Title
Structural and Mechanistic Insights into a Specialized Secretion System Necessary for M. tuberculosis Virulence

Author
Dovala, Dustin Leard

Publication Date
2014

Peer reviewed|Thesis/dissertation

Structural and Mechanistic Insights into a Specialized Secretion System Necessary for *M. tuberculosis* Virulence

by

Dustin Dovala

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

I would like to firstly acknowledge my parents, John and Pamela Dovala, my brother, John Travis Dovala, and my grandparents, Margaret and Barney Leard for their unending and unconditional support throughout my life as I have pursued my scientific education. I would have never made it this far without their help. And also to Elizabeth "Ginger Snap" Booth, who has been there for me for most of my graduate school career.

I am deeply indebted to my PI, Jeffery Cox, who took me into his lab and trained me how to be a scientist. He has always been extremely supportive, and has gone to bat for me on numerous occasions. I could not have asked for a better mentor. I would also like to thank Oren Rosenberg, who was instrumental for me to learn biochemistry and with whom I worked very closely on this project.

I'd also like to thank my friends – within UCSF and beyond – for their love and support these last several years.

Lastly, I'd like to thank the other members of my thesis committee, Hiten Madhani and Orion Weiner, for all of their ideas, support, and guidance.

Chapters 2 and 3 contain material previously submitted for publication:

Dustin Dovala*, Oren Rosenberg*, Xueming Li, Anastasia Bendebury, Janet Finer-Moore, James Holten, Yifan Cheng, Robert M. Stroud, and Jeffery S. Cox. Substrates Control Multimerization and Activation of the Multi-domain ATPase Motor of Type VII Secretion.

*Authors contributed equally to this work.

Abstract

Mycobacterium tuberculosis utilizes a specialized secretion system, known as Type VII Secretion (T7S) to translocate virulence factors into the host cell. T7S is absolutely necessary for virulence of *M. tuberculosis*, as well as other pathogens such as *S. aureus*. Intriguingly, T7S systems are broadly conserved amongst all Gram-positive bacteria, and likely play diverse roles in bacterial physiology. In this work, we focused on the two conserved components present in all T7S systems: the FtsK-like ATPase EccC and the WXG-100 substrate EsxB. We developed a biochemical model system to answer questions about the structure and molecular mechanisms of these two protein components, and confirm our results using genetics in *M. tuberculosis*. Using this system, we i) identified the substrate-binding pocket on EccC, ii) identified the non-canonical "signal sequence" on the EsxB substrate, iii) determined the molecular basis of substrate specificity, iv) solved the structure of the EccC ATPase both bound and unbound to signal sequence, as well as the EssC homolog from *Geobacillus thermodenitrificans*, v) determined a molecular mechanism of EccC autoinhibition, and vi) uncovered a mechanism by which substrates control activity of EccC through multimerization. Overall, this work generated the first biochemical model for T7S.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1: *M. tuberculosis* and the Type VII Secretion System

Introduction: The global health burden of *Mycobacterium tuberculosis*

Mycobacterium tuberculosis, the causative agent of tuberculosis disease, is a facultative intracellular pathogenic bacterium that primarily infects human macrophages. *M. tuberculosis* is wide-spread and highly contagious, and thus approximately one-third of the world population is latently infected with the bacteria. In 2013 alone, *M. tuberculosis* resulted in 9 million new instances of tuberculosis disease and 1.5 million deaths, making this bacterium the second highest killer due to a single infectious agent (behind HIV/AIDS) [1].

M. tuberculosis is primarily spread through aerosol, when actively-infected individuals cough, sneeze or speak and disseminate bacteria suspended in micro-droplets (<10 μm). Inhalation of as few as 1-10 bacteria is often sufficient to cause latent infection [2]. Of immuno-competent individuals latently infected with *M. tuberculosis*, approximately ninety percent will not develop tuberculosis disease over their lifetimes [1]. However, those with compromised immune systems, such as people with HIV/AIDS, have a much higher chance of developing active disease.

Tuberculosis disease primarily impacts third world and developing countries, and up to half of new cases may go undetected [1]. Persons suffering from untreated tuberculosis disease not only face a high mortality rate (seventy percent within ten years in HIV-negative individuals) [3], but will also infect ten to fifteen new individuals every year [4]. First line treatment usually consists of four antibiotics, ethambutol, pyrazinamide, rifampin, and isoniazid, and can last anywhere from six months to a year.

Unfortunately, poor patient adherence to drug regimens has led to the emergence of drug-resistant, multi-drug-resistant (MDR), and extensively-drug-resistant (XDR) strains of *M. tuberculosis* [1], [5].

New cases of MDR-TB have been reported in every country surveyed [1], [5]. These patients require a much longer course of antibiotics, often up to two years, including more expensive and less effective drugs, many of which are toxic. The economic, social, and global health burden of drug resistance in *M. tuberculosis* is only exacerbated by the fact that in the last forty years, only one new drug, bedaquiline, has been developed specifically to fight this disease. A much greater understanding of how *M. tuberculosis* causes disease and survives in the host cell must be achieved to generate new treatment methodologies to complement the antibiotics already available.

M. tuberculosis **life cycle**

The *M. tuberculosis* life cycle begins when bacteria, suspended in micro droplets, are inhaled into the lungs of a naïve host. Alveolar macrophages recognize the bacteria and phagocytose them. Phagocytosis elicits a number of transcriptional and post-translational modes of regulation within the bacterium, resulting in secretion of a number of substrates that serve to halt phagosome maturation, access the host cytosol, and elicit the type I interferon response. These virulence factors serve to transform the macrophage's typically toxic environment into a replicative niche, where the bacilli can grow and divide. Upon reaching a certain density, the *M. tuberculosis* bacilli promote macrophage cell-death and infect other nearby host cells. While the infection can become systemic, only bacteria in the lung are transmitted from one individual to

another. This occurs when an actively infected individual coughs, propelling bacteria suspended in micro droplets into the environment.

The Type VII Secretion System

Aside from the conserved sec system, many bacteria require specialized secretion systems to translocate proteins and other macromolecules across the cell membrane (Figure 1.1). In Gram-negative bacteria, these systems typically arise from the need to get substrates across both the inner and outer membrane. *M. tuberculosis*, despite being classified as Gram-positive [6], also utilizes a specialized secretion system (Type VII Secretion or T7S) in addition to the more conserved sec and tat systems to translocate proteins out of the cell [7]. The necessity for such a system may arise from the highly complex cell envelope of *M. tuberculosis* -- consisting of an inner membrane, a layer of peptidoglycans, a layer of arabanogalactans, and an outer membrane composed of tightly packed mycolic acids (Figure 1.2) [8].

Genome-wide analysis has revealed the presence of five T7S systems in *M. tuberculosis*, denoted ESX-1-5. ESX-1, the most widely studied, is absolutely required for virulence and is the primary genetic difference between virulent and avirulent mycobacteria [9]–[12]. The ESX-3 system is necessary for mycobactin-mediated iron uptake as well as zinc metabolism [13]–[16], while ESX-5 has been implicated in the secretion of PE/PPE proteins and also appears, along with ESX-1, to have a role in virulence [17]–[19]. The roles of ESX-2 and the likely ancestral ESX-4 system remain a mystery.

T7S systems are broadly conserved amongst all Gram-positive bacteria, with systems differing somewhat in composition between the actinobacteria (which include the mycobacteria) and the firmicutes (which include other pathogens such as *Staphylococcus aureus*) (Figure 1.3). The biological roles of these systems are widely variant and almost totally unknown for most systems, but include virulence (*M. tuberculosis* [18], [20]–[22]*, S. aureus* [23]), conjugation (*M. smegmatus* [24]), sporulation (*S. scabies* [25]), and nutrient uptake (*M. tuberculosis* [13], [15]).

T7S systems appear to be somewhat heterogeneous with respect to their protein compositions. Some, such as the ESX-1 system in *M. tuberculosis*, contain a very large number of proteins across multiple genetic loci; while others, such as ESX-4, are relatively simple and contain only a few proteins [11]. Within the actinobacteria, however, most systems contain, at least, a core set of six proteins (Figure 1.4): a subtilisin-like serine protease (MycP) [26], [27], an eleven-pass transmembrane protein (EccD), a membrane-associated protein of unknown function (EccB) [28], an FtsK-like ATPase (EccC), and the ~100 amino acid W-X-G family proteins EsxA and EsxB. The only proteins that are absolutely conserved between both actinobacteria and firmicutes, however, are EccC (called EssC in the firmicutes) and EsxB (sometimes annotated as EsxA in firmicute systems).

The ESX-1 system

The virulence-associated ESX-1 system is highly regulated in *M. tuberculosis*, perhaps due to the high immunogenicity of the EsxA and EsxB substrates [29]. Unusually, in this system an additional genetic locus, which encodes the espACD operon, is

necessary for secretion of all other substrates [30]. This system is under tight regulation by the ESX-1 encoded EspR transcription factor, which also binds elsewhere in the genome and may have roles as a nucleoid-associated protein [31]–[34]. Recent work has also uncovered a role for the protein EspI in the negative regulation of the ESX-1 system in response to low ATP levels [35], though the exact mechanisms of this regulation remain unknown.

The exact role for ESX-1 secreted substrates is unclear; however, aside from virulence the ESX-1 system is necessary for induction of the Type I Interferon response [36], host cytosol access [21], [37], recruitment of autophagy machinery [38], [39] and inhibition of phagosome maturation [40], [41]. Because ESX-1 is involved in so many aspects of *M. tuberculosis* virulence, it may represent an ideal non-conventional drug target. Thus, a detailed molecular-level understanding of how it works is necessary.

Figure 1.1

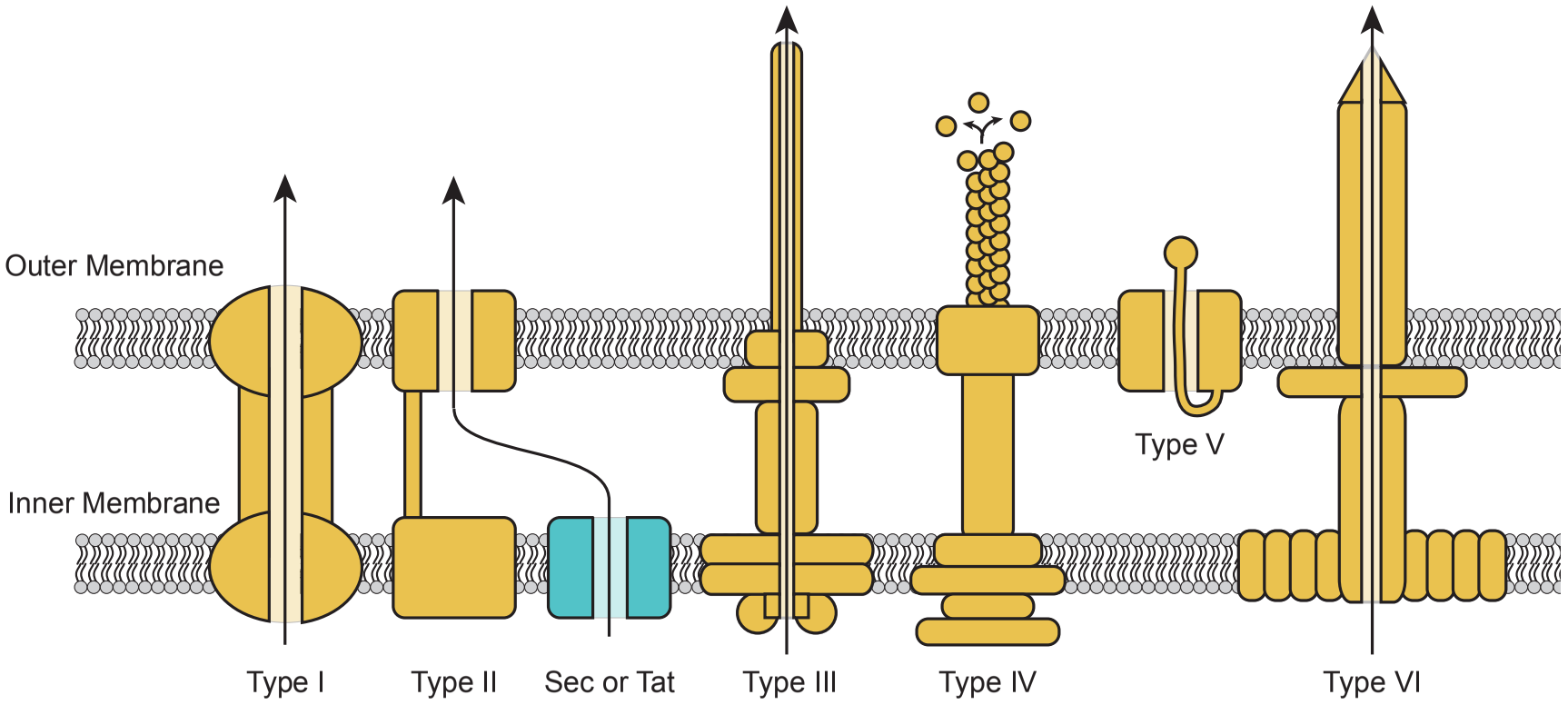

Outer Membrane
Inner Membrane
Type I
Type II
Sec or Tat
Type III
Type IV
Type V
Type VI

Figure 1.1: Schematic of other specialized secretion systems found in bacteria

Basic schematic of the Types I-VI secretion systems found in Gram-negative bacteria. The Type I, III, IV, and VI systems are independent of the conserved sec and tat secretion systems, while the Type II system requires substrates to be first secreted to the periplasm before secretion past the outer membrane. Likewise, the Type V autotransporter must itself be first transport to the periplasm by the sec system.

Figure 1.2

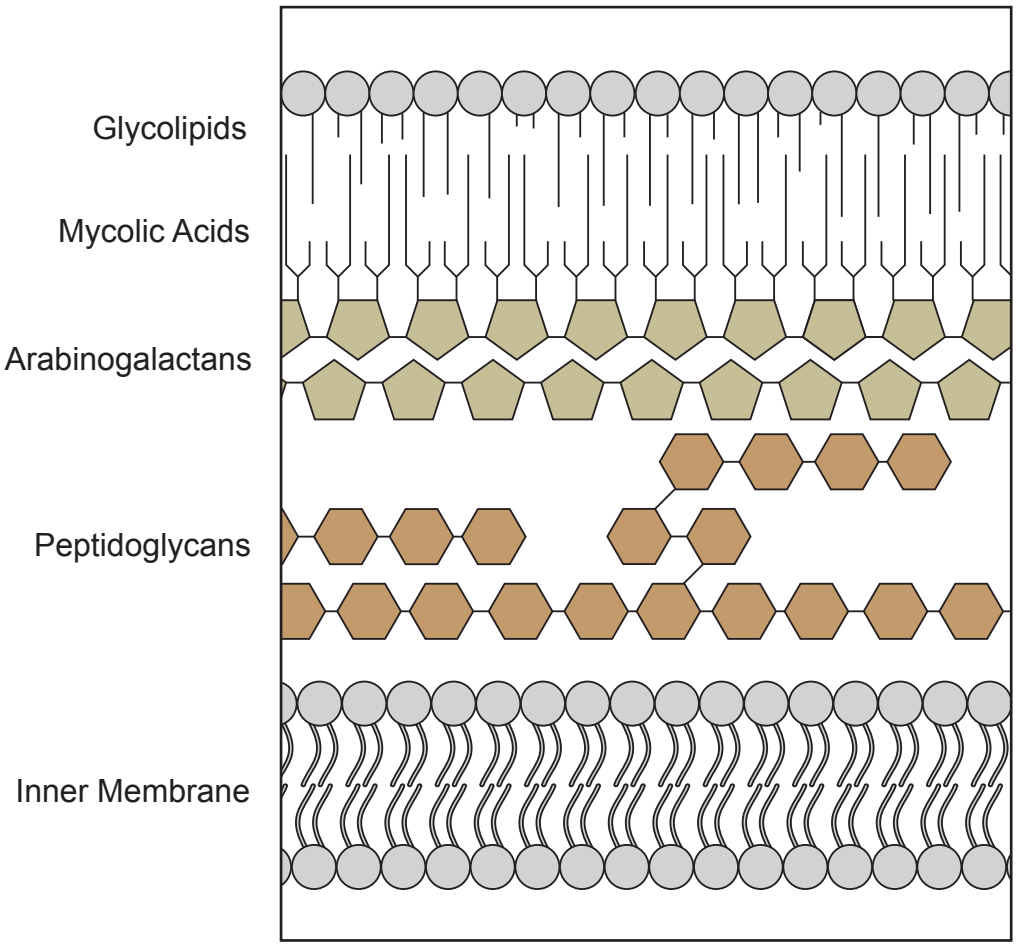

Glycolipids
Mycolic Acids
Arabinogalactans
Peptidoglycans
Inner Membrane

Figure 1.2: Schematic of the *M. tuberculosis* cell envelop

M. tuberculosis contains an inner membrane as well as an outer membrane composed primarily of mycolic acids and other glycolipids. Between the inner and outer membrane are thick layers of peptidoglycans and arabinogalactans.

Figure 1.3

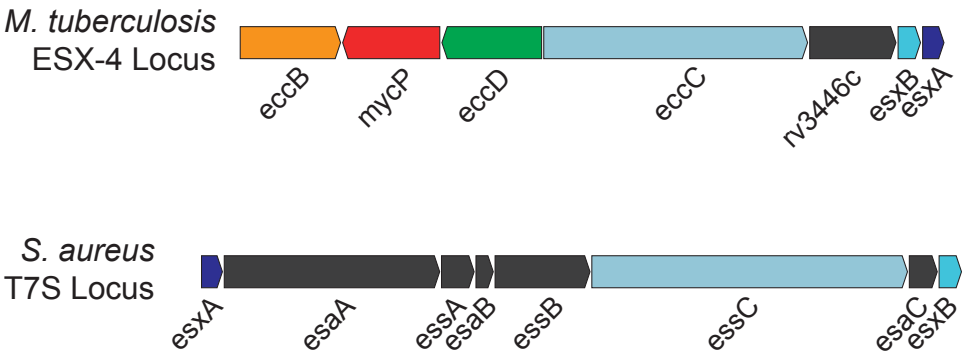

M. tuberculosis
ESX-4 Locus
eccB
mycP
eccD
eccC
rv3446c
esxB
esxA
S. aureus
T7S Locus
esxA
esaA
essA
esaB
essB
essC
esaC
esxB

Figure 1.3: Comparison of T7S systems in actinobacteria and firmicutes

Comparison of the ESX-4 T7S system from *M. tuberculosis* (a member of the actinobacteria) with the T7S system from *S. aureus* (a member of the firmicutes). Despite being common in the actinobacteria, the firmicute systems lack homologs of eccB, mycP, and eccD. Additionally, despite being present in the *S. aureus* system (although not in its typical locus downstream of EsxB), many firmicutes lack the W-X-G substrate EsxA. The essC gene in firmicutes is homologous to the eccC gene in actinobacteria (also known as yukBA in *Bacillus subtilis*), and along with the EsxB substrate (yukE in *B. subtilis*) are the only conserved proteins in all T7S systems.

Figure 1.4

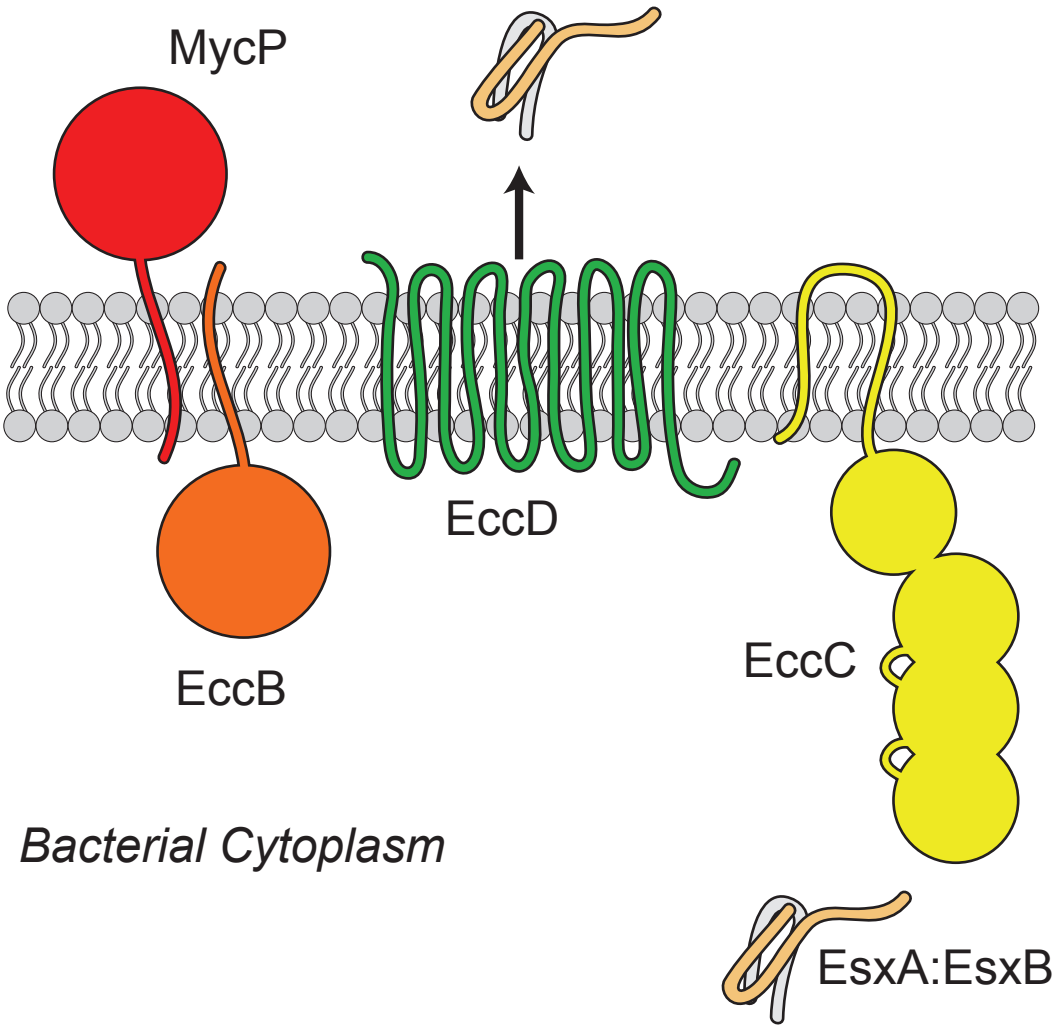


Figure 1.4: Schematic of major T7S components in actinobacteria

Most actinobacteria T7S systems contain these proteins. MycP is a serine protease; EccB is a membrane-bound protein of unknown function; EccD is an 11-pass transmembrane protein which contains a ubiquitin-like domain, also of unknown function; EccC is a membrane-bound ATPase belonging to the FtsK-like ATPase family; and EsxA and EsxB are secreted substrates belonging to the WXG-100 protein family.

Chapter 2: Molecular determinants of T7 signal sequence binding

Abstract

Type VII protein secretion (T7S) systems are present in many bacterial families and are important for their role in virulence of several pathogens, especially the ESX-1 system of *Mycobacterium tuberculosis*. The EccC secretion-type ATPase is a key component of all T7S systems and is thought to recognize and pump substrates out of the bacterial cell [11], [42]. Unfortunately, EccC is a large, multi-domain ATPase that has been extremely difficult to express and purify, which has hindered progress in understanding the mechanism of T7S. Here we report the robust expression and purification of EccC from a thermophilic bacterium, *Thermomonospora curvata*, and the use of this technological advent to develop an in vitro assay to study T7S signal sequence recognition. We show that the *T. curvata* T7S system is closely related to ESX-1 at the sequence level and likely functions analogously to the *M. tuberculosis* system as the known protein-protein contacts of EccC and its substrates are conserved. We have used this system to further examine the mechanism of substrate recognition by dissecting the interactions between EccC and its substrate complex EsxAB. We report the first structures of the C-terminal EsxB signal sequence that includes two regions thought to be important for substrate binding and recognition by EccC, and identify the binding pocket on EccC with co-crystal studies. Furthermore, we show that single amino acid substitutions in a hitherto unseen C-terminal helix on EsxB prevent binding to EccC, and that substrate-translocase specificity is completely coded in the C-terminal seven amino acids of EsxB.

Introduction

Recognition of secretory substrates by other secretion systems

Protein and oligonucleotide secretion systems rely on specific sequence or structural motifs to correctly identify and interact with secretory substrates. In bacteria, the conserved Sec system, which transports unfolded substrates across the membrane, is responsible for the secretion of most extracellular proteins. In this system the signal sequence, whose structure is conserved in all domains of life, is encoded on the N-terminus of protein substrates and can be recognized during or after translation [43]. In co-translational secretion, the signal recognition particle (SRP) recognizes and binds to the signal sequence during translation and brings the nascent secretory substrate:ribosome complex to the Sec translocase (SecYEG and SecA), where motor protein SecA accepts the peptide and powers its translocation (using ATP and the PMF) through a pore formed by SecYEG. During post-translational secretion, the signal sequence is recognized by chaperones, such as SecB, which keep substrates unfolded and deliver them to SecA directly. Upon translocation, the signal sequence is cleaved by a signal peptidase.

Alternatively, the conserved twin-arginine transporter (Tat) system translocates folded substrates across the membrane. The Tat system utilizes a similar, though typically longer, N-terminal signal sequence which also contains two almost invariant arginine residues [43]. This signal sequence is often recognized initially by chaperones which aid in substrate folding and help prevent erroneous targeting to the Sec system. Eventually, substrates are brought to the TatBC complex. In the presence of a PMF,

this stimulates the recruitment of TatA which is thought to form ring-shaped gated pores and possibly power translocation across the membrane.

In diderms, additional secretion apparatuses are required to translocate substrates across the second membrane (Figure 1.1) [7], [44], [45]. Some of these first require Sec or Tat secretion systems to bypass the inner membrane (Types II and V) [46], [47], while others are alternatives to Sec and Tat altogether (Type I, III, IV, and VI) [44], [48]–[50]. Each of these systems requires additional or alternative signal sequences and/or conformations, many of which are currently unknown.

The Type VII signal sequence

Unlike Sec and Tat substrates, T7S substrates lack an N-terminal signal sequence [51], [52]. In fact, these substrates lack any conserved sequences at the N- or C-termini. Attempts to identify the signal sequence have been limited to genetic experiments. To date, while there are many known substrates of T7S, only one, EsxB, has been shown to actually interact with the putative translocase EccC [52], while other substrates may rely on specific chaperones for targeting and secretion rather than direct binding [53]. In early work, regions of EsxB were removed and resultant proteins were tested for their ability to interact with EccC in a yeast-two-hybrid system [52]. These experiments revealed that the C-terminal-most residues of EsxB were absolutely required for interaction with EccC. However, attempts to identify a general signal sequence failed, as the C-terminal residues were not conserved amongst any of the other known substrates.

Later work in *Mycobacterium marinum* identified a motif consisting of Tyr - Xaa - Xaa - Xaa - Asp/Glu (or Y-x-x-x-D/E) that was present in most known T7S substrates [54]. Alteration of the tyrosine, acidic residue, or the spacing between them abrogated secretion of several ESX-1 and ESX-5 substrates in *M. marinum*. Because the sequence is highly conserved and necessary for secretion, it has been called a "general secretion signal" for T7S. However, it is unclear how this sequence, which is identical amongst substrates of most T7S systems, would be sufficient to drive the high degree of specificity observed in organisms with multiple T7S systems. In this same set of experiments, the authors also showed that in *M. marinum* swapping the C-termini of PE35 (an ESX-1 substrate) and PE25 (an ESX-5 substrate) did not switch system specificity, calling into doubt some of the earlier findings [54].

Results

T. curvata model system

As repeated attempts to express and purify *M. tuberculosis* EccC were unsuccessful, we screened a number of EccC and EssC homologs from a wide range of bacterial species for robust expression in *E. coli*. We expressed these proteins as His-tagged fusions that lacked the predicted transmembrane domains at the N-termini and found that EccC from *Thermonospora curvata* and EssC from *Geobacillus thermodenitrificans* were both highly expressed and soluble.

Because *T. curvata* is an actinobacteria and its T7S system more closely resembles that of *M. tuberculosis* (Figure 2.1A), we moved forward with this species. Indeed, analysis of the genomic context of eccC provided further evidence that this

gene is involved in T7S as this locus encodes a core set of T7S proteins, including EccB, EccD, MycP, and the dimeric substrate EsxAB (Figure 2.1A). In particular, the *T. curvata* T7S system most closely resembles the ESX-4 system from *M. tuberculosis*, which has been identified as the most ancestral amongst the five [55]. Importantly, in addition to EccC, we also successfully overexpressed and purified both EsxA and EsxB. These proteins were properly folded and stable as His-tagged fusions even without solubility tags, such as MBP, which have been necessary for purification of other EsxAB homologs [56].

In addition to biochemical isolation of Ecc membrane protein complexes of the ESX-5 system [57], individual protein-protein interactions have been identified between soluble components/substrates of ESX-1 by yeast two-hybrid [20]. To begin to determine if the *T. curvata* system functions similarly to ESX-1, we tested whether these protein-protein interactions are conserved in *T. curvata* T7S proteins. The previous yeast-two-hybrid studies showed that in *M. tuberculosis*, EccCa and EccCb interact strongly to make a complete EccC complex. To determine if the two halves of EccC from *T. curvata* interact in the absence of a continuous polypeptide chain, we tested the equivalents of EccCa ($EccC_{(200-720)}$) and EccCb ($EccC_{(721-1315)}$) in the same system. The split proteins interacted strongly (Figure 2.1B), indicating the presence of a conserved protein-protein interface between these domains. We also included EsxA and EsxB in this analysis and found that these proteins, like their counterparts in ESX-1, form a tight heterodimer (Figure 2.1B). Most importantly, the C-terminal domains of EccC interacted strongly with EsxB, but not EsxA, suggesting that the mechanism of substrate selection by EccC is also conserved (Figure 2.1B). Moreover, because these

interactions are analogous to those described for ESX-1, they may be conserved amongst all T7S systems.

EccC binds to the EsxAB complex through EsxB

We first wished to confirm the yeast-two-hybrid interactions directly with purified EccC, EsxA, and EsxB using size exclusion chromatography (SEC). In agreement with the previous studies, EccC directly bound to EsxB, but not EsxA, as evidenced by a dramatic shift in the elution profile (Figure 2.2A-B). Because EsxB interacts with both EsxA and EccC via yeast-two-hybrid, the three proteins are predicted to form a ternary complex. Indeed, addition of EccC, EsxA and EsxB proteins resulted in the formation of a complex that migrated as a single peak via SEC containing all three proteins (Figure 2.2C-D).

The structure of EsxAB reveals a signal sequence that is present on a C-terminal amphipathic helix

How EccC selects secretory substrates is a key unanswered question. Having purified proteins now allowed us to address this issue mechanistically. To address the question of substrate specificity by EccC, we first performed structural studies of the *T. curvata* EsxAB substrate heterodimer. We purified selenomethionine-derivitized EsxA and native EsxB separately and combined them in a 1:1 molar ratio to form the heterodimer EsxAB. This complex readily crystallized and we were able to determine the structure to 2.0 Å resolution (crystallographic information and statistics are shown in Table 2).

The *T. curvata* EsxAB heterodimer adopts a very similar structure to other W-X-G protein dimers, including *M. tuberculosis* $EsxAB_1$. Both EsxA and EsxB form alpha

helical structures containing two primary alpha helices broken by a short 3_{10} helix containing the W-X-G motif (Figure 2.3A). The proteins bind in an anti-parallel fashion to form a four helix bundle which buries 1535 $Å^2$ of hydrophobic surface.

Alignment of EsxB from the *T. curvata* structure with the ESX-1 EsxB homolog (PDBid 3FAV, chain a, rms = 1.96 Å) revealed that in *T. curvata*, EsxA binds to EsxB at an offset angle (27°) relative to the *M. tuberculosis* dimer. The physiologic relevance of this difference remains unclear (Figure 2.3D).

In contrast to all of the other EsxAB crystal structures available, the *T. curvata* EsxAB structure encompassed nearly the entire EsxB protein, with only the N-terminal three and very C-terminal amino acid missing. Importantly, this structure gave us the first glimpse of the Y-x-x-x-D/E motif and C-terminal signal sequence residues implicated in substrate recognition (Figure 2.3B-C). This is counter to other structural studies in which one or both of these motifs are unstructured [58], [59]. Surprisingly, the C-terminal residues of EsxB predicted to be recognized by EccC lie within an additional, previously unseen alpha helix (Figure 2.3C). This helix is connected to the rest of the protein via a small linker, suggesting there may be flexibility between the helix and the rest of the heterodimer.

The Y-x-x-x-D/E motif (YEARE in *T. curavata* EsxB), which is required for the secretion of some substrates through the T7S systems ESX-1 and ESX-5 [54], is situated near the end of the second major EsxB helix. While this portion of the helix does not appear structurally distinct from the rest of the protein, it is interesting that

there is an interaction between the tyrosine residue and the W-X-G motif of EsxA (Figure 2.3B).

The C-terminal alpha helix is competent to bind to EccC, and does so in a nucleotide-independent manner

To directly test binding of EccC to the C-terminal alpha helix, we created a fluorescently labeled probe and performed fluorescence anisotropy binding assays with purified EccC (Figure 2.4A). The probe, containing only the last ten residues of EsxB (the residues present on the C-terminal helix only), was competent to bind EccC with an apparent K_d of approximately 10 - 15 μM (Figure 2.4B). While not a particularly strong interaction, these values are on par with other substrate:translocase interactions, such as the Sec signal sequence binding to SecA [60].

