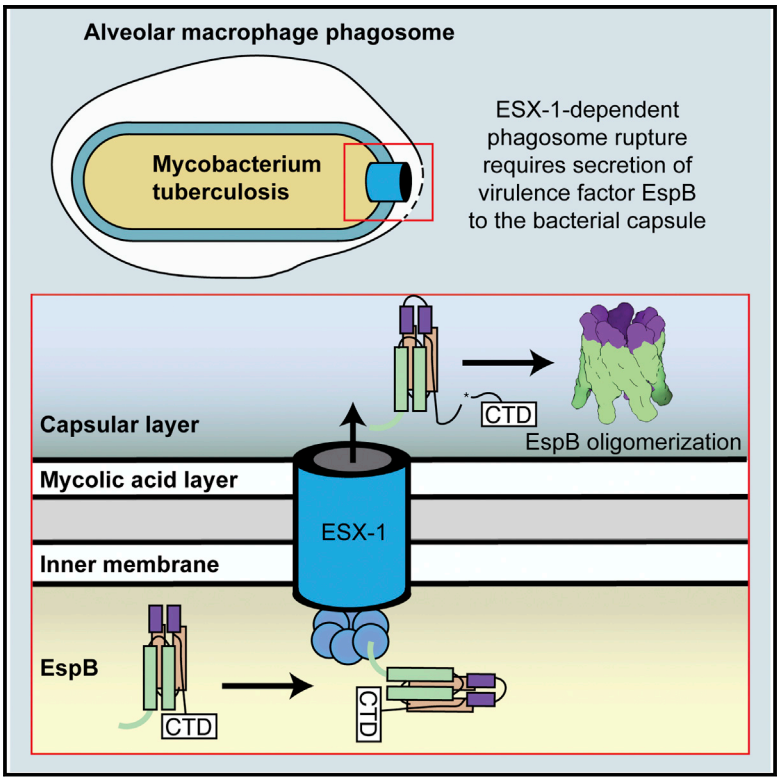
Article

Structure

# Structure of EspB from the ESX-1 Type VII Secretion System and Insights into its Export Mechanism

## Graphical Abstract




## Highlights

- The crystal structure of EspB reveals a fused PE/PPE homology domain
- EspB has a stabilized bipartite secretion signal that targets the EccCb1 ATPase
- EspB oligomerizes to form a ring-shaped heptamer
- A model of the heptamer was fit to EM density and crosslinking data

## Authors

Matthew Solomonson, Dheva Setiaputra, ..., Calvin K. Yip, Natalie C.J. Strynadka

## Correspondence

ncjs@mail.ubc.ca

## In Brief

*Mycobacterium tuberculosis* exports virulence factors to its surface using the ESX-1 secretion system, progressing the infection of human macrophages. Solomonson et al. show that one of these factors, EspB, adopts a PE/PPE-like fold and oligomerizes to form a barrel-shaped structure with heptameric symmetry.

## Accession Numbers

4WJ1
4WJ2
3J83

Solomonson et al., 2015, Structure *23*, 1–13
March 3, 2015 
http://dx.doi.org/10.1016/j.str.2015.01.002

CellPress

Please cite this article in press as: Solomonson et al., Structure of EspB from the ESX-1 Type VII Secretion System and Insights into its Export Mechanism, Structure (2015), http://dx.doi.org/10.1016/j.str.2015.01.002

Structure
Article



# Structure of EspB from the ESX-1 Type VII Secretion System and Insights into its Export Mechanism

Matthew Solomonson,[1,2] Dheva Setiaputra,[1] Karl A.T. Makepeace,[3] Emilie Lameignere,[1,2] Evgeniy V. Petrotchenko,[3] Deborah G. Conrady,[1,2] Julien R. Bergeron,[1,2] Marija Vuckovic,[1,2] Frank DiMaio,[4] Christoph H. Borchers,[3] Calvin K. Yip,[1] and Natalie C.J. Strynadka[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Life Sciences Centre, The University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

[2]Centre for Blood Research, Life Sciences Centre, The University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

[3]University of Victoria Proteomics Centre, 3101-4464 Markham Street, Victoria, BC V8Z 7X8, Canada

[4]Biochemistry Department, University of Washington, J557 Health Sciences Building, Seattle, WA 98195-7370, USA

*Correspondence: ncjs@mail.ubc.ca

http://dx.doi.org/10.1016/j.str.2015.01.002

## SUMMARY

***Mycobacterium tuberculosis* (*Mtb*) uses the ESX-1 type VII secretion system to export virulence proteins across its lipid-rich cell wall, which helps permeabilize the host's macrophage phagosomal membrane, facilitating the escape and cell-to-cell spread of *Mtb*. ESX-1 membranolytic activity depends on a set of specialized secreted Esp proteins, the structure and specific roles of which are not currently understood. Here, we report the X-ray and electron microscopic structures of the ESX-1-secreted EspB. We demonstrate that EspB adopts a PE/PPE-like fold that mediates oligomerization with apparent heptameric symmetry, generating a barrel-shaped structure with a central pore that we propose contributes to the macrophage killing functions of EspB. Our structural data also reveal unexpected direct interactions between the EspB bipartite secretion signal sequence elements that form a unified aromatic surface. These findings provide insight into how specialized proteins encoded within the ESX-1 locus are targeted for secretion, and for the first time indicate an oligomerization-dependent role for Esp virulence factors.**

## INTRODUCTION

Mycobacteria possess a unique hydrophobic cellular envelope that confers resistance to various stresses and contributes to the resilience of pathogens such as *Mycobacterium tuberculosis* (*Mtb*) (Niederweis et al., 2009). The mycobacterial cell wall is composed of an inner membrane, peptidoglycan/arabinogalactan layers, a mycolic acid-containing layer, and an outer capsular layer (Sani et al., 2010) (Figure 1A). To export proteins across these physical barriers, mycobacteria utilize the general secretion pathway, twin-arginine transporter, and up to five distinct ESX secretion systems (designated ESX-1 through ESX-5, referred to as the type VII secretion system), which variously function in virulence (ESX-1), iron acquisition (ESX-3), and cell surface decoration (ESX-5) (Champion and Cox, 2007). The ESX-1 system of pathogenic bacteria is of particular interest due to its involvement in essential pathogenesis steps including host phagosomal permeabilization (Watson et al., 2012), bacterial escape to cytosol (van der Wel et al., 2007), and macrophage killing (Simeone et al., 2012). Intriguingly, the ESX-1 locus involved in DNA conjugation in the non-pathogenic *Mycobacterium smegmatis* is nearly syntenic and to some extent interchangeable with the virulence locus (Wirth et al., 2011), making *M. smegmatis* ESX-1 a useful model system.

Despite its central role in pathogenesis, the structure of the ESX apparatus remains poorly characterized. The core inner membrane export machinery consists of an EccC ATPase, a membrane-bound MycP protease that processes secreted proteins (Ohol et al., 2010), and several other transmembrane apparatus proteins (Abdallah et al., 2007) (Figure 1A). Each locus also encodes a more variable set of components such as chaperones, ATPases, transcription factors, and, crucially, a specific set of secreted proteins (Houben et al., 2013).

