-dependent substrates.

SatS is required for growth of *M. tuberculosis* in macrophages

Prior TraSH/Tnseq analyses using pooled libraries of transposon mutants predicts SatS to be required during murine and macrophage infections; however, this prediction was never validated. In Chapter 3, using a $\Delta satS$ mutant and a complemented strain, we directly

demonstrated a role for SatS in *M. tuberculosis* growth in macrophages. Our results from these macrophage infections should be expanded to investigate a role for SatS in *M. tuberculosis* virulence in mice. SapM is known to be important early in infection, due to its role in preventing phagosome maturation arrest (5, 38). In contrast, cholesterol import by Mce4 is important during the persistence phase of infection (39). The requirement of SatS for the export of both of these proteins suggests that SatS may be important throughout infection, with a role in both acute establishment of disease as well as persistence in the host. Furthermore, the different effects on export of the $\Delta satS$ and $\Delta secA2$ mutants suggest that the phenotypes of these two strains during infection may be different, and teasing apart those differences may teach us more about *M. tuberculosis* pathogenesis.

Conclusions

By way of a genetic suppressor screen in *M. smegmatis*, we identified structural differences between SecA1 and SecA2, identified suppressors in SecA2 that map to SecY and substrate interacting sites, and characterized SatS as a new protein with roles in protein export and virulence in *M. tuberculosis*. The regions of structural difference between SecA2 and SecA1 and suppressor mutations identified in this dissertation represent exciting new directions for exploring the functional differences between SecA2 and SecA1 proteins. This work not only expands our understanding of the specialized SecA2 protein export pathway of mycobacteria but it provides important functional information for a previously uncharacterized *M. tuberculosis* protein that contributes to protein export and pathogenesis. Further, by identifying SatS as a protein export chaperone, our studies expand our appreciation of the diversity of chaperones in biological systems. Although chaperones have common functions and features, substantial structural diversity exists among these proteins, which is further highlighted by the new fold

revealed in the structure of SatS_C. Molecular chaperones are emerging as candidates for being new targets for drug therapies of microbial virulence (40, 41). The importance of SatS in mycobacterial pathogenesis presents us with an opportunity to exploit its functions in the search for novel and effective anti-mycobacterial drugs. Our whole cell SapM phosphatase activity assay is a high throughput approach to monitor secretion of SapM and could be used to identify inhibitory compounds of SapM, SatS, or other components of SecA2-dependent protein export through the high throughput screening of large chemical libraries. Taken together, the work presented in this dissertation allowed us to achieve our original goal: to improve our understanding of SecA2 and the SecA2 pathway. By identifying SatS, this work helped map molecular details of an *M. tuberculosis* pathway required for virulence. Our results reveal new questions that can be addressed in the future to continue improving our understanding of the mechanism of SecA2-dependent export and the role of the SecA2 pathway in pathogenesis.

Table 4.1. All suppressor mutations identified using *M. smegmatis* SecA2 K129R or *M. tuberculosis* SecA2 K115R

Isolate(s)	Effect on residue(s) in <i>MsSecA2</i>	Effect on residue(s) in <i>MsSecY</i>	Effect on residues in <i>MsSatS</i>
6S, 9S	Deletion of AA 182-185	DNS ^a	DNS
23S	Duplication of AA 182-185	DNS	DNS
2S	Asp326 → His	DNS	DNS
25S	Glu insertion at AA 364	DNS	DNS
34S	Thr459 → Ile	DNS	DNS
SSW8 ^b	Val528 → Gly*	NM ^c	NM
SSW11	Deletion of AA 651-658*	NM	NM
21B	Deletion of AA 734-741	DNS	DNS
38S	Deletion of AA 732-739	DNS	DNS
4S	NM	2 bp insertion at -137	DNS
24S	NM	C→G at -134	DNS
7S	NM	NM	Large chromosomal deletion including <i>satS</i>
20B	NM	NM	Large chromosomal deletion including <i>satS</i>
SSW2	NM	NM	Large chromosomal deletion including <i>satS</i>
10S	NM	NM	T → C at -6
33S	NM	NM	Frameshift at Trp25
33B	NM	NM	Cys123 → STOP
3S ^d	NM	NM	Gly134 → Asp
SSW7	DNS	NM	Frameshift at Trp168
SSW9	DNS	NM	Frameshift at Asp169
29S	NM	NM	Frameshift at Pro261
1S	NM	NM	Deletion of AA 269-270
SSW1 ^d	DNS	NM	Frameshift at Asp407
5S	NM (Low SecA2 via WB)	NM	NM
SSW10	NM (Low SecA2 via WB)	DNS	NM
13B	NM (Low SecA2 via WB)	NM	NM
7B	NM	NM	NM (No SatS via WB)
39S	NM	NM	NM

^a DNS – Did not sequence

^b SSW suppressors were isolated from an *M. smegmatis* strain expressing *M. tuberculosis secA2 K115R*, which is comparable to the *M. smegmatis secA2 K129R* allele.

^c NM – no mutations

^d SatS still detectable via WB

* These suppressors were isolated from *MtbSecA2 K115R* so the amino acids are listed for *MtbSecA2* with the correct start codon (start at amino acid 31)