Because EccC is an ATPase, we wondered whether or not binding to this "signal sequence" was nucleotide-dependent. Addition of ATP modestly increased binding affinity (Figure 2.4B), but this is likely due to an increase in protein stability rather than a conformational change. Addition of EDTA to remove residual nucleotide (by removing Mg^{2+}) had no effect, and mutation of catalytic residues in $ATPase_1$ or $ATPase_3$ also failed to affect binding (Figure 2.4B). An equivalent K_d was measured when binding the probe to just $ATPase_3$, suggesting that this domain contains the EsxB binding pocket.

Hydrophobic residues V98 and L102 are necessary for EccC binding, Y84 and E88 are not

We next sought to identify key residues involved in the EccC:EsxB interaction. To that end, we performed alanine scanning mutagenesis along the C-terminal ten residues of

EsxB, along with Y84 and E88 from the Y-x-x-x-D/E motif, and tested binding to EccC. Two mutations, V98A and L102A, rendered EsxB incapable of binding EccC in our SEC assay (data not shown). Interestingly, these two residues reside on the hydrophobic side of the amphipathic C-terminal helix -- suggesting the interaction with EccC is mediated through this half of the helix.

We confirmed the SEC results utilizing a fluorescence anisotropy competition assay, in which EccC was preloaded with fluorescent probe and then subjected to unlabelled EsxB protein (Figure 2.5A). EsxB binding would therefore compete off the probe and reduce anisotropy. While wildtype EsxB and EsxAB were able to compete off the fluorescent probe, $EsxB_{(V98A)}$ and $EsxB_{(L102A)}$ were unable to compete with the probe even at very high concentrations (Figure 2.5B). This confirms these two amino acid substitutions each render EsxB incompetent to bind EccC.

Interestingly, despite being implicated in secretion, the Y-x-x-x-D/E motif had no effect on EccC binding in either SEC or fluorescence anisotropy assays (Figure 2.5B). The role for this motif during secretion remains unclear, though it may be involved at a later step in the catalytic cycle.

Co-crystal structure reveals hydrophobic binding pocket on $ATPase_3$

To further characterize the interaction between EccC and EsxB, and to identify the location of the binding pocket on EccC, we solved the crystal structure of the second and third ATPase domains (equivalent to EccCb in *M. tuberculosis*) bound to a peptide containing the last 23 residues of EsxB – including the Y-x-x-x-D/E motif (Figure 2.6A-B, and crystallographic information and statistics are shown on Table 2). Importantly, we

found that the binding pocket was located far from the ATP binding site, and, although there was a saturating concentration of peptide present in the crystallization condition, EccC was still bound to ATP. Only the C-terminal eight amino acids of the peptide were structured, and they formed an equivalent alpha helix as seen with the full length substrate. Additionally, binding of the signal sequence did not alter the conformation of EccC in any detectable way when compared to the structure of the full length protein (Figure 2.6C, Information on the full length EccC structure in Chapter 3).

As expected from our previous results, the bulk of the interaction between EccC and the signal sequence was mediated by the hydrophobic residues V98 and L102 on EsxB (Figure 2.6B). These residues reach into a largely hydrophobic pocket on $ATPase_3$. Mutation of any of the hydrophobic residues lining this pocket, but not of a conserved polar residue, prevented binding of EsxB as detected by yeast-two-hybrid (Figure 2.6D). This suggests that the interactions between EccC and EsxB that have previously been detected via yeast-two-hybrid most likely occurred at this site.

The C-terminal seven residues of EsxB drive specificity for the EccC interaction

We next wished to examine the question of specificity amongst different T7S systems. We first tested whether this specificity exists at the level of binding of substrate to EccC. To answer this question, we used yeast-two-hybrid to determine the inherent cross-talk between the *T. curvata* T7S system and the *M. tuberculosis* ESX-1 system at the level of EccC - EsxB binding. Our results indicated that there was no cross-talk between the two systems when examining wild-type proteins and that the interactions were highly

specific (Figure 2.7). Furthermore, *Mt*EsxB1 did not interact with *Tc*EsxA, demonstrating there is also specificity in the EsxAB interaction.

To directly test the contribution of the C-terminal amino acids of EsxB towards specificity, we swapped seven of these residues from *Tc*EsxB with their equivalents from *Mt*EsxB$_1$ and tested this protein's ability to interact with *Mt*EccCb$_1$. Interestingly, this variant of *Tc*EsxB was fully competent to interact with *Mt*EccCb$_1$ but now failed to interact with *Tc*EccCb, suggesting that the specificity signal exists within the C-terminal seven amino acids (Figure 2.7). Moreover, the "tail-swapped" *Tc*EsxB still interacted strongly with its conventional binding partner *Tc*EsxA, suggesting that specificity amongst substrate:substrate interactions is distinct and not dependent on the C-terminal seven amino acid residues.

We next wished to determine which of these residues are the most important for specificity. To that end, we made a series of amino acid substitutions at various positions within the signal sequence (Figure 2.7). First, we substituted the two hydrophobic residues that extend into the binding pocket in the peptide-bound crystal structure to their *Mt*EsxB$_1$ equivalents (V98L L102M). However, while these substitutions attenuated the interaction with *Tc*EccCb they did not confer the ability to bind to *Mt*EccCb$_1$. We next substituted both of these residues along with additional C-terminal residues. Interestingly, switching four amino acids to their tuberculosis counterparts (V98L L102M N103G G104F) was sufficient to alter binding specificity to primarily bind *Mt*EccCb$_1$. Substitution of these four amino acids was sufficient to change the specificity of *Tc*EsxB such that it bound *Mt*EccCb$_1$, but not *Tc*EccCb. However, the interaction between this mutant and *Mt*EccCb$_1$ was not as strong as the

full tail swap -- suggesting that the other amino acids on the tail play some role in specificity. Indeed, substitution of the sequence between the two hydrophobic residues which insert into the binding pocket (QAL in *T. curvata* to SSQ from *M. tuberculosis*) greatly attenuates binding to *Tc*EccCb. Interestingly, one mutant, V98L L102M G104F, bound both EccC homologs, albeit more strongly to *Tc*EccCb. These data support the hypothesis that the C-terminal signal sequence helix is not only required for binding to EccC, but is also the primary specificity signal which ensures a substrate is secreted through the appropriate T7S translocon.

Conclusions and implications for future research

For this work, we generated a novel biochemical system to study Type VII Secretion system components utilizing the thermophilic actinobacterium *Thermomonospora curvata*. Using this system, we have identified the molecular determinants of signal sequence binding at both the EccC and EsxB level. EccC binds to EsxB at the C-terminal seven amino acid "signal sequence," which binds in a hydrophobic pocket on $ATPase_3$ of EccC. Identification of the binding pocket and specificity signal have important implications for drug and vaccine development and engineering artificial substrates and translocases for various applications. Furthermore, the development of this biochemical model system will allow us to further unravel the mechanisms that regulate T7S to ever greater detail.

The current live tuberculosis vaccine strain, *Mycobacterium bovis* BCG, is a naturally avirulent strain that lacks several large genetic loci when compared to virulent *M. tuberculosis*. The most important of these loci, RD1, encodes the ESX-1 secretion

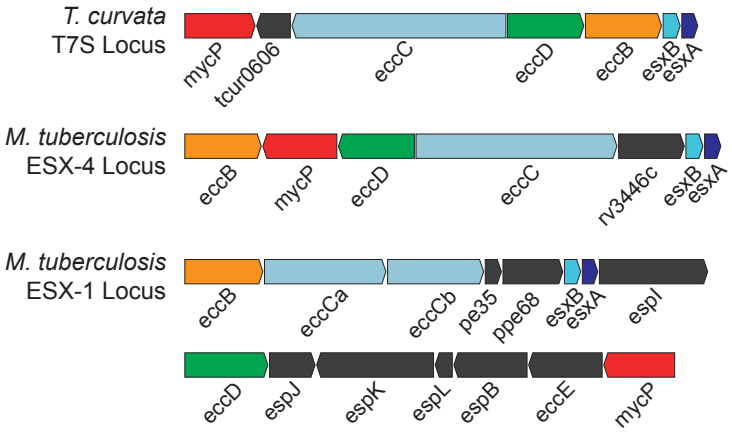
system. Thus, the vaccine strain does not secrete the virulence factor EsxA nor EsxB, which are both highly immunogenic antigens [29]. Using the targeting information discerned in this work, one intriguing possibility is to encode different signal sequences on these two proteins, facilitating their export through one of the other alternative ESX systems in the BCG vaccine strain. Ideally, lack of a complete ESX-1 system would maintain avirulence, while production and secretion of EsxA and EsxB would significantly bolster the immune response to the vaccine, hopefully improving its efficacy.

Additionally, it may be possible to target fluorescent proteins or other reporter proteins through different T7S systems to study the temporal dynamics of their activation in the context of live cell microscopy. Generation of novel T7S systems in certain bacteria may also be feasible by changing $ATPase_3$, which controls specificity on the ATPase, to that of an orthologous system, and then providing properly targeted substrates heterologously. This may be a way to get specific proteins secreted from Gram-positive bacteria with thick or otherwise impermeable membranes.

Lastly, through structural studies of the binding pocket on $ATPase_3$, it may be possible to target small molecules to prevent substrate binding and secretion. Ideally, this inhibition would phenocopy the T7S system knockout, which in *M. tuberculosis* and *S. aureus* has attenuated virulence.

Figure 2.1

A


T. curvata
T7S Locus
mycP
tcur0606
eccC
eccD
eccB
esxB
esxA
M. tuberculosis
ESX-4 Locus
eccB
mycP
eccD
eccC
rv3446c
esxB
esxA
M. tuberculosis
ESX-1 Locus
eccB
eccCa
eccCb
pe35
ppe68
esxB
esxA
espI
eccD
espJ
espK
espL
espB
eccE
mycP

B

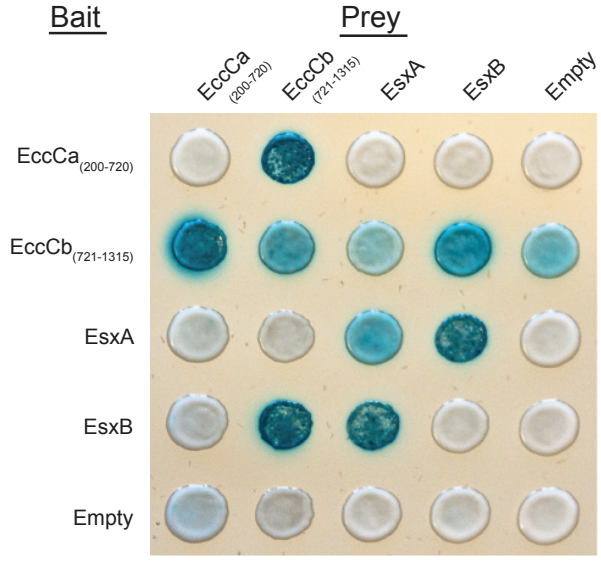

Bait
Prey
EccCa (200-720)
EccCb (721-1315)
EsxA
EsxB
Empty
EccCa(200-720)
EccCb(721-1315)
EsxA
EsxB
Empty

Figure 2.1: The *Thermomonospora curvata* T7S system

(A) The *T. curvata* T7S system locus as it compares with the *M. tuberculosis* ESX-1 and ESX-4 T7S systems. The *T. curvata* system is similar in composition to the ESX-4 system, which is predicted to be the most ancestral of the five ESX systems in *M. tuberculosis*.

(B) The known T7S system interactions are recapitulated in the *T. curvata* system. Specifically, the N- and C-terminal halves of EccC (denoted EccCa and EccCb) interact to form a complex, EsxB interacts with EccCb, and EsxB interacts with EsxA. The EccCb bait strain appeared to be more promiscuous and subject to false positives than the prey counterpart, interacting with the negative control.

Figure 2.2

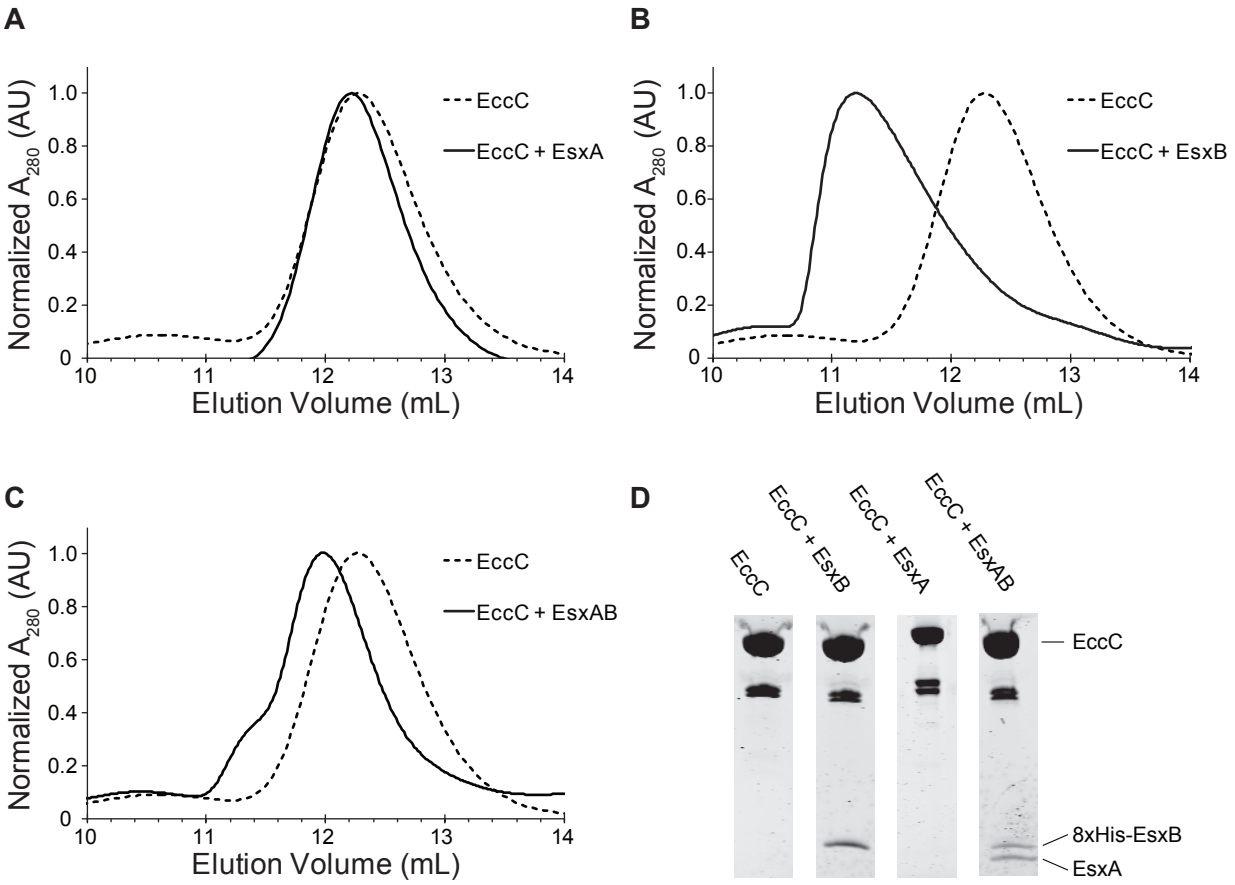

A
Normalized A_{280} (AU)
Elution Volume (mL)
EccC
EccC + EsxA
B
Normalized A_{280} (AU)
Elution Volume (mL)
EccC
EccC + EsxB
C
Normalized A_{280} (AU)
Elution Volume (mL)
EccC
EccC + EsxAB
D
EccC
EccC + EsxB
EccC + EsxA
EccC + EsxAB
EccC
8xHis-EsxB
EsxA

Figure 2.2: Examination of T7S system interactions with purified proteins

(A) As predicted by the yeast-two-hybrid experiment, EccC does not interact with the EsxA substrate by size exclusion chromatography.

(B) EccC forms a complex with the addition of EsxB, confirming the yeast-two-hybrid results.

(C) EccC forms a ternary complex with EsxA and EsxB. EsxB likely acts as the bridge between EccC and EsxA.

(D) SDS-PAGE gels indicating which proteins were found in the peak fractions from (A) – (C).

Figure 2.3

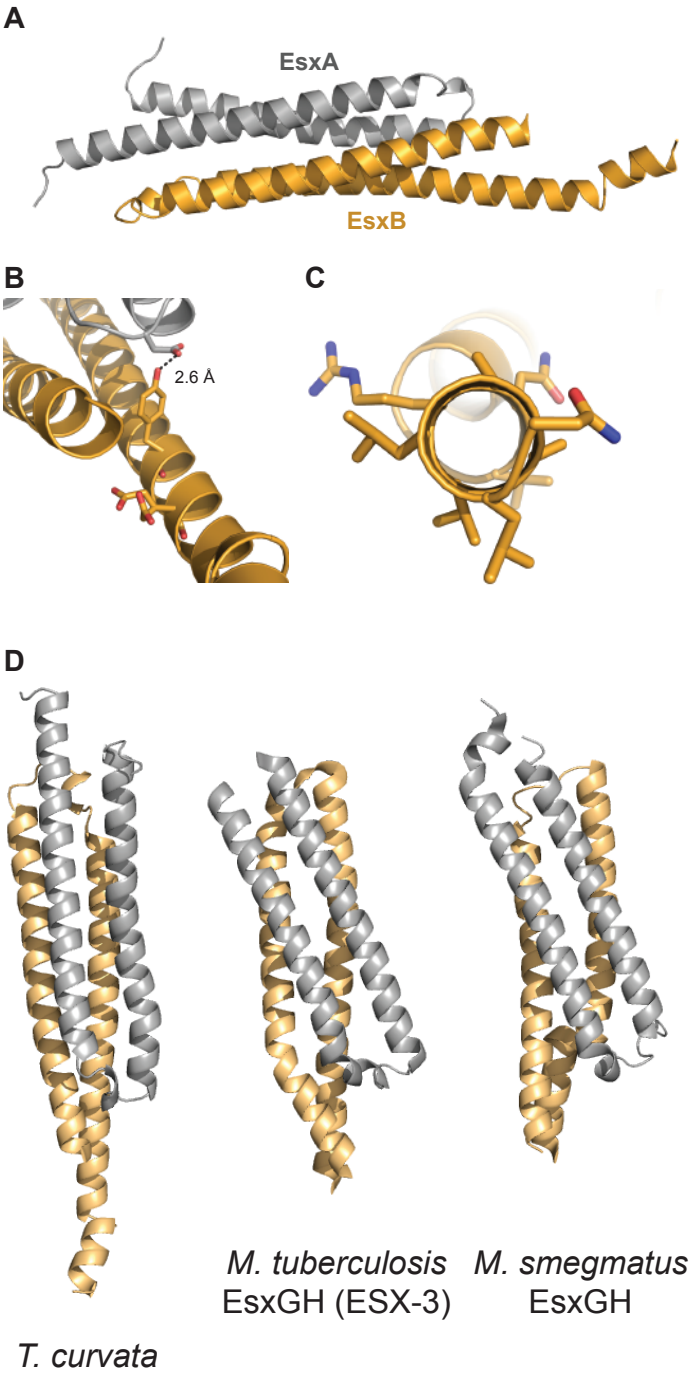

A
EsxA
EsxB
B
2.6 Å
C
D
T. curvata
M. tuberculosis
EsxGH (ESX-3)
M. smegmatus
EsxGH

Figure 2.3: Structure of the EsxAB substrate heterodimer

(A) Structure of the EsxAB heterodimer to 2.0 Å resolution. Importantly, the proteins are alpha-helical, with a small previously unseen C-terminal alpha helix on EsxB.

(B) A close up view of the Y-x-x-x-D/E motif implicated in signal sequence binding [54]. Importantly, the tyrosine residue forms a hydrogen bond with the conserved W-x-G motif of EsxA.

(C) A close up view of the C-terminal residues of EsxB which have been implicated in signal sequence binding [52]. These residues form a separate amphipathic helix in this structure.

(D) Comparison of the *T. curvata* EsxAB homolog with those from the *M. tuberculosis* [59] and *M. smegmatis* [56] ESX-3 system (designated EsxGH in these systems). Note the offset angle of EsxA relative to EsxB.

Figure 2.4

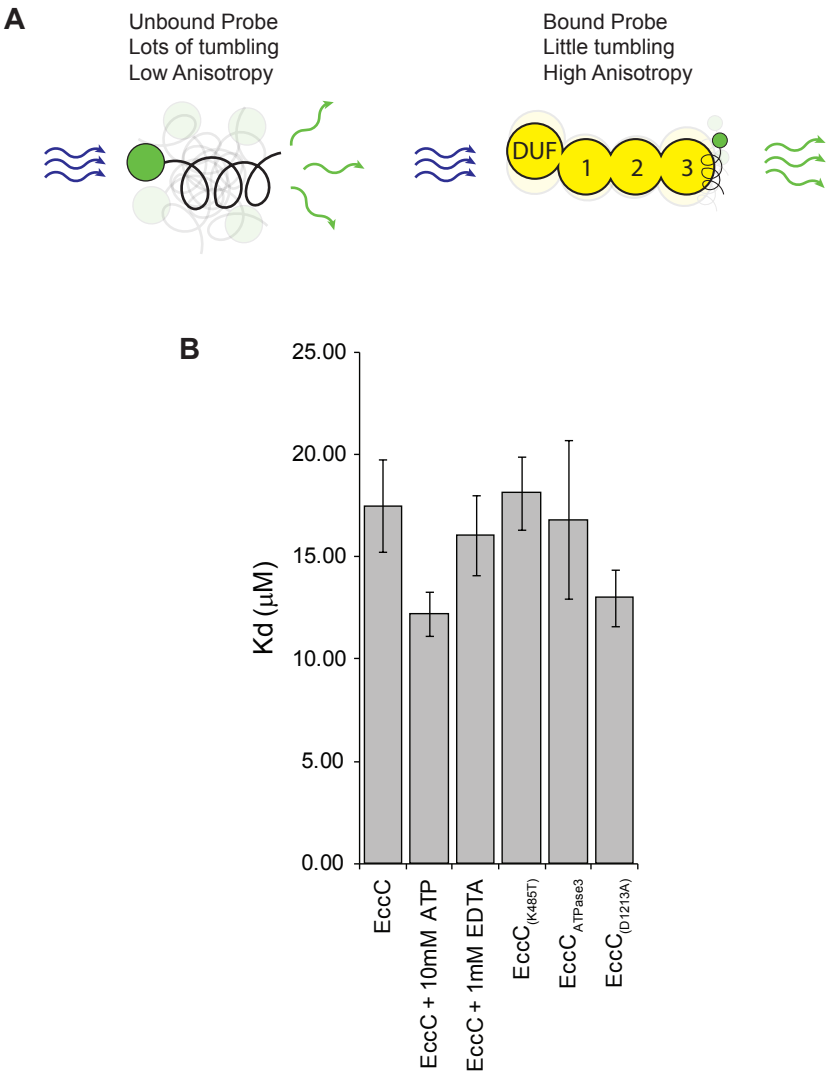

A
Unbound Probe
Lots of tumbling
Low Anisotropy
Bound Probe
Little tumbling
High Anisotropy
DUF
1
2
3
B
Kd (μM)
25.00
20.00
15.00
10.00
5.00
0.00
EccC
EccC + 10mM ATP
EccC + 1mM EDTA
EccC(K485T)
EccCATPase3
EccC(D1213A)

Figure 2.4: The EccC / EsxB interaction is nucleotide-independent

(A) Cartoon depicting the fluorescence anisotropy assay utilized for this experiment. The probe contains the last ten amino acids of EsxB conjugated to a fluorescein at the N-terminus. Binding of the probe to EccC results in a dramatic increase in fluorescence anisotropy, which can be used to calculate binding parameters.

(B) Dissociation constants for the EccC / EsxB-probe interaction under different conditions.

Figure 2.5

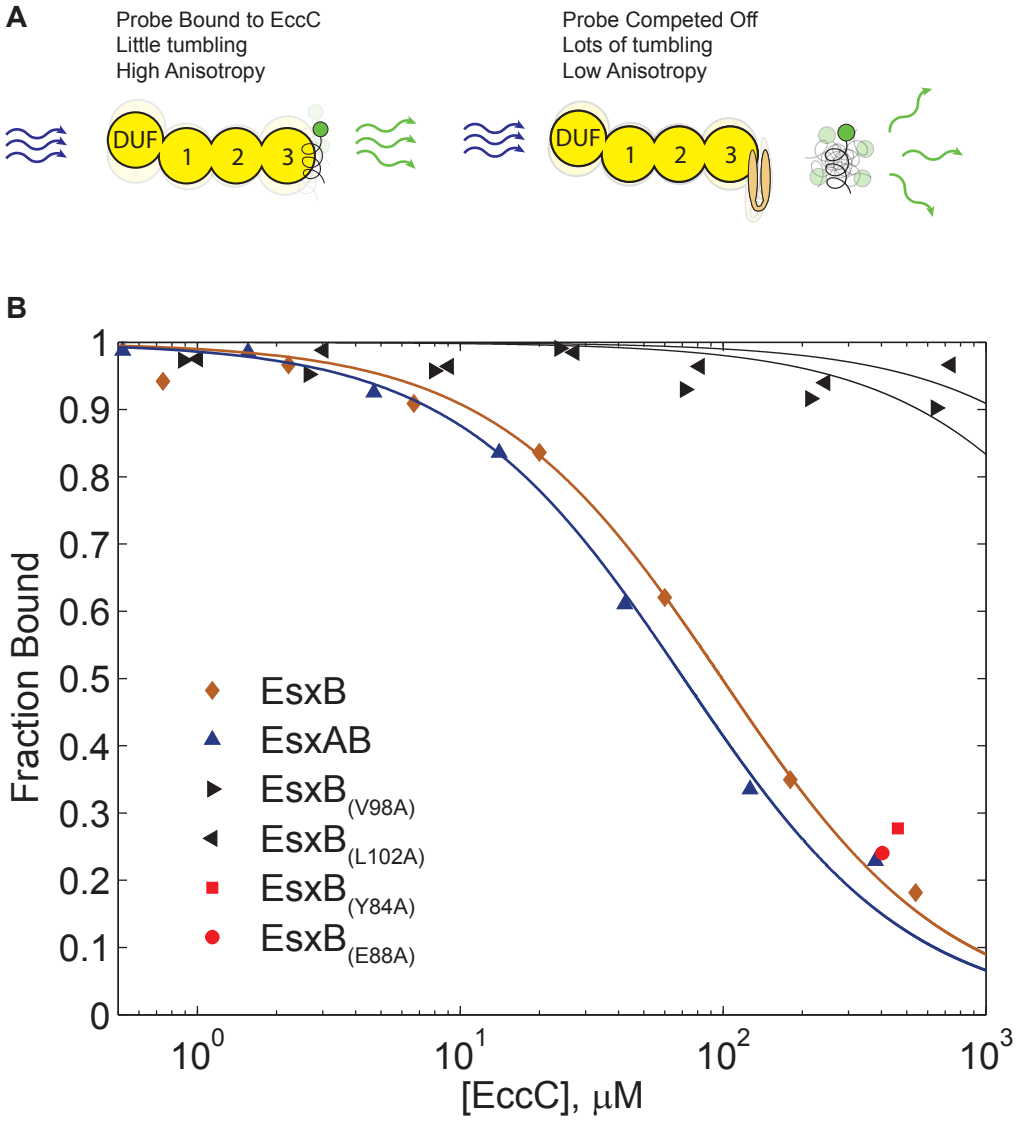

A
Probe Bound to EccC
Little tumbling
High Anisotropy
Probe Competed Off
Lots of tumbling
Low Anisotropy
DUF
1
2
3
DUF
1
2
3
B
Fraction Bound
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
EsxB
EsxAB
EsxB(V98A)
EsxB(L102A)
EsxB(Y84A)
EsxB(E88A)
10^0
10^1
10^2
10^3
[EccC], μM

Figure 2.5: V98 and L102 on EsxB are required for the EccC / EsxB interaction

(A) Cartoon depiction of the fluorescence anisotropy competition assay utilized to compare different EsxB constructs. Briefly, EccC is preloaded with fluorescent probe (same as in Figure 2.4) and full length, unlabeled EsxB is used to compete the probe off. Binding of EsxB thus decreases fluorescence anisotropy.

(B) EsxB, EsxAB, $EsxB_{(Y84A)}$, and $EsxB_{(E88A)}$ can robustly compete off probe, indicating these constructs are able to bind to EccC. $EsxB_{(V98A)}$ and $EsxB_{(L102A)}$ cannot compete off the probe, even at near mM concentrations, indicating these residues are absolutely required for binding to EccC.

Figure 2.6

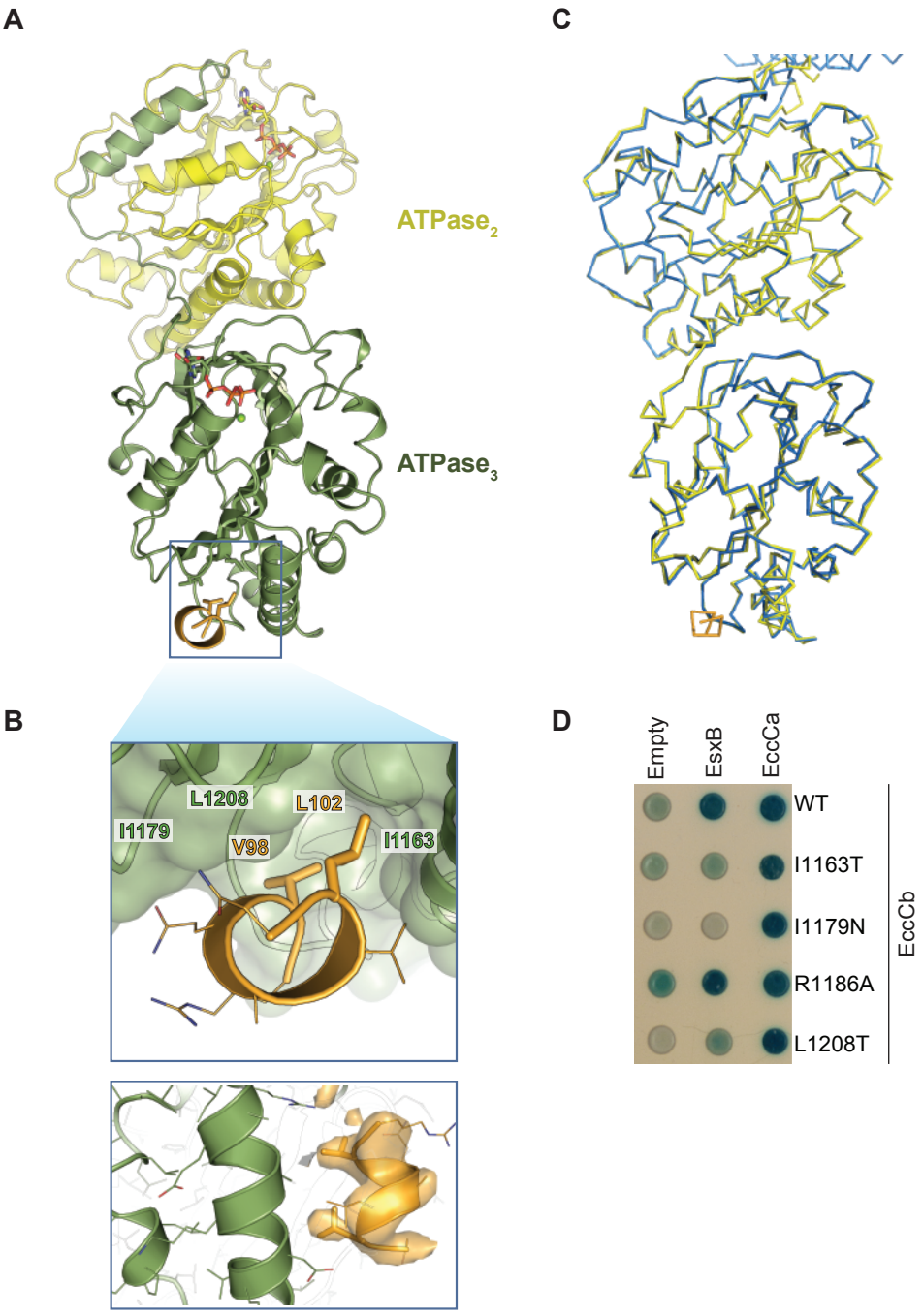

A
ATPase2
ATPase3
B
L1208
L102
I1179
V98
I1163
C
D
Empty
EsxB
EccCa
WT
I1163T
I1179N
R1186A
L1208T
EccCb

Figure 2.6: The structure of EccCb bound to signal sequence

(A) The crystal structure of the EccCb fragment from *T. curvata* bound to a peptide containing the C-terminal 23 amino acids of EsxB (including the Y-x-x-x-D/E motif). The peptide binds to the C-terminal ATPase domain ($ATPase_3$). Importantly, only the C-terminal 7-8 residues of the probe are structured (depending on the chain), suggesting that the C-terminal helix is the site of EccC / EsxB binding. Additionally, ATP is bound to both $ATPase_2$ and $ATPase_3$, suggesting that signal sequence binding does not stimulate ATP hydrolysis in these domains.

(B) A close up view of the binding pocket. The pocket is lined with hydrophobic residues, and V98 and L102 from EsxB mediate the bulk of the interaction. Also shown is the difference density map for the peptide.

(C) Comparison of the conformation of bound and unbound EccC showing that signal sequence binding has no significant effect on the conformation of $ATPase_2$ or $ATPase_3$ (more information on the unbound structure is in Chapter 3).