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

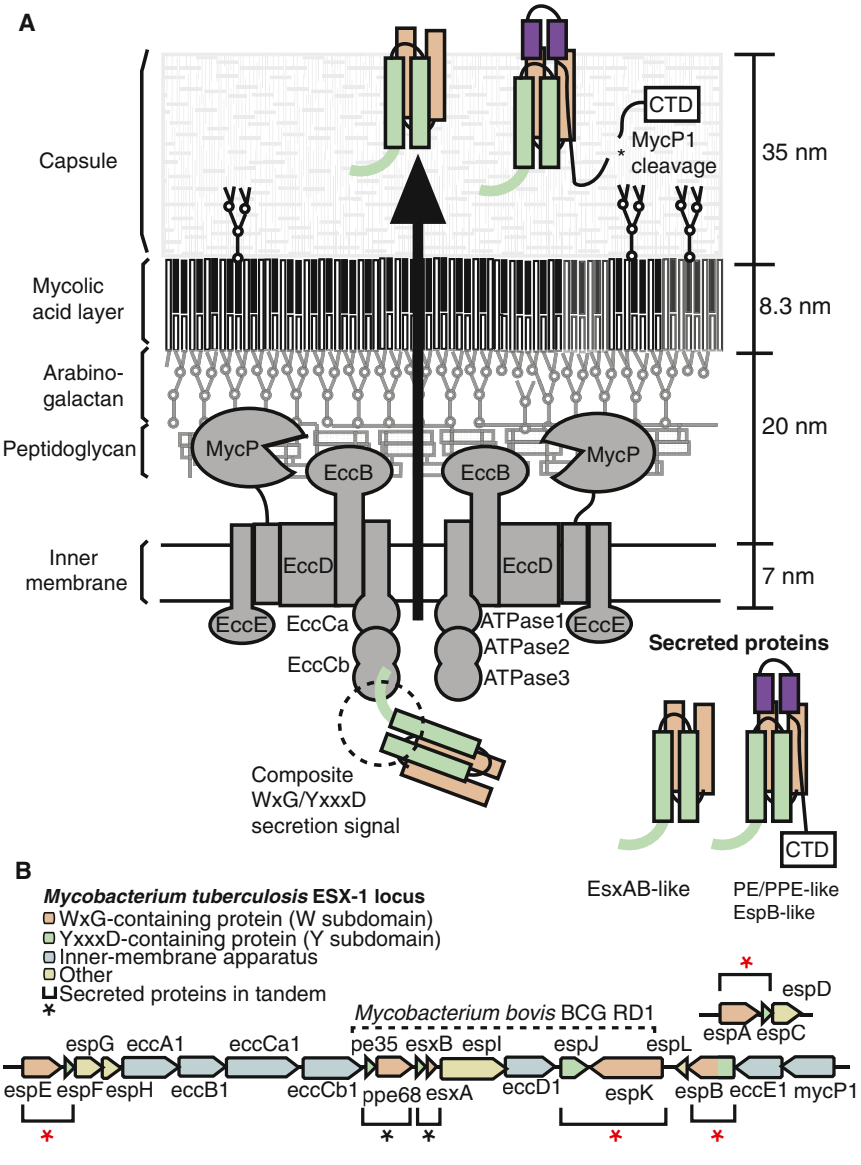



**Figure 1. Overview of ESX Secretion in Mycobacteria**

(A) Schematic summary of inner membrane secretion complex, secreted proteins, and mycobacterial cell wall composition and dimensions (schematic adapted from Brennan and Crick, 2007; Houben et al., 2012; Sani et al., 2010).

(B) The *M. tuberculosis* ESX-1 virulence locus, with tandemly organized secreted proteins that together possess the full composite secretion signal marked by asterisks.

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

EspB in particular is directly involved in ESX-1 membranolytic function, and strains lacking this secreted protein are as attenuated for virulence as those lacking the entire secretion apparatus (Gao et al., 2004). Knocking out EspB in the related pathogen *Mycobacterium marinum* abrogates red blood cell hemolysis (Gao et al., 2004), eliminates macrophage cytotoxicity (Gao et al., 2004), inhibits intracellular growth and phagosome maturation (McLaughlin et al., 2007; Xu et al., 2007; Huang and Bao, 2014), and completely attenuates virulence in zebrafish infection models (Gao et al., 2004). A transposon mutant of EspB also decreases *M. smegmatis* conjugation efficiency by 1000-fold, suggesting a role in conjugal transfer as well (Coros et al., 2008). EspB has a C-terminal domain (CTD) that is processed by the membrane-bound protease, mycosin-1 (MycP1), following secretion (McLaughlin et al., 2007; Ohol et al., 2010; Solomonson et al., 2013; Wagner et al., 2013; Xu et al., 2007) (Figure 1A). The purpose of this cleavage event is unknown but is required for full pathogenesis of *Mtb*, and has been reported to modulate the quantity of protein secreted by ESX-1 (Ohol et al., 2010) and/or regulate phospholipid binding by EspB (Chen et al., 2013b).

Despite this dual role in regulating and mediating ESX-1 virulence, structural data are lacking for EspB or any other secreted Esp protein in the ESX-1 locus. Moreover, the mechanism by which Esp proteins are secreted through ESX-1 is not entirely clear. Recent work has culminated in the identification of a general "secretion signal" that targets cytoplasmic ESX substrates, including EspB, to the secretion apparatus for export across the inner membrane (Champion et al., 2006; Daleke et al., 2012a) (Figure 1A). The structure of this secretion signal has primarily been studied in the simplest ESX-secreted substrates of the WXG100 family such as EsxA and EsxB (Arbing et al., 2010, 2013; Ilghari et al., 2011; Poulsen et al., 2014; Renshaw et al., 2005). WXG100 family members are secreted as homo- or heterodimeric complexes that adopt elongated, antiparallel helical bundles, where a WxG motif is present in at least one member of the heterodimer (labeled "W subdomain" and colored orange in Figures 1 and 2). This WxG motif forms one half of a bipartite composite secretion signal (Sysoeva et al., 2014) that is thought to target the heterodimeric complex to EccC for export (Champion et al., 2006) (Figure 1A). The second member of the heterodimer contributes the other half of this composite secretion signal: a flexible C-terminal "export arm" that harbors a characteristic Tyr-x-x-x-[Asp/Glu] (YxxxD) motif and at least one additional conserved hydrophobic residue seven positions downstream of the YxxxD motif

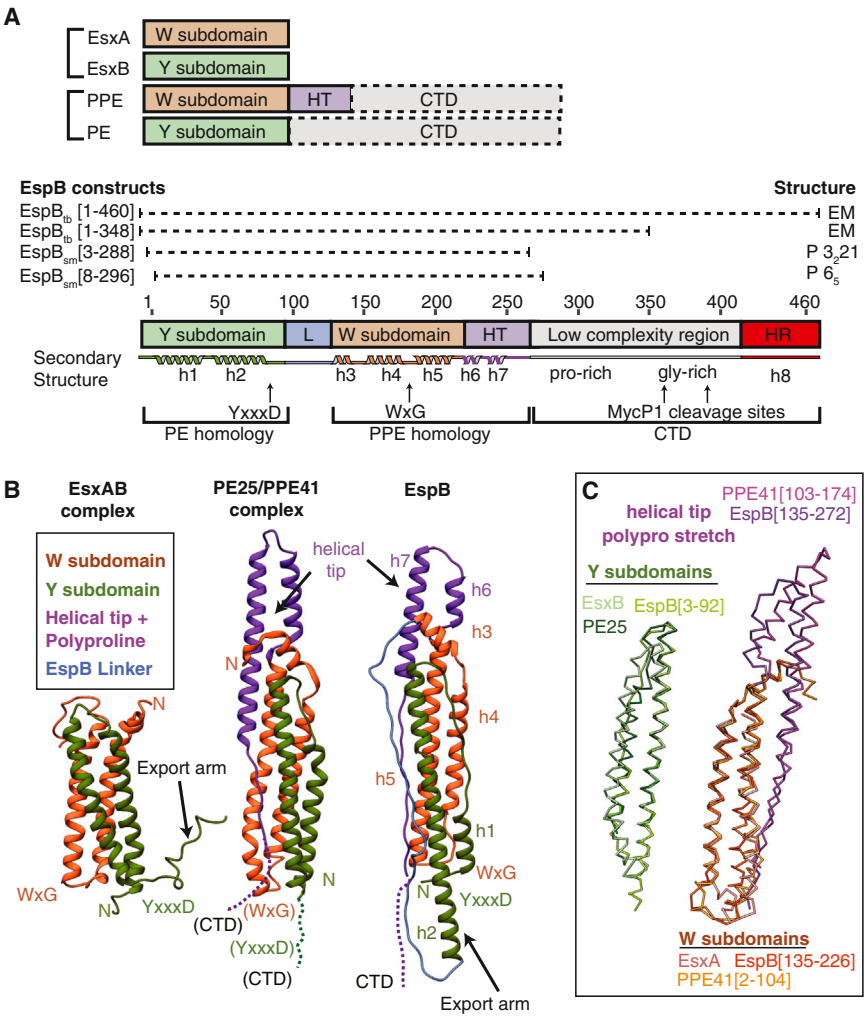



**Figure 2. Crystal Structure of *M. smegmatis* EspB Compared with Model ESX Substrates EsxAB and PE25/PPE41**

(A) Summary of ESX-secreted proteins discussed in this paper. W subdomain indicates presence of WxG motif, Y subdomain indicates presence of YxxxD motif. L, linker; HR, homology region; HT, helical tip.