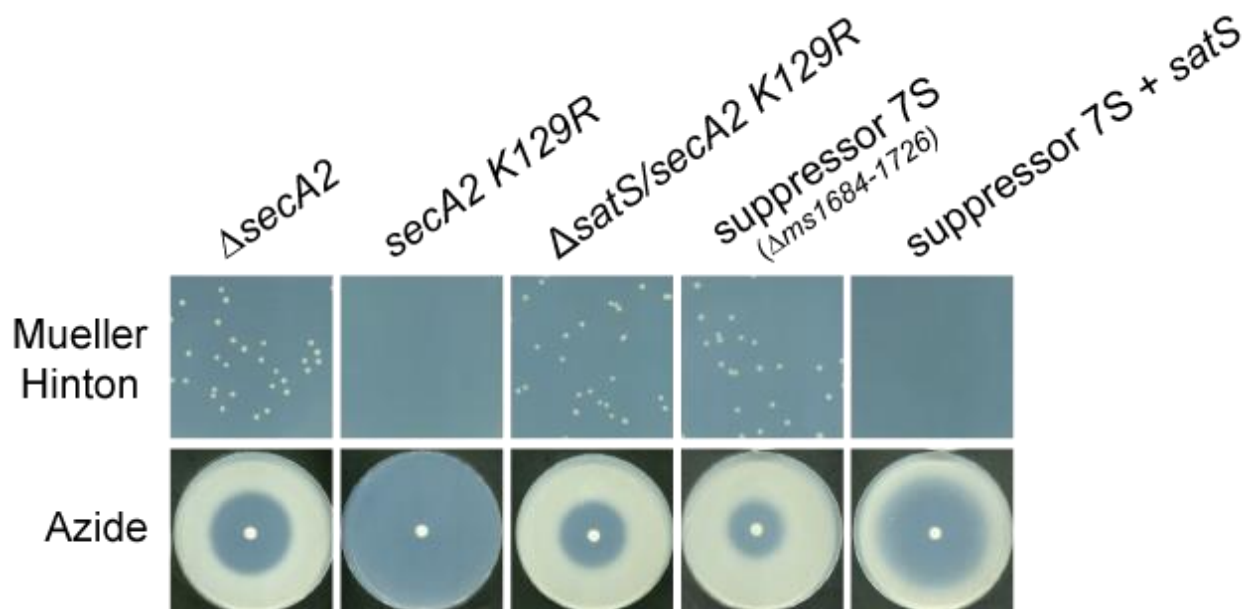


Figure 4.1. Adding back *satS* to the extragenic suppressor 7S (containing a large chromosomal deletion from *msmeg1684* to *msmeg1726*) is sufficient to restore SecA2 K129R phenotypes in this strain. Azide sensitivity and Muller-Hinton agar growth phenotypes for the indicated strains as described in Chapter 3.

```

mtSecA2      APKFYEEELSDKVSE----- 660
msSecA2      SPEFYAKLAEEIGEDAE----- 666
cgSecA2      APAFAAAELEDLQS----- 645
sgSecA2      ---FVIEYVIKQVAYEQSFENRADFYR FILHHFSNRAERIPQEFDIHSPKE----- 682
lmSecA2      ILREVAEYSFIHPSDIPEEELEIYYSRQKELGGTKFPISFDQVTLMEPRE----- 677
mtSecA1      MVRIVITAYVDC-ATGEGYAEDWDL DALWTALKTLYPVGITADSLTRKDHEFERDDL TRE 733
msSecA1      MLVIDVITAYVDC-ATAEGYAEDWDL ETLWTALKTLYPVGIDHRDLIDSDAVGEPGEL TRE 730
ecSecA       IREIVFKATIDAYIPPQSLEEMWDIPGLQERLKNDFDLPLIAEWLDKEPE-----LHEE 733
lmSecA1      MIQFTVNFIVSSNASSHEPEEAWN LQGIIDYVDANLLPEGTIT--LEDLQN-----RTSE 682

mtSecA2      -----ERLETICRQIMLYHLDRGWADHLAYLADIRESI 693
msSecA2      -----ERLEKICRLIMLYHLDRGWCEHLAFLADIRESI 699
cgSecA2      -----VREQAARDIMLYHLDYNWSEHLALMDDVRESI 677
sgSecA2      -VSHLLQAIAEQELLAKKS---Y LKSDKLYSQFQRLAILKAIDENWVEQVDYLQQLKSAL 738
lmSecA2      -----VVEEIVSWHKKER---NKFFPAETIAAIEREVYLNLMQMWVMHLDAMVQLREGI 728
mtSecA1      ELLEALLKDAERAYAAREAELEEIAGEGAMRQLERNVLLNVIDRKWREHLYEMDYLKEGI 793
msSecA1      ELLDALIKDAERAYAEREKQIEA IAGEGAMRQLERNVLLNVIDRKWREHLYEMDYLKEGI 790
ecSecA       TLRERILAQSIIEVYQRKE----EVVGAEMMRHFEKGVMLQTLDSLWKEHLAAMDYL RQGI 789
lmSecA1      DIQNLILDKIKAAAYDEKE----TLLPPEEFNEFEKVVLRLRVDTKWVDHIDAMDHLRDGI 738

                :   *   :*   *   ::   :   :...:

```

Figure 4.2. SecA2 orthologs with a truncated or missing HWD. Multiple sequence alignments of SecA2 and SecA2 orthologs, specifically the portion of the IRA1 domain (*Mtb*SecA2 residues 590-733) that contains the Helical Wing Domain (HWD) shown in red. The alignment was constructed with CLUSTALW 2.1. mt = *M. tuberculosis*, ms = *M. smegmatis*, cg = *C. glutamicum*, lm = *L. monocytogenes*, sg = *S. gordonii*, ec = *E. coli*. The residues deleted in suppressor SSW11 are highlighted in purple: AA 651-658 YEELSDK. * = fully conserved residue, : = conservation between strongly similar residues, . = conservation between weakly similar residues.