(D) Mutation of any of the hydrophobic residues lining the binding pocket, but not a conserved arginine, prevent binding of EsxB to EccC by yeast-two-hybrid.

Figure 2.7

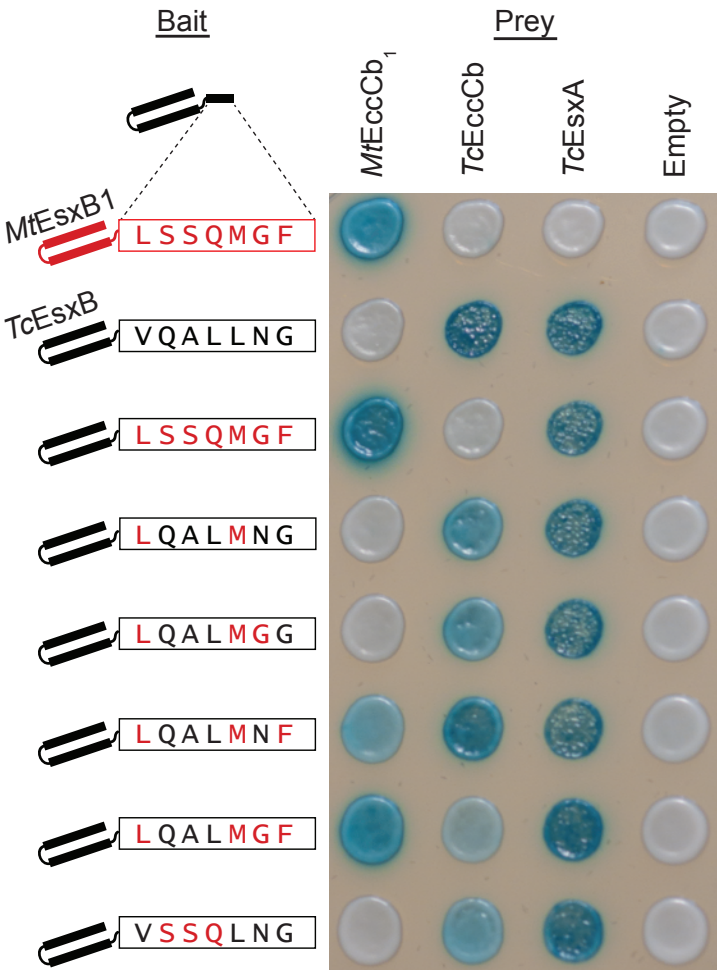

Bait
Prey
MtEsxB1
TcEsxB
LSSQMGF
VQALLNG
LSSQMGF
LQALMNG
LQALMGG
LQALMNF
LQALMGF
VSSQLNG
MtEccCb1
TcEccCb
TcEsxA
Empty

Figure 2.7: The C-terminal seven residues of EsxB encode EccC specificity

Two different EccC / EsxB pairs were tested for their specificity in the yeast-two-hybrid system. Under normal conditions, binding of each substrate is highly specific for its cognate EccC ATPase. Swapping the C-terminal seven amino acids of EsxB between the two paralogs is sufficient to completely switch specificity for EccC, while maintaining wild type specificity for EsxA. Switching subsets of residues within this region have intermediate effects on specificity.

Chapter 3: Activation of EccC by secretory substrates

Abstract

Protein secretion systems must be tightly regulated to avoid wasting energy and/or secretory substrates via futile cycles. To date, no information regarding the mechanisms of regulation of Type VII secretion have been uncovered at the biochemical level. Here, we show that EccC exists in an auto-inhibited form, but that displacement of a conserved linker between $ATPase_1$ and $ATPase_2$ activates the enzyme's ATPase activity and makes it sensitive to the addition of substrates. We further show that the substrate EsxB activates EccC by acting as a dimerization domain -- forming EccC:EsxB:EsxB:EccC complexes, which then further assemble into active, higher order multimers. Additionally, EsxA, another substrate of T7S, can compete with EsxB on the dimerization surface, breaking apart the complex and forming EccC:EsxB:EsxA ternary complexes, which are inactive. Lastly, we show that EccC relies on an arginine finger mechanism both *in vivo* and *in vitro*, suggesting that multimerization is a physiologically relevant aspect of the enzyme's function. These data suggest a novel role for EsxB as not only an important secretory substrate, but also an integral component of the secretion apparatus itself -- explaining the co-dependent secretion observed with T7S.

Introduction

The P-loop NTPase family

EccC belongs to a super-family of proteins called P-loop NTPases -- specifically within the Additional Strand Catalytic Glutamate (ASCE) family, with strong homology to FtsK/SpoIIIE [55]. NTPases utilize energy released from the hydrolysis of the high energy phosphodiester bond between the β and γ phosphates of NTPs to perform biologically important tasks, such as exerting work on substrates. P-loop NTPases vary greatly in sequence, structure, and function, but are characterized generally by the presence of two conserved motifs. The first motif, known as the Walker A motif, is located on the phosphate-binding loop (P-loop) and interacts with the β and γ phosphates of the NTP. Typically this motif consists of GXXXXGK(T/S), with the lysine invariant [61]. Mutation of the lysine to a polar residue (such as threonine) results in loss of NTP binding. The second motif, the Walker B, consists of hhhh(D/E)E (h represents a hydrophobic residue) [61]. The first acidic residue is present in all P-loop NTPases and coordinates a Mg^{2+} ion that is involved with NTP hydrolysis. The second acidic residue coordinates a water molecule for nucleophilic attack on the γ phosphate of the NTP (Figure 3.1) [62]. Mutation of the Walker B residues results in loss of catalysis, but not binding, of the NTP.

Some multimeric ATPases and GTPases also contain one or more conserved arginine residues that extend from a neighboring subunit into the active site near the γ phosphate [62]. The charges on the arginine, typically referred to as an “arginine finger” stabilize the transition state of ATP or GTP during hydrolysis -- dramatically increasing

the rate of catalysis [62]. It is by this mechanism that many GTPase Activating Proteins (GAPs) activate small GTPases. Mutations in the arginine finger typically phenocopy mutations in the Walker A motif with respect to enzymatic activity; however, it has been shown that the arginine is not necessary for nucleotide binding [62].

Regulation and structure of ASCE NTPases

Regulation of NTPases is critical to avoid futile loss of energy and substrates. Coupled with the structural and functional diversity of ASCE NTPases is a wide array of regulatory mechanisms that have so far been uncovered. Often these involve N- or C-terminal accessory domains. In SpoIIIE, for example, the γ-domain inhibits the ATPase activity of the motor domain in the absence of substrate (DNA), but activates it when bound to a specific DNA sequence [63]. Substrate binding can also directly activate or inhibit NTPase activity. This often works through a "glutamate switch" mechanism in which a nearby residue (usually an asparagine) forms a hydrogen bond with the catalytic glutamate in the Walker B motif, holding it in a position not competent for NTP hydrolysis [64]. Upon substrate binding, the glutamate is released and can catalyze hydrolysis.

Typically NTPase motors are only functional as part of a multimeric assembly, often a ring in which the substrate is passed through the central pore [62]. In many of these cases, such as with FtsK, the protein complex must multimerize around the substrate first before becoming active [65]. This makes sense in the context of DNA motors, as from the point of view of the motor the substrate is infinitely long and thus the only way to thread the substrate through the pore is to form around it. The biochemical

composition of these multimers can vary greatly from one system to the next -- ranging from homo-hexamers (eg: Ftsk) to hexameric assemblies containing six different proteins (eg: Mcm2-7; the minichromosome maintenance complex [66]). Aside from hexamers, which are the most common arrangement, tetramers (eg: MgsA [67]), pentamers (eg: the clamp loader [68]) and heptamers (eg: NtrC1 [69]) are also observed occasionally.

Structure and mechanism of EccC

Prior to the current study there was no information on the structure and/or mechanism of EccC or EssC. Similarity to SpoIIIE/FtsK has led most to speculate that the protein forms a hexameric assembly of ATPase domains. However, due to the unique nature of EccC, this could mean either two molecules forming a single layered hetero-hexameric ring, or six molecules forming a three layered homo-hexameric ring. Although currently there are no other known enzymes containing three ATPase domains to help shed light on EccC's potential structure, both modes of assembly have been observed with other proteins. The microtubule-binding motor protein dynein contains six ATPase domains (two with inactivating mutations in their Walker A motifs) on a single polypeptide chain [70]. These domains arrange into a single asymmetric ring-shaped assembly. Conversely, ClpB, which contains two tandem ATPase domains, forms a two-layered hexameric complex [71]. In the case of the ESX-1 homolog of EccC, the split nature of the protein could yield further, more complicated arrangements as well. In both cases, there is some division of labor amongst the domains -- with some acting as regulators, providing structural support, or assisting assembly.

Results

EccC is held in an autoinhibited state

To begin to understand the molecular mechanisms of EccC regulation, we established an *in vitro* assay to measure ATPase activity in the presence or absence of EsxB substrate. We found that neither the cytoplasmic domain of EccC (referred hereafter as just "EccC") nor the EccCb construct possessed any measureable ATPase activity even in the presence of up to 10 μM EsxB (Figure 3.2A-B). Thus, we wondered if the ATP-binding (Walker A) and ATP-hydrolysis (Walker B) catalytic residues were conserved in EccC, or if the protein had diverged away from an ATPase role to fulfill some other function for the translocase. Alignment of 142 unique EccC homologs revealed that $ATPase_1$ has highly conserved Walker A and Walker B motifs, indicating an evolutionary pressure to maintain FtsK-like activity in that domain (Figure 3.2C). In contrast, $ATPase_2$ and $ATPase_3$, while maintaining a high degree of conservation of the Walker A motif, both displayed divergent Walker B residues. Specifically, the Walker B catalytic glutamate is completely divergent in $ATPase_2$, and is predominantly an aspartate in $ATPase_3$. These changes are predicted to eliminate or vastly reduce ATP hydrolysis in both domains [62], consistent with the presence of ATP in both active sites in the substrate-bound structure (Figure 2.6A). Because of the high degree of conservation in $ATPase_1$ compared to the other two ATPase domains, we hypothesized that $ATPase_1$ likely represents the active motor domain for the EccC protein, while $ATPase_2$ and $ATPase_3$ likely diverged to fulfill other roles (such as regulatory and substrate-binding roles).

To better understand how $ATPase_1$ differed from $ATPase_2$ and $ATPase_3$, we solved the structure of the full cytoplasmic domain of EccC (Figure 3.3A, crystallographic statistics shown on Table 3). The structure reveals the three ATPase domains are held rigidly together in direct translation (note: despite being present in the crystallization construct, the DUF was disordered in the crystal structure). Each domain was connected to the next via similar linker / pocket interactions. The structure of $ATPase_2$ and $ATPase_3$ from the *Geobacillus thermodinetrificans* EssC homolog shows a similar structure (Figure 3.3B), suggesting these features are likely conserved in firmicutes.

Despite the ESX-1 EccC homolog being split into two polypeptides, it too relied on a homologous linker emanating from $ATPase_2$ to form the complete EccCab complex (Figure 3.4). Interestingly, in contrast to $ATPase_2$ and $ATPase_3$, which were bound to ATP as in the signal sequence-bound structure, $ATPase_1$ was bound to a sulfate ion despite a high concentration of ATP in the crystallization condition. While it is possible that $ATPase_1$ could have rapidly hydrolyzed the ATP that binds to that domain, the ATPase assays we conducted would have been able to detect that activity. Thus, we hypothesized that despite the high degree of conservation in the Walker A and Walker B motifs, $ATPase_1$ was unable to bind nucleotide.

To better understand why $ATPase_1$ was unbound to ATP, we examined the active site of $ATPase_1$ in greater detail. We found that the catalytic residues in this domain were held in a strikingly different conformation than the other two ATPase domains (Figure 3.5). Specifically, the Walker A lysine was in an unfavorable rotamer and making a hydrogen bond with the Walker B aspartate, preventing the binding of

magnesium and water which are required for coordinating and hydrolyzing ATP. This conformation was very reminiscent of the ATP "empty" (β_E) subunit of the F_1-ATPase (Figure 3.5) which is known to have a very low affinity for nucleotide [72].

We next tested whether ATP binding in $ATPase_1$ was required for secretion *in vivo*. If binding to nucleotide was required for function of the protein in the context of the physiologic translocase, then we could conclude that EccC was being held in an inhibited state *in vitro*, which must somehow be overcome during the catalytic cycle. To answer this question, we created an eccC knockout strain of *M. tuberculosis* and complemented the knockout with either wildtype eccC or eccC in which different Walker A lysines had been mutated to threonine, a mutation that is known to prevent ATP binding [62]. As expected, the knockout strain was completely deficient for secretion of EsxB (Figure 3.6), and complementation with the wild type protein on an integrating plasmid using the native promoter restored secretion of EsxB to normal levels. However, mutation of any of the three Walker A lysines to threoine completely blocked secretion, suggesting that each domain is necessary for secretion, and that the domains do not act independently. Furthermore, this result suggests that $ATPase_1$ must switch into an ATP binding competent state during the catalytic cycle, and that EccC is otherwise held in an inhibited conformation.

Release of autoinhibition can be achieved by disrupting the linker$_2$ – pocket$_1$ interaction

Since ATPases often utilize N- and C-terminal appendages to regulate ATPase activity [63], [73], we hypothesized that $ATPase_1$ was being inhibited by its close association

with $ATPase_2$. Furthermore, in ESX-1 EccC is split into two polypeptides between $ATPase_1$ and $ATPase_2$, further suggesting this could be a site of regulation.

Unfortunately, $ATPase_1$ of EccC was not stable in isolation. However, we noticed that much of the interface between $ATPase_1$ and $ATPase_2$ is mediated by a highly conserved arginine in $pocket_1$, R543, which interacts with the conserved residues W762 and L763 on $linker_2$ (Figure 3.7A-B). We figured that disruption of this interaction may emulate binding of an allosteric regulator at this site and activate activity. Indeed, mutation of R543 to alanine resulted in the appearance of measurable ATPase activity (Figure 3.7C). Further mutation of the Walker B catalytic residue in $ATPase_1$, E592Q, completely prevented the activation, suggesting the activation is indeed dependent on $ATPase_1$ (Figure 3.7C) and not due to purification of a contaminating ATPase. Supporting the hypothesis that $ATPase_1$ represents the active motor domain for EccC, mutation of an equivalent residue between $ATPase_2$ and $ATPase_3$ had little effect on activity (Figure 3.7C).

We hypothesized that circumventing the regulatory mechanisms involved in EccC autoinhibition represents a dysregulation which could be expected to block secretion *in vivo*. Indeed, mutation of R543 to alanine in the *M. tuberculosis* $eccCa_1$ gene resulted in a substantial, but not total, loss of EsxB secretion, confirming the physiologic importance of this regulatory residue (Figure 3.7D-E).

We next wished to determine whether or not the linker-displaced (referred to as “activated”) EccC was sensitive to the addition of EsxB *in vitro*. Indeed, in the context of the R543A mutation, addition of EsxB increased the catalytic rate approximately five-

fold (Figure 3.8A), with a Hill coefficient of 1.836 ± 0.738 -- suggesting cooperativity. Addition of $EsxB_{(V98A)}$, which cannot bind EccC due to a mutation in the signal sequence, did not result in an increase in catalysis -- revealing that binding to EccC is a prerequisite to further activation of the ATPase activity (Figure 3.8B).

Interestingly, mutation of the Walker B catalytic glutamate in $ATPase_1$ completely abrogated ATPase activity, while mutation of catalytic residues in either $ATPase_2$ or $ATPase_3$ only reduced activity by half (Figure 3.8A). The combinatorial mutation in both $ATPase_2$ and $ATPase_3$ resulted in a further reduction in activity by half. These data further suggest that the primary role of $ATPase_2$ and $ATPase_3$ is regulatory, and that $ATPase_1$ is likely the primary source of ATP hydrolysis.

Importantly, as opposed to mutation of the signal sequence residue V98 to alanine, which prevents binding to and activation of EccC, mutation of residues within the Y-X-X-X-D/E motif on EsxB have no effect on substrate-based activation (Figure 3.8C).

Close examination of the linker-pocket interactions revealed them to be very similar to the interactions between the EsxB signal sequence and the binding pocket identified on $ATPase_3$ (Figure 3.9). It is intriguing to speculate that *in vivo*, other substrates or protein components bind to these pockets and displace the linkers, functionally emulating the effects of the R543A mutation.

Substrates control EccC activity by controlling multimerization

Our next objective was to determine the mechanisms by which EsxB activates EccC. Some ATPases, including FtsK, only assemble into multimers and activate in the

presence of their substrate [65]. We tested whether EccC also forms multimers in the presence of EsxB. To probe the multimeric state of EccC and of the active state of EccC, we took multiple approaches. First, we examined the hydrodynamic radius of EccC with or without bound EsxB, as well as a fusion protein (to guarantee a 1:1 molar ratio of EccC to EsxB). Indeed, addition or fusion of EsxB to EccC resulted in a dramatic shift in the elution profile indicative of multimer formation (Figure 3.10A).

Next, we tested the catalytic rate of EccC as a function of ATPase concentration. Monomeric enzymes will catalyze the same number of reactions per enzyme molecule per unit time, regardless of their concentration. However, if an enzyme must multimerize to become active, then it will be more active at higher concentrations as multimer formation is promoted. We found that without substrate, the activated EccC does not exhibit concentration-dependent activity, suggesting that the R543A mutation simply permits a basal level of ATPase activity (Figure 3.10B). However, pre-incubation with EsxB or fusion of EccC to EsxB reveals strong concentration-dependence, which is abrogated via mutation of the Walker B glutamate of $ATPase_1$. These data were highly suggestive that the EsxB-induced activity relied on multimerization.

We next attempted to capture higher order multimers via glutaraldehyde crosslinking. When we added EsxB to either active or wild type EccC, we observed the appearance of a very high molecular weight band (labeled “multimer”) (Figure 3.11A-C). While the band was weakly present in the absence of EsxB, it did not increase with addition of $EsxB_{(V98A)}$, suggesting that binding of EsxB is necessary to greatly stabilize this oligomeric state (Figure 3.11D-E). Densitometry revealed that the concentration of the multimer band mirrored the ATPase activity measured under the same conditions

with the activated EccC. While correlative, this is highly suggestive that the active form of EccC is represented by the high molecular weight multimer captured with our crosslinking assay. However, because the presence of the multimer band did not differ significantly when comparing activated (R543A) and wild type EccC, we conclude that multimerization is a necessary, but not sufficient aspect of the catalytic cycle. Thus, *in vitro* EccC likely adopts a multimeric conformation at a low rate, but is unable to hydrolyze ATP until an additional conformational change (displacement of $linker_2$) occurs.

Lastly, we tested the reliance of EccC on an arginine finger mechanism for activity *in vitro* and secretion *in vivo*. FtsK and other AAA+ ATPases utilize an arginine finger mechanism for catalysis of ATP. The arginine finger is a catalytically-necessary residue (typically an arginine) that extends from one subunit into the active site of an adjacent subunit and, along with the Walker A and B motifs, helps coordinate ATP for binding and hydrolysis. Importantly, this residue only holds significance in the context of an ATPase that forms larger multimers. Alignment of EccC homologs with FtsK revealed that $ATPase_1$ contains a completely conserved arginine finger (Figure 3.12A). Supporting this finding, structural alignment showed that R616 was positioned identically with a known arginine finger from TrwB (another FtsK-like ATPase that is found in Type IV secretion) (Figure 3.12C).

Supporting a functional role for this residue, mutation of the *T. curvata* $ATPase_1$ arginine finger to glutamine (R616Q) in the context of the activated EccC-EsxB fusion drastically reduced ATPase activity *in vitro* (Figure 3.12B,D), but had no effect on multimerization of the fusion protein (Figure 3.14G-H). Importantly, R616 exists on a

loop on the surface of $ATPase_1$ and is thus likely not required for structural integrity of the protein. Interestingly, mutation of the arginine finger reduced ATPase activity to the basal levels seen without substrate. Thus we conclude that the substrate-induced activation of EccC works through a mechanism involving multimerization and utilization of an arginine finger. These data are also consistent with the concentration-independence seen with the activated EccC sans substrate.

Interestingly, equivalent arginine residues are absent in $ATPase_2$ and $ATPase_3$ suggesting that this mode of multimerization is either mediated through a different mechanism that evolved separately from the FtsK arginine finger, or that multimerization is not part of the reaction mechanism of $ATPase_2$ and $ATPase_3$. The lack of conservation in the catalytic residues, including the arginine finger, suggests positive selective pressure on $ATPase_1$ to maintain an FtsK-like function, while this function diverged in $ATPase_2$ and $ATPase_3$. Additionally, mutation of the predicted arginine finger in *M. tuberculosis* to alanine completely abrogated secretion of EsxB *in vivo*, suggesting utilization of the arginine finger is a general feature of all EccC homologs (Figure 3.12E-F).

We next wondered how EsxB induces multimerization. The simplest hypothesis is that EsxB acts directly as a multimerization domain, forming homo-oligomers and presenting multiple signal sequences. To begin to test this, we measured the molecular weight of EsxB in solution using analytical ultracentrifugation. We found that EsxB predominantly forms homodimers in solution (Figure 3.13A). Fusion of EsxB to EccC also resulted in a construct that was dimeric, further supporting EsxB's role as a dimerization domain (Figure 3.13B).

Altogether, based on these data we speculate that EsxB acts as a dimerization domain to form EccC:EsxB:EsxB:EccC complexes, which can then further associate into higher order, active multimers (in a similar system, the ATPase ClpX contains a dimerization domain which has been shown to stabilize the formation of the active hexamer [74]). The formation of higher order multimers appears to occur at low frequency *in vitro*, but is likely further stabilized by interactions with other proteins in the context of the *in vivo* translocase. An additional substrate or other protein component then binds to the pocket on $ATPase_1$ and displaces the linker emanating from $ATPase_2$, activating the complex by releasing the autoinhibition on $ATPase_1$. The now active EccC complex utilizes an arginine finger mechanism to hydrolyze ATP and presumably perform work on substrates.

EsxA disrupts multimerization and activity

Because the EsxAB hetero-dimer appears more stable than the EsxB homo-dimer, we wondered if EsxA, which also forms a homo-dimer in isolation (Figure 3.14A), could compete with EsxB to break apart the EccC-EsxB multimers. Indeed, addition of EsxA resulted in a drastic decrease in protein complex size as determined by gel filtration and analytical ultracentrifugation (Figure 3.14B-D). Molecular weight determination by AUC revealed that EsxA reduces complexes from primarily dimers and trimers in solution to almost exclusively monomers (Figure 3.14D). During crosslinking experiments with an activated EccC-EsxB fusion, EsxA addition drastically and cooperatively decreased the presence of the high molecular weight multimer band (Figure 3.14E-F). This decrease in multimer, as determined by gel densitometry, was mirrored almost exactly by a concomitant decrease in ATPase activity--again suggesting the catalytically active form

of EccC is multimeric. Importantly, mutation of the arginine finger residue did not affect formation of the multimer (Figure 3.14G-H).

Furthermore, the Hill coefficient for EsxA inhibition was calculated to be 2.272 ± 0.726. In other words, one EsxA molecule inhibits *at least* three active EccC enzymes. This provides further evidence that the active form of EccC is a higher order multimer.

We were interested to see if other proteins that bind to the EsxB dimerization surface could also inhibit EccC activity in the same way. To test this, we titrated $EsxB_{(V98A)}$ in with activated EccC-EsxB fusion. $EsxB_{(V98A)}$ is identical to EsxB except that it cannot bind to EccC, even at high concentrations. Thus, the interface between EsxB and the binding-incompetent form of EsxB should be the same as the homo-dimerization surface of wild type EsxB. As expected, this protein also led to a decrease in activity (Figure 3.15). However, unlike EsxA, the inhibition from $EsxB_{(V98A)}$ was non-cooperative (Hill coefficient = 1.015 ± 0.274) (Figure 3.15). Even if we expect $EsxB_{(V98A)}$ to bind less tightly to EsxB than would EsxA, the shape of the binding curve, determined by the Hill coefficient, should not change. Thus, we can conclude that either EsxB homo-dimerization is fundamentally different than EsxAB hetero-dimerization (such as parallel versus anti-parallel binding), or EsxA binding elicits other conformational changes that facilitates a cooperative disassembly of active EccC complex.

Conclusions and model for EccC regulation by secretory substrates

In this work we've sought to gain insight into EccC and, more generally, T7S through genetic, biochemical, and structural approaches. The unique structure of EccC, specifically the presence of three ATPase domains held in a tandem array, combined

with genetic data showing the necessity of nucleotide binding at each domain, suggests a division of labor mechanism with separate roles for each ATPase domain. As we've shown, ATPase$_1$ differs greatly from ATPase$_2$ and ATPase$_3$, and shares a much higher degree of homology with FtsK, including conservation of the catalytic Walker B residues and the arginine finger. Additionally, we've shown that mutation of catalytic residues in ATPase$_1$, but not ATPases$_{2\text{-}3}$, completely abrogate ATPase activity *in vitro*. For these reasons, we believe that ATPase$_1$ serves the primary role of exerting force on T7S substrates, while ATPases$_{2\text{-}3}$ serve other regulatory roles. This hypothesis is further supported by our data showing that disruption of a residue involved in the interaction between ATPase$_1$ and ATPase$_2$ releases an auto-inhibitory mechanism, activating the ATPase. In addition to regulating activity of ATPase$_1$, we also showed (Chapter 2) that ATPase$_3$ serves as the primary site for substrate selection and binding. Thus, while much remains unknown, we have begun to piece apart specific roles for each domain.

Recently, a similar study focusing on the roles of ATP-binding at each ATPase domain was conducted in the *yuk* system in *Bacillus subtilis*, a firmicute [75]. Interestingly, in this T7S system only ATP-binding in ATPase$_1$ of YukBA (the EssC homolog) is required for secretion of the substrate YukE (EsxB homolog). This data supports our notion that ATPase$_1$ is the functional motor domain while ATPases$_2$ and ATPase$_3$ have diverged to fulfill other roles. It is unknown whether the difference in the requirement for ATP-binding in ATPases$_{2\text{-}3}$ in YukBA versus EccC$_1$ stems from fundamental differences between T7S in the firmicutes and actinobacteria, or results from the split nature of EccC in the ESX-1 system.

Our work has also uncovered an important role for EsxB as an integral part of the secretion apparatus. Because EsxB is necessary for multimerization and activation of EccC, it could help explain the co-dependent nature of substrate secretion unique to T7S systems. The fact that the EsxAB heterodimer is much more stable, however, raises the important question of whether or not EsxA and EsxB exist *in vivo* in forms other than the EsxAB heterodimer. Currently there is no data that addresses this question.

However, the *yuk* system, as well as systems in other firmicutes, lack an EsxA homolog and encode only EsxB, suggesting a physiologic role for EsxB homo-dimers [76], [77]. Additionally, it has been shown using specific *in vivo* cross-linking in *B. subtilis* that YukE (EsxB) is recognized and secreted as a homo-dimer [78]. Further, in agreement with our *in vitro* inhibition experiments using EsxA and $EsxB_{(V98A)}$, introduction of exogenous YukE in which the C-terminal seven amino acids are removed exhibits a dominant-negative effect on secretion *in vivo* [78]. These data, from a firmicute system, fully support our model for EccC activation and validate some of our results *in vivo* -- suggesting it may be general for all T7S systems.

The exact structure of the fully activated and multimeric EccC is still unknown. However, our findings suggest it is unlikely that EccC forms an assembly of six ATPase domains from two EccC molecules. First, the ATPase domains of EccC are held rigidly together, and it would necessitate a large conformational change in the protein to position them such that the arginine finger of $ATPase_1$ would fit into the active site of an adjacent protomer. Second, the activity of EccC correlated with the presence of a much

higher molecular weight band than would be predicted for a dimer. Finally, cross-linking and AUC reveal several higher order multimers that form in solution.

While EccC shares a great amount of homology with FtsK/SpoIIIE, there are also several structural differences (eg: several additional helices and sheets). Interestingly, however, the R543 residue on EccC lies within a helix (often designated α2) that is typically associated with detecting the binding of substrates. Additionally, the pocket in which R543 resides, as well as the equivalent pocket on $ATPase_2$, shares structural homology with the signal sequence binding pocket on $ATPase_3$ (Figure 3.9). Thus, it is intriguing to speculate that additional proteins may bind to these pockets, allowing for coordinated secretion of multiple T7S substrates.

Overall, our results allow us to posit a model in which secretory substrates play an important role in T7S regulation (Figure 3.16). In the absence of substrate, EccC is monomeric and tightly inactivated via interactions between $ATPase_1$ and $ATPase_2$. Binding of the EsxB substrate to $ATPase_3$ drives EccC multimerization, but is not sufficient for activation. Allosteric interactions through displacement of $linker_2$ from its binding groove in $ATPase_1$, which relieves the inhibitory interaction with $ATPase_2$, are also required to permit full activation of EccC ATPase activity. While we do not yet know the nature of this second activation signal, we suspect that other substrates and/or Type VII components bind to the $ATPase_1$ pocket, creating an “AND” logic gate by which secretion of multiple substrates are coordinated, a well-established phenomenon unique to Type VII secretion. Alternatively, tight control of EccC may provide a mechanism to prevent futile ATPase cycles and directly control the levels of secretion substrates in response to multiple regulatory signals. Our model also

suggests that EsxB homodimers, in addition to EsxAB heterodimers, play an important role in Type VII secretion, a notion supported by the observation that ancestral T7-like systems, such as those found in the phylum firmicutes, lack EsxA homologs and EsxB exists solely as a homo-dimer. Moreover, this model suggests that the two most prominent T7 substrates, EsxA and EsxB, play a direct role in regulating secretion of all T7 substrates via their stimulatory and inhibitory effects on EccC multimerization and activation, although experimental validation of this hypothesis will require further examination of the interaction between EsxA, EsxB and EccC in the living cell.

Figure 3.1

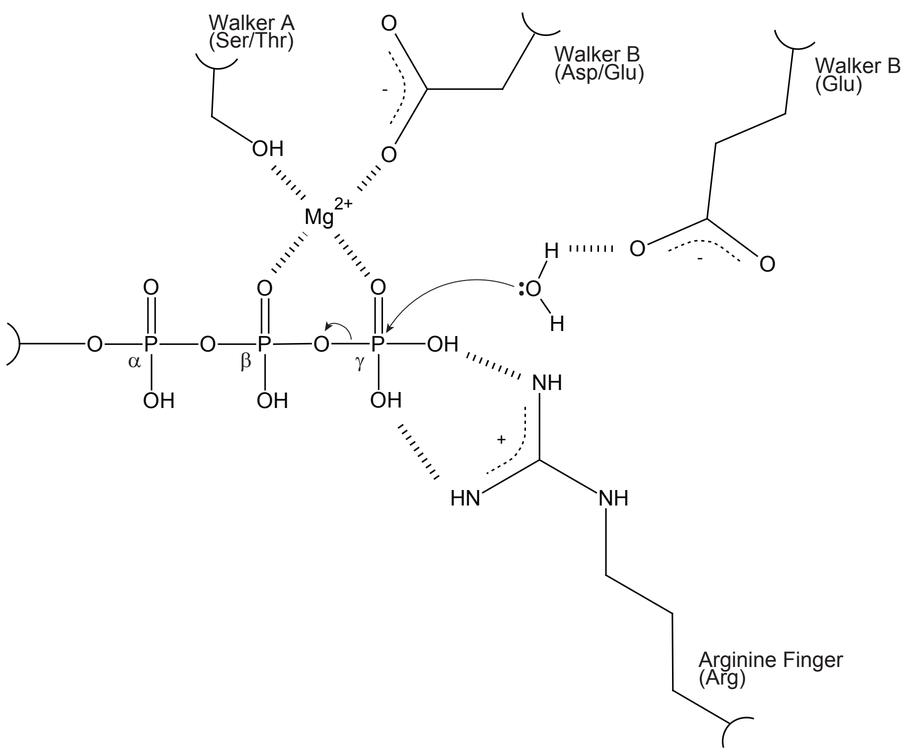

Walker A
(Ser/Thr)
Walker B
(Asp/Glu)
Walker B
(Glu)
OH
O
O
Mg2+
H
O
O
:O
H
O
O
O
O
P
O
P
O
P
OH
α
β
γ
OH
OH
OH
NH
+
HN
NH
Arginine Finger
(Arg)

Figure 3.1: Schematic of some NTPase active site residues coordinating hydrolysis of NTP

The active site of NTPases are typically composed of two motifs: the Walker A motif (consisting of GXXXXGK(T/S), required for NTP binding) and the Walker B motif (consisting of hhhh(D/E)E, required for NTP hydrolysis). Magnesium is a required cofactor necessary for coordination of the β and γ phosphates, and is itself coordinated by the Walker A serine/ threonine (serine shown) and the first Walker B acidic residue (aspartate shown). The second Walker B acidic residue, glutamate, primes a water molecule for nucleophilic attack on the γ phosphate, which leaves NDP and Pi after the reaction takes place. An additional residue, the arginine finger, extends from an adjacent subunit into the active site and helps coordinate the NTP. Not shown is the invariant Walker A lysine residue, which also interacts with the β and γ phosphates.