(B) Structures of representative members of Esx (1WA8) and PE/PPE (2G38) families for comparison with EspB, color-coded by feature.

(C) Structural superposition of isolated Y and W subdomains across ESX-secreted protein classes demonstrates conserved modularity.

(D) Surface representation of EspB Y and W subdomains, with the linker, helical tip, polyproline stretch, and aromatic patch regions shown as ribbon and/or sticks.

(Poulsen et al., 2014) (this member is labeled "Y subdomain" and colored green in Figures 1 and 2). Previous nuclear magnetic resonance (NMR) (Renshaw et al., 2005), crystallographic (Arbing et al., 2013; Poulsen et al., 2014), and two-hybrid studies (Champion et al., 2006) indicate that this dynamic export arm is stabilized such that residues align on the same face to interact with the EccC ATPase, representing a key phase of the export cycle and a potential determinant of ESX substrate specificity (Champion et al., 2009).

The final class of ESX-secreted protein relevant to the work presented here are the PE and PPE protein subfamilies, named for the characteristic Pro-Glu or Pro-Pro-Glu motifs located in the N-terminal regions of the respective variants (Sampson, 2011). Like EspB, PE and PPE proteins also possess CTDs, which encode for low-complexity repetitive structures hypothesized to promote antigenic variation at the cell surface. Crystallographic analysis of the PE25/PPE41 complex from ESX-5 revealed a helical bundle reminiscent of the EsxAB heterodimer (Strong et al., 2006) (Figure 2B).

Here, we describe for the first time the crystal structure of EspB from *M. smegmatis* ($EspB_{sm}$), which reveals structural elements common to both the EsxAB and PE/PPE family of proteins, with a structured view of the composite secretion signal that clearly classifies it as a member of the WXG100 superfamily. $EspB_{sm}$ adopts a distinct homomeric helical bundle and an additional customized appendage that, like the PE/PPE heterodimers, replaces one of the dual signal sequences present in canonical ESX substrates. Our light scattering and negative stain electron microscopy (EM) data further show that EspB from *Mtb* ($EspB_{tb}$) has a propensity to oligomerize, revealing the potential for EspB to serve as structural subunits in the construction of cell wall-associated architectures of mycobacteria as widely hypothesized (Champion et al., 2009; Fortune et al., 2005).

## RESULTS

### Crystal Structure of EspB

The X-ray crystallographic structure of the N-terminal helical domain of EspB from *M. smegmatis* was determined in two crystal forms, one encompassing residues 3–288 (space group $P3_221$, PDB code 4WJ1) and a second crystal form encompassing residues 8–296 (space group $P6_5$, PDB code 4WJ2 (Table 1). Due to predicted disorder, the CTD was not included in the construct used to grow the $P3_221$ form (Figure 2A), and this CTD was also not observed in the $P6_5$ form even though it was included in the construct, likely due to proteolytic degradation

**Table 1. Data Collection and Refinement Statistics**

| | *M. smegmatis* EspB [3–288] Se-Met SAD | *M. smegmatis* EspB [8–296] Native MR |
|---|---|---|
| **Data Collection** | | |
| Space group | P $3_2$ 2 1 | P $6_5$ |
| Cell dimensions | | |
| *a*, *b*, *c* (Å) | 125.75, 125.75, 75.06 | 148.29, 148.29, 49.69 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å)[a] | 61.8–2.41 (2.51–2.41) | 48.5–2.80 (2.95–2.8) |
| $CC_{1/2}$ | 0.999 (0.972) | 0.997 (0.845) |
| $R_{pim}$ | 0.026 (0.137) | 0.054 (0.309) |
| $R_{meas}$ | 0.121 (0.627) | 0.145 (0.852) |
| $I/\sigma I$ | 22.8 (5.6) | 12.9 (2.3) |
| Completeness (%) | 100.0 (100.0) | 97.6 (99.3) |
| Redundancy | 21.8 (20.7) | 7.1 (7.4) |
| **Refinement** | | |
| Resolution (Å) | 61.8–2.41 | 48.5–2.8 |
| No. of reflections | 26,584 | 15,287 |
| $R_{work}/R_{free}$[b] | 0.1755/0.2011 | 0.2378/0.2718 |
| No. of atoms | | |
| Protein | 2162 | 2039 |
| Ligand/ion | 0 | 0 |
| Water | 152 | 41 |
| B factors | | |
| Protein | 48 | 57 |
| Ligand/ion | 0 | 0 |
| Water | 51 | 64 |
| Rms deviations | | |
| Bond lengths (Å) | 0.008 | 0.010 |
| Bond angles (°) | 1.033 | 1.015 |
| Ramachandran | | |
| % Favored | 99 | 96 |
| % Allowed | 1 | 4 |
| % Outliers | 0 | 0 |
| Molprobity | | |
| Clashscore | 6 | 10 |

[a]Values in parentheses represent the highest-resolution shell.
[b]10% of reflections were excluded from refinement and used to calculate $R_{free}$.

or disorder within the large crystal solvent channels of this crystal form (~80% solvent). Although highly similar in overall architecture (0.69 Å backbone root-mean-square [rms] deviation), the two structures display significant differences in functionally relevant regions (discussed below). Unless noted, we describe structural features based on the P$3_2$21 form, as this structure was refined to higher resolution (2.4 Å).

$EspB_{sm}$ (residues 3–288, Figure 2A) is composed of long, antiparallel coiled-coil helices characteristic of a PE/PPE heterodimer (Figure 2B, right). Unlike other ESX-associated proteins, $EspB_{sm}$ is not a heterodimer but rather adopts a PE/PPE-like fold due to an apparent genetic fusion of heterodimeric members into a single open reading frame by a 35-residue linker (Figure 2A). A side-by-side comparison of $EspB_{sm}$ with the crystal structures of the ESX-secreted substrates EsxAB and PE25/PPE41 reveals the modular nature of ESX-secreted proteins (Figures 2B and 2C). The helix-turn-helix domains of PE25, EsxB, and the N-terminal domain of $EspB_{sm}$ residues 1–92 are roughly superimposable, each with a characteristic YxxxD motif at the turn (this subdomain is labeled "Y subdomain" and colored green in Figures 1 and 2.) $EspB_{sm}$ residues 135–226, EsxA, and PPE41 residues 2–104 are also roughly superimposable, each with a WxG motif at the turn of the second helix-turn-helix subdomain (labeled "W subdomain" and colored orange in Figures 1 and 2). In $EspB_{sm}$, the Y and W subdomains form an extensive interface of non-covalent interactions, and the linker connecting them adopts an extended conformation that anchors loosely to the helical bundle through van der Waals interactions (Figure 2D)

Compared with PE25/PPE41, $EspB_{sm}$ has a shorter h1 helix, with residues in this position instead adopting an extended conformation (Figure 2B). $EspB_{sm}$ and PE25/PPE41 possess features that are not found in EsxAB. Notably, helix 3, helix 6, and helix 7 of $EspB_{sm}$ (residues 231–269) create a "helical tip" that is rich in solvent-exposed hydrophobic/aromatic/acidic residues (Figure 2D, purple). This helical tip is slightly smaller in $EspB_{sm}$ than in PE25/PPE41. The tip ends in an extended polyproline stretch (residues 270–288) that sandwiches between the Y and W subdomains through hydrophobic interactions provided by prolines and aliphatic side chains (Figure 2D). Sequence comparisons suggest that in $EspB_{sm}$ and many PPE family proteins, this helical tip is connected via the polyproline stretch to the CTD of varying sequence (Figure 2D).