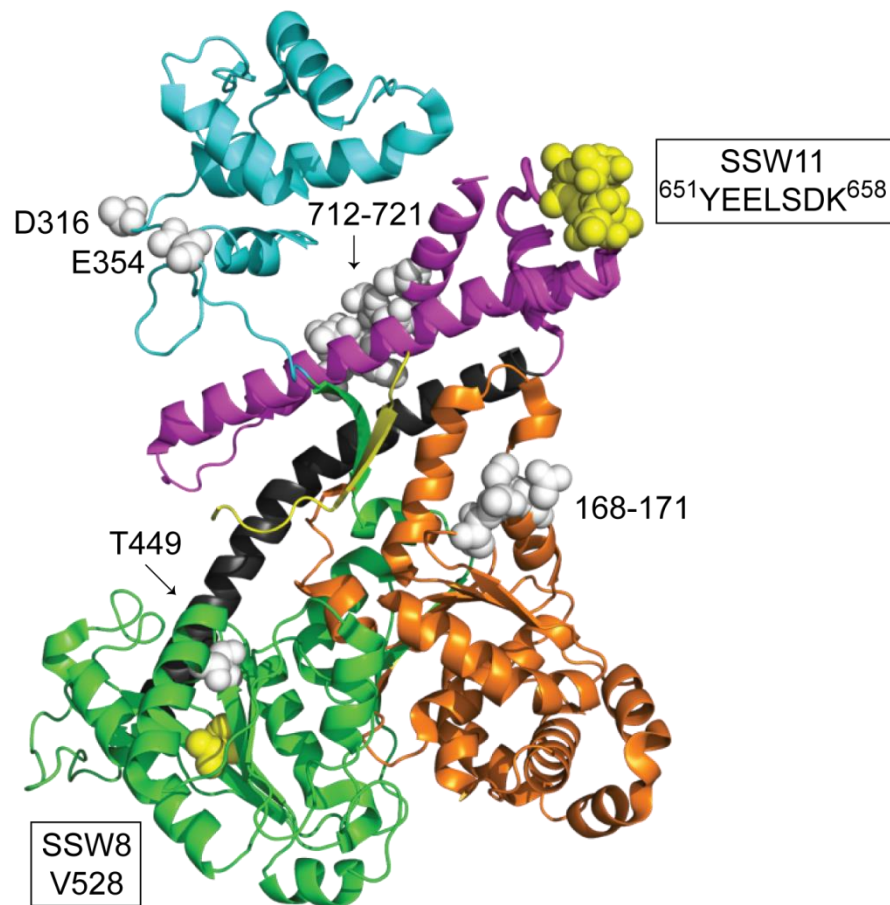


Figure 4.3. Intragenic suppressor mutations of *M. tuberculosis* SecA2 K115R mapped onto *M. tuberculosis* SecA2 structure using Pymol. Mutations of the two new suppressors are shown as yellow spheres and are labelled with boxes. SSW11 is a seven amino acid deletion at the tip of the IRA1 domain, in a region that usually contains the HWD (*M. tuberculosis* SecA lacks the HWD). SSW8 is a single amino acid substitution of a valine for a glycine located near a previously identified suppressor in the NBD2 domain. Mutations identified in suppressors of *M. smegmatis* SecA2 K139R in chapter 2 are shown in white spheres.

Table 4.2. Protein export chaperone features shared by SatS

Generalized Export Chaperone Feature ^a	SatS
Highly acidic (pI<5.0)	pI = 3.8
Cytoplasmic location	SatS is exclusively in the cytoplasm
Required for export of a subset of proteins of a specific pathway	SatS is required for SapM and Mce export but not for SecA2-exported SBPs or PknG
T3SCs and EspG commonly located in an operon with substrates	<i>sapM-satS</i> are in an operon
Interacts with substrate	SatS interacts with SapM as shown by Co-IP from mycobacteria and effects in the aggregation assay
Protects substrate from inappropriate interactions (aggregation, folding/misfolding, degradation)	SatS protects SapM precursor from aggregation <i>in vitro</i> . SatS maintains cellular levels of SapM and Mce proteins, consistent with a role preventing degradation
Targets substrates to machinery	Unknown, but the G134D mutation indicates at least two separable roles in export

^a Specifically SecB, Type 3 Secretion Chaperones (T3SCs) and EspG

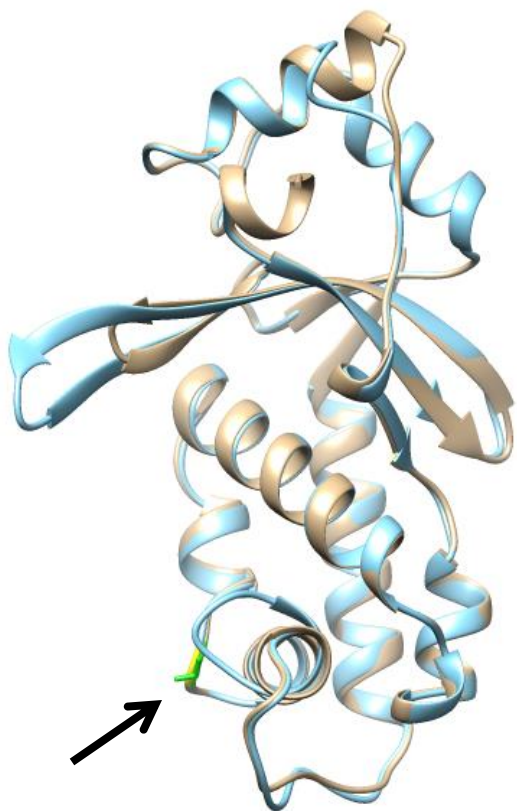


Figure 4.4. SatS_N structure is predicted to be remarkably similar to SatS_C structure. Tan ribbon structure is the solved SatS_C domain (AA 242-420). Modeled on top of the SatS_C structure is the predicted tertiary structure of the SatS_N domain (AA 1-178) in blue. The residue mutated in suppressor 3S (G134) is indicated in yellow. The equivalent residue on the C domain (T375) is indicated in green. Both residues are indicated by a black arrow. T375 is exposed to the surface of SatS_C and G134 is predicted to be surface exposed on SatS_N.