Figure 3.2

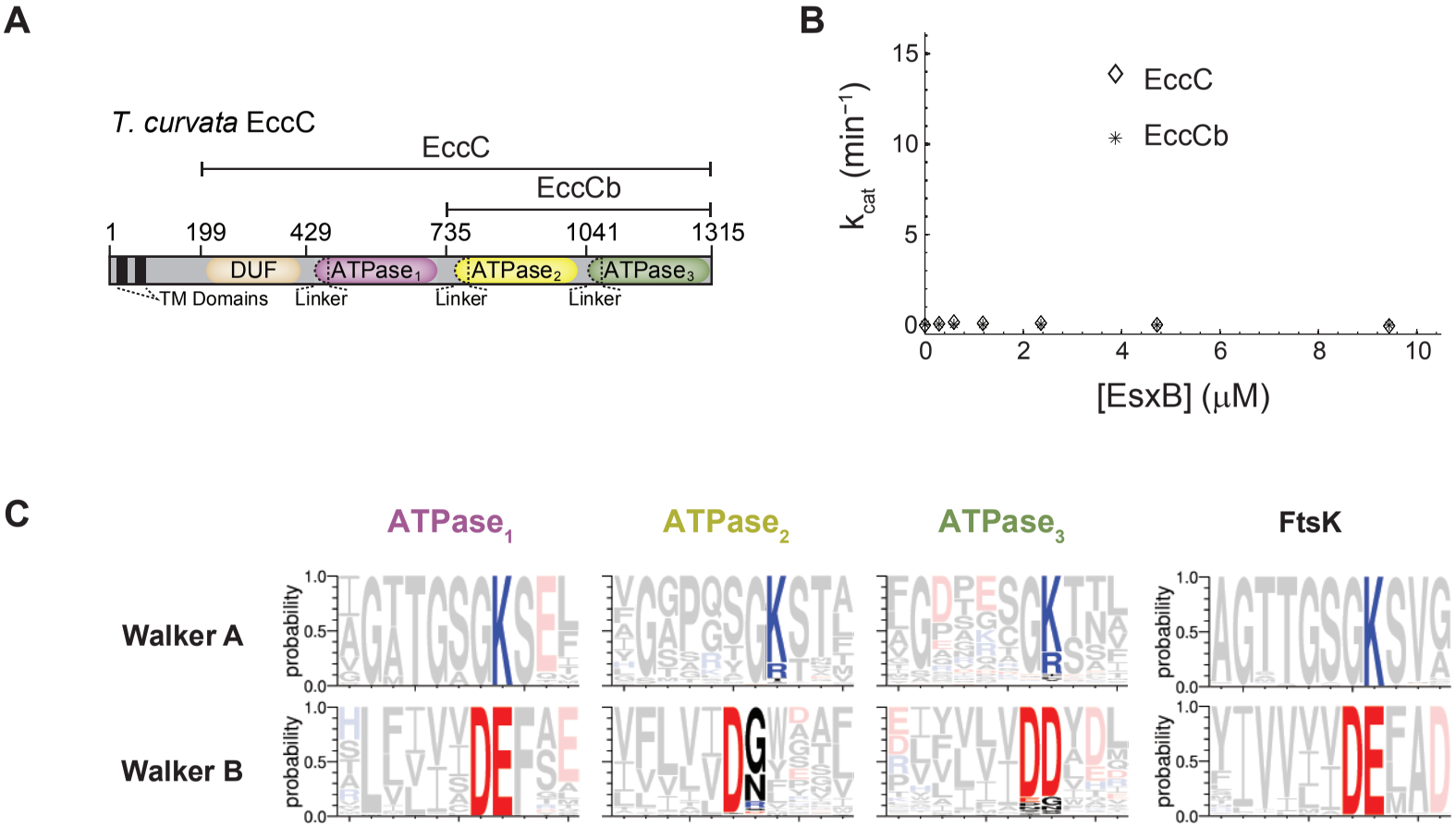

A
T. curvata EccC
EccC
EccCb
1
199
429
735
1041
1315
DUF
ATPase1
ATPase2
ATPase3
TM Domains
Linker
Linker
Linker
B
kcat (min−1)
15
10
5
0
EccC
EccCb
0
2
4
6
8
10
[EsxB] (μM)
C
ATPase1
ATPase2
ATPase3
FtsK
Walker A
Walker B
probability
1.0
0.5
0.0

Figure 3.2: EccC does not hydrolyze ATP *in vitro*

(A) Domain schematic of the *T. curvata* EccC protein with markers indicating the residues utilized in two protein constructs. The first, simply labeled "EccC" contains the cytoplasmic domain of the protein, including the DUF and all three ATPase domains. The second, "EccCb," contains $ATPase_2$ and $ATPase_3$, and was generated by aligning the *T. curvata* EccC with $EccCb_1$ from the *M. tuberculosis* ESX-1 system.

(B) Neither EccC nor EccCb hydrolyze ATP, even with increasing amounts of EsxB substrate.

(C) Conservation of the Walker A and Walker B motifs in each ATPase domain compared to FtsK. The Walker B catalytic glutamate is highly divergent in $ATPase_2$, and has been changed to an aspartate in $ATPase_3$, while these residues are highly conserved in $ATPase_1$. The Walker A lysine is well conserved in all three domains.

Figure 3.3

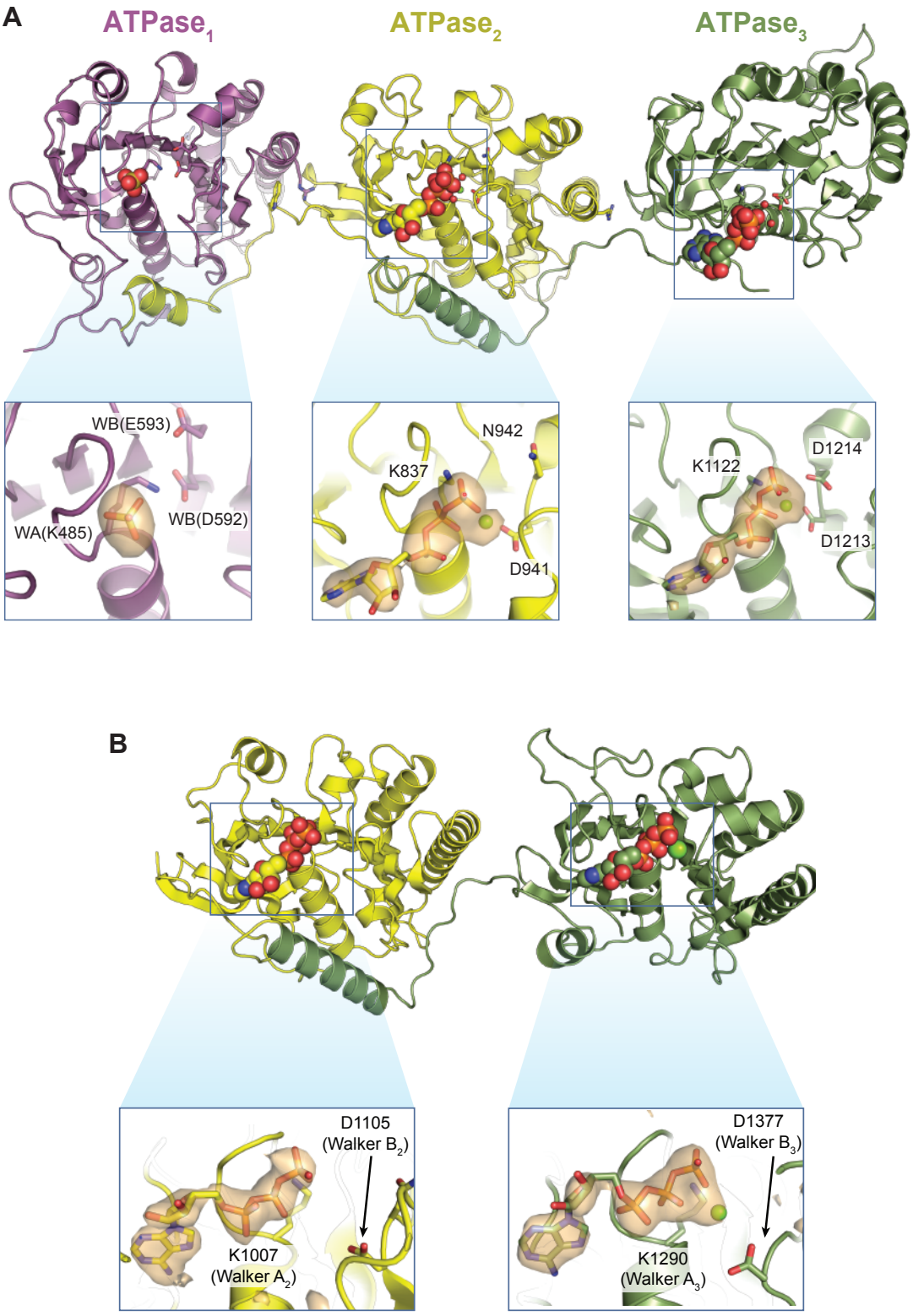

A
ATPase1
ATPase2
ATPase3
WB(E593)
WA(K485)
WB(D592)
N942
K837
D941
K1122
D1214
D1213
B
D1105
(Walker B2)
K1007
(Walker A2)
D1377
(Walker B3)
K1290
(Walker A3)

Figure 3.3: Structure of the cytoplasmic domain of EccC

(A) The structure of the cytoplasmic domain (construct "EccC," Figure 3.2) of EccC is shown with each ATPase domain represented by a different color. Each ATPase domain is held in direct translation with its neighbor through linker/pocket interactions. In each domain, the active site has been magnified and the positions of catalytically important residues shown. The nucleotides are shown with their difference density map. Importantly, while $ATPase_2$ and $ATPase_3$ are bound to ATP, $ATPase_1$ has a sulfate ion in the active site.

(B) The structure of the second and third ATPase domains of EssC from *Geobacillus thermodenitrificans*, a firmicute. The same linker-pocket architecture is observed in this homolog.

Figure 3.4

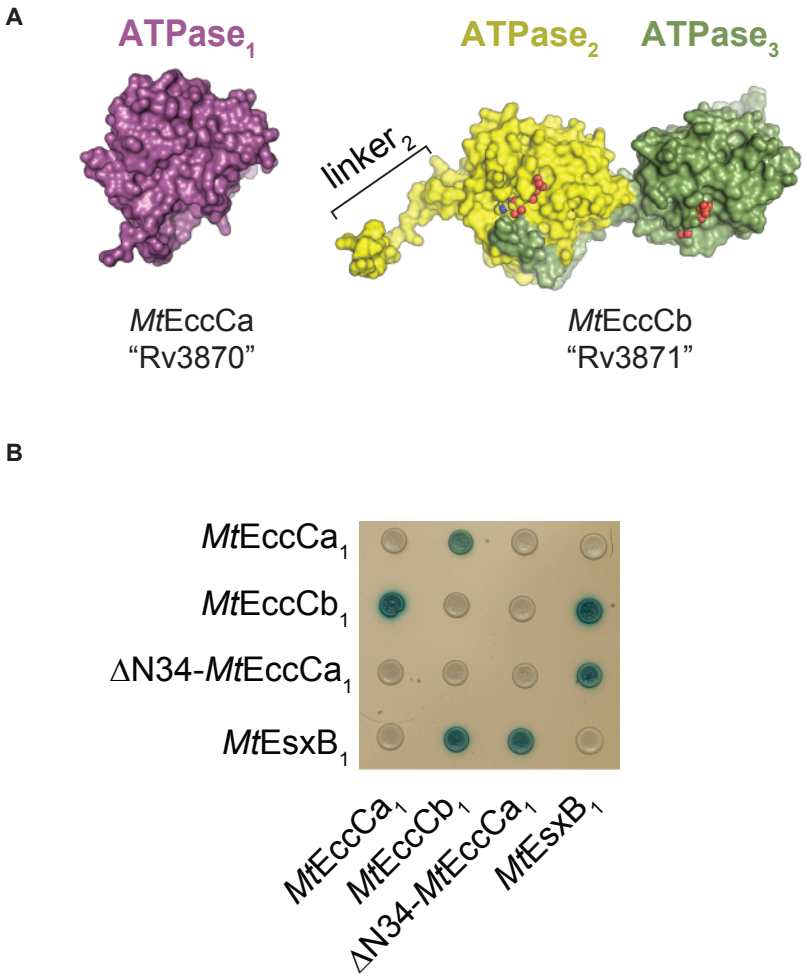

A
ATPase1
ATPase2
ATPase3
linker2
MtEccCa
"Rv3870"
MtEccCb
"Rv3871"
B
MtEccCa1
MtEccCb1
ΔN34-MtEccCa1
MtEsxB1
MtEccCa1
MtEccCb1
ΔN34-MtEccCa1
MtEsxB1

Figure 3.4: The linker / pocket interactions between ATPase domains are conserved in *M. tuberculosis*

(A) A model of $EccCa_1$ and $EccCb_1$ from the *M. tuberculosis* ESX-1 system based on the *T. curvata* EccC structure highlighting the presence of $linker_2$ extending from EccCb.

(B) Yeast-two-hybrid analysis shows that removal of linker2 disrupts the interaction between $EccCb_1$ and $EccCa_1$. Importantly, removal of these residues does not interfere with the interaction between EccCb1 and $EsxB_1$, suggesting that the structure and function of the protein are maintained in the deletion mutant.

Figure 3.5

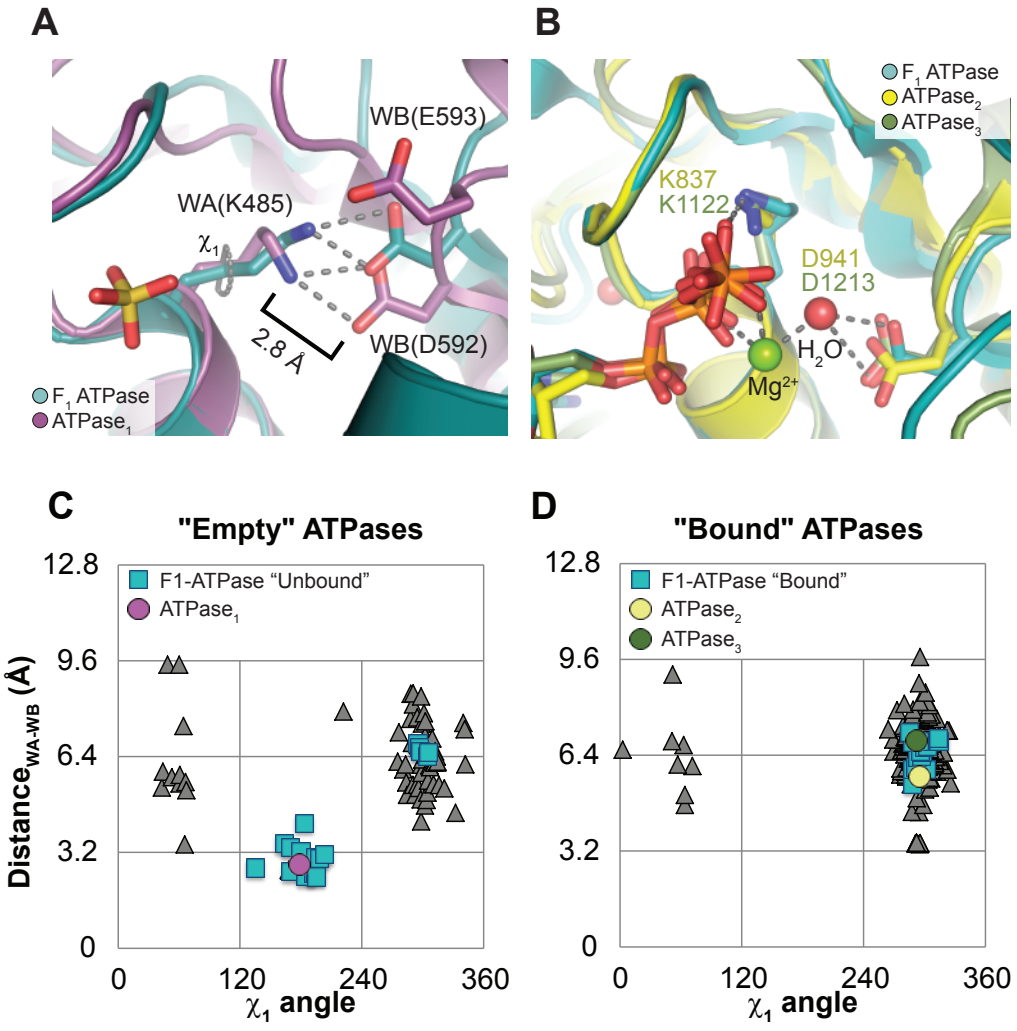

A
WB(E593)
WA(K485)
χ1
2.8 Å
WB(D592)
F1 ATPase
ATPase1
B
F1 ATPase
ATPase2
ATPase3
K837
K1122
D941
D1213
H2O
Mg2+
C
"Empty" ATPases
F1-ATPase "Unbound"
ATPase1
Distance WA-WB (Å)
12.8
9.6
6.4
3.2
0
0
120
240
360
χ1 angle
D
"Bound" ATPases
F1-ATPase "Bound"
ATPase2
ATPase3
12.8
9.6
6.4
3.2
0
0
120
240
360
χ1 angle

Figure 3.5: The $ATPase_1$ active site residues are held in an unusual conformation

(A) The active site of $ATPase_1$ (purple), with the "empty" subunit of F_1-ATPase (PDB 1H8H) overlaid in cyan. Importantly, the Walker A lysine is in an unusual conformation and forms a hydrogen bond with the Walker B aspartate, preventing magnesium binding and thus binding to ATP.

(B) The active site of $ATPase_2$ (yellow) and $ATPase_3$ (green) overlaid with the AMP-PNP-bound subunit of F_1-ATPase (cyan) from 1H8H.

(C) A graph representing the distance between the Walker A lysine amino group and the closest Walker B carboxylate oxygen, as a function of the rotameric position of the Walker A lysine. Each triangle represents one of 311 pdb chains of an ATP bound, P-loop ATPase identified by our protocol (see methods).

(D) A similar graph to (C) except the triangle represents the residues of "empty" ATPases from PDB entries that contain both a bound and unbound P-loop ATPase domain in the same file.

Figure 3.6

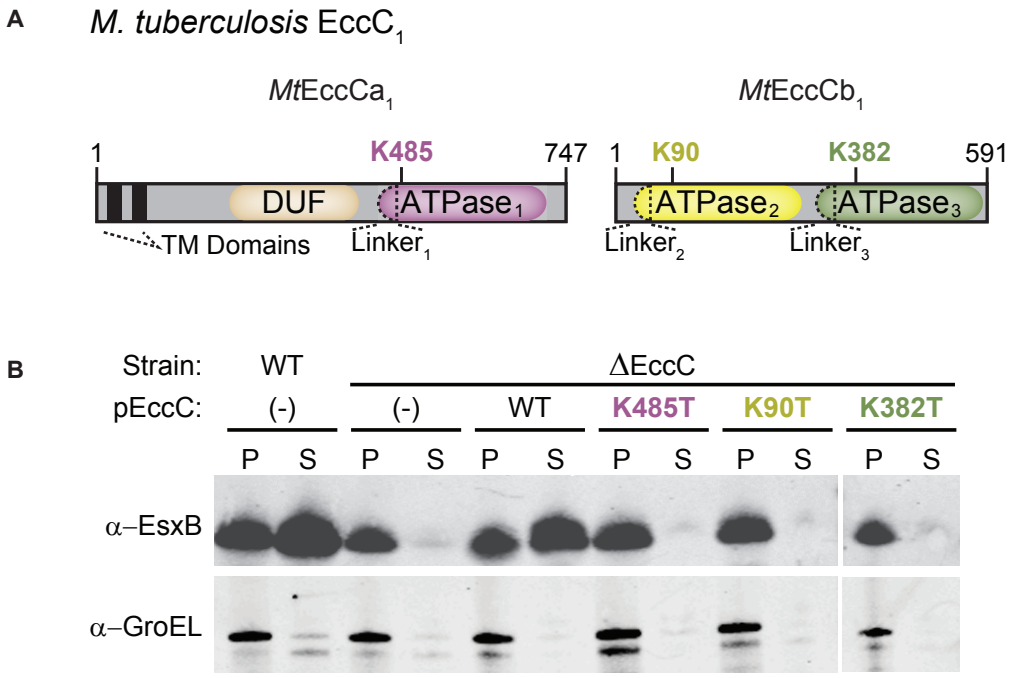

A
M. tuberculosis EccC1
MtEccCa1
MtEccCb1
1
K485
747
DUF
ATPase1
TM Domains
Linker1
1
K90
K382
591
ATPase2
ATPase3
Linker2
Linker3
B
Strain:
WT
ΔEccC
pEccC:
(-)
(-)
WT
K485T
K90T
K382T
P S P S P S P S P S P S
α–EsxB
α–GroEL

Figure 3.6: ATP binding is required at each ATPase domain of EccC for secretion of EsxB *in vivo*

(A) Domain map of $EccCa_1$ and $EccCb_1$ from *M. tuberculosis*. The Walker A lysine residues are indicated.

(B) Secretion assay with wild type, ΔeccC, ΔeccC::peccC, and ΔeccC::peccC(Walker A mutations). Wild type *M. tuberculosis* secretes robust quantities of EsxB, which depends completely on the presence of the eccC gene. Complementation of the knockout restores secretion to near wild type levels. However, complementation with eccC containing mutation of any of the three Walker A lysine residues fails to restore secretion.

Figure 3.7

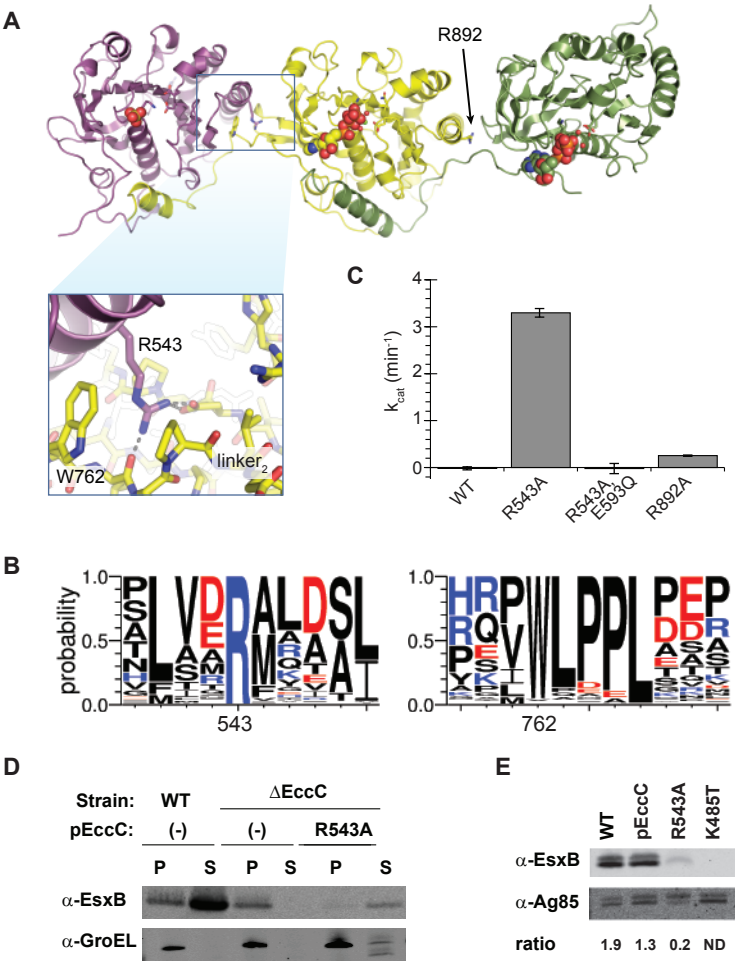

A
R892
R543
W762
linker2
C
kcat (min-1)
4
3
2
1
0
WT
R543A
R543A E593Q
R892A
B
probability
1.0
0.5
0.0
543
762
D
Strain:
WT
ΔEccC
pEccC:
(-)
(-)
R543A
P
S
P
S
P
S
α-EsxB
α-GroEL
E
WT
pEccC
R543A
K485T
α-EsxB
α-Ag85
ratio
1.9
1.3
0.2
ND

Figure 3.7: Autoinhibition of EccC can be relieved through disruption of the linker$_2$ / pocket$_1$ interaction

(A) Structure of EccC with magnification of the ATPase$_1$ / ATPase$_2$ interface. R543, which extends from pocket$_1$, mediates a large amount of the interaction with linker$_2$. An equivalent arginine between ATPase$_2$ and ATPase$_3$, R892, is also shown.

(B) The R543 is one hundred percent conserved, and the linker residues it binds are also highly conserved.

(C) Mutation of R543 to alanine results in a dramatic increase in ATPase activity, which is dependent on the catalytic glutamate of the ATPase$_1$ Walker B motif. Mutation of R892 to alanine results in a much smaller increase in activity.

(D) Secretion assay with wild type, ΔeccC, and ΔeccC::peccC(R543A) strains of *M. tuberculosis*. Here, the R543A mutation dramatically decreases, but does not eliminate totally, secretion of EsxB.

(E) Secretion assay with wild type, ΔeccC::peccC, ΔeccC::peccC(R543A), and ΔeccC::peccC(K485T). Here the secretion of EsxB is compared to the secretion of a control protein, Ag-85, and the ratio quantified below.

Figure 3.8

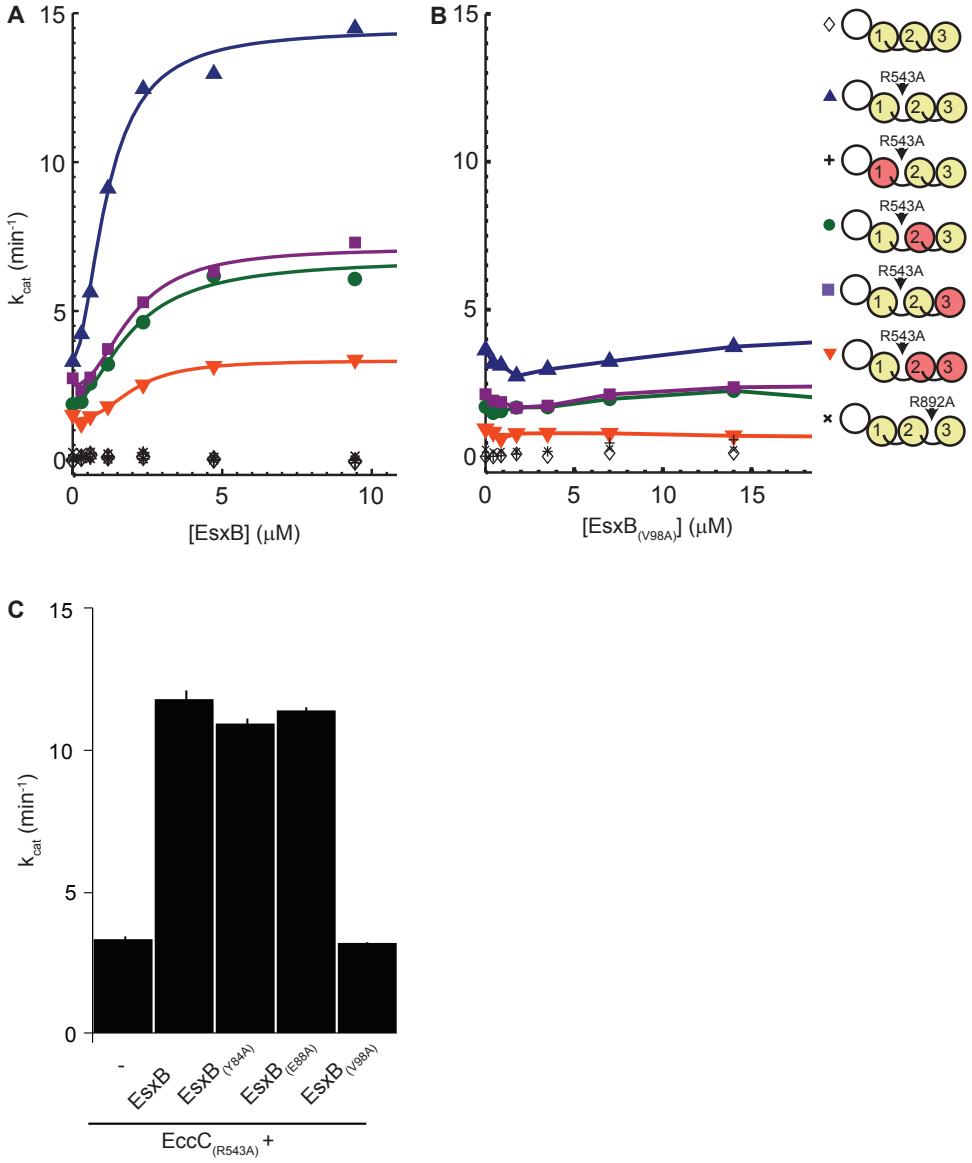

A
B
C
15
10
5
0
k_{cat} (min^{-1})
[EsxB] (μM)
[EsxB$_{(V98A)}$] (μM)
R543A
R892A
1
2
3
-
EsxB
EsxB$_{(Y84A)}$
EsxB$_{(E88A)}$
EsxB$_{(V98A)}$
EccC$_{(R543A)}$ +

Figure 3.8: Activated EccC is sensitive to the addition of substrates

(**A**) Titration of EsxB activates $EccC_{(R543A)}$. Mutation of Walker B residues in $ATPase_2$ or $ATPase_3$ decrease, but do not eliminate, ATPase activity, while mutation of the Walker B residue in $ATPase_1$ abrogates all activity. This suggests a regulatory role for $ATPase_2$ and $ATPase_3$.

(**B**) Titration of $EsxB_{(V98A)}$, which does not bind EccC, does not activate EccC(R543A) or any other EccC protein.

(**C**) The maximal increase in k_{cat} stimulated by the addition of EsxB is eliminated with mutation of the signal sequence residue V98, but is not affected by changes in the Y-XXX-D/E region.

Figure 3.9

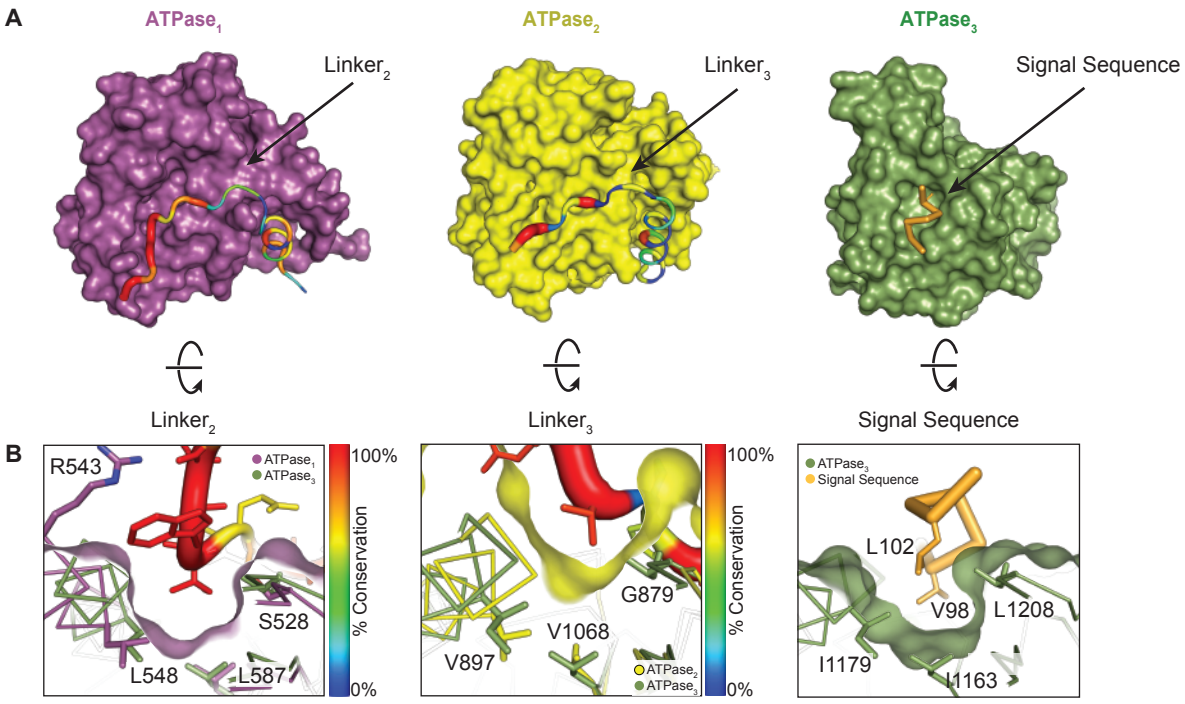

A
ATPase1
ATPase2
ATPase3
Linker2
Linker3
Signal Sequence
B
Linker2
Linker3
Signal Sequence
R543
S528
L548
L587
ATPase1
ATPase3
100%
% Conservation
0%
G879
V1068
V897
ATPase2
ATPase3
Signal Sequence
L102
V98
L1208
I1179
I1163

Figure 3.9: Residues in linker$_2$ and linker$_3$ mimic the substrate and bind to pocket$_1$ and pocket$_2$ on EccC

(**A**) The individual ATPase domains are shown and have been rotated to reveal the path of the linker across the ATPase domain. The linker is colored and weighted in diameter according to the degree of conservation across 142 unique EccC sequences.

(**B**) The surface has been rotated to highlight the linker groove. ATPase$_3$ and the pocket residues overlay ATPase$_1$ and ATPase$_2$ to highlight the homologies in the linker binding and signal sequence binding pockets. Note that the ATPase$_2$ pocket is significantly shallower than ATPase$_1$ and ATPase$_3$.