## EspB Has a Bipartite Signal Sequence that Targets the EccCb1 ATPase

The $EspB_{sm}$ crystal structure also reveals the unique relative disposition of multiple characteristic ESX sequence motifs, including a structured view of the helical export arm (Figure 3A, middle). In this work, we define the export arm as $EspB_{sm}$ residues 78–93, corresponding to the appendage that is dynamic in the EsxAB NMR structure (Figure 3A, left) and lacking electron density in the PE25/PPE41 structure (Figure 3A, right). The prevalent view is that residues from the export arm align on the same helical face upon binding the EccCb1 ATPase, targeting ESX-secreted substrates for secretion (Callahan et al., 2009; Champion et al., 2006; Daleke et al., 2012a). In keeping with this hypothesis, residues within the export arm of EspB are identical in sequence to those shown to be essential for $EsxB_{tb}$ interaction with EccCb1 (Champion et al., 2006), and either mutation of Y78 in the YxxxD motif (Daleke et al., 2012a) or deletion of the EccCb1 ATPase (McLaughlin et al., 2007) similarly disrupts EspB secretion. In the $EspB_{sm}$ P$3_2$21 structure, this arm takes the form of a fully stabilized helix where export arm residues Y78, D81, L89, S90, and M93 indeed align on the same helical face to interact with a symmetry-related molecule in the P$3_2$21 crystal, representing a disposition that we suggest mimics the EccCb ATPase-bound state (Figure 3A, middle and Figure S1A). The α-helical pattern is similar to that seen in recent Esx structures (Arbing et al., 2013; Poulsen et al., 2014), in particular to the *Streptococcus agalactiae* homodimeric EsxA structure determined by Poulsen and colleagues (Figure S1B). Notably, the portion of this arm containing L89/S90/M93 was not resolved in the electron density of the P$6_5$ structure of $EspB_{sm}$, which has

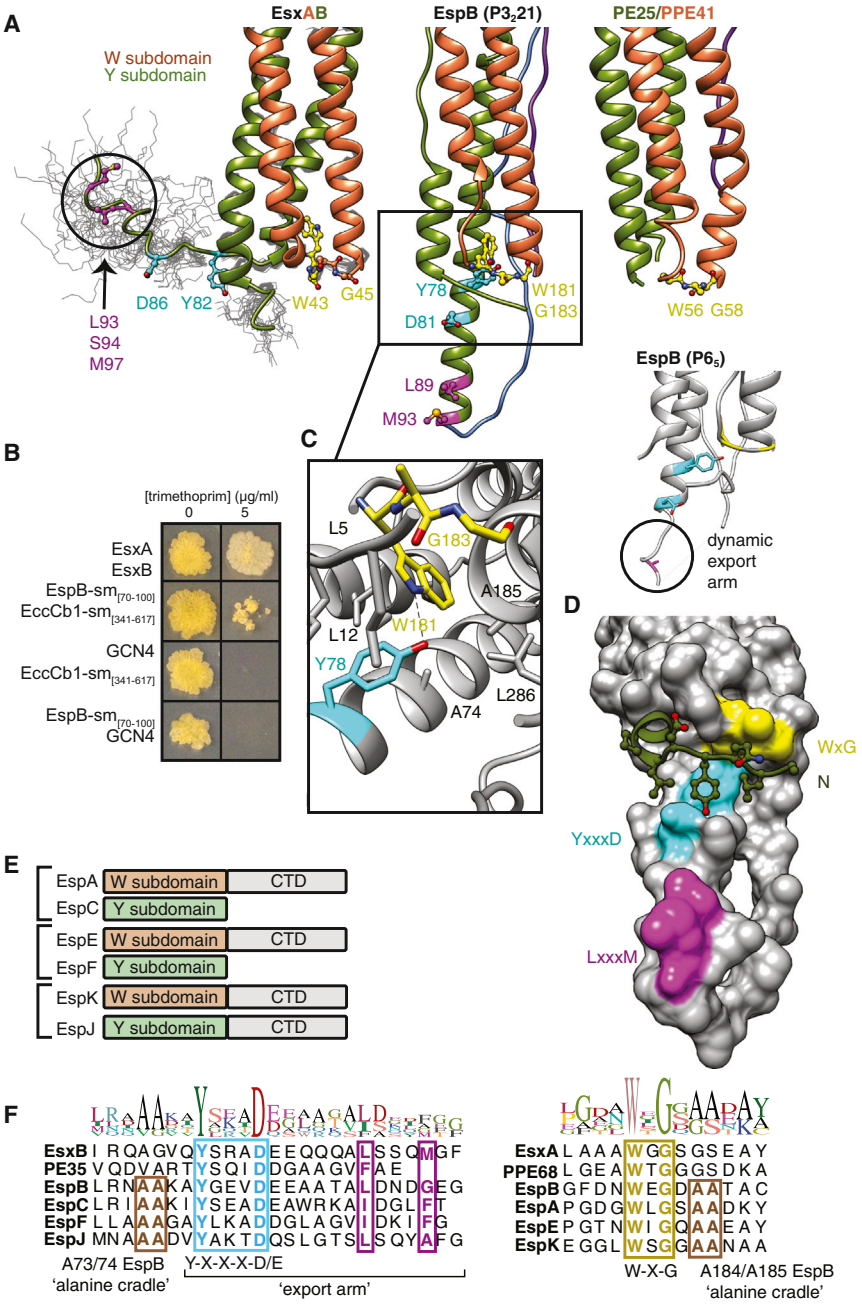



**Figure 3. Structure of $EspB_{sm}$ Stabilized Bipartite Composite Secretion Signal**

(A) Comparison of secretion signal structures across the major ESX-secreted protein classes. 28 structures of the EsxAB NMR ensemble (1WA8) and PE25/PPE41 (2G38) crystal structures are shown.

(B) *M. smegmatis* murine dihydrofolate reductase protein fragment complementation assay demonstrates in vivo interaction between $EspB_{sm-[70–100]}$ "export arm" and the second FtsK-homology domain of $EccCb1_{sm,}$ residues (341–617).

(C) Compared with EsxAB, EspB has a flipped WxG motif that allows the W180 indole nitrogen to form a direct hydrogen bond with the Y78 side chain hydroxyl. The interaction is stabilized by conserved hydrophobic residues that form an "alanine cradle."

(D) The composite secretion signal creates a continuous hydrophobic surface that interacts with the N-terminal portion of the construct in the $P3_221$ structure.

(E) Presence of secretion signal motifs in other Esp proteins of the ESX-1/EspACD locus.

(F) Multiple sequence alignment of the bipartite composite signal sequence-containing proteins in the *M. tuberculosis* ESX-1/EspACD locus. See also Figure S1.

no such stabilizing crystal contacts, indicating that in the unbound conformation, a portion of the $EspB_{sm}$ export arm adopts a dynamic state similar to that of EsxB (Figure 3A, lower right, Figure S1C).