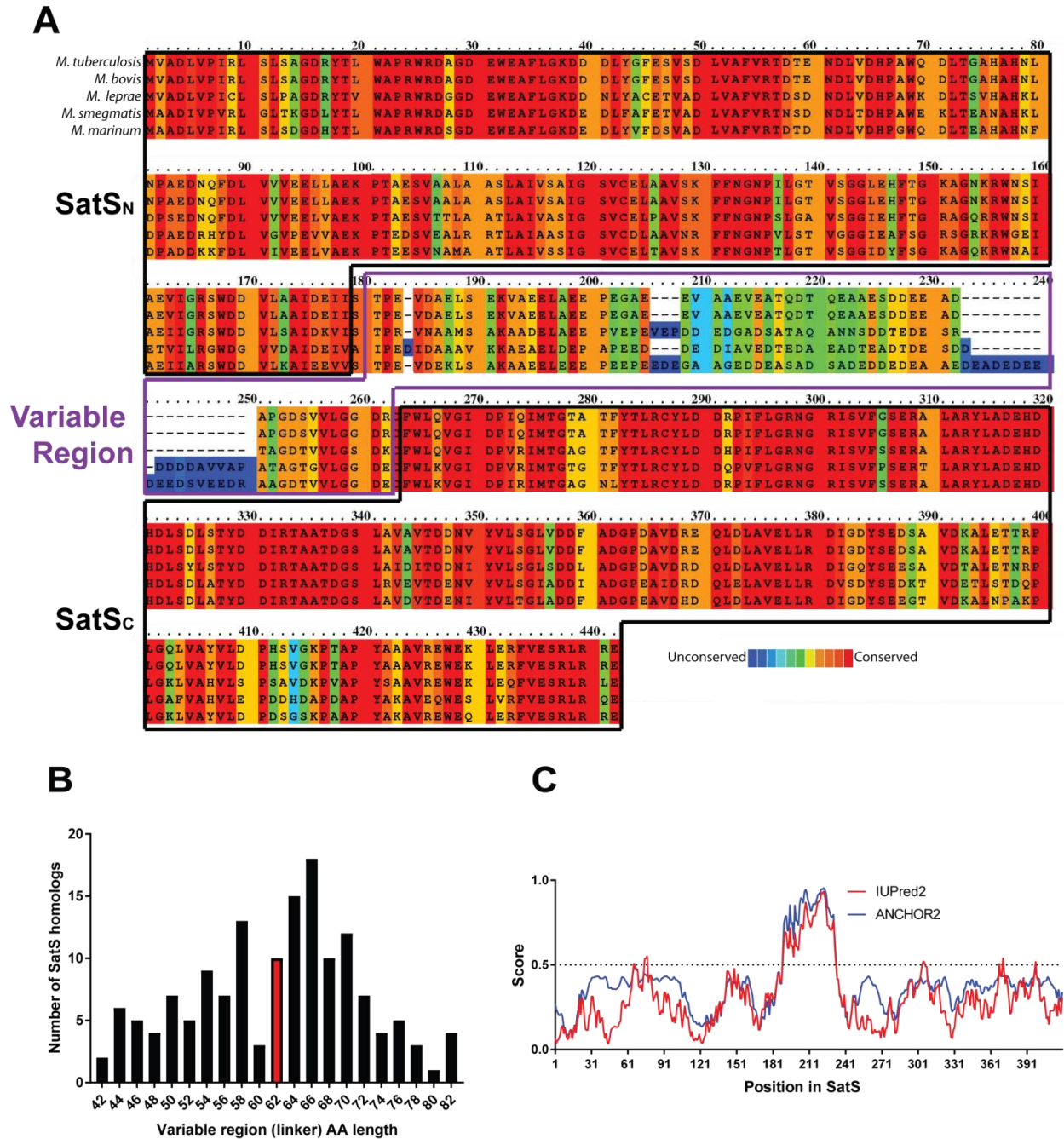


Figure 4.5. SatS_N and SatS_C domains are separated by a variable linker. A) PRALINE alignments of five closely related SatS homologs from mycobacteria showing residue conservation of the SatS_N and SatS_C domains (outlined in black) and the variable region (outlined in purple). Highly conserved residues are highlighted in red and unconserved residues are highlighted in blue. B) Histogram of variable region lengths in 150 SatS homologs. Homolog list was generated using Consurf. The red bar includes the length for *M. tuberculosis* SatS variable region (42 amino acids). C) Prediction of disordered regions in SatS generated by IUPred. A score over 0.5 is predicted to be an intrinsically disordered region.