Figure 3.10

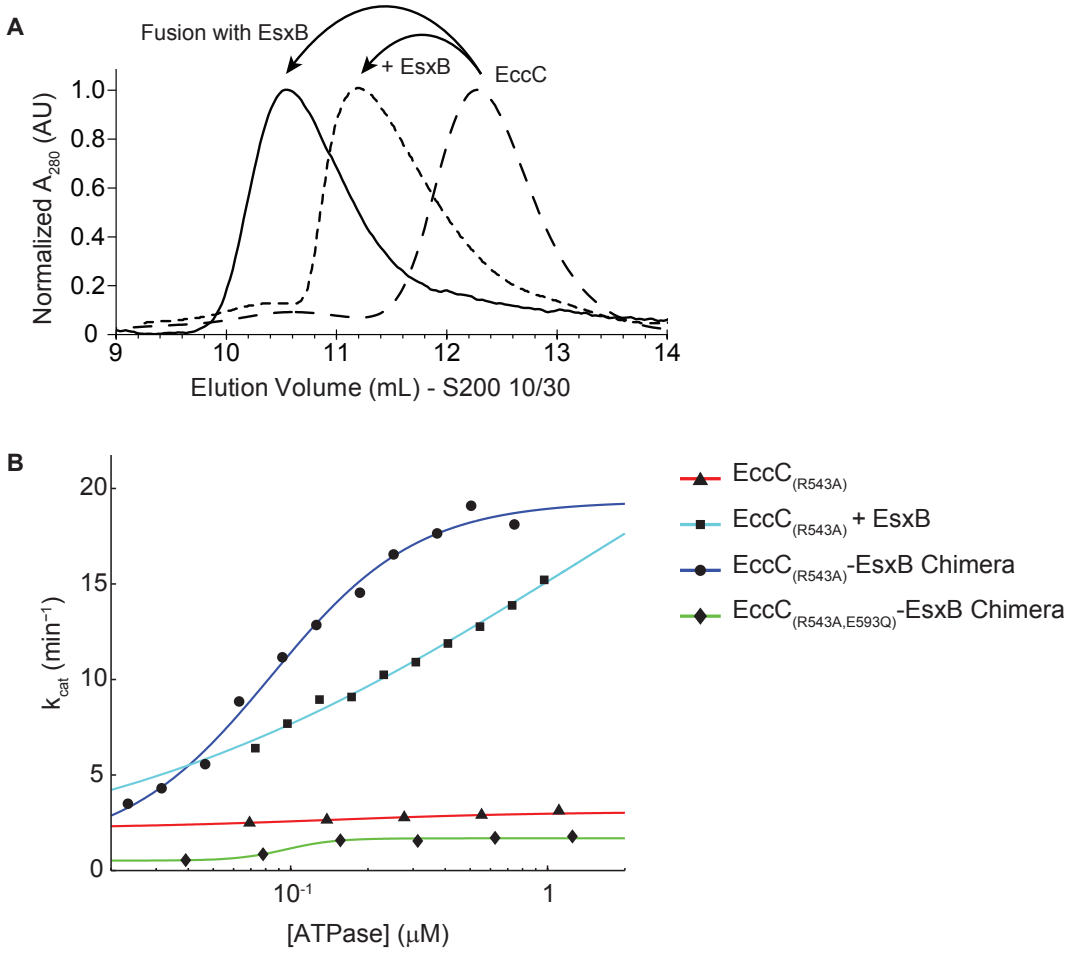

A
Fusion with EsxB
+ EsxB
EccC
Normalized A_{280} (AU)
1.0
0.8
0.6
0.4
0.2
0
9
10
11
12
13
14
Elution Volume (mL) - S200 10/30
B
20
15
10
5
0
k_{cat} (min^{-1})
10^{-1}
1
[ATPase] (μM)
$EccC_{(R543A)}$
$EccC_{(R543A)}$ + EsxB
$EccC_{(R543A)}$-EsxB Chimera
$EccC_{(R543A,E593Q)}$-EsxB Chimera

Figure 3.10: EsxB increases EccC complex size and results in concentration-dependent activity

(**A**) Gel filtration chromatograms showing the increase in hydrodynamic radius of EccC with addition of EsxB or fusion to EsxB via a flexible linker. The change in elution volume is substantially greater than would be expected for the increase in size from the substrate alone, which is only approximately 15 kDa.

(**B**) ATPase activity as a function of enzyme concentration. Without substrate, $EccC_{(R543A)}$ does not display concentration-dependent activity. However, fusion or addition of the EsxB substrate results in strong concentration dependence, suggesting that the active complex requires multimer formation. Importantly, the bulk of the activity and concentration dependence is contingent on a functional Walker B motif on $ATPase_1$.

Figure 3.11

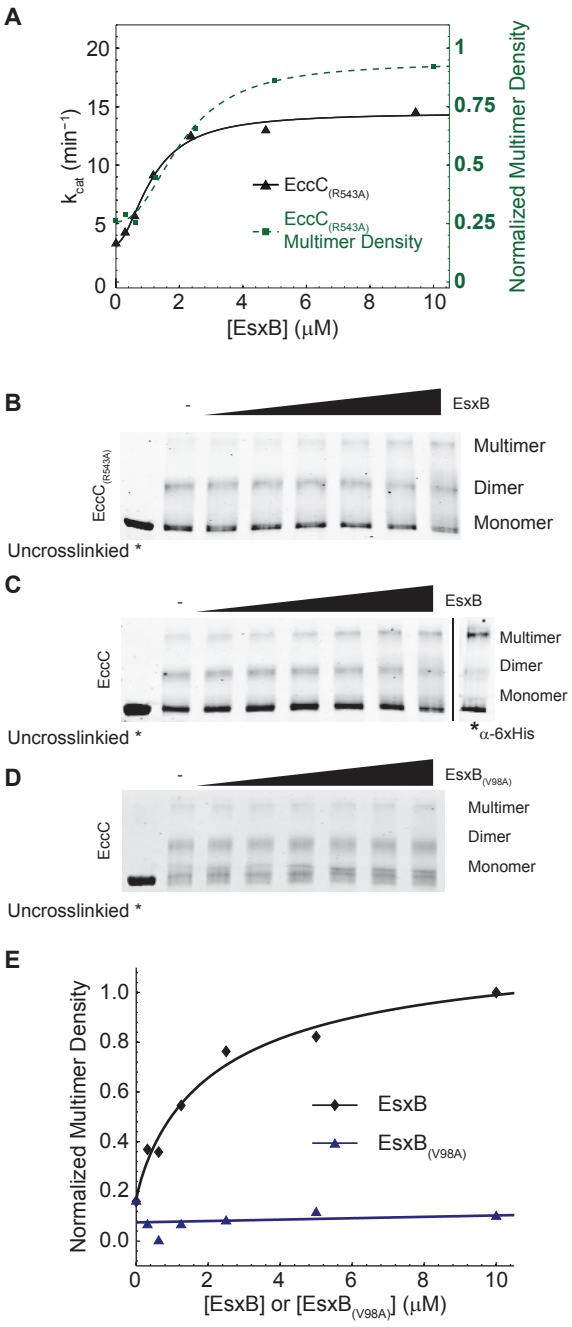

A
20
15
10
5
0
k_{cat} (min^{-1})
0
2
4
6
8
10
[EsxB] (μM)
1
0.75
0.5
0.25
0
Normalized Multimer Density
EccC(R543A)
EccC(R543A) Multimer Density
B
-
EsxB
EccC(R543A)
Multimer
Dimer
Monomer
Uncrosslinkied *
C
-
EsxB
EccC
Multimer
Dimer
Monomer
Uncrosslinkied *
*α-6xHis
D
-
EsxB(V98A)
EccC
Multimer
Dimer
Monomer
Uncrosslinkied *
E
Normalized Multimer Density
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
[EsxB] or [EsxB(V98A)] (μM)
EsxB
EsxB(V98A)

Figure 3.11: Glutaraldehyde crosslinking reveals higher order oligomers that correlate highly with activity

(A) Activity and presence of higher order multimer (see B) as a function of EsxB concentration.

(B) Glutaraldehyde crosslinking of $EccC_{(R543A)}$ with increasing concentrations of EsxB. Samples were run on a 10% gel and stained with InstantBlue (a coomassie-based detection method).

(C) Similar to (B) but with wild type EccC. A similar increase in multimer can be seen despite the lack of activity in this construct. Also shown is a western positively identifying each band as EccC.

(D) Similar to (C) but with titration of $EsxB_{(V98A)}$ and shown at higher contrast. While some multimer exists naturally, addition of the mutated substrate has no effect on its concentration.

(E) A graph quantifying the results of (C) and (D).

Figure 3.12

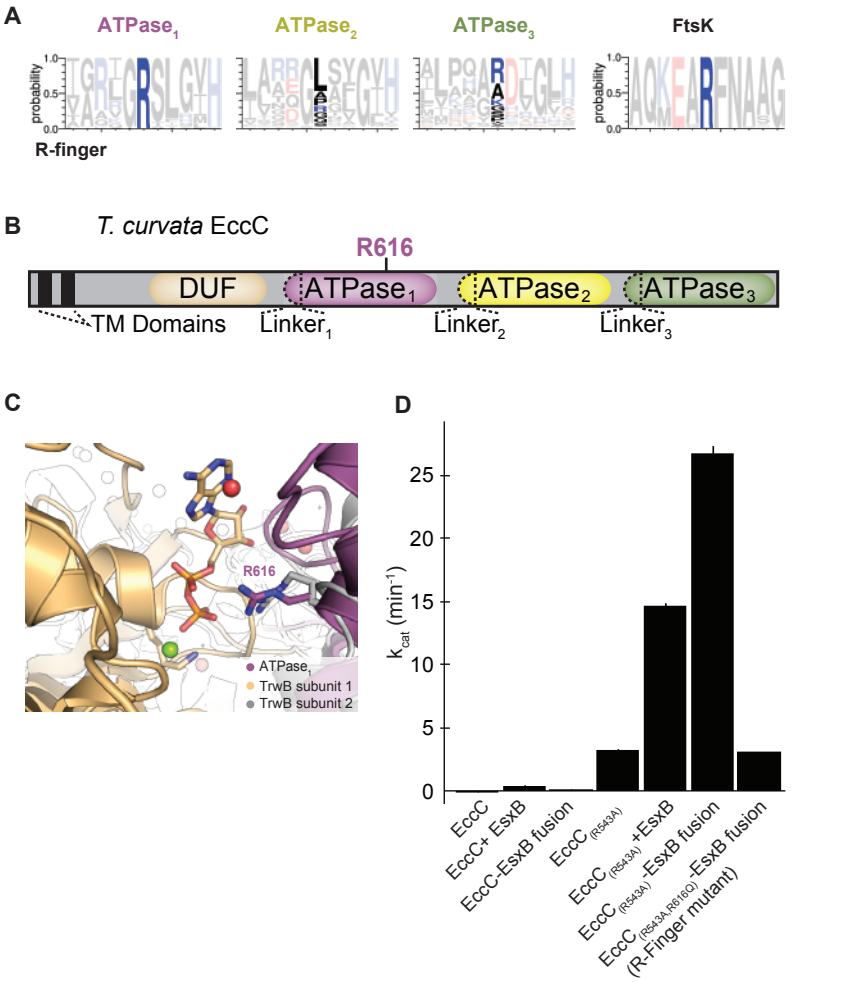

A
ATPase1
ATPase2
ATPase3
FtsK
probability
R-finger
B
T. curvata EccC
R616
DUF
ATPase1
ATPase2
ATPase3
TM Domains
Linker1
Linker2
Linker3
C
R616
ATPase1
TrwB subunit 1
TrwB subunit 2
D
kcat (min-1)
25
20
15
10
5
0
EccC
EccC+ EsxB
EccC-EsxB fusion
EccC(R543A)
EccC(R543A) +EsxB
EccC(R543A) -EsxB fusion
EccC(R543A-R616Q) -EsxB fusion
(R-Finger mutant)
E
M. tuberculosis EccC1
R623
DUF
ATPase1
TM Domains
Linker1
MtEccCa1
ATPase2
ATPase3
Linker2
Linker3
MtEccCb1
F
Strain:
WT
Δ EccC
pEccC:
(-)
(-)
WT
R623A
P
S
P
S
P
S
P
S
α-EsxB
α-GroEL
M. tuberculosis

Figure 3.12: EccC utilizes an arginine finger mechanism

(A) Conservation of the arginine finger residue in each ATPase domain as well as in FtsK. $ATPase_1$, but not $ATPase_2$ or $ATPase_3$, contains a highly conserved arginine finger residue.

(B) Domain map of EccC from *T. curvata* indicating the position of the arginine finger.

(C) Structural alignment of EccC with TrwB, showing that the predicted arginine finger residue on EccC possesses the same relative position of the arginine finger residue from a related ATPase.

(D) Activity of several EccC constructs. Mutation of the arginine finger residue drastically reduces activity of the activated EccC-EsxB fusion protein.

(E) Domain map of $EccC_1$ from *M. tuberculosis* indicating the position of the arginine finger.

(F) Secretion assay showing that mutation of the arginine finger residue in *M. tuberculosis* prevents secretion of EsxB.

Figure 3.13

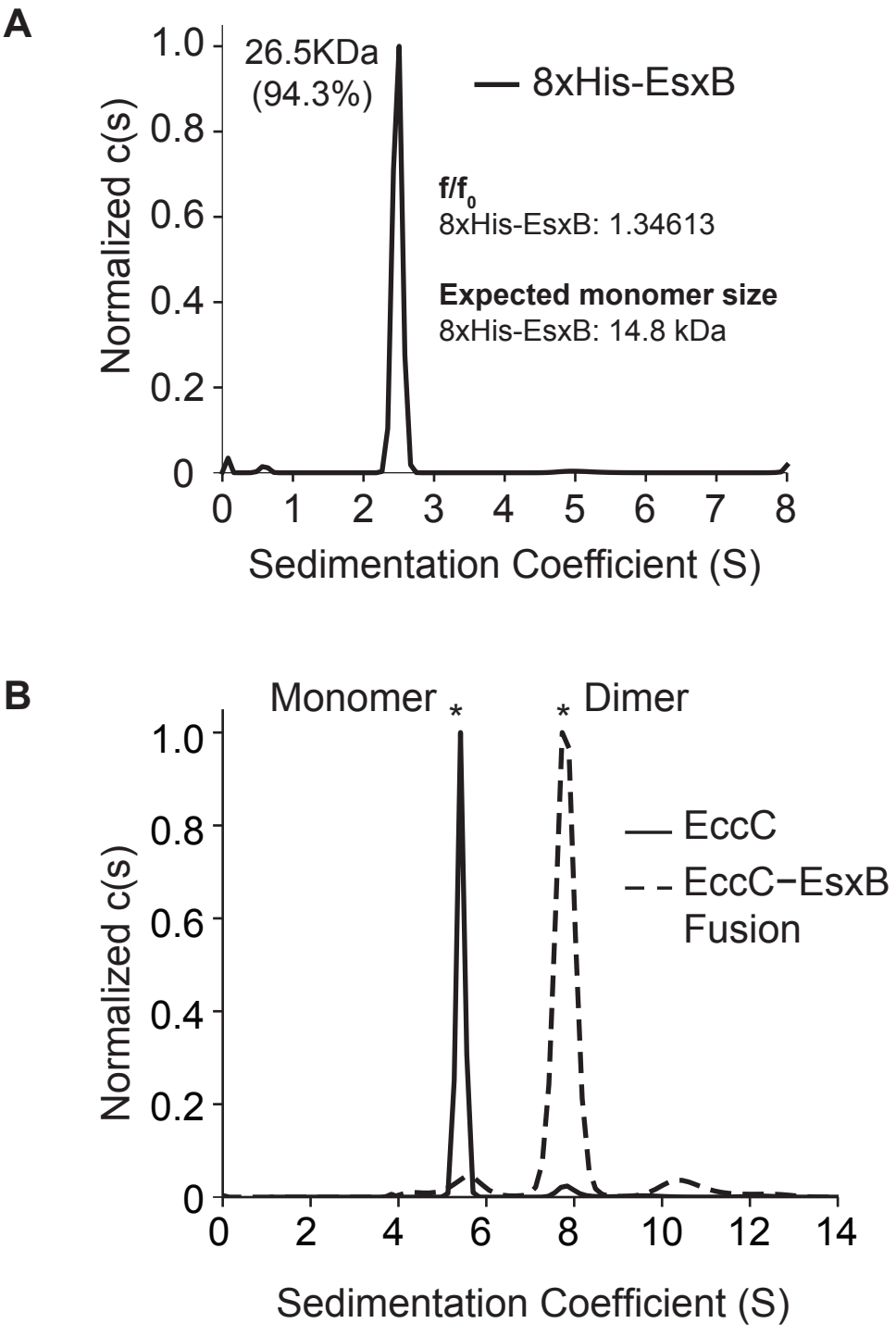

A
26.5KDa
(94.3%)
8xHis-EsxB
f/f0
8xHis-EsxB: 1.34613
Expected monomer size
8xHis-EsxB: 14.8 kDa
Normalized c(s)
Sedimentation Coefficient (S)
B
Monomer
Dimer
EccC
EccC−EsxB
Fusion
Normalized c(s)
Sedimentation Coefficient (S)

Figure 3.13: EsxB acts as a dimerization domain

(A) Analytical ultracentrifugation indicating the mass of EsxB in solution corresponds to a dimmer.

(B) Fusion of EsxB to EccC results in dimerization of the construct as determined by AUC.

Figure 3.14

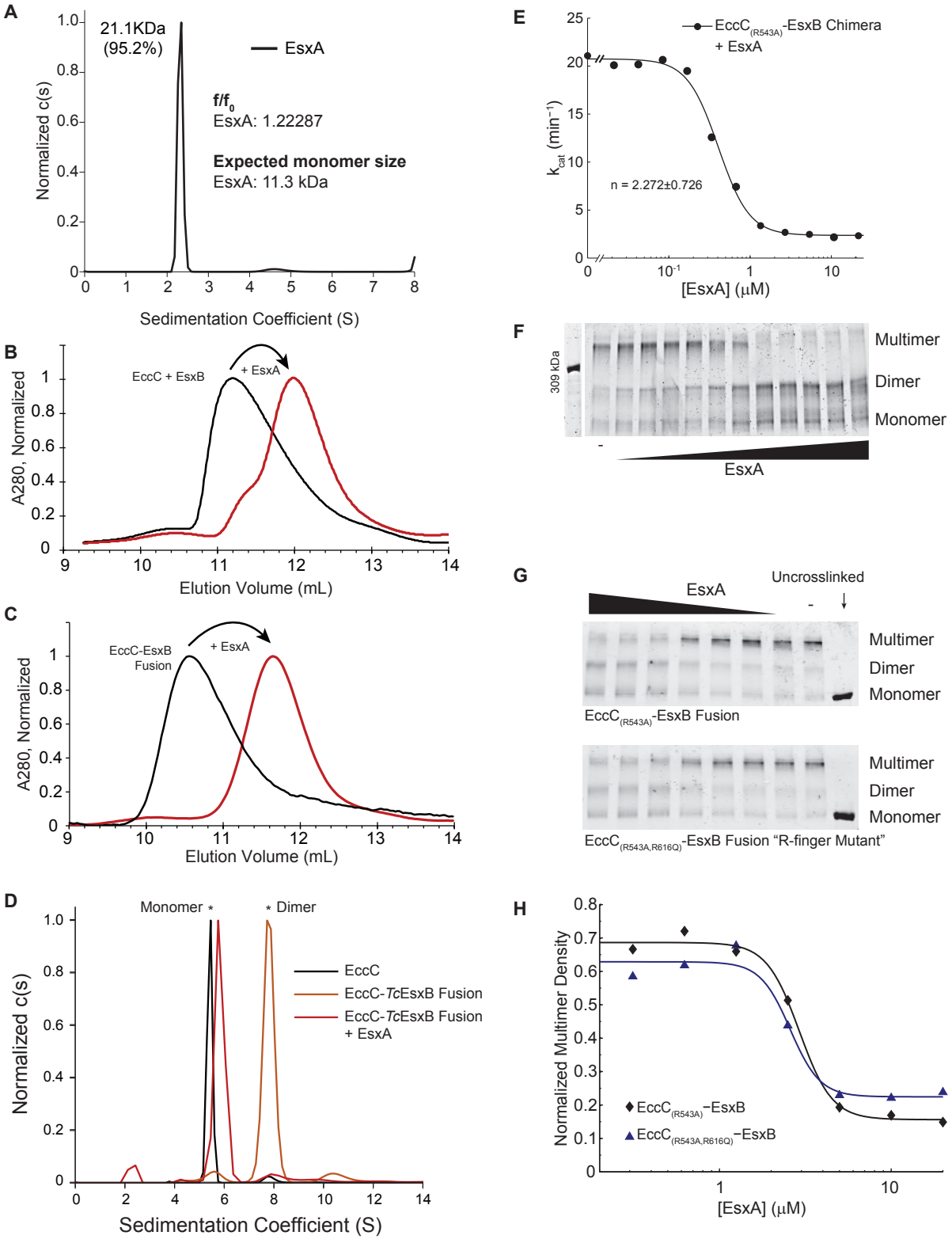

A
21.1KDa
(95.2%)
EsxA
f/f0
EsxA: 1.22287
Expected monomer size
EsxA: 11.3 kDa
Normalized c(s)
Sedimentation Coefficient (S)
B
EccC + EsxB
+ EsxA
A280, Normalized
Elution Volume (mL)
C
EccC-EsxB
Fusion
+ EsxA
A280, Normalized
Elution Volume (mL)
D
Monomer
Dimer
EccC
EccC-TcEsxB Fusion
EccC-TcEsxB Fusion
+ EsxA
Normalized c(s)
Sedimentation Coefficient (S)
E
EccC(R543A)-EsxB Chimera
+ EsxA
kcat (min-1)
n = 2.272±0.726
[EsxA] (μM)
F
309 kDa
Multimer
Dimer
Monomer
-
EsxA
G
EsxA
Uncrosslinked
-
Multimer
Dimer
Monomer
EccC(R543A)-EsxB Fusion
Multimer
Dimer
Monomer
EccC(R543A,R616Q)-EsxB Fusion "R-finger Mutant"
H
Normalized Multimer Density
EccC(R543A)–EsxB
EccC(R543A,R616Q)–EsxB
[EsxA] (μM)

Figure 3.14: EsxA disrupts EccC multimers and decreases ATPase activity

(A) Analytical ultracentrifugation indicating the mass of EsxA in solution corresponds to a dimmer.

(B) Gel filtration chromatogram indicating a substantial decrease in hydrodynamic radius of the EccC + EsxB complex with addition of EsxA.

(C) Gel filtration chromatogram indicating a substantial decrease in hydrodynamic radius of the EccC- EsxB fusion protein with addition of EsxA.

(D) Analytical ultracentrifugation showing addition of EsxA results in loss of dimerization of the EccC-EsxB fusion protein.

(E) Rapid and cooperative decrease in ATPase activity of the active EccC-EsxB fusion protein with addition of EsxA.

(F) Glutaraldehyde crosslinking of the samples used in (E) showing a rapid decrease in the concentration of the multimer which mirrors the loss in activity.

(G) Mutation of the arginine finger does not affect the ability of the EccC-EsxB fusion to multimerize as determined by glutaraldehyde crosslinking.

(H) Quantification of (G) by gel densitometry.

Figure 3.15

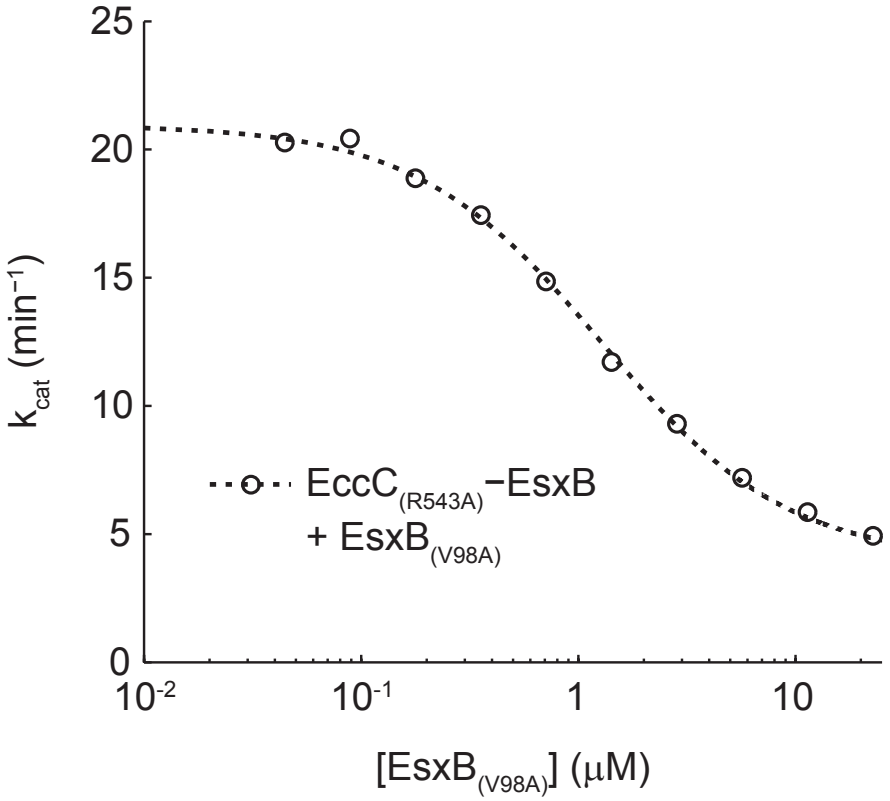

25
20
15
10
5
0
k_{cat} (min^{-1})
10^{-2}
10^{-1}
1
10
$EccC_{(R543A)}$−EsxB
+ $EsxB_{(V98A)}$
[$EsxB_{(V98A)}$] (μM)

Figure 3.15: Competition for the EsxB dimerization surface inhibits EccC-EsxB

Addition of the signal sequence mutant $EsxB_{(V98A)}$ disrupted activity similarly to addition of EsxA, suggesting that EsxA inhibits the EccC-EsxB complex by breaking the EsxB:EsxB dimerization interface. However, inhibition by EsxB(V98A) does not exhibit the same cooperativity observed with EsxA.

Figure 3.16

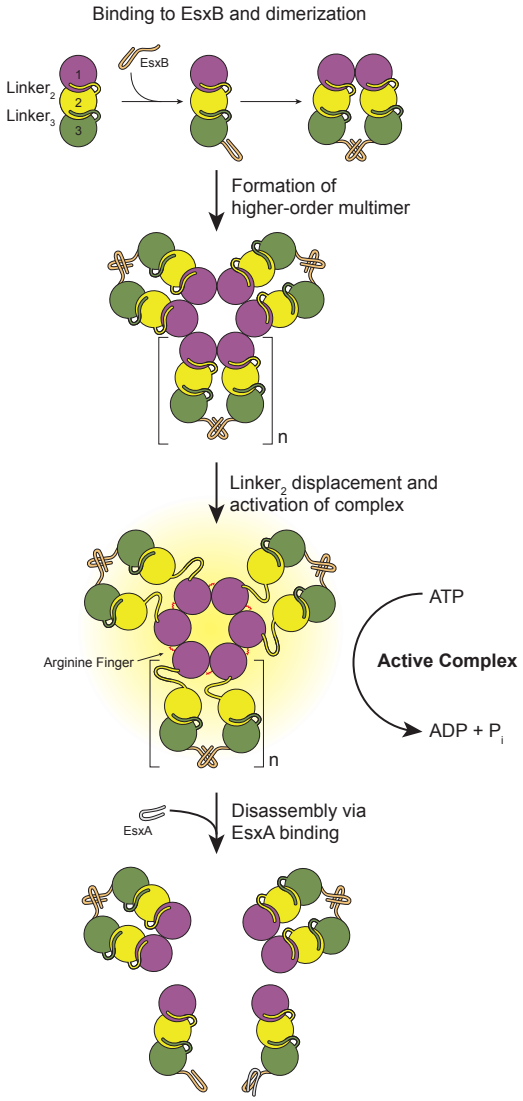

Binding to EsxB and dimerization
EsxB
Linker2
Linker3
1
2
3
Formation of
higher-order multimer
n
Linker2 displacement and
activation of complex
Arginine Finger
ATP
Active Complex
ADP + Pi
n
EsxA
Disassembly via
EsxA binding

Figure 3.16: Model for EccC activation and inhibition by substrates

In the absence of substrates EccC is monomeric. Interaction with EsxB leads to dimerization of the ATPase and higher order multimerization, but cannot activate the enzyme. In this study we used the R543A mutation to disrupt the interaction between ATPase1 and ATPase2 although in vivo this role is likely played by other proteins that bind to ATPase1 analogously to the binding of EsxB to ATPase3. Once ATPase1 is displaced, EccC is activated further by multimerization mediated by a conserved arginine finger. EsxA can disrupt the EsxB:EsxB interaction, resulting in disassembly of multimeric EccC and loss of activity.

Chapter 4: In progress projects and preliminary data

Characterization of the four MutT homologs from *Mycobacterium tuberculosis*

Reactive oxygen species (ROS) represent a major hazard that all cells must face. ROS exposure results in the accumulation of oxidated and damaged nucleotides, which can incorporate into DNA and cause mutation and double strand breaks [79]. Nudix hydrolases represent one line of defense which serves to clear the cell of damaged nucleotides [80]. In *E. coli*, the nudix hydrolase MutT rapidly hydrolyzes 8-oxo-dGTP into 8-oxo-dGMP, preventing its incorporation into DNA [81]. 8-oxo-dGTP represents a major hazard for cells, as it can readily form base pairs with either dCTP or dATP (Figure 4.1).

There are four MutT homologs in *Mycobacterium tuberculosis*, Rv2985 (MutT1), Rv1160 (MutT2), Rv0413 (MutT3), and Rv3908 (MutT4) [82]. Transposon insertion in mutT1 or mutT4 result in increased rates of mutation *in vivo*, highlighting the importance of these proteins for maintaining DNA integrity [82].

Recent work has found that ROS are generated during antibiotic treatment and may represent an important mode of killing [83], [84]. It is possible that inhibition of the MutT proteins in *M. tuberculosis* may increase the efficacy of other antibiotics by preventing the repair of some of the damage caused by the ROS. Indeed, a similar strategy has proven successful in the destruction of cancer cells by inhibition of the human Mth1 gene, a nudix hydrolase [85].

I used a phosphate release assay (Figures 4.2 and 4.3) combined with a library consisting of 39 nucleotides to try to identify substrates for each of the four MutT

homologs. Ideally, identification of specific substrates along with the phosphate release assay will facilitate the screening for small molecule inhibitors of each protein. To that end, I found each MutT had one or two specific substrates whose rate of hydrolysis was several times higher than the plate average (normalized data shown in Figure 4.4).

Similar to the *E. coli* MutT, both MutT1 and MutT2 hydrolyze 8-oxo-dGTP. However, MutT2 primarily hydrolyzes dCTP (Figures 4.4 and 4.6A), and MutT3 and MutT4 hydrolyze substrates not traditionally thought of in the context of DNA damage. Specifically, MutT3 primarily hydrolyzes dADP and MutT4 works on the substrate guanosine-5'-tetraphosphate (ppppG). Previous work on MutT1 and MutT2 confirm these findings using orthologous techniques [86], [87].

To determine which phosphodiester bond was being cleaved by each enzyme, I measured enzyme kinetics in the presence and absence of inorganic pyrophosphatase (PPase). I found that MutT1 was not reliant on PPase for detection of phosphate in the reaction solution, demonstrating that MutT1 cleaved only a single phosphate from 8-oxo-dGTP (ie: the γ-phosphate), producing 8-oxo-dGDP + Pi (Figure 4.5A). Unlike MutT1, phosphatase activity was undetectable with MutT2 in the absence of PPase. This shows that MutT2 cleaves pyrophosphate from 8-oxo-dGTP, producing 8-oxo-dGMP + PPi (Figure 4.5B). Thus, *M. tuberculosis* utilizes two different nudix hydrolases to clear 8-oxo-dGTP from the nucleotide pool which act on the substrate at two different locations (Figure 4.5C).

MutT4 intriguingly hydrolyzes the molecule ppppG. Measuring kinetics both in the presence and absence of PPase, I determined that MutT4 hydrolyzes the δ

phosphate to produce GTP and Pi (Figure 4.6C). Interestingly, when expressing MutT4 in *E. coli*, even before induction by IPTG the rate of growth of the bacteria was severely attenuated, suggesting even small quantities of this protein can have a major growth effect. MutT4 is also positively regulated in *M. tuberculosis* under hypoxic conditions [88]. While almost nothing is known of ppppG, it is intriguing to speculate that it could serve as an important bacterial second messenger. MutT3 also hydrolyzed ppppG, although at a much reduced rate (Figure 4.6B).

Examination of the effects of PDIM and EsxA on membrane disruption

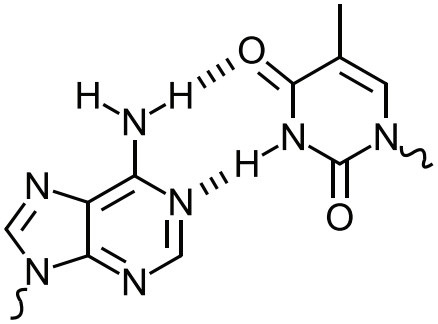
Currently almost nothing is known of the activity of individual virulence factors secreted by *M. tuberculosis*. The ESX-1 substrate EsxA is a highly immunogenic protein that is necessary for virulence and development of tuberculosis disease. Some work has suggested that EsxA may serve to permeabilize membranes, facilitating cytosolic access [89], [90]. However, these studies only saw permeabilization effects with very high concentrations of protein (>10 µg/mL), and utilized detergents in the purification of EsxA. I wondered whether the virulence lipid phthiocerol dimycocerosate (PDIM), which is found in virulent mycobacteria and is necessary for growth of *M. tuberculosis* in lung tissue [91], acted synergistically with EsxA to disrupt membranes.