To determine if the export arm of $EspB_{sm}$ interacts with $EccCb1_{sm}$ in vivo, we carried out a murine dihydrofolate reductase (mDHFR) protein fragment complementation assay (Singh et al., 2006). $EspB_{sm}$ residues 70–100, which encompass the export arm, were fused to the mDHFR-F[1,2] fragment. The second FtsK-like ATPase domain of EccCb1, residues 341–617, was fused to the mDHFR-F[3] fragment. If the fused proteins interact, the F[1,2] and F[3] halves of mDHFR associate and enzyme activity is reconstituted, allowing *M. smegmatis* to survive on media containing the antibiotic trimethoprim. A strain expressing known interacting proteins, EsxB-F[1,2] and EsxA-F[3], served as a positive control, and growth was observed (Figure 3B). For negative controls, $EspB_{sm}$-F[1,2] and EccCb1-F[3] were co-expressed with GCN4-F[3] and GCN4-F[1,2], respectively. GCN4 is a protein not associated with ESX-1 and no interaction is expected, and indeed no growth for these strains was observed (Figure 3B). The experimental strain expressing $EspB_{sm}$-F[1,2] and EccCb1-F[3] did exhibit growth, indicating an interaction, but to a lesser extent than the positive control (Figure 3B). It is established that in the mDHFR assay, growth is proportionate to the strength of the interaction (Singh et al., 2006), suggesting that the export arm associates weakly with the EccC. The structural and in vivo interaction data are thus compelling evidence that EspB and EsxB are secreted through ESX-1 by a similar export mechanism involving recognition of a conserved stabilized helix by the EccCb1 ATPase.

Recent secretion/mutagenesis experiments with the *Bacillus subtilis* homodimeric ESX substrate YukE suggest that the WxG motif works in conjunction with the YxxxD motif adjacent to the stabilized export arm (Sysoeva et al., 2014). The WxG motif has also been shown to be indispensable for EspA secretion (Chen et al., 2013a), and this is presumably also the case for EspB. Remarkably, in the $EspB_{sm}$ crystal structure, Y78 from the YxxxD motif and W181 from the WxG motif are seen interacting directly

(Figure 3C). A survey of Esx protein structures deposited in the PDB indicates that this is the first observation in which both elements of the Esx bipartite signal sequence are in direct contact (Figure S1D). In our $EspB_{sm}$ structure, the WxG motif is uniquely flipped towards the YxxxD motif within the export arm such that the Y78 side chain hydroxyl forms a direct hydrogen bond with the W180-Nε side chain indole nitrogen. Furthermore, the flip of the WxG is dependent on G182 within the motif, adopting dihedral angles that can only be assumed by glycine (Figure 3C). The interaction between Y78 and W180 is specifically stabilized by surrounding hydrophobic residues, primarily through a conserved "alanine cradle" formed by side chains of L12, A73, and A74 which interact with the Y78 aromatic ring, and L12, A184, A185, and L286 which interact with the indole ring atoms of W181 (Figure 3C). These surrounding hydrophobic residues appear to stabilize a configuration in which the Y78 and W181 side chains form a perpendicular, aromatic surface that is observed in both crystal forms, independent of crystal contact influence (Figure 3D). When structures from the two $EspB_{sm}$ crystal forms are compared, this aromatic surface is variably bound to hydrophobic side chains of Y6 in $P3_221$ structure or L294 in the $P6_5$ structure, suggesting that this surface may have indiscriminately bound the adjacent free N- and C-terminal regions of the constructs that were crystallized (Figure 3D; Figure S1E). In a physiological context, it is possible that rather than binding these termini, the stabilized export arm and WxG loop may operate as a unified protein-protein interaction surface that is poised to interact with the EccCb ATPase (Sysoeva et al., 2014).

We next investigated whether this composite secretion signal is a general feature of other Esp proteins in the ESX-1/EspACD loci. Using the structural homology server PHYRE2 (Kelley and Sternberg, 2009) we established that EspC, EspF, and EspJ each have predicted Y subdomains, and EspA, EspE, and EspK each have predicted W subdomains fused to CTDs (Figure 3E). Furthermore, the residues forming the alanine cradle that stabilize the interaction between the WxG and YxxxD motifs in the $EspB_{sm}$ structure are conserved across these Esp proteins (Figure 3F). As EspA and EspC, EspE and EspF, and EspJ and EspK are tandemly arrayed in a similar manner as known ESX heterodimers (Figure 1B), it is tempting to speculate that these pairs also associate with similarly positioned export features as EspB, and that these other Esp proteins likely belong to the WXG100 superfamily.

## Quaternary Structure of EspB

While our structural studies of the N-terminal domain of *M. smegmatis* $EspB_{sm}$ provide insight into the molecular basis of ATPase-mediated export, subsequent analysis of our full-length *Mtb* EspB ($EspB_{tb\text{-}[1–460]}$) revealed that EspB has the capability of forming higher order oligomeric states. Immediate clues of EspB oligomerization arose during purification, with protein eluting on a gel filtration column at multiple retention times, suggesting that this protein can self-associate in a concentration-dependent manner. To more accurately determine the subunit stoichiometry of the higher molecular species, we carried out size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis. The SEC-MALS experiment detected a monodisperse peak of molecular weight 52 kDa (corresponding to mass of a single $EspB_{tb\text{-}[1–460]}$ monomer) as well as a polydisperse high molecular weight peak with a molecular mass of 351 kDa at its center, equivalent to seven EspB copies (Figure 4A; Figure S2). In agreement with the SEC-MALS data, our analytical ultracentrifugation sedimentation velocity experiments identified a low molecular weight 2.1S species (44 ± 5 kDa) and a high molecular weight species with a sedimentation coefficient of 10.5S (317 ± 41 kDa) (Figure 4B).

To characterize the structural properties of the $EspB_{tb\text{-}[1–460]}$ oligomer, we next subjected the peak fraction from our gel filtration run to negative stain EM analysis. Raw images show ring-shaped structures that were evenly distributed across the grid (Figure 4C). Indeed, more than 90% of the two-dimensional (2D) class averages obtained from reference-free classification of 2,733 particles show an overall circular ring shape, with several displaying unfused or incomplete rings (Figure 4D; Figure S3A). To determine which part of EspB was responsible for ring formation, we generated a truncated construct $EspB_{tb\text{-}[1–348]}$ that approximates the previously determined position of the mycosin protease (MycP1) cleavage site (Figure 2A) and thus mimics the physiologically processed form (Solomonson et al., 2013). This truncation mutant, which encompasses the N-terminal PE/PPE homology domain and a portion of the low-complexity region of the CTD, still produced ring-shaped structures, indicating that it is the PE/PPE homology region of EspB forming the ring (Figure 4D; Figure S3B). However, the truncated $EspB_{tb\text{-}[1–348]}$ shows a more diverse set of orientations compared with full-length $EspB_{tb\text{-}[1–460]}$, with 20% of the class averages depicting a tubular shape that appears to represent a side view, as its width is consistent with that estimated from the top-view averages (Figure 4D). Even with no imposed symmetry, projection averaging of the "ring" views clearly showed seven distinct "nodes" (Figure 4D, red arrows), which is fully consistent with our SEC-MALS and analytical ultracentrifugation data. Notably, a distinct appendage was seen protruding from the core of the ring in some of the top-view $EspB_{tb\text{-}[1–460]}$ averages (Figure 4D, blue arrow). This structure was absent in the $EspB_{tb\text{-}[1–348]}$ averages, suggesting that this region corresponds to EspB residues 350–460. However, the $EspB_{tb\text{-}[1–348]}$ averages did reveal some unidentified density located at one end of the putative side view, which may correspond to the residual 50 CTD amino acids (Figure 4D, green arrows).

We next determined the 3D reconstruction of the $EspB_{tb\text{-}[1–348]}$ multimer using an approach that involves generating an initial model by rotating a side view class average 360° about the *z* axis and iteratively refining this initial model against untilted data containing top and side views. The final map, with 7-fold symmetry imposed, has a resolution of 30 Å as measured by the Fourier shell correlation (FSC) function using the 0.5 FSC criterion (EM Data Bank accession code EMD-6120) (Figure 5A). This EM structure shows that seven copies of the $EspB_{tb\text{-}[1–348]}$ PE/PPE homology domain associate to form an overall cylindrical shape of 100 Å by 80 Å with a central pore of 50 Å in diameter.