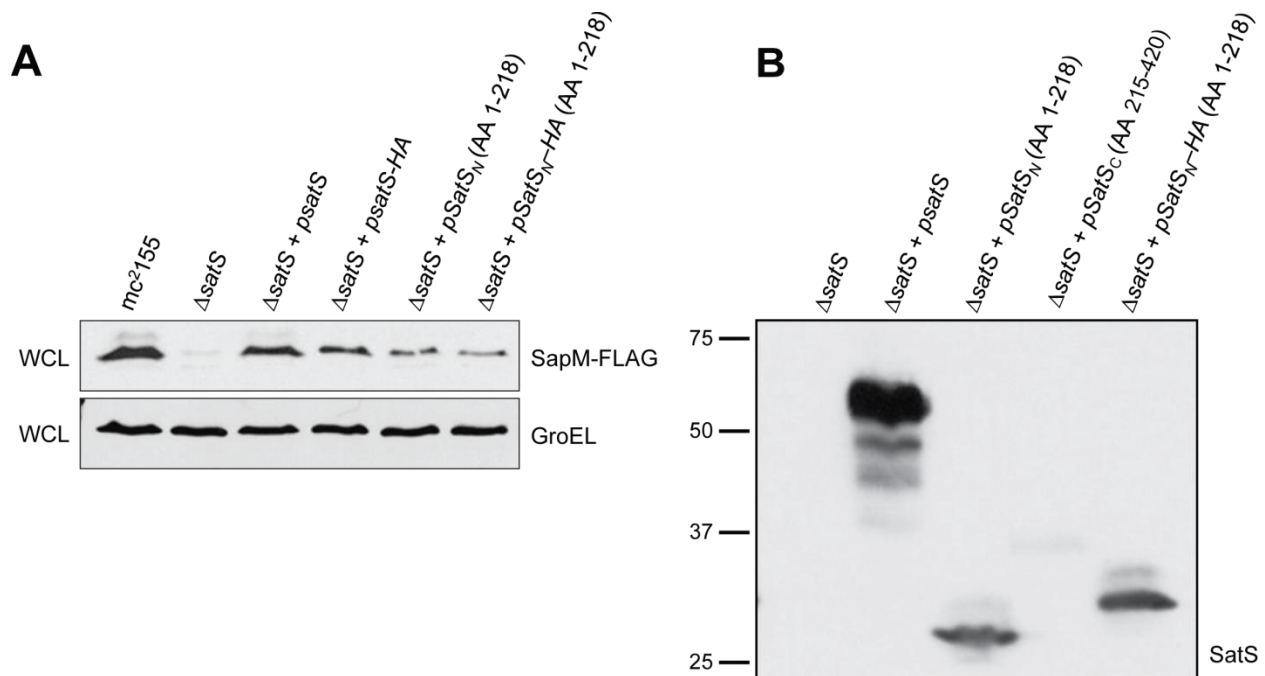


Figure 4.6. SatS_N domain can restore stability of SapM in a $\Delta satS$ mutant. A) Equal protein from whole cell lysates (WCL) of *M. smegmatis* *mc²155*, $\Delta satS$, $\Delta satS$ expressing wild-type SatS ($\Delta satS + psatS$) with or without an HA tag, and $\Delta satS$ expressing SatS_N ($\Delta satS + psatS_N$) with or without an HA tag were examined for levels of SapM-FLAG and GroEL by immunoblot. Note, this construct contains some of the disordered linker between SatS_N and SatS_C (AA 179-218). A similar construct made for the C domain containing some of the disordered linker (AA 215-420) does not express and was not used in these experiments as shown in panel B. B) Equal protein from WCL of *M. smegmatis* $\Delta satS$ expressing various SatS single domain constructs as labelled were examined for SatS expression by immunoblotting with SatS antibodies.

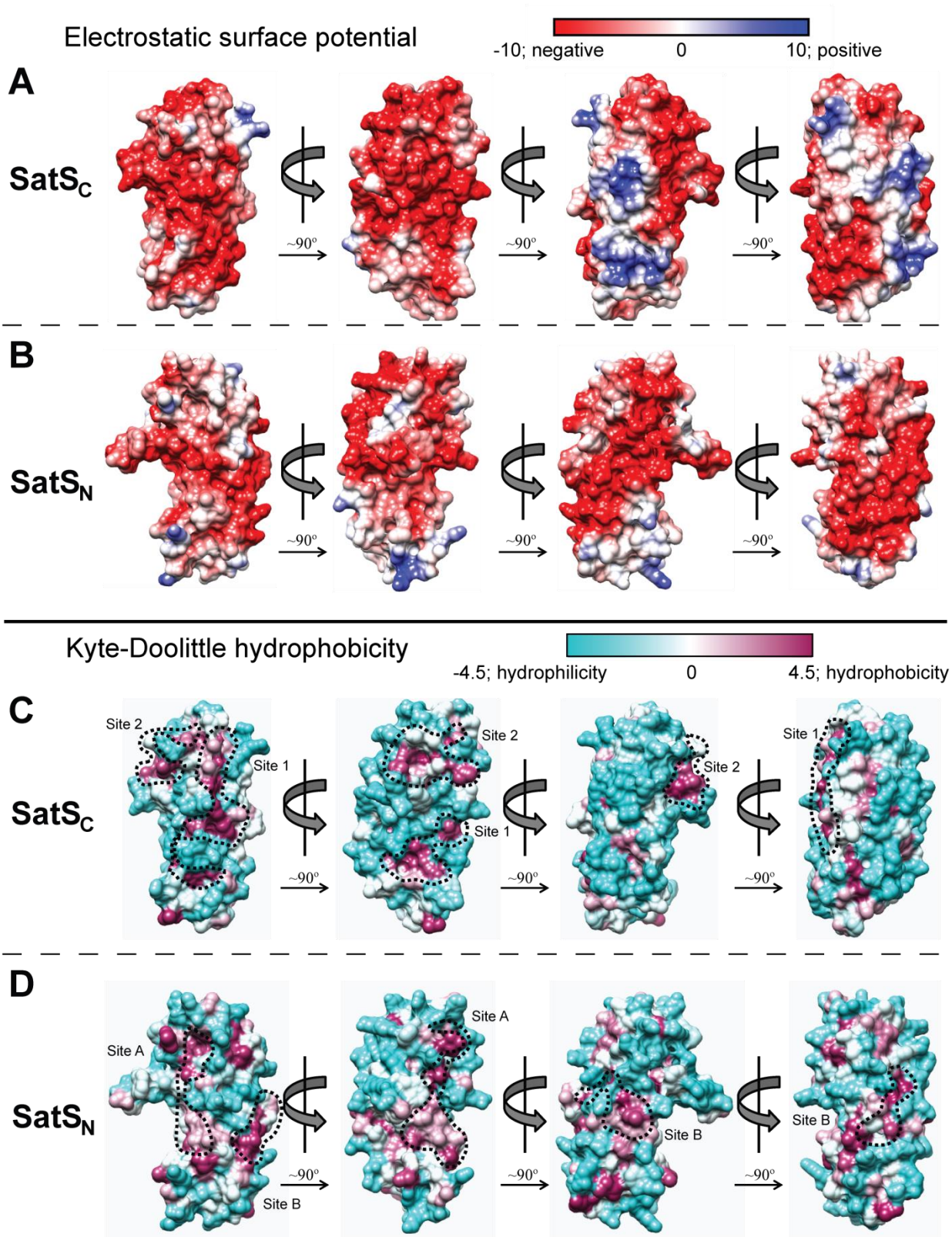


Figure 4.7. The N domain and C domain of SatS_{Mtb} display different surface characteristics. A) SatS_C surface colored according to electrostatic surface potential with red corresponding to a negative and blue to a positive potential. B) SatS_N model structure surface colored according to electrostatic surface potential. C) SatS_C surface colored according to hydrophobicity with purple corresponding to hydrophobic and cyan to hydrophilic residues. Sites with hydrophobic grooves circled in dashed lines and labelled Site 1 and Site 2. D) SatS_N as shown in Figure 4.3 surface colored according to hydrophobicity as described for panel A. SatS_N has two hydrophobic sites (Site A and Site B) outlined. The four images in each panel are 90° rotations of the two SatS domains. All images were generated with USCF Chimera (29).