To test this hypothesis, I measured viability of A549 cells as a proxy for membrane integrity in the presence or absence of both PDIM and EsxA, both highly purified. I found that while growth on PDIM decreased cell viability, presence of EsxA had only modest effects even at 50 µg/mL (Figure 4.7).

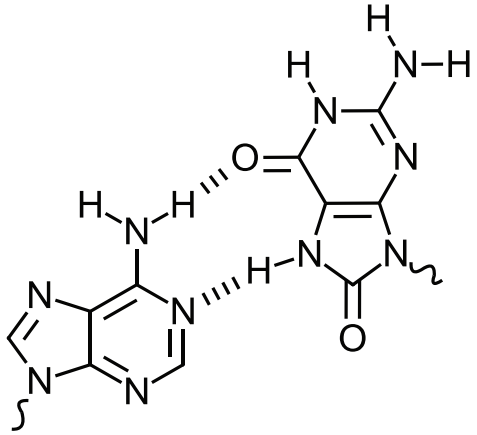
It has been shown that *M. tuberculosis* bacilli have membrane lysing activity that is dependent on ESX-1 secretion [92]. It is possible that EsxA does act as a lysis-promoting agent in the context of the living cell, but not when purified away from other protein or lipid components. However, previous studies using purified EsxA should be considered carefully due to the use of harsh detergents in the washing of inclusion bodies.

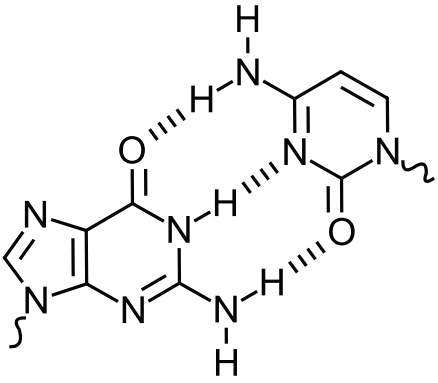
Figure 4.1

A:T

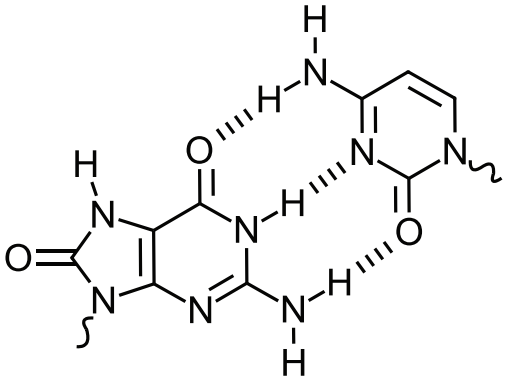


A:8-oxo-G





G:C



8-oxo-G:C

Figure 4.1: Base-pairing of 8-oxo-G to A or C

Diagram showing the hydrogen bonds formed during the A:T, A:8-oxo-G, G:C, and 8-oxo-G:C. Pairing between 8-oxo-dG and A requires a rotation of the bond between the base and the ribose (indicated by a ~).

Figure 4.2

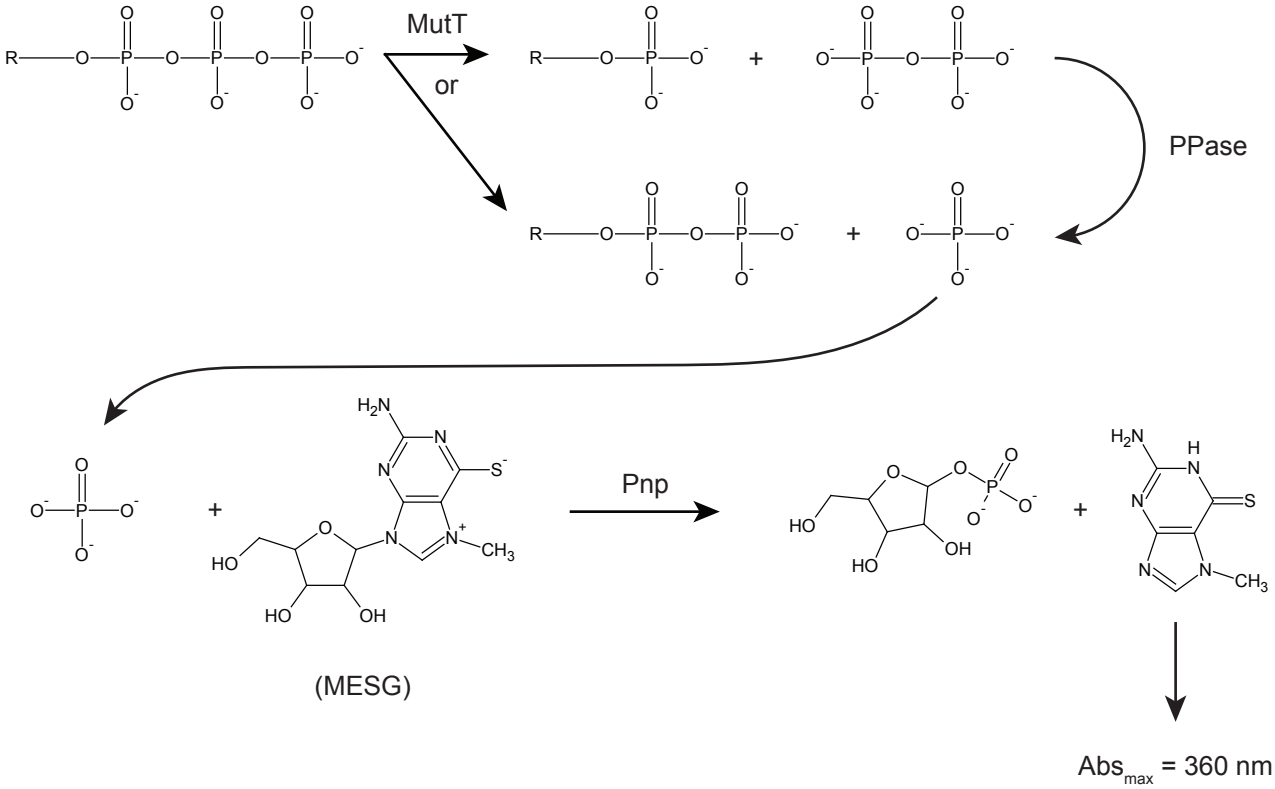
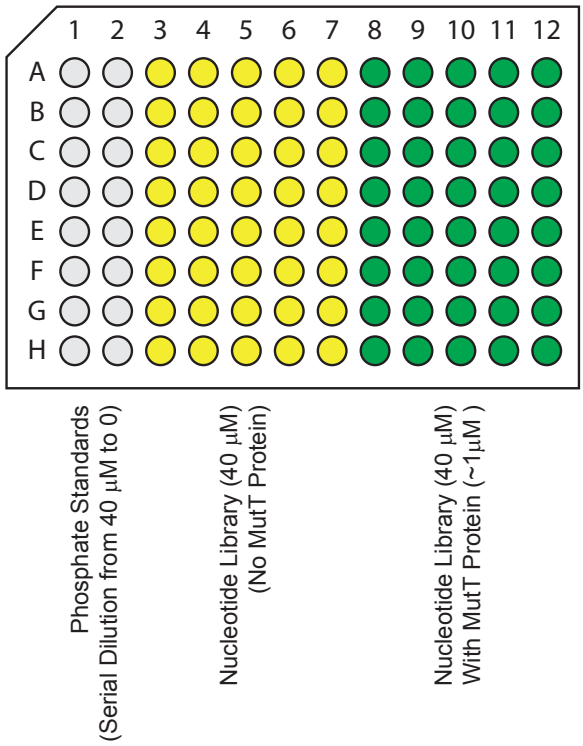


Figure 4.2: Molecular basis for the phosphate release assay used to detect activity of MutT homologs

MutT cleaves either one or two phosphates from a nucleotide substrate. Because pyrophosphate (PPi) is not detectable with this assay, pyrophosphatase (PPase) is added to convert PPi to inorganic phosphate (Pi). The protein purine nucleoside phosphorylase (Pnp) acts on Pi and the substrate 2-amino-6-mercapto-7-methyl riboside (MESG) to produce ribose-1-phosphate and 2-amino-6-mercapto-7-methylpurine, which has a maximum absorbance of 360 nm. Abs_{360} is used as a measure of activity. Use of phosphate standards (not shown) allow for conversion of absorbance to mass of Pi generated per unit time.

Figure 4.3

A


1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
Phosphate Standards
(Serial Dilution from 40 µM to 0)
Nucleotide Library (40 µM)
(No MutT Protein)
Nucleotide Library (40 µM)
With MutT Protein (~1µM)

B

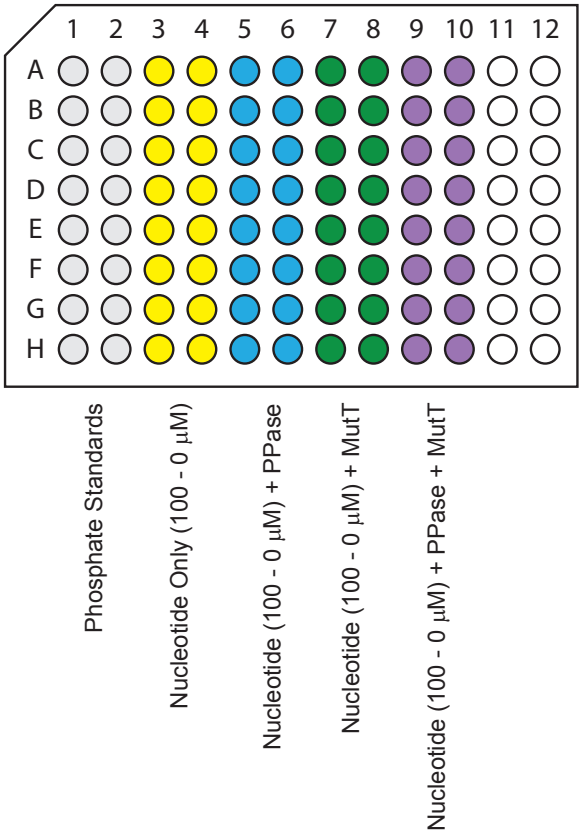

1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
Phosphate Standards
Nucleotide Only (100 - 0 µM)
Nucleotide (100 - 0 µM) + PPase
Nucleotide (100 - 0 µM) + MutT
Nucleotide (100 - 0 µM) + PPase + MutT

Figure 4.3: 96-well plate format for screening the nucleotide library and for measuring kinetic parameters for individual nucleotides

(A) The nucleotide screen assay was set up as in this diagram. Two rows were allotted for phosphate standards (gray), followed by five rows (yellow) of substrates without the MutT protein. These wells were used as blanks for the +MutT reactions, shown in green. Concentrations used for this screen: ~40 µM substrate; ~1 µM MutT.

(B) Nucleotides identified in the screen were tested against specific MutT proteins at varying concentrations to generate activity curves. Importantly, each nucleotide was tested with or without PPase to ascertain the location of hydrolysis. Again, phosphate standards were used to quantify the results.

Figure 4.4

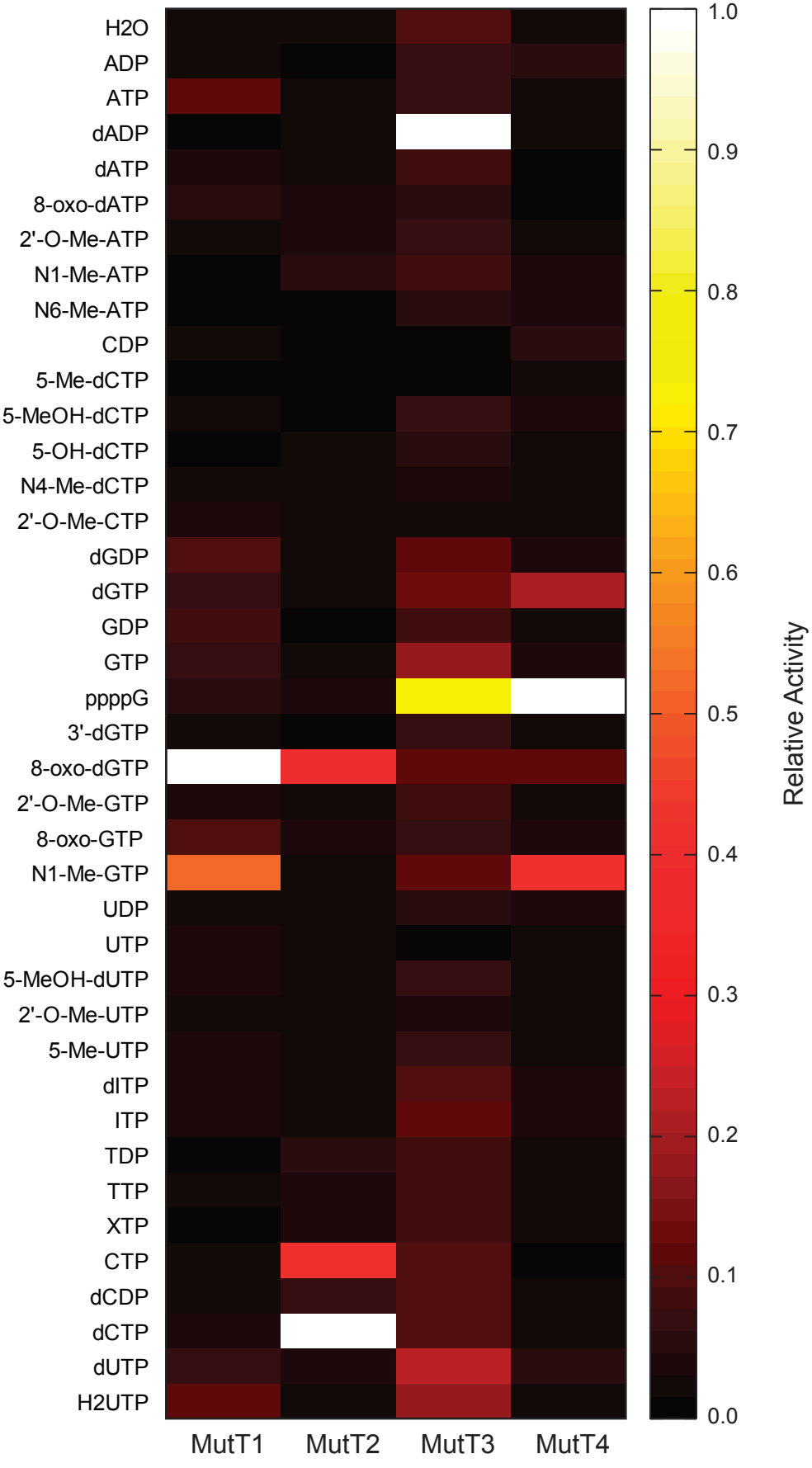

H2O
ADP
ATP
dADP
dATP
8-oxo-dATP
2'-O-Me-ATP
N1-Me-ATP
N6-Me-ATP
CDP
5-Me-dCTP
5-MeOH-dCTP
5-OH-dCTP
N4-Me-dCTP
2'-O-Me-CTP
dGDP
dGTP
GDP
GTP
ppppG
3'-dGTP
8-oxo-dGTP
2'-O-Me-GTP
8-oxo-GTP
N1-Me-GTP
UDP
UTP
5-MeOH-dUTP
2'-O-Me-UTP
5-Me-UTP
dITP
ITP
TDP
TTP
XTP
CTP
dCDP
dCTP
dUTP
H2UTP
MutT1
MutT2
MutT3
MutT4
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Relative Activity

Figure 4.4: Nucleotide-specificity for MutT1, MutT2, MutT3, and MutT4 from *Mycobacterium tuberculosis*

39 nucleotides were screened against the four MutT homologs from *M. tuberculosis*. As expected, MutT1 hydrolyzes predominantly 8-oxo-dGTP. However, an additional substrate for MutT1, N1-methyl-GTP, was also identified. MutT2 also has activity against 8-oxo-dGTP, but is primarily a CTPase. MutT3 mainly hydrolyzed dADP, but also had activity against guanosine-5'-tetraphosphate (ppppG). MutT4 primarily hydrolyzed ppppG, but had secondary activity against N1-methyl-GTP.

Figure 4.5

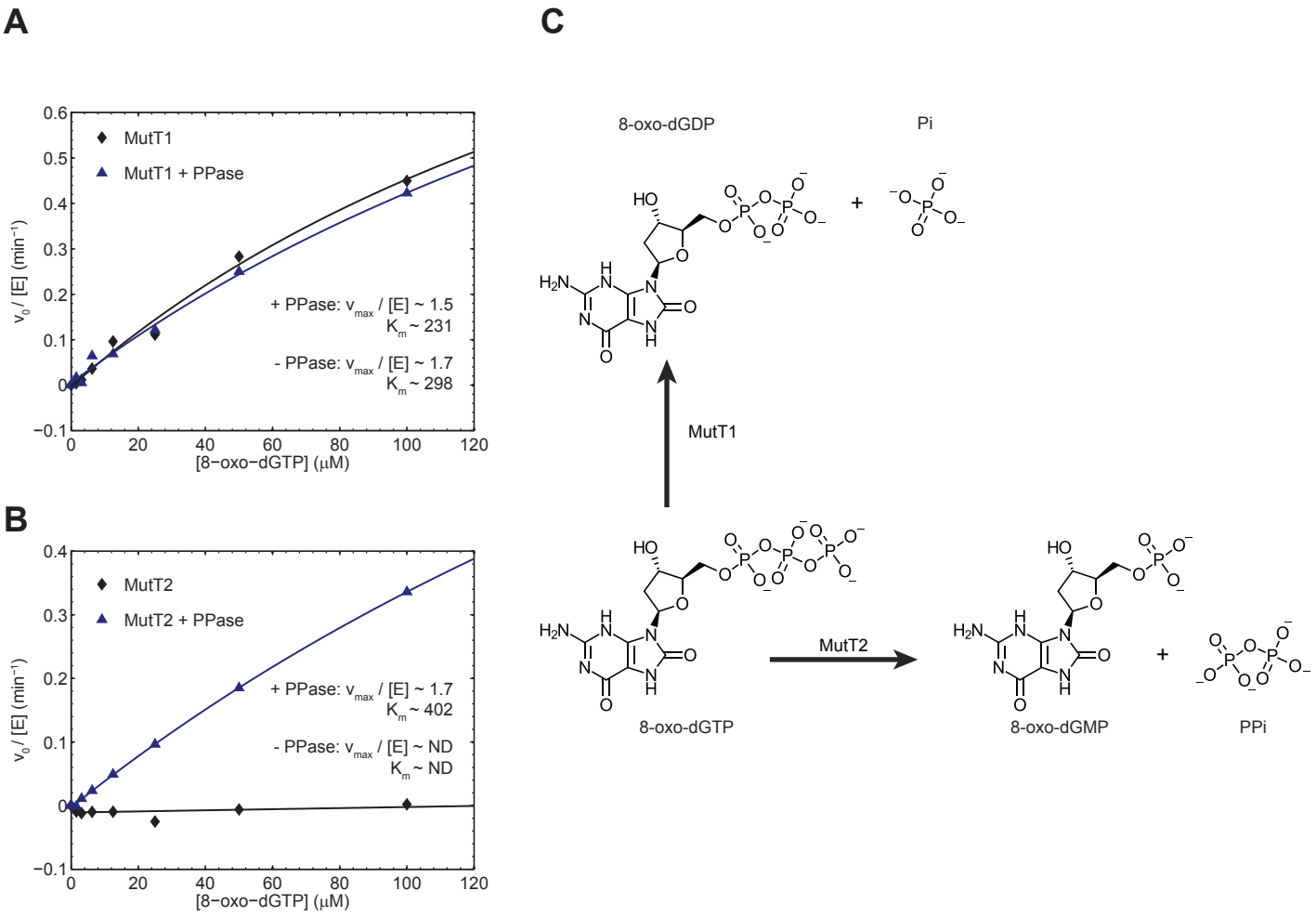

A
B
C
MutT1
MutT1 + PPase
+ PPase: v_max / [E] ~ 1.5
K_m ~ 231
- PPase: v_max / [E] ~ 1.7
K_m ~ 298
v_0 / [E] (min^-1)
[8−oxo−dGTP] (μM)
MutT2
MutT2 + PPase
+ PPase: v_max / [E] ~ 1.7
K_m ~ 402
- PPase: v_max / [E] ~ ND
K_m ~ ND
8-oxo-dGDP
Pi
MutT1
8-oxo-dGTP
MutT2
8-oxo-dGMP
PPi

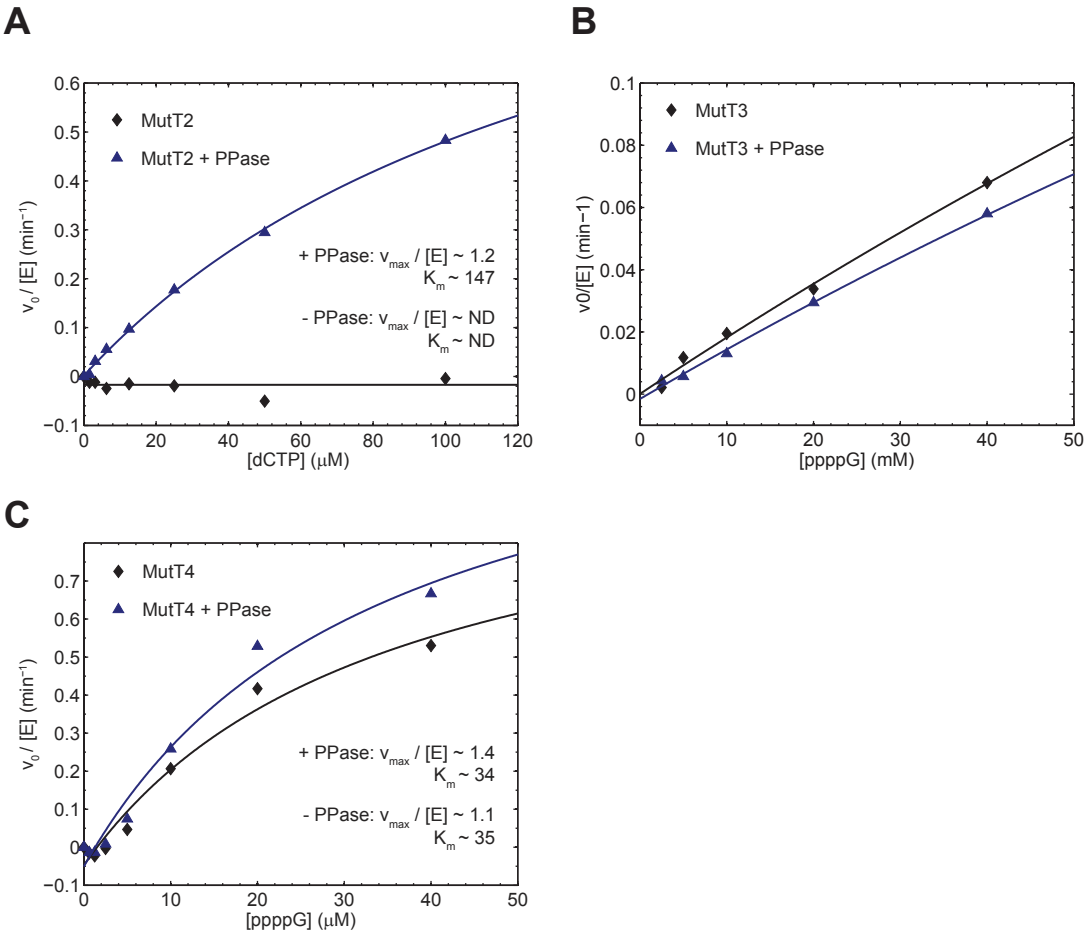
Figure 4.5: MutT1 and MutT2 hydrolyze 8-oxo-dGTP at different phosphates

(A) Activity curve for MutT1 compared to increasing concentrations of 8-oxo-dGTP in the absence (black diamonds) or presence (blue triangles) of PPase. The activity of MutT1 is identical in each condition, indicating that the reaction produces inorganic phosphate, and thus must be hydrolyzing at the γ phosphate.

(B) Activity curve for MutT2 compared to increasing concentrations of 8-oxo-dGTP in the absence (black diamonds) or presence (blue triangles) of PPase. There is no detectable activity in the absence of PPase, but robust activity with PPase, indicating that the reaction produces inorganic pyrophosphate, and thus must be hydrolyzing at the β phosphate.

(C) Schematic showing the MutT1 and MutT2 reactions.

Figure 4.6


A
MutT2
MutT2 + PPase
+ PPase: v_max / [E] ~ 1.2
K_m ~ 147
- PPase: v_max / [E] ~ ND
K_m ~ ND
v_0 / [E] (min^-1)
[dCTP] (μM)
B
MutT3
MutT3 + PPase
v0/[E] (min−1)
[ppppG] (mM)
C
MutT4
MutT4 + PPase
+ PPase: v_max / [E] ~ 1.4
K_m ~ 34
- PPase: v_max / [E] ~ 1.1
K_m ~ 35
v_0 / [E] (min^-1)
[ppppG] (μM)

D

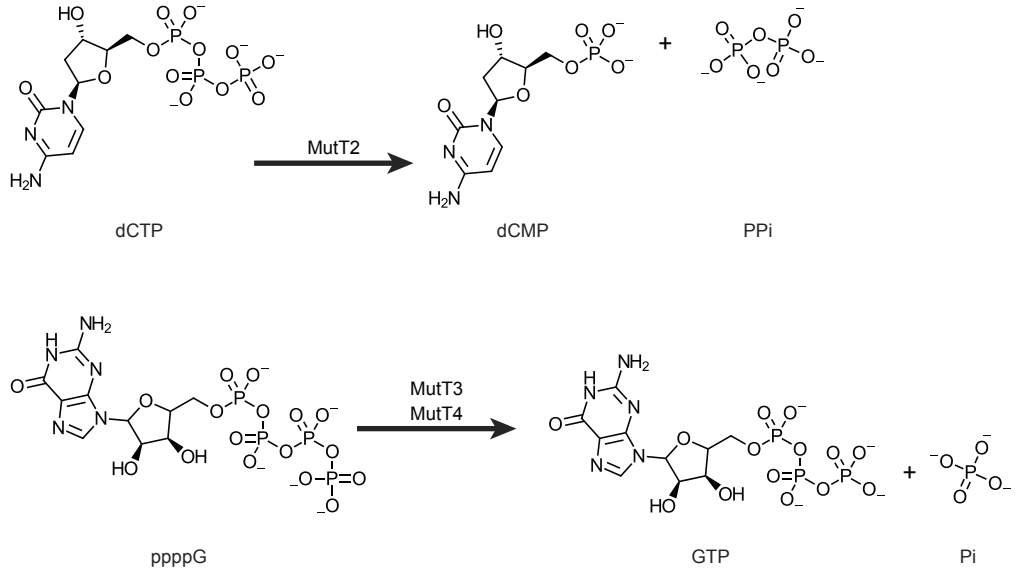

MutT2
dCTP
dCMP
PPi
MutT3
MutT4
ppppG
GTP
Pi

Figure 4.6: Analysis of other MutT reactions

(A) Activity curve for MutT2 compared to increasing concentrations of dCTP in the absence (black diamonds) or presence (blue triangles) of PPase. There is no detectable activity in the absence of PPase, but robust activity with PPase, indicating that the reaction produces inorganic pyrophosphate, and thus must be hydrolyzing at the β phosphate of dCTP.

(B) Activity curve for MutT3 compared to increasing concentrations of ppppG in the absence (black diamonds) or presence (blue triangles) of PPase. There is no detectable activity in the absence of PPase, but robust activity with PPase, indicating that the reaction produces inorganic pyrophosphate, and thus must be hydrolyzing at the δ phosphate. It should be noted, however, that the reaction rate was very low.

(C) Activity curve for MutT4 compared to increasing concentrations of ppppG in the absence (black diamonds) or presence (blue triangles) of PPase. There is no detectable activity in the absence of PPase, but robust activity with PPase, indicating that the reaction produces inorganic pyrophosphate, and thus must be hydrolyzing at the δ phosphate.

(D) Schematics showing the reactions from (A) – (C).

Figure 4.7

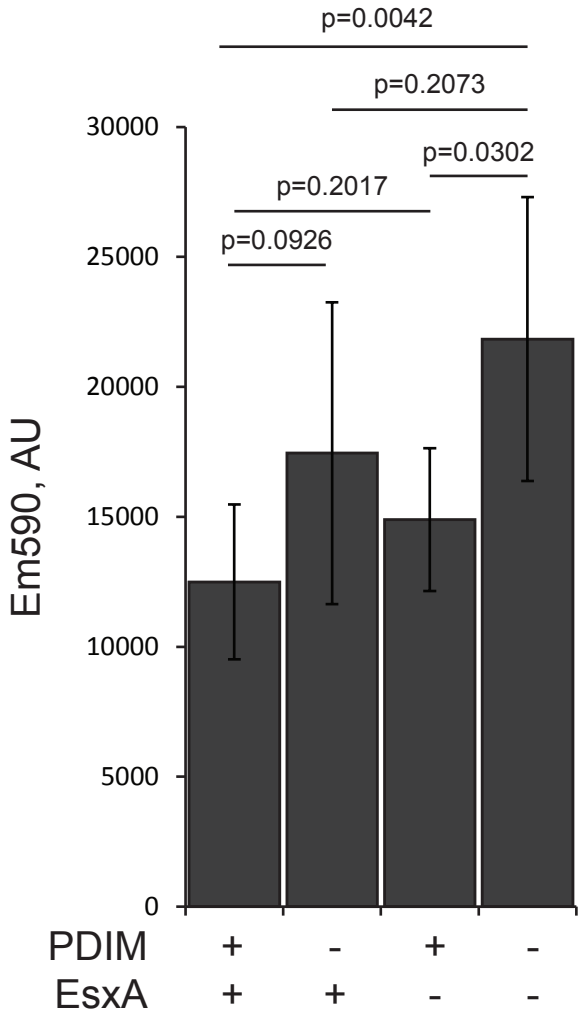

p=0.0042
p=0.2073
p=0.0302
p=0.2017
p=0.0926
30000
25000
20000
15000
10000
5000
0
Em590, AU
PDIM + - + -
EsxA + + - -

Figure 4.7: Effect of PDIM and EsxA on membrane stability

Viability of A549 cells was decreased with addition of PDIM but not EsxA. EsxA had mild effects on cell viability that were not statistically significant. EsxA was used at 50 µg/mL. Error bars represent one standard deviation above and below the mean.

Chapter 5: Conclusions and future directions

Mycobacterium tuberculosis remains the number one bacterial killer in the world, and continues to be a major global health, economic, and social burden. This pathogen, and others such as *S. aureus*, requires the Type VII Secretion system to establish infection and promote virulence. Thus, understanding the mechanisms of T7S is vital for the development of desperately needed therapeutics to combat these diseases.

This work represents the first biochemical and structural study of the conserved EccC and EsxB components of the Type VII Secretion system. We have made several important discoveries that have broad implications not only for understanding the biology of Type VII Secretion, but also for structure-based drug design, engineering novel secretion systems and substrates, and designing improved vaccine strains.

In the first part of this project, outlined in chapter two, we succeeded in creating a novel model system utilizing proteins from *Thermomonospora curvata* to study the structure and biochemistry of conserved T7S system proteins for the first time. We initially focused on the substrate / ATPase interaction, and using biochemistry and structural biology managed to characterize this interaction at the atomic level. We found that the substrate bound to the third ATPase domain of EccC in a deep hydrophobic groove. Importantly, we firmly established that the “signal sequence” for EccC binding resides on the C-terminal seven residues of EsxB, and that the Y-x-x-x-D/E motif implicated in secretion [54] is not necessary for this step. Furthermore, we found that the C-terminal seven residues of EsxB were sufficient to control substrate specificity.

Identification of the binding pocket has implications for structure-based drug design. Small molecules can be screened *in silico* for their ability to bind to the substrate-binding pocket on EccC and then tested for their ability to block EsxB binding and thus substrate secretion.

Additionally, with knowledge of the specificity signal for T7S, we can begin to design and build artificial substrates and translocases for a variety of purposes. For example, by changing the signal sequence for EsxB from the ESX-1 system to that of another T7S in the BCG vaccine strain, we could create an improved vaccine that elicits a much greater immune response in the patient.

In the second part of this project, outlined in chapter three, we developed a model for regulation of EccC by secretory substrates. We found that EccC exists in an autoinhibited state that is relieved upon displacement of a specific linker region bridging $ATPase_1$ with $ATPase_2$. Upon relief of inhibition, EccC is activated through substrate-induced multimerization. Importantly, EsxB acts as a dimerization domain, bringing together two EccC molecules which can then associate into higher order multimers, while EsxA can compete for the EsxB dimerization surface and inhibit formation of these complexes.

The details of regulation we have uncovered are suggestive of several potential avenues for drug development. First, small molecules can be screened that bind to and stabilize the autoinhibited conformation of EccC. Second, small molecules could be screened for their ability to block EccC multimerization, which we have shown is necessary for activation of the protein. And lastly, drugs could be screened for their

ability to block ATP binding at any of the three ATPase domains, which we've shown *in vivo* is necessary for secretion.

While this work has provided the first mechanistic details of how the Type VII Secretion system works, there are still many unanswered questions. For instance, if EsxA acts as a negative regulator of EccC, then how does it get secreted? A number of possibilities come to mind. EsxA may get secreted simultaneously upon inhibition of the complex, possibly as a heterodimer with EsxB. Alternatively, the inhibitory effect of EsxA may be attenuated in the context of the physiologic translocase, where additional stabilizing elements may be present. Furthermore, because these substrates play different roles in the regulation of EccC, it would be useful to know their relative concentrations in the cell and see if that changes throughout an infection cycle.