## Modeling the EspB Oligomer

To determine the molecular details of how the monomeric EspB structure assembles into a heptameric complex, we examined the $P3_221$ and $P6_5$ $EspB_{sm}$ crystal lattices for potential interaction interfaces. The $P3_221$ crystal exhibits a remarkable

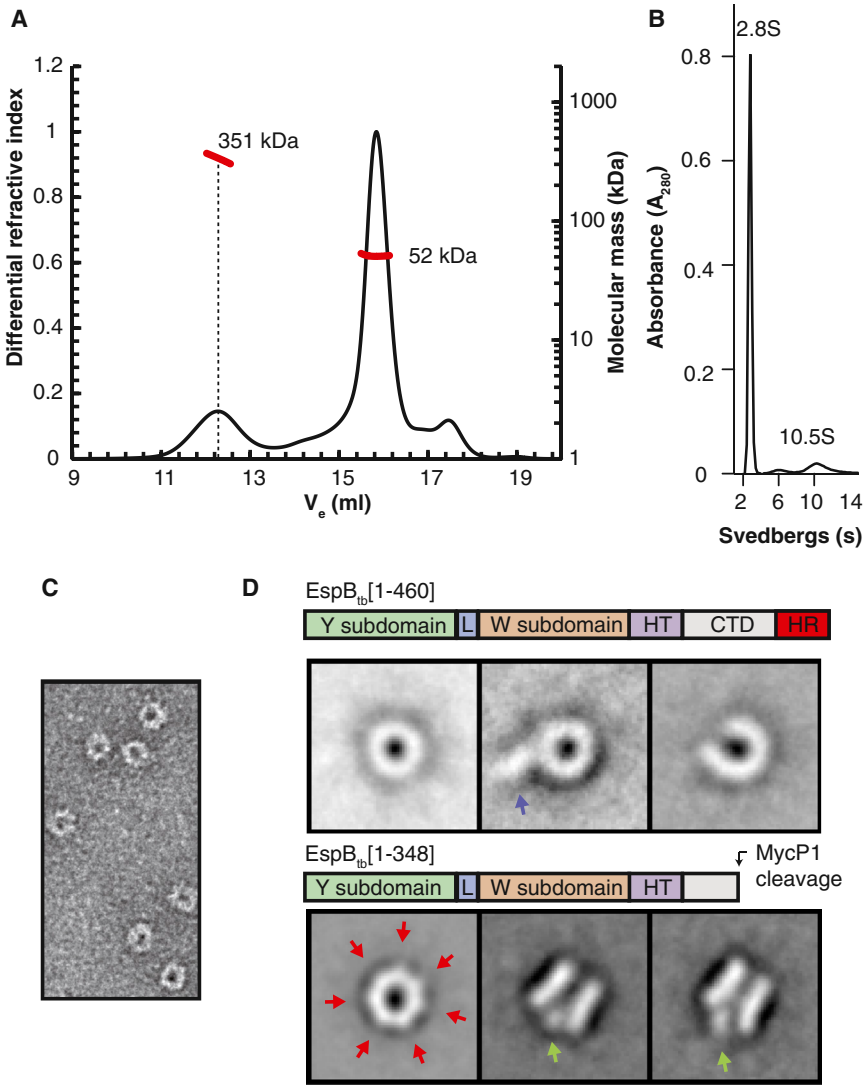



**Figure 4. Quaternary Structure of Purified $EspB_{tb}$**

(A) SEC-MALS analysis of $EspB_{tb\text{-}[1–460]}$ shows presence of higher order species.

(B) Analytical ultracentrifugation analysis of EspB.

(C) Representative negative stain EM micrograph of full-length $EspB_{tb}$.

(D) 2D class averages of $EspB_{tb\text{-}[1–460]}$ and truncated $EspB_{tb\text{-}[1–348]}$. Red arrows draw attention to distinct nodes in the unsymmetrized averages that provide evidence for heptameric stoichiometry. Blue arrow indicates appendage protruding from the core of the ring found only in $EspB_{tb\text{-}[1–460]}$ averages. Green arrows indicate unexplained density in the side view. See also Figures S2 and S3.

hexagonal lattice in which EspB monomers interlock in a herringbone pattern formed by tight side-by-side stacking at the hexagonal edges and tip-to-tail contacts between export arm residues and the helical tip. However, the potential oligomeric structures generated by these symmetries do not match the dimensions estimated from the $EspB_{tb}$ EM structure. We next carried out chemical crosslinking coupled to mass spectrometry (CXMS) using the isotopically coded cleavable affinity-purifiable crosslinker cyanur-biotin-dimercapto-propionyl-succinimide (CBDPS). In multiple independent experiments, we identified CBDPS crosslinks between K259-K259′ and K259-K267′ peptides (Table 2), indicating that these residues are located within 25 Å from one another. Our CXMS data definitively show that intermolecular interactions based on crystal contacts are different from the ones mediating EspB oligomerization. As such, we decided to carry out *in silico* Rosetta-based symmetric modeling with an $EspB_{tb\text{-}[11–286]}$ homology model using an approach we have previously developed (DiMaio et al., 2011; Bergeron et al., 2013). Simulations with a stoichiometric constraint of seven resulted in model convergence to several low-energy clusters (Figure 5B). The fourth-lowest energy cluster was selected as the final model, as it satisfies constraints from our CXMS data (density-fitted model PDB code: 3J83, Figure 5C). The calculated model-map correlation was 0.9, indicating good agreement between the heptameric $EspB_{tb}$ model and the EM reconstruction (Figure 5A). Regions of EspB lying outside the EM density correspond to the dynamic export arm, which, according to the $P6_5$ EspB crystal structure, are unlikely to be well ordered in the EM reconstruction. The model places the C-terminal region of the helical domain of $EspB_{tb}$ in close proximity to the unexplained density located at one end of the multimer, which is consistent with this density corresponding to the residual 50 residues of the CTD. While further experiments will be necessary to refine the exact orientation and molecular interactions involved, this model provides a framework for delineating how $EspB_{tb}$ oligomerizes in the context of the ESX secretion apparatus.

## DISCUSSION

The EspB crystal structure provides new insight into the common structural characteristics shared by ESX substrates, and the elevated complexity of ESX-secreted proteins found in mycomembrane-encapsulated bacteria. Ancestral ESX-secreted proteins in Gram-positive bacteria likely originated as symmetric homodimers with WxG and YxxxD motifs located at both ends, and this is still observed in homodimeric substrates such as *B. subtilis* YukE (Sysoeva et al., 2014) (Figure 6A). However, organisms such as *Mtb* encode more complicated heterodimeric ESX substrates, with each "monomer" possessing either a WxG motif or a YxxxD motif, but not necessarily both. This arrangement places the composite signal sequence at just one end of the helical bundle (Figure 6B). In the case of EspB, a linker has merged the heterodimer into a single open reading frame. Furthermore, in EspB and other PPE proteins of mycobacteria,