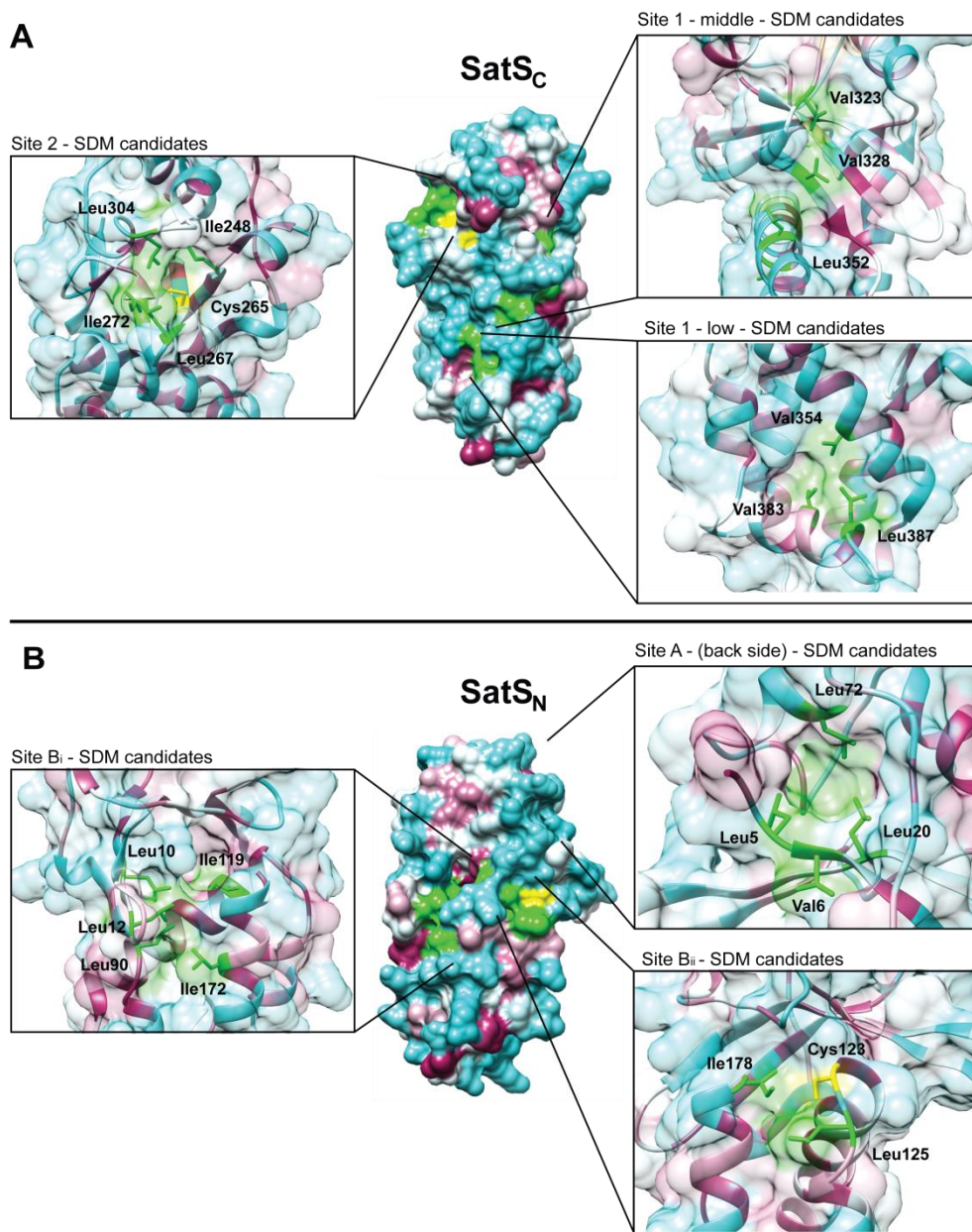



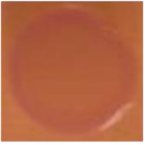

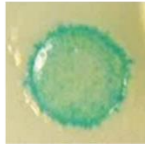
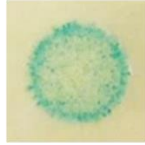



Figure 4.8. Hydrophobic grooves in SatS with candidate residues for mutagenesis to test for roles in substrate binding. A) The solved structure of SatS_C surface colored according to hydrophobicity with purple corresponding to hydrophobic and cyan to hydrophilic residues. Candidate residues occupying hydrophobic grooves are highlighted in green and labelled in the pop-out figures. The conserved cysteine residue is highlighted in yellow. B) The modeled structure of SatS_N surface colored according to hydrophobicity. Candidate residues occupying hydrophobic grooves are highlighted in green and labelled in the pop-out figures. The conserved cysteine residue is highlighted in yellow. All images were generated with USCF Chimera (29).

A

MacConkey Maltose Media		
	SatS-T18	T18-SatS
SatS-T25		
T25-SatS		

B

LB X-gal Media		
	SatS-T18	T18-SatS
SatS-T25		
T25-SatS		

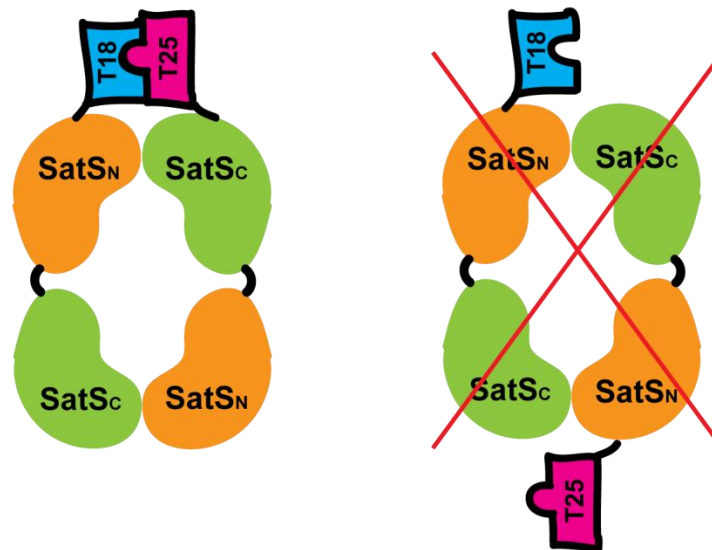
C

Figure 4.9. SatS forms an orientation-specific multimer in a bacterial two-hybrid assay. A and B) SatS-CyaA domain fusions are defined by the CyaA domain fused (T18 or T25) and the order of SatS and the domain. BTH101 *E. coli*, lacking an endogenous CyaA, is transformed with two expression vectors each containing a domain of SatS fused to T18 or T25, as indicated in the table. 2 μ L of culture are spotted onto (A) MacConkey media supplemented with 2% maltose or (B) LB media containing 40 μ g/mL X-gal. C) Cartoon of a SatS multimer model that could account for the results in panels A and B. SatS_N and SatS_C refer to the two domains of SatS.