Another important unanswered question is what disrupts linker$_2$ *in vivo* to activate EccC. Obvious candidates in actinobacteria include EccB and EccD, and fragments of these proteins should be tested for their ability to activate EccC. It is also possible that other substrates may bind to this pocket, providing a mechanism that supports coordinated secretion of multiple T7S substrates. Phage display with peptide libraries may represent a useful technique to uncover what sequences can bind to each pocket, and allow for a directed search for additional EccC activators.

While this work has revealed distinct roles for each of the three ATPase domains of EccC, almost nothing is known of the role the DUF plays in secretion. Deciphering a role for the DUF may uncover more targets for drug development or reveal additional

modes of regulation for T7S. Other regions of EccC, such as the transmembrane domains, should also be examined more thoroughly as well.

Overall, while many questions remain about EccC and T7S as a whole, this work has provided the first biochemical model from which numerous hypotheses can be generated and tested. Hopefully future work will incorporate other T7S proteins in the model to build a more complete picture of this biologically and medically important secretion system.

Chapter 6: Methods

Strain and plasmid construction

Plasmids were cloned using Sequence and Ligation-Independent Cloning (SLIC). Vector backbones and gene inserts were digested for 35 minutes at room temperature with LIC-qualified T4 polymerase from Novagen as per manufacturer's instructions. The T4 polymerase was heat-killed at 75°C for 20 minutes, and digested fragments were combined with a 2:1 insert:vector ratio. The reaction was then heated to 70°C for 30 seconds and then cooled at 0.1°C/s to 40°C. The entire reaction volume was then transformed into chemically competent E. coli (strain DH5-α). Vectors used for protein expression were further isolated and transformed into either Rosetta 2 cells or C41(DE3) cells. Yeast strains were made via lithium acetate transformation.

Yeast-two-hybrid

Directed yeast-two-hybrid was performed by inserting genes of interest into both bait (pEG202) and prey (pJSC401) vectors via SLIC cloning. The resulting plasmids were transformed into yeast strains EGY48 and W303-1a. A vector containing a lacZ reporter system (pSH18) was also transformed into EGY48. Yeast were first grown separately in selective media, then mated by mixing both strains in non-selective media. Mated yeast were then plated onto double-selective media containing 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal). Images of plates were obtained using a Nikon D700 camera.

Shape-independent determination of protein size via light scattering

The molecular weight and monodispersity of EccC was determined using a Superdex 200 10/300 column and the Malvern-Viscotek TDAMax system, which has four detectors: static light scattering, refractive index, UV absorbance and an on-line viscometer. After calibration of the system with a standard protein, purified ovalbumin, the concentration of EccC was determined using either the RI (Snell's Law) or UV detectors (Beer-Lambert Law). The molecular weight from a known concentration of EccC was then calculated by static light scattering using Rayleigh Equation for small molecules.

$C_{UV} = \frac{UV}{K_{UV^*}{}^{dA}/_{dc}}$; Beer-Lambert Law

$C_{RI} = RI * \frac{RI_{sol}}{K_{RI^*}{}^{dn}/_{dc}}$; Snell's Law

$M_{avg} = \frac{K_{LS}K_{opt}LS_{90^\circ}}{RI_{sol}^2 * C * \left({}^{dn}/_{dc}\right)^2}$; Rayleigh Equation for Small Molecules

Native protein expression and purification

Bacterial cultures were inoculated to an initial OD_{600} of 0.015 from overnight starter cultures, and grown at 37°C while shaking at 225 rpm. Cells were grown in Terrific Broth to an OD of 0.4 before temperature was reduced to 20°C. Upon reaching temperature (approximately 20 minutes later), cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to grow for an additional 16 hours. Cultures were collected and centrifuged at 4000 x g for 15 minutes. Bacteria were resuspended in lysis buffer containing 2.5xPBS (3.7 mM KH_2PO_4, 38 mM Na_2HPO_4,

300 mM NaCl, 7 mM KCl) and 10 mM imidazole, then lysed with 3-5 passes through an Avestin Emulsiflex C3. Lysate was cleared by centrifugation at 235,000 x g for 1 hr. Cleared lysate was then subjected to immobilized metal affinity chromatography (IMAC) using Ni-NTA resin from Qiagen. Protein was washed with 3 column volumes of lysis buffer, 5 column volumes of wash buffer (equivalent to lysis buffer but containing 40 mM imidazole), and eluted five times with 1 column volume of elution buffer (lysis buffer / 500 mM imidazole). Elutions were checked using SDS-PAGE and coomassie staining and peak fractions were pooled and concentrated. Protein was then further purified with size exclusion chromatography (SEC). SEC was performed using a Superdex 200 16/600 column from GE Life Sciences attached to an AKTA FPLC machine. The column was pre-equilibrated with SEC buffer (20 mM Tris pH 8.0 at room temperature, 100 mM NaCl). The protein was run at a flow rate of 1.0 mL/min, and 1 mL fractions were collected. Fractions were analyzed using SDS-PAGE, and chosen fractions were collected, pooled, and concentrated. Protein was then diluted to a final volume of 50 mL with QA buffer (20 mM Tris pH 8.0 at room temperature, 15 mM NaCl). The protein was then further purified using a 1 mL Q HP column from GE Life Sciences. Protein was bound and eluted with a continuous gradient of NaCl up to 500 mM. 250 µL fractions were collected, and analyzed using SDS-PAGE. Fractions of high homogeneity were pooled, concentrated, and stored at -80°C.

For cleavage of the 8xHis tag (when necessary), 3C protease was used. The protein buffer was supplemented with 1 mM β-mercaptoethanol and 2 µg of 3C protease was added per mg of protein to be cleaved. Cleavage took place at room temperature for 1 hr or at 4°C overnight. The protease, which contains a 6xHis tag, was

removed by running the reaction over a nickel column and collecting the flow through. Cleaved protein was then re-purified with anion exchange (Q HP column) as previously described, and checked for correct size by SDS-PAGE.

SEC binding assays

SEC binding and peak shift assays were performed using a Superdex 200 10/300 column from GE Life Sciences. 1-2 mg of EccC was added to a 4-fold molar excess of either $EsxB_{Tc}$, $EsxA_{Tc}$, or $EsxAB_{Tc}$ and brought to a final volume of 250 µL in SEC buffer (20 mM Tris pH 8.0, 100 mM NaCl). The whole volume was loaded onto a 500 µL loop and injected onto the column with a flow rate of 0.5 mL/min. 250 µL fractions were collected and the peak fraction was analyzed via SDS-PAGE. Gels were stained with Coomassie and scanned on an Odyssey IR scanner from LI-COR Biosciences at 700 nm.

Mycobacterial secretion assays

Mycobacterial cultures were grown in 7H9 media to mid-log phase, centrifuged, then re-suspended in Sauton's I media to an OD_{600} of 0.05. Cultures were grown 5 days and then centrifuged and re-suspended in Sauton's II media to an OD_{600} of 0.05. Cultures were grown in Sauton's II for 5 days. Cultures were centrifuged at 3,000 x g and both pellets and supernatants were collected. Pellets were boiled for 20 minutes before removal from the BSL3, while supernatants were twice-filtered through a 0.22 µm filter. Pellets were lysed in lysis buffer through sonication, and supernatants were concentrated 100-fold. Lysate from the pellet and supernatants were run on SDS-PAGE gels and blotted onto nitrocellulose membranes. Proteins were visualized using

antibodies against $EsxB_{1Mt}$ (Rb polyclonal, Abcam ab45074). We controlled for lysis in the samples by immunoblotting for GroEL (Rb polyclonal, Stressgen ADI-SPS-875-D).

Glutaraldehyde crosslinking

Protein was first buffer exchanged into PBS using a PD-10 column (GE Healthcare Life Sciences). 2-5 µg of total protein was then incubated with 0.2% EM-grade glutaraldehyde for 10 minutes in a final volume of 25 µL. Reactions were quenched with addition of 2.5 µL of 1 M Tris pH 8.0 and run on a 7.5% or 10% polyacrylamide gel at 200 V for 33 minutes. Gels were stained with Coomassie and scanned on an Odyssey IR scanner from LI-COR Biosciences at 700 nm.

Selenomethionine-derivatized expression and purification

1L cultures were inoculated to an initial OD_{600} of 0.015 from overnight saturated starter cultures. Cells, including the starter cultures, were grown in M9 minimal media and shaken at 225 rpm at 37°C. An amino acid mix containing 50mg/L SeMet was added to growing cultures when they reached an OD_{600} of 0.5. After fifteen minutes of further growth, the temperature was reduced to 20°C. Once cells reached this temperature (approximately 20 minutes later), the bacteria were induced with 1 mM IPTG and allowed to grow for an additional 16 hours. Purification then preceded the same as the native protein, with the exception that all buffers also contained 1 mM TCEP to prevent oxidation of the selenium.

X-ray Crystallography

For the EccC crystals, fractions from the gel filtration column (without a Q column step) were concentrated to 14 mg/ml and adjusted 10 mM ATP from a 100 mM stock (Roche). The protein was mixed with a 1:1 volume of 0.1 M Tris pH 8.0, 0.6-1.1 M $(NH_4)_2SO_4$, 0.1 M glycine, 0.020 mM $MgCl_2$, 0.1 M NaCl,, equilibrated in a vapor diffusion experiment at 20°C over 500 mL of the mother liquor. The crystals grew to their full size in about 2 weeks. They were cryoprotected in 10-20% glycerol in mother liquor and a complete data set was collected at ID-24 (APS-NECAT, Argonne,IL) using the microfocus beam to collect along the long axis of the crystal in a continuous helix. Ta_6Br_{12} and Potassium tetrachloroplatinate derivative crystals were obtained by soaking the crystals for 30 min in a saturated solution of Ta_6Br_{12}, 10 mM mercury acetate or 10 mM potassium tetrachloroplatinate prepared in mother liquor and then transferring them to 20% glycerol in mother liquor for cryoprotection. SeMet crystals emerged in the same conditions as the native protein and they were treated equivalently for cyroprotection (10-20% Glycerol) and mounting. All derivative data was collected at ALS beamline 8.3.1.

For the EccCb-signal sequence crystals, peak fractions from the gel filtration were concentrated to 0.25 mM and mixed in a 1:1 molar ratio with a synthetic peptide consisting of the last 23 amino acids of the *T. curvata* EsxB homolog (Tcurv_0610) (ITYEAREEAAQQSVNRVQALLNG, Elim Biopharmaceuticals) and adjusted to 10 mM ATP with a 100 mM stock (Roche). The protein was mixed with a 1:1 volume of 0.1 M citric acid pH 5.1-5.2, 1.28-1.4 M LiCl equilibrated in a vapor diffusion experiment at 20°C over 500 mls of the mother liquor. The crystals grew to their full size in about 1

week. They were cryoprotected in 0.544 M magnesium acetate in mother liquor and a complete data set was collected at ALS beamline 8.3.1 and processed using HKL2000 [93].

For the *Gb*EssCb crystals, the selenomethionine protein was concentrated to 34 mg/ml and then adjusted to 10 mM ATP with a 100 mM stock. The protein was mixed with a 1:1 volume of 30% PEG 300, 0.1 M sodium acetate, pH 4.6 and 20% glycerol and equilibrated in a vapor diffusion experiment at 20°C over 500 mL of the mother liquor. The crystals grew to their full size in about 2 weeks. They were cryoprotected in 0.544 M magnesium acetate in mother liquor and two complete data sets were collected at ALS beamline 8.3.1 and processed using HKL2000.

For the *Tc*EsxAB crystals, selenomethionine-derivitized DDX3 and cleaved DDX4 were combined in a 1:1 molar ratio and the *Tc*EsxAB complex was purified via SEC as described previously (in the presence of 1 mM TCEP). The best crystals formed over three to four days in 70 mM sodium acetate pH 4.6, 5.6% (w/v) PEG 4000, and 30% glycerol. This condition also proved to be cryoprotective. Diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2 and processed with XDS [94] and Scala [95].

Structure Solution and Computational Methods

The structure of tantalum bromide cluster (Ta_6Br_{12}) derivative EccC was solved at 7.5 Å resolution with SAD, as implemented in SHELX C/D/E [96], which showed two molecules in the asymmetric unit and clear alpha-helices. Because of the high solvent content of the crystals, the solvent flattening protocol was very effective and provided

excellent contrast. This initial map allowed for placement of 6 FtsK (2IUU) [65] chains into the asymmetric unit using the real-space fitting algorithm of Molrep [97], though it was impossible to distinguish between the chains at this resolution. The initial model was used to calculate an anomalous difference map between the native data at 3.1 Å and the potassium tetrachloroplatinate, mercury acetate, or Ta_6Br_{12} derived crystals a 4.0 Å, 3.79 Å and 7.5 Å resolution, respectively. These maps allowed for unambiguous placement of 28 platinum atoms, 19 mercury atoms and 28 Ta_6Br_{12} clusters. Using these heavy atom positions, MIRAS phases were calculated in the program Sharp [98] and the phases were extended to 3.1 Å using 500 cycles of Solomon [99] with no mask. The protein structure was refined through sequential rounds of building in Coot [100] and refinement in Phenix [101] with assistance from a low resolution SeMet difference map to place the methionine residues. Initial refinement of the structure was limited by the poor quality of the starting model.

The structure of SeMet EssCb was determined using SAD data in Autosharp with subsequent building and refinement in Coot and Phenix to a current Rfree of 26.33%. Manual comparison of models of the three EccC ATPase domains based on the higher resolution EssCb allowed for correction of register errors in the lower resolution structure and continuation of refinement. We found that inclusion of data with a good $CC_{1/2}$ score, despite very high R_{sym} values improved the electron density maps [102].

The structure of the EccC fragment bound to the signal sequence was solved with molecular replacement of $ATPase_2$-$ATPase_3$ and placement of the C-terminal residues from the *T. curvata* EsxB crystal structure into the F_o-F_c difference density map. There are four molecules in the asymmetric unit, but molecules C and D have

much higher temperature factors and so post-refinement analysis of the structure proceeded with molecule A (which was better ordered than B) and its associated peptide.

SAD phases for EsxAB were calculated using the SHELX C/D/E. The solvent flattened maps were used to build an initial model using Coot and refined using Phenix. Secondary structure was assigned using Stride [103]. All structural figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.6 Schrödinger, LLC).

Bioinformatics

The clan of P-loop NTPases (CL0023) containing 198 member families was mapped onto the PDB using the tools available on the Pfam website (http://pfam.sanger.ac.uk/). With filtering for duplicates, this resulted in the identification of 2061 unique PDB files containing an annotation of a P-loop NTPase. We then applied a filter to each PDB to identify the position of the Walker A and Walker B motifs. In detail, the sequence of each PDB chain was extracted to single character text and searched using the program awk for the pattern “G....GK” and for the pattern "[ACFGHILMNPQSTVWY][ACFGHILMNPQSTVWY][ACFGHILMNPQSTVWY]D" to capture the Walker A and Walker B motifs on a single chain. Each residue in the PDB was analyzed for its secondary structure propensity by assigning Ramachandran angles to each residue and then calculating a score for helical or sheet propensity on a per residue basis. We calculated a score representing the average Ramachandran angles in the sequence directly downstream of the Walker A motif, where there is a conserved

α-helix, and directly upstream of the Walker B motif where there is a conserved β sheet. The cut off for determining the existence of helical and sheet secondary structure based on this score was determined empirically and may over-filter some highly divergent P-loop NTPases. This procedure resulted in capturing 4557 instances of the Walker A motif in the PDB data set. We calculated a series of measurements on these PDB chains using programs developed in an ad hoc manner by the authors, and which are available on request. First we calculated the χ-dihedral angles of the Walker A lysine. We next calculated the distance between the Nζ of Walker A lysine and whichever carboxylate oxygen was closest and between the unweighted center of mass of the Walker A lysine and the nearest heavy atom. To limit the search to PDB chains containing only ATP bound subunits we searched for chains containing an ATP ligand within 20 Å of the Walker A lysine (311 chains). To find "empty" ATPases, the original set of pdb files were filtered for files that contained a ligand with the wildcard "ade*" in a HETATM less than 20 Å from any Walker A residue. These were then sorted by distance of the ligand from the Walker A residue of a particular chain. We then chose those chains with a "ade*" HETATM in the PDB file that did not contain a "ade*" HETATM within 20 Å of the Walker A of that particular chain, resulting in a set of ATPase chains that contained a chains with and without adenine nucleotide in the same PDB file. Obvious non-ATPases were removed by hand (83 chains).

Surface area and hydrogen bonding interactions were calculated using the Pisa server [104]. T7 secretion system proteins were identified in *T. curvata* using protein BLAST [105]. EccC sequence alignments were calculated using the Concise Microbial Blast Tool (https://www.ncbi.nlm.nih.gov/genomes/prokhits.cgi) and hand curated to

separate EccC family members from FtsK family members. Proteobacteria FtsK sequence alignments were taken from the Pfam website. Logos were calculated using the WebLogo 3 web interface (http://weblogo.threeplusone.com/create.cgi).

Fluorescence Anisotropy

Fluoresence anisotropy experiments were performed using a K2 fluorometer at room temperature. In addition to specific additives mentioned in the text, samples were diluted into a buffer containing 20 mM Tris pH 8.0 and 100 mM NaCl. Anisotropy measurements were done using a 5-carboxy-fluorescein-labeled peptide with the sequence 5-FAM-VNRVQALLNG. EccC was serially diluted and added to fixed concentrations of peptide (approximately 1 μM), and each reaction was allowed to reach equilibrium before anisotropy was measured. Anisotropy of the peptide alone was also measured and used to compute anisotropy change (the arithmetic difference between measured anisotropy and the anisotropy of the free fluorescent peptide). Binding data were fit to the equation:

$$\theta = \frac{\theta_{max}\,[EccC]}{[EccC] + Kd},$$

where θ = anisotropy change, θ_{max} = maximum anisotropy change, and Kd = the dissociation constant; using the MATLAB lsqcurvefit function.

Competition assays were performed by mixing fixed concentrations of fluorescent peptide and EccC (approximately 25 μM) with varying concentrations of competitor and measuring anisotropy once equilibrium was reached. For these experiments, data were fit to the equation:

$$\theta = \theta_{max}\left[1 - \left(\frac{[Competitor]}{[Competitor] + IC_{50}}\right)\right]$$

and values for the parameters were again determined with MATLAB lsqcurvefit. In all assays, the fluorescein-bound peptide was excited with polarized 485 nm light and emission was detected at 525 nm.

Analytical Ultracentrifugation

All sedimentation experiments were performed using a Beckman Coulter analytical ultracentrifuge equipped with a sole absorption optical scanner (Optima XLA). Specific volume for each protein and buffer viscosity were estimated using Sednterp [106]. Data were analyzed using sedfit and graphed using GUSSI.

ATPase Assays

ATPase activity was measured in a coupled-ATPase assay, at saturating ATP concentration (saturation determined with the wildtype enzyme, data not shown). For measurement of the concentration dependent activity of the EccC constructs and complexes, a concentrated stock of the protein was serially diluted and the exact protein concentrations were measured using A280 absorbance. The protein dilutions (30 µl) were then added to 70 µl ATPase assay mix (containing 50 mM Tris Acetate pH 8.5, 10 mM magnesium chloride, 150 mM potassium acetate, 0.6 mM phosphoenolpyruvate, 10 mM ATP, 1 mM DTT, 0.3 mM NADH, ~30 U pyruvate kinase, ~40 U lactate dehydrogenase, and 0.2 mg/mL BSA) at room temperature and then mixed and heated to 37° C over 1 min. The A_{340} was measured in a plate reader for one hour and the slope of the line was used to calculate the rate of ATP consumption in the assay using

the following equation: $rate = -\frac{1}{\epsilon_{ATP} * l * c_{ATPase}} \frac{dA_{340}}{dt}$, where ϵ_{ATP} = 6220 $M^{-1}cm^{-1}$, c_{ATPase} = the concentration of ATPase, and l = the path length of the sample cell. The resulting number was divided by the concentration of enzyme to obtain the k_{cat}. For the titrations of EsxA and EsxB, serial dilutions of each protein were prepared and mixed with concentrations near 1 μM of the various EccC mutants (the exact concentrations as determined by A280 were used to calculate k_{cat}).

Fitting of the k_{cat} values to determine different biophysical parameters was done in MATLAB using the curve-fitting toolbox. For substrate activation, data were fit to the equation a*(x^n / (x^n + b^n)) + C; where x is substrate concentration, b represents the EC_{50}, n is the Hill coefficient, and a and C are scaling and offset parameters, respectively. For the EsxA inhibition studies, the data were fit to the equation a*(1-(x^n / (x^n + b^n))) + C; with the same parameter definitions as above.

Purification of MtEsxA

*Mt*EsxA (ESAT-6) was cloned into the pET21a expression plasmid in the NdeI and SalI sites (vector pDDX59). Another construct containing an N-terminal 8xHis tag (pH3C / SmaI) was also created, but the protein failed to express in this context. Thus, the protein was purified without an affinity tag.

1 L of cells were grown at 37°C to an OD600 of approximately 0.8, at which point they were induced with 1 mM IPTG and allowed to grow for an additional 4 hours. Cells were then spun down (4,000 rpm / 20 minutes) and resuspended in 40 mL lysis buffer (20 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM EDTA). Cells were lysed by sonication and the lysate was centrifuged (40,000 rcf / 15 minutes). The pellet, which contained the

*Mt*EsxA as inclusion bodies, was washed 1 x with wash buffer (20 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA). The pellet was then solubilized with solubilization buffer (6 M Gu-HCl, 300 mM NaCl, 1 mM EDTA, 1 x PMSF) in the cold room for approximately 1 hour using a 15 mL falcon tube and stirbar. This solution was then centrifuged in eppendorf tubes at 4°C (14,000 rpm / 10 minutes). The supernatant was removed and filtered through a 0.22 µm filter. Filtered supernatant was then placed in a 3,500 MWCO Slide-a-lyzer dialysis cassette and dialized into refolding buffer (20 mM Bis-Tris pH 6.5, 100 mM NaCl, 1 mM EDTA). Buffer was changed twice and then allowed to dialyze overnight at 4°C. Upon dialysis into refolding buffer, there was a considerable amount of precipitation. However, SDS-PAGE analysis of this precipitant reveals that it contains only non-*Mt*EsxA proteins. The soluble fraction was removed from the dialysis cassette and spun down in the cold room (14,000 rpm / 10 minutes). The supernatant was diluted to 50 mL in Buffer A (20 mM Bis-Tris pH 6.5, 15 mM NaCl, 1 mM EDTA) and run on a Q column. The protein was eluted in a continuous gradient with Buffer B (20 mM Bis-Tris pH 6.5, 500 mM NaCl, 1 mM EDTA) and was >99% pure by SDS-PAGE.

Purification of PDIM

PDIM was purified from cultures of *M. tuberculosis* Erdman strain. 900 mL cultures (grown in 7H9 media) were grown in bottles with vented caps and pre-autoclaved stir bars (stir bars set such that the top of the media was agitated). These cultures were grown to mid log phase (OD600 ~ 0.8) and then harvested via centrifugation (3,000 rpm / 5 minutes). Pellets were resuspended in ~20 mL of methanol in a glass vial and wiped out of the BSL3. The samples were then rocked for 12 hours at room temperature. The

PDIM in methanol was then centrifuged (2,700 rpm / 5 minutes) and the methanol discarded, followed by three additional washes with 10 mL methanol, and then two extractions with 10 mL petroleum ether (1 hour while rocking). The petroleum ether extracts were then transferred to a glass vial and dried under N_2 gas (can store here at -20 °C). PDIM was then resuspended in cobalt precipitation buffer (4 mL hexanes, 3 mL MeOH) and shook for 5 minutes, followed by vigorous vortexing. 1 mL of cobalt thiocyanate (3 g cobalt nitrate, 20 g ammonium thiocyanate, 100 mL water) was then added, followed by more vigorous vortexing. The sample was then centrifuged (2,700 rpm / 5 minutes) and the organic layer was saved (top layer). 4 mL hexanes were added to the aqueous layer and extracted again. Both organic phases were combined, transferred to a glass vial, and dried under N_2 gas (can store here at -20 °C). Finally, samples were further purified via TLC. PDIM was resuspended in 100 µL hexanes and spotted along a pre-scored TLC plate (Analtech Silica Gel G 2000 micron), then run in a 98:2 petroleum ether:acetone system. A scored section was broken off and dipped in a bath of 1:1 H_2SO_4:H_2O, then baked for approximately fifteen minutes. This section was aligned with the TLC plate, and the PDIM containing region was scraped off into a glass bottle, extracted with chloroform, then transferred to a pre-weighed bottle. The purified PDIM was then dried under N_2 gas and weighed.

Measurement of cell viability after treatment with PDIM and/or EsxA

96 well tissue culture plates were treated with either PDIM suspended in hexanes or hexanes alone and allowed to dry via evaporation. A549 cells were then plated and allowed to adhere and grow for 24 hours. Media was then replaced with DMEM (-phenol red)/FBS/Pen/Strep with or without EsxA at 50 µg/mL. After 24 hours, wells

were treated with 10% PresoBlue (Life Technologies) and incubated at 37°C for 20 minutes. Cell viability was measured using fluorescence (560ex / 590em).

Appendix: Supplemental Tables

Table 1: Strains and plasmids

Strain/Plasmd	Description	Resistance	Source/ref.
E. coli C41(DE3)	Expression strain	-	[107]
E. coli EL350	DH10B [lcl857 (*cro–bioA*), araC–P_{BAD}*cre*]		[108]
M. tuberculosis – Erdman	Wild-type	-	W. R. Jacobs, Jr.
M. tuberculosis Strain 13.2 (ΔeccC)	Erdman Δ*eccCa*$_1$-Δ*eccCb*$_1$		This study
pH3C	pET vector modified with an N-terminal His_8 tag- followed by a 3C protease cleavage and a T7 promoter	kan^R	[109]
pHM3C	pET vector modified with an N-terminal His_6 tag- followed by an MBP fusion with a 3C protease cleavage and a T7 promoter	kan^R	[109]
pORX2	*Thermomonaspora curvata* EccC (Tcur_0607) in pBluescript	amp^R	This study
pORX87(*Tc*EccC$_{(cyto)}$)	pH3C-*Tc*EccC$_{(199-1315)}$	kan^R	This study
pORX31 (*Tc*EccCb)	pH3C -*Tc*EccC$_{(759-1315)}$	kan^R	This study
pORX76 (*Gb*EssC$_{(cyto)}$)	pH3C-*Gb*EssC$_{(436-1479)}$	kan^R	This study
pORX71 (*Gb*EssCb)	pH3C-*Gb*EssC$_{(1204-1479)}$	kan^R	This study
pORX222	pH3C-*Tc*EccC$_{(199-1315,R543A)}$	kan^R	This study
pORX223	pH3C-*Tc*EccC$_{(199-1315,R543A,E593Q)}$	kan^R	This study
pORX225	pH3C-*Tc*EccC $_{(199-1315,R543A,D941A)}$	kan^R	This study
pORX227	pH3C-*Tc*EccC $_{(199-1315,R543A,D1213A)}$	kan^R	This study
pORX229	pH3C-*Tc*EccC $_{(199-1315,R543A,D941A,D1213A)}$	kan^R	This study
pORX230	pH3C-*Tc*EccC $_{(199-1315,R892A)}$	kan^R	This study
pMV306	Integrates at *attB* site on the *M. tuberculosis* chromosome	kan^R or hyg^R	[110]
pMSG360	attB-integrating empty vector	kan^R	[111]

pORG12 (p*Mt*EccC)	pMV306- 480 bp region up-stream of *rv3867* - *rv3870-rv3871*	kan^R	This study
pORG16	p*Mt*EccC-*Mt*EccCa$_{(K485T)}$	kan^R	This study
pORG17	p*Mt*EccC-*Mt*EccCb$_{(K90T)}$	kan^R	This study
pORG19	p*Mt*EccC- *Mt*EccCb$_{(K382T)}$	kan^R	This study
pDDG1	p*Mt*EccC-*Mt*EccCb $_{(R623A)}$	kan^R	This study
pDDG4	p*Mt*EccC-*Mt*EccCa$_{(R543A)}$	kan^R	This study
Cre pYO11	*cre* recombinase under control of *groEL* promoter in mycobacterial shuttle vector	zeo^R	This study
EGY48	Yeast-two-hybrid bait background strain: MATα	MATα; trp1, his3, ura3, lexAop-LEU2	R. Brent
W303A	Yeast-two-hybrid prey background strain: MATa	MATa; ura3-1; leu2-3,-112; his3-11,-15; trp1-1; ade2-1; can1-100	P. Walter
pEG202	LexA bait plasmid, 2µm	HIS3; amp^R	[112]
pJSC401	Prey plasmid, GAL1prom-GAL4AD, 2µm	TRP1;amp^R	[20]
pSH18	8 x lexAop-lacZ reporter, 2µm	URA3;amp^R	[112]
pDD907	*Tc*EccCa Bait	HIS3;amp^R	This study
pDD908	*Tc*EccCb Bait	HIS3; amp^R	This study
pDD909	*Tc*EccCa Prey	TRP1;amp^R	This study
pDD910	*Tc*EccCb Prey	TRP1;amp^R	This study
pDD911	*Tc*EsxB Bait	HIS3; amp^R	This study
pDD912	*Tc*EsxA Bait	HIS3; amp^R	This study
pDD913	*Tc*EsxB Prey	TRP1;amp^R	This study
pDD914	*Tc*EsxA Prey	TRP1;amp^R	This study
pDD929	*Tc*EsxB (LSSQMGF) Bait	HIS3; amp^R	This study
pDD921-TNPV	*Mt*EccCb Δ34 Bait	HIS3; amp^R	This study
pDD928-I1163T	*Tc*EccCb I1163T Bait	HIS3; amp^R	This study
pDD928-I1179T	*Tc*EccCb I1179T Bait	HIS3; amp^R	This study
pDD928-R1186A	*Tc*EccCb R1186A Bait	HIS3; amp^R	This study

pDD928-L1208T	*Tc*EccCb L1208T Bait	HIS3; amp^R	This study
pDDX1	pHM3C-*Tc*EsxB	kan^R	This study
pDDX2	pHM3C-*Tc*EsxA	kan^R	This study
pDDX3	pH3C-*Tc*EsxB	kan^R	This study
pDDX4	pH3C-*Tc*EsxA	kan^R	This study
pDDX5	pHM3C-Tcur_0612	kan^R	This study
pDDX9	pH3C-*Tc*EccB	kan^R	This study
pDDX20	pH3C-*Tc*EccC$_{(199-1315)}$-linker-*Tc*EsxB	kan^R	This study
pDDX23	pH3C-*Tc*EsxB$_{(Y84A)}$	kan^R	This study
pDDX24	pH3C-*Tc*EsxB$_{(E88A)}$	kan^R	This study
pDDX27	pH3C-*Tc*EsxB$_{(V95A)}$	kan^R	This study
pDDX28	pH3C-*Tc*EsxB$_{(N96A)}$	kan^R	This study
pDDX29	pH3C-*Tc*EsxB$_{(R97A)}$	kan^R	This study
pDDX30	pH3C-*Tc*EsxB$_{(V98A)}$	kan^R	This study
pDDX31	pH3C-*Tc*EsxB$_{(Q99A)}$	kan^R	This study
pDDX32	pH3C-*Tc*EsxB$_{(A100T)}$	kan^R	This study
pDDX33	pH3C-*Tc*EsxB$_{(L101A)}$	kan^R	This study
pDDX34	pH3C-*Tc*EsxB$_{(L102A)}$	kan^R	This study
pDDX35	pH3C-*Tc*EsxB$_{(N103A)}$	kan^R	This study
pDDX36	pH3C-*Tc*EsxB$_{(G104T)}$	kan^R	This study
pDDX38	pH3C-*Tc*EsxB$_{(+5\ Ala\ at\ C\text{-}term)}$	kan^R	This study
pDDX52	pH3C-*Tc*EccC$_{(199-1315,R543A)}$-linker-*Tc*EsxB	kan^R	This study
pDDX53	pH3C-*Tc*EccC$_{(199-1315,R543A,E593Q)}$-linker-*Tc*EsxB	kan^R	This study
pDDX57	pH3C-*Tc*EccC$_{(199-1315,R543A,R616Q)}$-linker-*Tc*EsxB	kan^R	This study
pDDX59	pET23a-*Mt*EsxA	amp^R	This study
pDDX60	pET23a-*Mt*EsxA$_{(Q55A)}$	amp^R	This study
pDDX61	pET23a-*Mt*EsxA$_{(Y51A)}$	amp^R	This study
pDDX62	pET23a-*Mt*EsxA$_{(W43A)}$	amp^R	This study
pDDX63	pH3C-*Mt*MutT1	kan^R	This study
pDDX64	pH3C-*Mt*MutT2	kan^R	This study

pDDX65	pH3C-*Mt*MutT3	kan^R	This study
pDDX66	pH3C-*Mt*MutT4	kan^R	This study

Table 2: Data Collection and Refinement Statistics for *Tc*EsxAB, *Gb*EssC, and *Tc*EccCb:*Tc*EsxB$_{(Last23)}$

Data Set	***Tc*EsxA: *Tc*EsxB$_{Tc}$ – SeMet (Peak)**	***Gb*EssC$_{(929-1479)}$ – SeMet (Peak)**	***Tc*EccCb:*Tc*EsxB(Last23)**
Additive		5 mM ATP	5 mM ATP
Space group	$C222_1$	$P2_1$	$P4_12_12$
Cell dimension a (Å)	56.25	52.34	216.15
Cell dimension b (Å)	69.81	62.97	216.15
Cell dimension c (Å)	128.41	83.88	186.51
β-angle (°)		107.47	
Molecules/asymmetric unit	1	1	4
Data Collection			
X-ray source	SSRL 12-2	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	0.9797	0.9797	1.1158
Resolution (Å)	34.9 - 2.0 (2.05 - 2.00)	49.93 - 2.45 (2.537 - 2.45)	49.15 - 3.24 (3.356 - 3.24)
R_{sym}(%)	8.6 (95)	8.6 (74)	31 (209)
I/σI	13.57 (3.32)	8.88 (1.37)	10.28 (1.32)
Completeness (%)	99.20 (99.71)	97.41 (93.79)	100 (100)
Multiplicity	9.9 (9.8)	4.2(3.6)	15.7(15.5)
$CC_{1/2}$	0.999 (.958)	0.996 (0.601)	0.997 (0.6)
Refinement			
PDB ID	4LWS	4LYA	4N1A
Number of unique reflections	17335 (1708)	18846 (1797)	70460 (6942)
R_{work} (%)	19.00 (29.98)	20.74 (31.65)	22.62 (32.71)
R_{free} (%)	23.80 (30.05)	26.33 (36.48)	26.60 (35.74)
Average B factors (Å^2)	30.60	27.10	95.7
Number of atoms (nonhydrogen)	1646	4311	17460
Stereochemistry			
Rmsd bond lengths (Å)	0.017	0.002	0.004
Rmsd bond angles (°)	1.42	0.67	0.76
Ramachandran outliers (%)	0	0	0.09
Ramachandran favored (%)	99.5	95.0	95.0

Table 3: Data Collection and Refinement Statistics for *Tc*EccC$_{(cyto)}$

Data Set (all with 5 mM ATP)	**Native**	**Ta_6Br_{12}**	**$K_2Pt\ Cl_4$**	**HgAcetate**	**SeMet SAD**
Space group	C_2	C_2	C_2	C_2	C_2
Cell dimension a (Å)	251.816	255.562	251.816	251.705	253.023
Cell dimension b (Å)	116.316	114.986	116.316	115.551	115.432
Cell dimension c (Å)	174.11	174.322	174.1	177.347	173.871
β-angle (°)	102.7	102.23	102.7	103.554	102.67
Molecules/asymmetric unit	2	2	2	2	2
Data Collection					
X-ray source	APS 24-ID	ALS 8.3.1	ALS 8.3.1	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	0.9796	1.2551	1.07216	1.00840	0.9796
Resolution (Å)	48.93 -2.9 (3.00-2.90)	50-7.56	50-4.07	50-3.79	50-4.17
R_{sym}(%)	14.2 (223)	60.1 (179)	22.3 (85.8)	21.0 (82.1)	25.5 (81.6)
I/σI	6.9 (0.6)	9.2 (1.6)	7.23 (1.04)	5 (1.2)	4.97 (1.25)
Completeness (%)	99.33 (98.72)	89.8 (91.3)	96.27 (75.11)	98.7 (98.7)	95.46 (75.88)
Multiplicity	6.9 (6.4)	2.0 (2.0)	6.8 (3.9)	3.2 (3.2)	6.4 (4.0)
$CC_{1/2}$	0.998 (0.65)				
Phasing statistics					
Sites (occ)		20	28	19	
Phasing power – acentric (iso/ano)/centric		(0.760/0.74)/0.081	(1.517/1.047)/0.321	(0.447/0.255)/0.255	
Rcullis – acentric (iso/ano)/centric		(0.633/0.954)/0.659	(0.645/0.935)/0.691	(0.842/0.996)/0.846	
Overall FoM - acentric/centric		0.13582/0.23895			
Refinement					
PDB ID	4NH0				
Number of unique reflections	108032 (10657)				
R_{work} (%)	0.2321 (0.4034)				
R_{free} (%)	0.2507 (0.4172)				
Average B factors (Å^2)	83.50				
Number of atoms (nonhydrogen)	13462				
Stereochemistry					
Rmsd bond lengths (Å)	0.005				
Rmsd bond angles (°)	0.88				
Ramachandran outliers (%)	0.18				
Ramachandran favored (%)	94				

References

[1] World Health Organization, “Global Tuberculosis Report,” 2014.