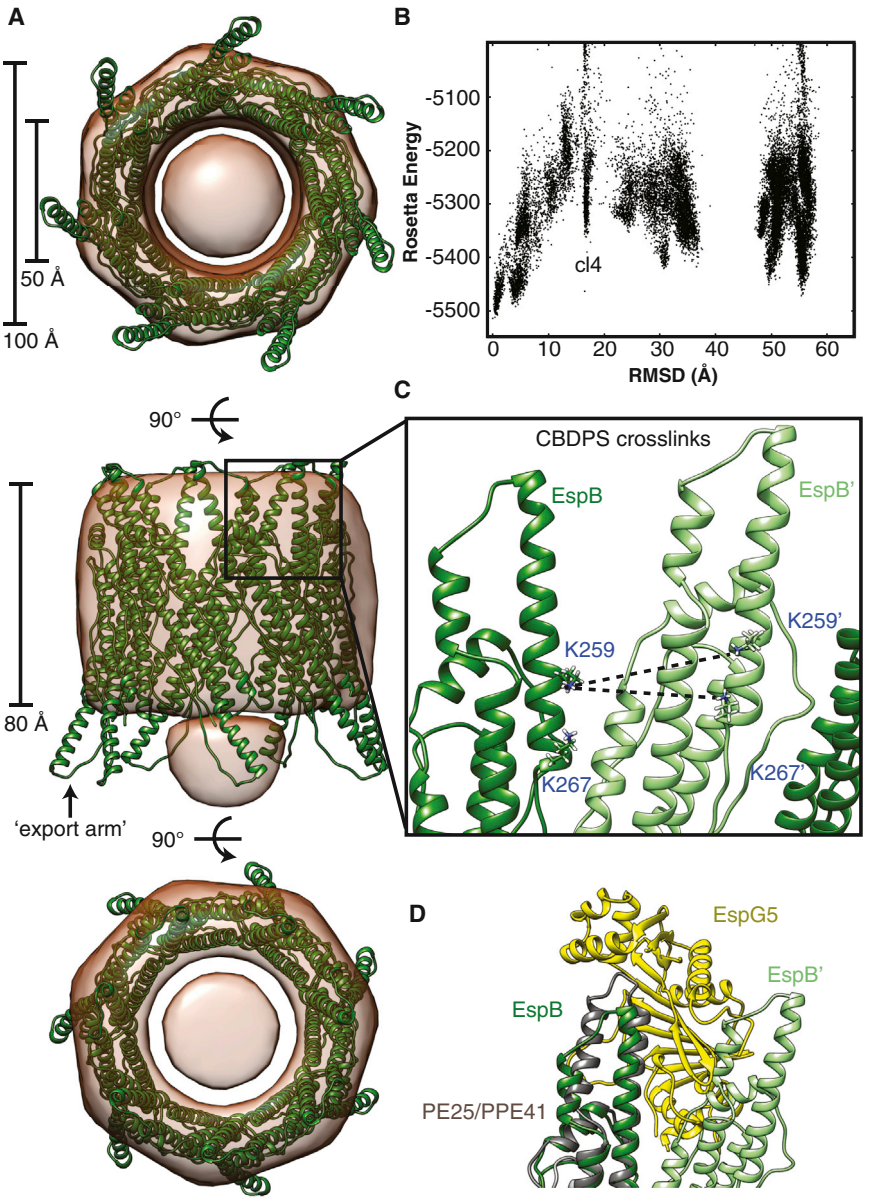



**Figure 5. Heptameric $EspB_{tb}$ Model Fit to Negative Stain 3D Reconstruction Density with Location of Mass Spectrometry-Derived Crosslinks Mapped to Adjacent Subunits**

(A) EspB negative stain 3D reconstruction with heptameric $EspB_{tb}$ model fit to the EM density, approximate dimensions indicated.

(B) Top Rosetta energies for $EspB_{tb}$ heptameric model.

(C) Locations of CBDPS-crosslinked lysine residues mapped to the proposed model.

(D) Overlay of the PE25/PPE41/EspG5 ternary chaperone complex (4KXR) on the heptameric EspB model.

the canonical motifs on one end of the helical bundle have been phased out and replaced by a "helical tip" (Figure 6B).

The EspB structure reveals an unexpected hydrogen-bonded interaction between the defining tyrosine and tryptophan side chains in the YxxxD and WxG motifs positioned in our structure on the same helical face as conserved acidic and hydrophobic residues of the stabilized export arm. We note the similarity of the stabilized export arm to the signal peptides of the secretory pathway, which also adopt a helical conformation when bound to lipids or translocon components (Briggs et al., 1986; Chou and Gierasch, 2005). We suspect the unique conformation of the WxG and YxxxD motifs observed in our crystal structure represents a functional disposition that is adopted by all ESX substrates at some critical point during their translocation. The conserved alanine cradle serves to stabilize this direct interaction, perhaps favoring a secretion-competent conformation of the composite signal. Our analysis also shows that the ESX-1-associated EspA, EspE, and EspK have WxG motifs and EspC, EspF, and EspJ have YxxxD motifs, and all have residues of the "alanine cradle" that potentially support their direct interaction. We predict EspAC, EspEF, and EspJK to be heterodimeric complexes that associate to ensure the presence of both secretion motifs in the complex with a similar structural configuration observed in EspB, perhaps allowing some or all of these unique ESX-1 substrates to be secreted through the same pathway as the well-studied EsxAB heterodimer. Thus, we suggest that these proteins be included in the WXG100 family, resulting in a total of five heterodimeric WXG100 pairs in addition to one fused homomer (EspB) in the ESX-1 locus, which in pairwise combination possess the complete composite secretion signal (Figures 1B and 3F).

Of particular interest are the mycobacteria-specific features of ESX-secreted proteins, such as the helical tip we observe in EspB and PPE proteins (Figure 6B). The helical tip is relatively hydrophobic, raising the possibility of its involvement in traversing the mycomembrane or perhaps in facilitating interaction with an unidentified component within the mycomembrane that aids this process. This appendage may also have a role in oligomerization, particularly in light of recent studies showing that EspG chaperones bind directly to the helical tip of PPE41 from the PE25/PPE41 complex, preventing self-aggregation (Daleke et al., 2012b; Ekiert and Cox, 2014; Korotkova et al., 2014). When the PE25/PPE41/EspG5 ternary complex (4KXR) is aligned to EspB in the context of our heptameric model, the EspG5 chaperone is situated at an interface identical to that of the crosslinked lysines identified by our CXMS experiments (Figure 5D). This raises the possibility that EspB may also have a yet to be identified chaperone that prevents premature self-association in the *Mtb* cytoplasm.

**Table 2. Identified Inter-Protein CBDPS Crosslinks of $EspB_{tb\text{-}[1–348]}$ and $EspB_{tb\text{-}[1–460]}$**

| Deconvoluted mass (MH+) (L) | Δ, ppm | Charge | Residue numbers (1) | Crosslinked residue (1) | Peptide sequence (1) | Residue numbers (2) | Crosslinked residue (2) | Peptide sequence (2) | Fragment ion score | Intensity score | CID-cleavage products | Modification | Digestion | $^{14}N/^{15}N$ inter/intra |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **$EspB_{tb\text{-}[1–348]}$** | | | | | | | | | | | | | | |
| 3157.43919 | −0.1 | 2 | 257–267 | 259 | (R)SEKVLTEYNNK(A) | 257–267 | 259 | (R)SEKVLTEYNNK(A) | 85 | 42 | 4 | CBDPS | Tr | Inter |
| 1242.55321 | 0.7 | 2 | 259–261 | 259 | (E)KVL(T) | 265–267 | 267 | (Y)NNK(A) | 188 | 9 | 4 | CBDPS | PK | Inter |
| 1543.74107 | 1.2 | 2 | 257–262 | 259 | (R)SEKVLT(E) | 259–261 | 259 | (E)KVL(T) | 107 | 16 | 4 | CBDPS | PK | Inter |
| 1644.79063 | 0 | 2 | 257–262 | 259 | (R)SEKVLT(E) | 259–262 | 259 | (E)KVLT(E) | 150 | 18 | 4 | CBDPS | PK | Inter |
| **$EspB_{tb\text{-}[1–460]}$** | | | | | | | | | | | | | | |
| 3173.43207 | 0.3 | 4 | 257–267 | 259 | (R)SEKVLTEYNNK(A) | 257–267 | 259 | (R)SEKVLTEYNNK(A) | 185 | 50 | 4 | CBDPS+ 1Ox | Tr | Inter |

L, light isotopic form of the crosslink. Δ, mass difference between theoretical and observed crosslink masses. Peptide sequences are shown with preceding and following residues in the protein sequence in parentheses. Fragment ion, intensity scores, and CID-cleavage products number reflect crosslink MS/MS spectrum match quality. PK, Proteinase K; Tr, trypsin. Reconfirmation of crosslinks as intra- or inter-protein origin with $^{14}N/^{15}N$ crosslinking experiments is indicated.