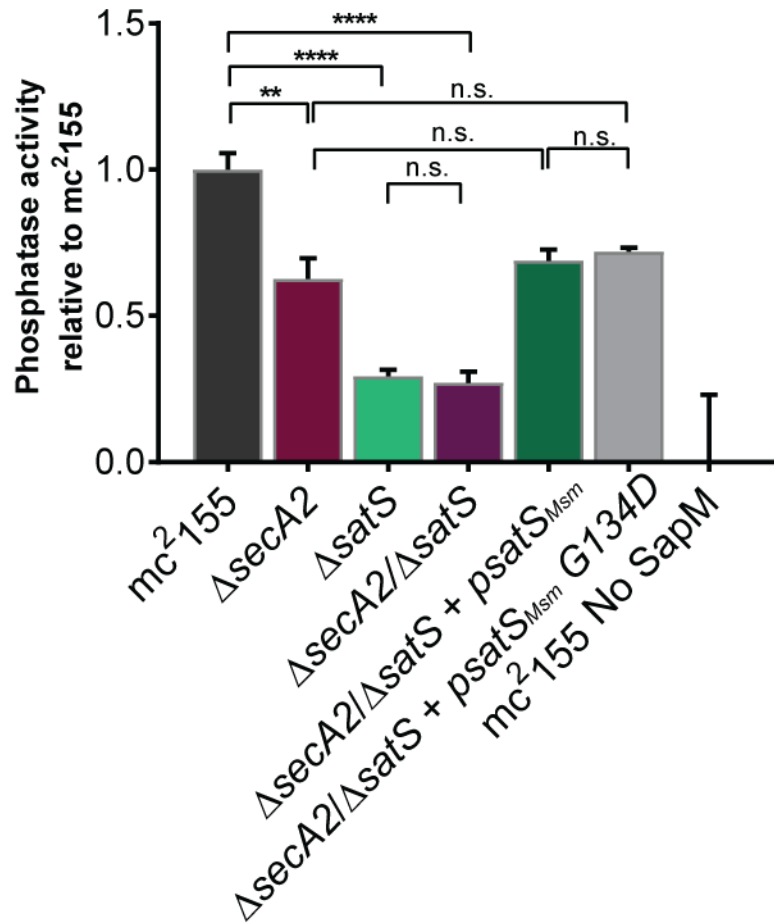


Figure 4.10. SatS G134D affects a SecA2-specific role of SatS. Whole cell phosphatase activity assay using *M. smegmatis* mc²155, ΔsecA2, ΔsatS, the ΔsecA2ΔsatS double mutant, ΔsecA2ΔsatS expressing wild-type SatS (ΔsatS+psatS_{Msm}), and ΔsecA2ΔsatS expressing the SatS G134D mutant (ΔsatS+psatS_{Msm}G134D). All strains are expressing SapM or an empty vector as indicated (mc²155 No SapM). Triplicate wells were grown in a 96 well plate for 24 hours at 37°C before measuring phosphatase activity by quantifying cleavage of pNPP. Rates were normalized as described in Chapter 3. Data is representative of five independent experiments. Error bars represent standard deviation of the mean of three independent replicates for each strain. n.s. – no significant difference; **, p<0.01; ****, p < 0.0001 by ANOVA and Tukey's post hoc test.



Figure 4.11. SatS and Mce4E interact in *M. smegmatis*. Lysates from *M. smegmatis* $\Delta secA2/\Delta satS$ expressing SatS_{Mtb} and Mce4E_{Msm}-HA or only expressing SatS_{Mtb} were used for co-immunoprecipitation using anti-HA conjugated beads. Lysates (left) and immunoprecipitations (right) for each strain were probed with SatS antibody and rabbit HA antibody for Mce4E.

REFERENCES

1. Global tuberculosis report 2017. Geneva: World Health Organization, 2017 Contract No.: CC BY-NCSA 3.0 IGO.
2. Braunstein M, Espinosa BJ, Chan J, Belisle JT, R. Jacobs W. 2003. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 48(2):453-64.
3. Kurtz S, McKinnon KP, Runge MS, Ting JPY, Braunstein M. 2006. The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. *Infection and Immunity* 74(12):6855-64.
4. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, et al. 2015. Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & Cellular Proteomics* 14(6):1501-16.
5. Zulauf KE, Sullivan JT, Braunstein M. 2018. The SecA2 pathway of *Mycobacterium tuberculosis* exports effectors that work in concert to arrest phagosome and autophagosome maturation. *PLOS Pathogens* 14(4):e1007011.
6. Rigel NW, Gibbons HS, McCann JR, McDonough JA, Kurtz S, Braunstein M. 2009. The accessory SecA2 system of mycobacteria requires ATP binding and the canonical SecA1. *The Journal of Biological Chemistry* 284(15):9927-36.
7. Bieker-Brady K, Silhavy TJ. 1992. Suppressor analysis suggests a multistep, cyclic mechanism for protein secretion in *Escherichia coli*. *The EMBO Journal* 11(9):3165-74.
8. Rigel N. 2008. Characterization of the SecA2 protein export pathway of mycobacteria [Dissertation]. Chapel Hill, NC: University of North Carolina at Chapel Hill.
9. Ligon LS, Rigel NW, Romanchuk A, Jones CD, Braunstein M. 2013. Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of mycobacteria. *J Bacteriol* 195(19):4456-65.
10. Hou JM, D'Lima NG, Rigel NW, Gibbons HS, McCann JR, Braunstein M, et al. 2008. ATPase activity of *Mycobacterium tuberculosis* SecA1 and SecA2 proteins and its importance for SecA2 function in macrophages. *J Bacteriol* 190(14):4880-7.
11. DeJesus M, Sacchettini JC, Ioerger TR. 2013. Reannotation of translational start sites in the genome of *Mycobacterium tuberculosis*. *Tuberculosis* 93(1):18-25.
12. Chen Y, Bauer BW, Rapoport TA, Gumbart JC. 2015. Conformational changes of the clamp of the protein translocation ATPase SecA. *Journal of Molecular Biology* 427(14):2348-59.
13. Bhanu MK, Zhao P, Kendall DA. 2013. Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility method. *J Bacteriol* 195(20):4709-15.
14. Das S, Grady LM, Michtav J, Zhou Y, Cohan FM, Hingorani MM, et al. 2012. The variable subdomain of *Escherichia coli* SecA functions to regulate SecA ATPase activity and ADP release. *J Bacteriol* 194(9):2205-13.