[2] R. Riley, C. Mills, W. Nyka, N. Weinstock, P. Storey, L. Sultan, M. Riley, and W. Wells, “AERIAL DISSEMINATION OF PULMONARY TUBERCULOSIS A TWO-YEAR STUDY OF CONTAGION IN A TUBERCULOSIS WARD,” *Am. J. Hyg.*, vol. 70, pp. 185–196, 1959.

[3] E. W. Tiemersma, M. J. van der Werf, M. W. Borgdorff, B. G. Williams, and N. J. D. Nagelkerke, “Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review.,” *PLoS One*, vol. 6, no. 4, p. e17601, Jan. 2011.

[4] World Health Organization, “Tuberculosis Fact Sheet N°104.” 2014.

[5] Centers for Disease Control, “Tuberculosis with Extensive Resistance to Second-Line Drugs,” *Morb. Mortal. Wkly. Rep.*, vol. 55, no. 11, 2006.

[6] L. M. Fu and C. S. Fu-Liu, “Is Mycobacterium tuberculosis a closer relative to Gram-positive or Gram–negative bacterial pathogens?,” *Tuberculosis*, vol. 82, no. 2, pp. 85–90, Dec. 2002.

[7] P. a D. Champion and J. S. Cox, “Protein secretion systems in Mycobacteria.,” *Cell. Microbiol.*, vol. 9, no. 6, pp. 1376–84, Jun. 2007.

[8] M. Sani, E. N. G. Houben, J. Geurtsen, J. Pierson, K. de Punder, M. van Zon, B. Wever, S. R. Piersma, C. R. Jiménez, M. Daffé, B. J. Appelmelk, W. Bitter, N. van der Wel, and P. J. Peters, “Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins.,” *PLoS Pathog.*, vol. 6, no. 3, p. e1000794, Mar. 2010.

[9] G. G. Mahairas, P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover, “Molecular Analysis of Genetic Differences between Mycobacterium bovis BCG and Virulent M . bovis,” vol. 178, no. 5, pp. 1274–1282, 1996.

[10] A. S. Pym, P. Brodin, R. Brosch, M. Huerre, and S. T. Cole, “Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti.,” *Mol. Microbiol.*, vol. 46, no. 3, pp. 709–17, Nov. 2002.

[11] W. Bitter, E. N. G. Houben, D. Bottai, P. Brodin, E. J. Brown, J. S. Cox, K. Derbyshire, S. M. Fortune, L.-Y. Gao, J. Liu, N. C. Gey van Pittius, A. S. Pym, E. J. Rubin, D. R. Sherman, S. T. Cole, and R. Brosch, “Systematic genetic

nomenclature for type VII secretion systems.," *PLoS Pathog.*, vol. 5, no. 10, p. e1000507, Oct. 2009.

[12] A. Pym, P. Brodin, L. Majlessi, R. Brosch, C. Demangel, A. Williams, K. Griffiths, G. Marchal, C. Leclerc, and S. Cole, "Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis," *Nat. Med.*, pp. 533–539, 2003.

[13] M. S. Siegrist, M. Steigedal, R. Ahmad, A. Mehra, M. S. Dragset, B. M. Schuster, J. a Philips, S. a Carr, and E. J. Rubin, "Mycobacterial Esx-3 requires multiple components for iron acquisition.," *MBio*, vol. 5, no. 3, pp. e01073–14, Jan. 2014.

[14] M. S. Siegrist, M. Unnikrishnan, M. J. McConnell, M. Borowsky, T.-Y. Cheng, N. Siddiqi, S. M. Fortune, D. B. Moody, and E. J. Rubin, "Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 44, pp. 18792–7, Nov. 2009.

[15] A. Serafini, D. Pisu, G. Palù, G. M. Rodriguez, and R. Manganelli, "The ESX-3 secretion system is necessary for iron and zinc homeostasis in Mycobacterium tuberculosis.," *PLoS One*, vol. 8, no. 10, p. e78351, Jan. 2013.

[16] A. Serafini, F. Boldrin, G. Palù, and R. Manganelli, "Characterization of a Mycobacterium tuberculosis ESX-3 conditional mutant: essentiality and rescue by iron and zinc.," *J. Bacteriol.*, vol. 191, no. 20, pp. 6340–4, Oct. 2009.

[17] A. M. Abdallah, J. Bestebroer, N. D. L. Savage, K. de Punder, M. van Zon, L. Wilson, C. J. Korbee, A. M. van der Sar, T. H. M. Ottenhoff, N. N. van der Wel, W. Bitter, and P. J. Peters, "Mycobacterial secretion systems ESX-1 and ESX-5 play distinct roles in host cell death and inflammasome activation.," *J. Immunol.*, vol. 187, no. 9, pp. 4744–53, Nov. 2011.

[18] D. Bottai, M. Di Luca, L. Majlessi, W. Frigui, R. Simeone, F. Sayes, W. Bitter, M. J. Brennan, C. Leclerc, G. Batoni, M. Campa, R. Brosch, and S. Esin, "Disruption of the ESX-5 system of Mycobacterium tuberculosis causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation.," *Mol. Microbiol.*, vol. 83, no. 6, pp. 1195–209, Mar. 2012.

[19] A. M. Abdallah, T. Verboom, E. M. Weerdenburg, N. C. Gey van Pittius, P. W. Mahasha, C. Jiménez, M. Parra, N. Cadieux, M. J. Brennan, B. J. Appelmelk, and W. Bitter, "PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5.," *Mol. Microbiol.*, vol. 73, no. 3, pp. 329–40, Aug. 2009.

[20] S. Stanley, S. Raghavan, W. W. Hwang, and J. S. Cox, "Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 22, pp. 13001–6, Oct. 2003.

[21] D. Houben, C. Demangel, J. van Ingen, J. Perez, L. Baldeón, A. M. Abdallah, L. Caleechurn, D. Bottai, M. van Zon, K. de Punder, T. van der Laan, A. Kant, R. Bossers-de Vries, P. Willemsen, W. Bitter, D. van Soolingen, R. Brosch, N. van der Wel, and P. J. Peters, “ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria.,” *Cell. Microbiol.*, vol. 14, no. 8, pp. 1287–98, Aug. 2012.

[22] J. I. Aguilo, H. Alonso, S. Uranga, D. Marinova, a Arbués, a de Martino, a Anel, M. Monzon, J. Badiola, J. Pardo, R. Brosch, and C. Martin, “ESX-1-induced apoptosis is involved in cell-to-cell spread of Mycobacterium tuberculosis.,” *Cell. Microbiol.*, vol. 15, no. 12, pp. 1994–2005, Dec. 2013.

[23] M. L. Burts, W. a Williams, K. DeBord, and D. M. Missiakas, “EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of Staphylococcus aureus infections.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 4, pp. 1169–74, Jan. 2005.

[24] A. Coros, B. Callahan, E. Battaglioli, and K. M. Derbyshire, “The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis.,” *Mol. Microbiol.*, vol. 69, no. 4, pp. 794–808, Aug. 2008.

[25] J. K. Fyans, D. Bignell, R. Loria, I. Toth, and T. Palmer, “The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen Streptomyces scabies.,” *Mol. Plant Pathol.*, vol. 14, no. 2, pp. 119–30, Feb. 2013.

[26] Y. M. Ohol, D. H. Goetz, K. Chan, M. U. Shiloh, C. S. Craik, and J. S. Cox, “Mycobacterium tuberculosis MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence.,” *Cell Host Microbe*, vol. 7, no. 3, pp. 210–20, Mar. 2010.

[27] M. Solomonson, P. F. Huesgen, G. a Wasney, N. Watanabe, R. J. Gruninger, G. Prehna, C. M. Overall, and N. C. J. Strynadka, “Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system.,” *J. Biol. Chem.*, vol. 288, no. 24, pp. 17782–90, Jun. 2013.

[28] K. V. Korotkov and T. J. Evans, “Structure of EccB1, a core component of the ESX-1 secretion system from Mycobacterium tuberculosis,” *To be Publ.*

[29] C. S. L. Arlehamn, J. Sidney, R. Henderson, J. a Greenbaum, E. a James, M. Moutaftsi, R. Coler, D. M. McKinney, D. Park, R. Taplitz, W. W. Kwok, H. Grey, B. Peters, and A. Sette, “Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ).,” *J. Immunol.*, vol. 188, no. 10, pp. 5020–31, May 2012.

[30] J. a MacGurn, S. Raghavan, S. Stanley, and J. S. Cox, “A non-RD1 gene cluster is required for Snm secretion in Mycobacterium tuberculosis.,” *Mol. Microbiol.*, vol. 57, no. 6, pp. 1653–63, Sep. 2005.

[31] S. Raghavan, P. Manzanillo, K. Chan, C. Dovey, and J. S. Cox, “Secreted transcription factor controls Mycobacterium tuberculosis virulence.,” *Nature*, vol. 454, no. 7205, pp. 717–21, Aug. 2008.

[32] B. Blasco, M. Stenta, L. Alonso-Sarduy, G. Dietler, M. D. Peraro, S. T. Cole, and F. Pojer, “Atypical DNA recognition mechanism used by the EspR virulence regulator of Mycobacterium tuberculosis.,” *Mol. Microbiol.*, vol. 82, no. 1, pp. 251–64, Oct. 2011.

[33] B. Blasco, J. M. Chen, R. Hartkoorn, C. Sala, S. Uplekar, J. Rougemont, F. Pojer, and S. T. Cole, “Virulence regulator EspR of Mycobacterium tuberculosis is a nucleoid-associated protein.,” *PLoS Pathog.*, vol. 8, no. 3, p. e1002621, Jan. 2012.

[34] O. S. Rosenberg, C. Dovey, M. Tempesta, R. a Robbins, J. S. Finer-Moore, R. M. Stroud, and J. S. Cox, “EspR, a key regulator of Mycobacterium tuberculosis virulence, adopts a unique dimeric structure among helix-turn-helix proteins.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 33, pp. 13450–5, Aug. 2011.

[35] M. Zhang, J. M. Chen, C. Sala, J. Rybniker, N. Dhar, and S. T. Cole, “EspI regulates the ESX-1 secretion system in response to ATP levels in Mycobacterium tuberculosis.,” *Mol. Microbiol.*, pp. 1–9, Jul. 2014.

[36] S. Stanley, J. E. Johndrow, P. Manzanillo, and J. S. Cox, “The Type I IFN Response to Infection with Mycobacterium tuberculosis Requires ESX-1-Mediated Secretion and Contributes to Pathogenesis,” *J. Immunol.*, vol. 178, no. 5, pp. 3143–3152, Feb. 2007.

[37] P. Manzanillo, M. Shiloh, D. Portnoy, and J. Cox, “Mycobacterium tuberculosis activates the DNA-dependent cytosolic surveillance pathway within macrophages,” *Cell Host Microbe*, vol. 11, no. 5, pp. 469–480, 2012.

[38] R. O. Watson, P. S. Manzanillo, and J. S. Cox, “Extracellular M. tuberculosis DNA targets bacteria for autophagy by activating the host DNA-sensing pathway.,” *Cell*, vol. 150, no. 4, pp. 803–15, Aug. 2012.

[39] M. C. Lerena and M. I. Colombo, “Mycobacterium marinum induces a marked LC3 recruitment to its containing phagosome that depends on a functional ESX-1 secretion system.,” *Cell. Microbiol.*, vol. 13, no. 6, pp. 814–35, Jun. 2011.

[40] J. Xu, O. Laine, M. Masciocchi, J. Manoranjan, J. Smith, S. J. Du, N. Edwards, X. Zhu, C. Fenselau, and L.-Y. Gao, “A unique Mycobacterium ESX-1 protein co-

secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation.," *Mol. Microbiol.*, vol. 66, no. 3, pp. 787–800, Nov. 2007.

[41] J. MacGurn and J. S. Cox, "A genetic screen for Mycobacterium tuberculosis mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system.," *Infect. Immun.*, vol. 75, no. 6, pp. 2668–78, Jun. 2007.

[42] C. Das, T. S. Ghosh, and S. S. Mande, "Computational analysis of the ESX-1 region of Mycobacterium tuberculosis: insights into the mechanism of type VII secretion system.," *PLoS One*, vol. 6, no. 11, p. e27980, Jan. 2011.

[43] P. Natale, T. Brüser, and A. J. M. Driessen, "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms.," *Biochim. Biophys. Acta*, vol. 1778, no. 9, pp. 1735–56, Sep. 2008.

[44] J. M. Silverman, Y. R. Brunet, E. Cascales, and J. D. Mougous, "Structure and regulation of the type VI secretion system.," *Annu. Rev. Microbiol.*, vol. 66, pp. 453–72, Jan. 2012.

[45] M. Sandkvist, "Type II Secretion and Pathogenesis MINIREVIEW Type II Secretion and Pathogenesis," vol. 69, no. 6, 2001.

[46] K. V Korotkov, M. Sandkvist, and W. G. J. Hol, "The type II secretion system: biogenesis, molecular architecture and mechanism.," *Nat. Rev. Microbiol.*, vol. 10, no. 5, pp. 336–51, May 2012.

[47] J. C. Leo, I. Grin, and D. Linke, "Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 367, no. 1592, pp. 1088–101, Apr. 2012.

[48] P. Delepelaire, "Type I secretion in gram-negative bacteria.," *Biochim. Biophys. Acta*, vol. 1694, no. 1–3, pp. 149–61, Nov. 2004.

[49] G. R. Cornelis, "The type III secretion injectisome.," *Nat. Rev. Microbiol.*, vol. 4, no. 11, pp. 811–25, Nov. 2006.

[50] R. Fronzes, P. J. Christie, and G. Waksman, "The structural biology of type IV secretion systems.," *Nat. Rev. Microbiol.*, vol. 7, no. 10, pp. 703–14, Oct. 2009.

[51] S. M. Fortune, a Jaeger, D. a Sarracino, M. R. Chase, C. M. Sassetti, D. R. Sherman, B. R. Bloom, and E. J. Rubin, "Mutually dependent secretion of proteins required for mycobacterial virulence.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 30, pp. 10676–81, Jul. 2005.

[52] P. D. Champion, S. Stanley, M. M. Champion, E. J. Brown, and J. S. Cox, “C-terminal signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis.,” *Science*, vol. 313, no. 5793, pp. 1632–6, Sep. 2006.

[53] M. H. Daleke, A. D. van der Woude, A. H. a Parret, R. Ummels, a M. de Groot, D. Watson, S. R. Piersma, C. R. Jiménez, J. Luirink, W. Bitter, and E. N. G. Houben, “Specific chaperones for the type VII protein secretion pathway.,” *J. Biol. Chem.*, vol. 287, no. 38, pp. 31939–47, Sep. 2012.

[54] M. H. Daleke, R. Ummels, P. Bawono, J. Heringa, C. M. J. E. Vandenbroucke-Grauls, J. Luirink, and W. Bitter, “General secretion signal for the mycobacterial type VII secretion pathway.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 28, pp. 11342–7, Jul. 2012.

[55] N. C. G. Van Pittius, J. Gamieldien, W. Hide, D. Gordon, R. J. Siezen, and A. D. Beyers, “The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G + C Gram-positive bacteria,” pp. 1–18, 2001.

[56] M. a Arbing, S. Chan, L. Harris, E. Kuo, T. T. Zhou, C. J. Ahn, L. Nguyen, Q. He, J. Lu, P. T. Menchavez, A. Shin, T. Holton, M. R. Sawaya, D. Cascio, and D. Eisenberg, “Heterologous expression of mycobacterial Esx complexes in Escherichia coli for structural studies is facilitated by the use of maltose binding protein fusions.,” *PLoS One*, vol. 8, no. 11, p. e81753, Jan. 2013.

[57] E. N. G. Houben, J. Bestebroer, R. Ummels, L. Wilson, S. R. Piersma, C. R. Jiménez, T. H. M. Ottenhoff, J. Luirink, and W. Bitter, “Composition of the type VII secretion system membrane complex.,” *Mol. Microbiol.*, vol. 86, no. 2, pp. 472–84, Oct. 2012.

[58] P. S. Renshaw, K. L. Lightbody, V. Veverka, F. W. Muskett, G. Kelly, T. a Frenkiel, S. V Gordon, R. G. Hewinson, B. Burke, J. Norman, R. a Williamson, and M. D. Carr, “Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6.,” *EMBO J.*, vol. 24, no. 14, pp. 2491–8, Jul. 2005.

[59] D. Ilghari, K. L. Lightbody, V. Veverka, L. C. Waters, F. W. Muskett, P. S. Renshaw, and M. D. Carr, “Solution structure of the Mycobacterium tuberculosis EsxG·EsxH complex: functional implications and comparisons with other M. tuberculosis Esx family complexes.,” *J. Biol. Chem.*, vol. 286, no. 34, pp. 29993–30002, Aug. 2011.

[60] L. M. Grady, J. Michtavy, and D. B. Oliver, “Characterization of the Escherichia coli SecA signal peptide-binding site.,” *J. Bacteriol.*, vol. 194, no. 2, pp. 307–16, Jan. 2012.

[61] P. I. Hanson and S. W. Whiteheart, “AAA+ proteins: have engine, will work.,” *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 7, pp. 519–29, Jul. 2005.

[62] P. Wendler, S. Ciniawsky, M. Kock, and S. Kube, “Structure and function of the AAA+ nucleotide binding pocket.,” *Biochim. Biophys. Acta*, vol. 1823, no. 1, pp. 2–14, Jan. 2012.

[63] M. Besprozvannaya, V. L. Pivorunas, Z. Feldman, and B. M. Burton, “SpoIIIE protein achieves directional DNA translocation through allosteric regulation of ATPase activity by an accessory domain.,” *J. Biol. Chem.*, vol. 288, no. 40, pp. 28962–74, Oct. 2013.

[64] X. Zhang and D. B. Wigley, “The ‘glutamate switch’ provides a link between ATPase activity and ligand binding in AAA+ proteins.,” *Nat. Struct. Mol. Biol.*, vol. 15, no. 11, pp. 1223–7, Nov. 2008.

[65] T. H. Massey, C. P. Mercogliano, J. Yates, D. J. Sherratt, and J. Löwe, “Double-stranded DNA translocation: structure and mechanism of hexameric FtsK.,” *Mol. Cell*, vol. 23, no. 4, pp. 457–69, Aug. 2006.

[66] M. J. Davey, C. Indiani, and M. O’Donnell, “Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture.,” *J. Biol. Chem.*, vol. 278, no. 7, pp. 4491–9, Feb. 2003.

[67] A. N. Page, N. P. George, A. H. Marceau, M. M. Cox, and J. L. Keck, “Structure and biochemical activities of Escherichia coli MgsA.,” *J. Biol. Chem.*, vol. 286, no. 14, pp. 12075–85, Apr. 2011.

[68] K. R. Simonetta, S. L. Kazmirski, E. R. Goedken, A. J. Cantor, A. Brian, R. Mcnally, S. N. Seyedin, D. L. Makino, M. O. Donnell, and J. Kuriyan, “The mechanism of ATP-dependent primer-template recognition by a clamp loader complex,” vol. 137, no. 4, pp. 659–671, 2009.

[69] S.-Y. Lee, A. De La Torre, D. Yan, S. Kustu, B. T. Nixon, and D. E. Wemmer, “Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains.,” *Genes Dev.*, vol. 17, no. 20, pp. 2552–63, Oct. 2003.

[70] J. P. Erzberger and J. M. Berger, “Evolutionary relationships and structural mechanisms of AAA+ proteins.,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 35, pp. 93–114, Jan. 2006.

[71] S. Lee, M. E. Sowa, Y. Watanabe, P. B. Sigler, W. Chiu, M. Yoshida, and F. T. F. Tsai, “The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state.,” *Cell*, vol. 115, no. 2, pp. 229–40, Oct. 2003.

[72] R. I. Menz, J. E. Walker, and A. G. W. Leslie, “Structure of Bovine Mitochondrial F 1 -ATPase with Nucleotide Bound to All Three Catalytic Sites : Implications for the Mechanism of Rotary Catalysis,” vol. 106, pp. 331–341, 2001.

[73] A. Peña, J. Ripoll-Rozada, S. Zunzunegui, E. Cabezón, F. de la Cruz, and I. Arechaga, “Autoinhibitory regulation of TrwK, an essential VirB4 ATPase in type IV secretion systems.,” *J. Biol. Chem.*, vol. 286, no. 19, pp. 17376–82, May 2011.

[74] T. a Baker and R. T. Sauer, “ClpXP, an ATP-powered unfolding and protein-degradation machine.,” *Biochim. Biophys. Acta*, vol. 1823, no. 1, pp. 15–28, Jan. 2012.

[75] T. L. Ramsdell, L. a Huppert, T. a Sysoeva, S. M. Fortune, and B. M. Burton, “Linked Domain Architectures Allow for Specialization of Function in the FtsK/SpoIIIE ATPases of ESX Secretion Systems.,” *J. Mol. Biol.*, Jun. 2014.

[76] R. Sundaramoorthy, P. K. Fyfe, and W. N. Hunter, “Structure of Staphylococcus aureus EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein.,” *J. Mol. Biol.*, vol. 383, no. 3, pp. 603–14, Nov. 2008.

[77] L. a Huppert, T. L. Ramsdell, M. R. Chase, D. a Sarracino, S. M. Fortune, and B. M. Burton, “The ESX system in Bacillus subtilis mediates protein secretion.,” *PLoS One*, vol. 9, no. 5, p. e96267, Jan. 2014.

[78] T. a Sysoeva, M. a Zepeda-Rivera, L. a Huppert, and B. M. Burton, “Dimer recognition and secretion by the ESX secretion system in Bacillus subtilis.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 21, pp. 7653–8, May 2014.

[79] M. Valko, M. Izakovic, M. Mazur, C. J. Rhodes, and J. Telser, “Role of oxygen radicals in DNA damage and cancer incidence,” *Mol. Cell. Biochem.*, vol. 266, no. 1/2, pp. 37–56, Nov. 2004.

[80] a G. McLennan, “The Nudix hydrolase superfamily.,” *Cell. Mol. Life Sci.*, vol. 63, no. 2, pp. 123–43, Jan. 2006.

[81] H. Maki and M. Sekiguchi, “MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis,” *Nature*, vol. 355, no. 6357, pp. 273–275, Jan. 1992.

[82] T. Dos Vultos, J. Blázquez, J. Rauzier, I. Matic, and J. Bla, “Identification of Nudix Hydrolase Family Members with an Antimutator Role in Mycobacterium tuberculosis and Mycobacterium smegmatis Identification of Nudix Hydrolase Family Members with an Antimutator Role in Mycobacterium tuberculosis and Mycobacterium s,” 2006.

[83] M. A. Kohanski, D. J. Dwyer, B. Hayete, C. A. Lawrence, and J. J. Collins, “A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics,” *Cell*, vol. 130, no. 5, pp. 797–810, Dec. 2014.

[84] D. Dwyer, D. Camacho, and M. Kohanski, “Antibiotic-induced bacterial cell death exhibits physiological and biochemical hallmarks of apoptosis,” *Mol. Cell*, vol. 46, no. 5, pp. 561–572, 2012.

[85] H. Gad, T. Koolmeister, A.-S. Jemth, S. Eshtad, S. a Jacques, C. E. Ström, L. M. Svensson, N. Schultz, T. Lundbäck, B. O. Einarsdottir, A. Saleh, C. Göktürk, P. Baranczewski, R. Svensson, R. P. Berntsson, R. Gustafsson, K. Strömberg, K. Sanjiv, M.-C. Jacques-Cordonnier, M. Desroses, A.-L. Gustavsson, R. Olofsson, F. Johansson, E. J. Homan, O. Loseva, L. Bräutigam, L. Johansson, A. Höglund, A. Hagenkort, T. Pham, M. Altun, F. Z. Gaugaz, S. Vikingsson, B. Evers, M. Henriksson, K. S. a Vallin, O. a Wallner, L. G. J. Hammarström, E. Wiita, I. Almlöf, C. Kalderén, H. Axelsson, T. Djureinovic, J. C. Puigvert, M. Häggblad, F. Jeppsson, U. Martens, C. Lundin, B. Lundgren, I. Granelli, A. J. Jensen, P. Artursson, J. a Nilsson, P. Stenmark, M. Scobie, U. W. Berglund, and T. Helleday, “MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool.,” *Nature*, vol. 508, no. 7495, pp. 215–21, Apr. 2014.

[86] A. G. G. Patil, P. B. Sang, A. Govindan, and U. Varshney, “Mycobacterium tuberculosis MutT1 (Rv2985) and ADPRase (Rv1700) proteins constitute a two-stage mechanism of 8-oxo-dGTP and 8-oxo-GTP detoxification and adenosine to cytidine mutation avoidance.,” *J. Biol. Chem.*, vol. 288, no. 16, pp. 11252–62, Apr. 2013.

[87] P. B. Sang and U. Varshney, “Biochemical properties of MutT2 proteins from Mycobacterium tuberculosis and M. smegmatis and their contrasting antimutator roles in Escherichia coli.,” *J. Bacteriol.*, vol. 195, no. 7, pp. 1552–60, Apr. 2013.

[88] D. Sherman, M. Voskuil, D. Schnappinger, R. Liao, M. Harrell, and G. Schoolnik, “Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding α-crystallin,” *Proc. …*, vol. 98, no. 26, 2001.

[89] J. Smith, J. Manoranjan, M. Pan, A. Bohsali, J. Xu, J. Liu, K. L. McDonald, A. Szyk, N. LaRonde-LeBlanc, and L.-Y. Gao, “Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in Mycobacterium marinum escape from the vacuole.,” *Infect. Immun.*, vol. 76, no. 12, pp. 5478–87, Dec. 2008.

[90] A. Kinhikar and I. Verma, “Potential role for ESAT6 in dissemination of M. tuberculosis via human lung epithelial cells,” *Mol. …*, vol. 75, no. 1, pp. 92–106, 2010.

[91] J. S. Cox, B. Chen, M. McNeil, and W. R. Jacobs, “Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice,” *Nature*, vol. 402, no. 6757, pp. 79–83, Nov. 1999.

[92] J. Rybniker, J. M. Chen, C. Sala, R. C. Hartkoorn, A. Vocat, A. Benjak, S. Boy-Röttger, M. Zhang, R. Székely, Z. Greff, L. Őrfi, I. Szabadkai, J. Pató, G. Kéri, and S. T. Cole, “Anticytolytic Screen Identifies Inhibitors of Mycobacterial Virulence Protein Secretion,” *Cell Host Microbe*, vol. 16, no. 4, pp. 538–548, Oct. 2014.

[93] Z. Otwinowski and W. Minor, “Processing of X-ray Diffraction Data Collected in Oscillation Mode ", Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, pp.307-326, C.W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).” 1997.

[94] W. Kabsch, “Integration, scaling, space-group assignment and post-refinement.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 133–44, Feb. 2010.

[95] H. R. Leslie, A.G., and Powell, “Evolving Methods for Macromolecular Crystallography. R.J. Read, and J.L. Sussman, eds. pp. 41–51.” 2007.

[96] G. M. Sheldrick, “Phase annealing in SHELX-90: direct methods for larger structures,” *Acta Crystallogr. Sect. A Found. Crystallogr.*, vol. 46, no. 6, pp. 467–473, Jun. 1990.

[97] A. Vagin and A. Teplyakov, “Molecular replacement with MOLREP.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 1, pp. 22–5, Jan. 2010.

[98] G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, and W. Paciorek, “Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 59, no. 11, pp. 2023–2030, Oct. 2003.

[99] J. P. Abrahams and a G. Leslie, “Methods used in the structure determination of bovine mitochondrial F1 ATPase.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 52, no. Pt 1, pp. 30–42, Jan. 1996.

[100] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 486–501, Apr. 2010.

[101] P. D. Adams, P. V Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, “PHENIX: a comprehensive Python-based system for macromolecular structure solution.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 213–21, Feb. 2010.

[102] P. A. Karplus and K. Diederichs, “Linking crystallographic model and data quality.,” *Science*, vol. 336, no. 6084, pp. 1030–3, May 2012.

[103] M. Heinig and D. Frishman, “STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins.,” *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W500–2, Jul. 2004.

[104] E. Krissinel and K. Henrick, “Inference of macromolecular assemblies from crystalline state,” *J. Mol. Biol.*, vol. 43, 2007.

[105] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[106] P. Schuck, “Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling,” vol. 78, no. March, pp. 1606–1619, 2000.

[107] B. Miroux and J. E. Walker, “Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.,” *J. Mol. Biol.*, vol. 260, no. 3, pp. 289–98, Jul. 1996.

[108] E. C. Lee, D. Yu, J. Martinez de Velasco, L. Tessarollo, D. a Swing, D. L. Court, N. a Jenkins, and N. G. Copeland, “A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA.,” *Genomics*, vol. 73, no. 1, pp. 56–65, Apr. 2001.

[109] J. Hammon, D. V Palanivelu, J. Chen, C. Patel, and D. L. Minor, “A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms.,” *Protein Sci.*, vol. 18, no. 1, pp. 121–33, Jan. 2009.

[110] C. K. Stover, V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs, and B. R. Bloom, “New use of BCG for recombinant vaccines,” *Nature*, vol. 351, no. 6326, pp. 456–460, Jun. 1991.

[111] D. Barkan, V. Rao, G. D. Sukenick, and M. S. Glickman, “Redundant function of cmaA2 and mmaA2 in Mycobacterium tuberculosis cis cyclopropanation of oxygenated mycolates.,” *J. Bacteriol.*, vol. 192, no. 14, pp. 3661–8, Jul. 2010.

[112] E. A. Golemis, I. Serebriiskii, R. L. Finley, M. G. Kolonin, J. Gyuris, and R. Brent, “Interaction Trap/Two-Hybrid System to Identify Interacting Proteins,” in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., 2001.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

18 December 2014

Author Signature

Date