- 15.D'Lima NG, Teschke CM. 2014. ADP-dependent conformational changes distinguish *Mycobacterium tuberculosis* SecA2 from SecA1. *The Journal of Biological Chemistry* 289(4):2307-17.
- 16.Vrontou E, Karamanou S, Baud C, Sianidis G, Economou A. 2004. Global co-ordination of protein translocation by the SecA IRA1 switch. *Journal of Biological Chemistry* 279(21):22490-7.
- 17.Huang C, Rossi P, Saio T, Kalodimos CG. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature* 537:202.
- 18.Ekiert DC, Cox JS. 2014. Structure of a PE–PPE–EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. *PNAS* 111(41):14758-63.
- 19.Stebbins CE, Galán JE. 2001. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* 414:77.
- 20.Daleke MH, van der Woude AD, Parret AHA, Ummels R, de Groot AM, Watson D, et al. 2012. Specific chaperones for the type VII protein secretion pathway. *Journal of Biological Chemistry* 287(38):31939-47.
- 21.Knoblauch NTM, Rüdiger S, Schönfeld H-J, Driessen AJM, Schneider-Mergener J, Bukau B. 1999. Substrate specificity of the SecB chaperone. *Journal of Biological Chemistry* 274(48):34219-25.
- 22.Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, et al. 2014. Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25–PPE41 dimer. *Mol Microbiol* 94(2):367-82.
- 23.Letzelter M, Sorg I, Mota LJ, Meyer S, Stalder J, Feldman M, et al. 2006. The discovery of SycO highlights a new function for type III secretion effector chaperones. *The EMBO Journal* 25(13):3223-33.
- 24.Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biology Direct* 7:12
- 25.Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, et al. 2014. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Research* 42(Database issue):D297-D303.
- 26.Mészáros B, Erdős G, Dosztányi Z. 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research* 46(W1):W329-W37.
- 27.Knodler LA, Bertero M, Yip C, Strynadka NCJ, Steele-Mortimer O. 2006. Structure-based mutagenesis of SigE verifies the importance of hydrophobic and electrostatic residues in type III chaperone function. *Mol Microbiol* 62(4):928-40.
- 28.Tsirigotaki A, De Geyter J, Šoštaric' N, Economou A, Karamanou S. 2016. Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology* 15:21.
- 29.Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, et al. 2004. UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25(13):1605-12.

30. Bordes P, Cirinesi A-M, Ummels R, Sala A, Sakr S, Bitter W, et al. 2011. SecB-like chaperone controls a toxin–antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *PNAS* 108(20):8438-43.
31. Armstrong DJ, Roman A. 1993. The anomalous electrophoretic behavior of the human papillomavirus type 16 E7 protein is due to the high content of acidic amino acid residues. *Biochemical and Biophysical Research Communications* 192(3):1380-7.
32. Kumamoto CA, Francetić O. 1993. Highly selective binding of nascent polypeptides by an *Escherichia coli* chaperone protein in vivo. *J Bacteriol* 175(8):2184-8.
33. Matsumoto H, Young GM. 2009. Essential role of the SycP chaperone in type III secretion of the YspP effector. *J Bacteriol* 191(5):1703-15.
34. Fekkes P, De Wit JG, Van Der Wolk JPW, Kimsey HH, Kumamoto CA, Driessen AJM. 1998. Preprotein transfer to the *Escherichia coli* translocase requires the co-operative binding of SecB and the signal sequence to SecA. *Mol Microbiol* 29(5):1179-90.
35. Thomas NA, Ma I, Prasad ME, Rafuse C. 2012. Expanded roles for multicargo and class 1B effector chaperones in type III secretion. *J Bacteriol* 194(15):3767-73.
36. Perkowski EF, Miller BK, McCann JR, Sullivan JT, Malik S, Allen IC, et al. 2016. An orphaned Mce-associated membrane protein of *Mycobacterium tuberculosis* is a virulence factor that stabilizes Mce transporters. *Mol Microbiol* 100(1):90-107.
37. Sapriel G, Wandersman C, Delepelaire P. 2003. The SecB chaperone is bifunctional in *Serratia marcescens*: SecB is involved in the Sec pathway and required for HasA secretion by the ABC transporter. *J Bacteriol* 185(1):80-8.
38. Puri RV, Reddy PV, Tyagi AK. 2013. Secreted acid phosphatase (SapM) of *Mycobacterium tuberculosis* is indispensable for arresting phagosomal maturation and growth of the pathogen in guinea pig tissues. *PLoS ONE* 8(7):e70514.
39. Pandey AK, Sasseti CM. 2008. Mycobacterial persistence requires the utilization of host cholesterol. *PNAS* 105(11):4376-80.
40. Henderson B, Allan E, Coates ARM. 2006. Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. *Infection and Immunity* 74(7):3693-706.
41. Neckers L, Tatu U. 2008. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host & Microbe* 4(6):519-27.