In the context of the ESX system as a whole, the finding that $EspB_{tb}$ can oligomerize is particularly exciting, as it indicates that the helical domains of ESX-secreted proteins may serve a structural role in addition to, or as part of, their role in facilitating export. Our proposed model places the helical tips of EspB on one side of the multimer, opposite the composite secretion signals and CTD (Figure 6C). This creates an arrangement of seven closely spaced hydrophobic helical tips, which in combination with the pore generated by the multimer, appears to lend itself to contribute to the membrane-puncturing activity required for export across the mycolic acid layer and/or carrying out ESX-1-mediated phenotypes such as phagosome permeabilization within *Mtb*-infected macrophages (de Jonge et al., 2007; Hsu et al., 2003; Smith et al., 2008) or conjugal transfer of DNA in *M. smegmatis*. Does EspB promote these activities in isolation or in concert with other components of ESX-1, such as the remaining five WXG100 pairs encoded by the locus and/or the inner membrane apparatus? It is currently thought that the inner membrane-spanning portion of the ESX apparatus forms a large molecular weight complex of probable symmetric composition (Houben et al., 2012), and EspB could contribute to the extracellular portion of this, perhaps anchoring to a tip of a membrane-spanning filament. That secreted proteins of ESX-1 are extracellular components of the secretion apparatus is supported by the co-localization of apparatus and secreted components to discrete poles of the bacteria (Carlsson et al., 2009; Wirth et al., 2011), and is further supported by the finding that many Esx/Esp proteins are co-dependent for secretion (Champion et al., 2009; Fortune et al., 2005). Moreover, expression of EspB with its endogenous promoter has been shown to be important in recovering the wild-type ESX phenotype, suggesting that stoichiometric expression levels of EspB is key (Xu et al., 2007). It should perhaps not be ruled out that in vivo, EspB may form structures of a higher order than cyclic heptamer to extend its span to the extracellular space, for example through helical assembly. The extent to which ESX-secreted proteins structures resemble the filamentous protein building blocks from other secretion systems has been noted (Pallen, 2002).

Finally, our 2D averages of top-view $EspB_{tb\text{-}[1–460]}$ show a region of density that we propose corresponds to the $EspB_{tb}$ CTD cargo domain. Previous reports suggest that this CTD is unstructured (Wagner et al., 2013), but our $EspB_{tb\text{-}[1–460]}$ data indicate it may become structured in the context of the multimer. This structured CTD is presumably driven by self-association of the conserved homology region of the CTD (Figure 6C), which was absent in the truncated $EspB_{tb\text{-}[1–348]}$ construct. Indeed, the top-view averages obtained using the $EspB_{tb\text{-}[1–348]}$ construct exhibited no such density extending from the ring. This suggests that MycP1-mediated proteolytic cleavage is an essential maturation event (Ohol et al., 2010) that has the potential to visibly alter the length of the CTD in context of the $EspB_{tb}$ multimer. However, the $EspB_{tb\text{-}[1–348]}$ EM averages do exhibit a region of density located at the mouth of the pore in the side view averages. This observation is intriguing, particularly in light of secretion experiments showing that MycP1 negatively regulates ESX-1 secretion by cleaving EspB. If EspB indeed contributes to the secretion system channel through which other ESX substrates are also secreted, the density we observe adjacent to the pore of processed EspB may function as a "plug" that carries out

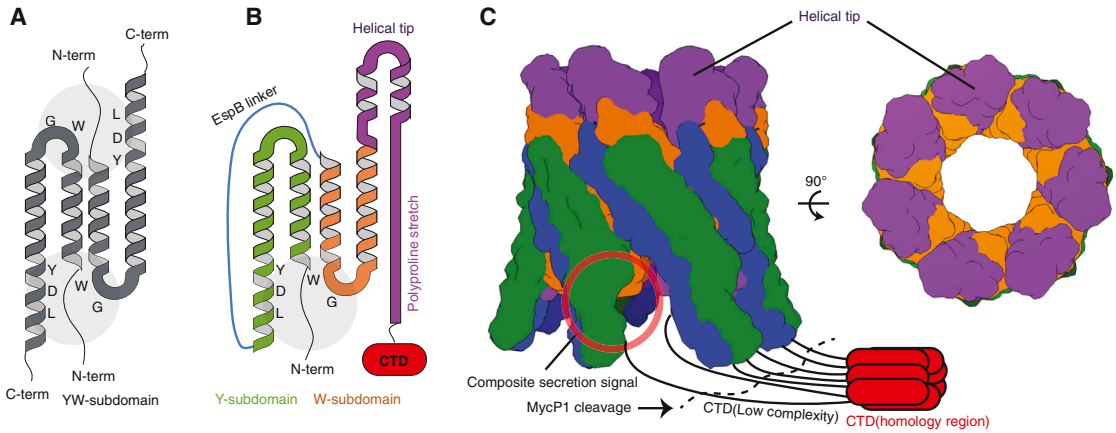



**Figure 6. Schematic Summary of ESX-Secreted Protein Specialization in Mycobacteria**

(A) Ancestral ESX substrates were homodimeric and displayed essential export motifs on both sides of the helical bundle with N/C termini on separate sides.

(B) Mycobacterial ESX proteins such as EspB, which are located in the cell walls of these organisms, have become increasingly asymmetric, losing one of the bipartite signal elements and gaining a helical appendage.

(C) Proposed model of EspB quaternary structure, placing the helical tip and secretion signals on opposite sides.

this negative regulation. Alternatively, perhaps this density corresponds to a domain that mediates interaction with other secretion system components, allowing assembly to be regulated by MycP1 processing.

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

## EXPERIMENTAL PROCEDURES

### Purification, Crystallization, and Structure Determination of EspB

EspB coding sequences from *M. smegmatis* (MSMEG_0076) and *Mtb* (Rv3881c) were cloned into pET28a(+), in-frame with an N-terminal cleavable histidine affinity tag with the sequence MGSSHHHHHHHHHHSSGLVPRGSH. Plasmids were transformed into *Escherichia coli* BL21 codon-plus (DE3)-RIPL-competent cells (Stratagene). Induction was initiated with 1 mM isopropyl 1-thio-β-D-galactopyranoside at $A_{600}$ = 0.6 followed by growth for 20 hr at 20°C. Cells expressing Se-Met derivative $EspB_{sm\text{-}[1–292]}$ were grown as previously described (Larrson, 2009). Prior to purification, cells were resuspended in 50 mM HEPES (pH 7.5), 500 mM NaCl, 10 mM imidazole, and protease inhibitor tablets (Roche cOmplete), and lysed with a C5 homogenizer (Avisten). Lysate was spun at 45,000 rpm for 1 hr, and the supernatant was passed over HisPur Ni-NTA resin (Thermo, 88222) and washed with buffer containing 25 mM imidazole, and finally eluted in buffer containing 250 mM imidazole. Eluate was concentrated and further purified on a Superose 6 10/300 GL column equilibrated with 20 mM HEPES (pH 7.5) and 150 mM imidazole. $EspB_{tb}$ protein corresponding to the oligomeric peak was concentrated and flash frozen at −80°C prior to EM, light scattering, and analytical ultracentrifugation analyses. The $EspB_{sm}$ variants were digested overnight with thrombin and purified on a Superdex 200 column prior to crystallization.

Diffracting Se-Met derivative $P3_221$ crystals grew using the $EspB_{sm\text{-}[1–292]}$ construct at 80 mg/ml in 100 mM HEPES (pH 6.8), 540 mM $MgCl_2$, and 15% PEG 6000, and reached maximum size in 5 days. Crystals of the $P6_5$ space group were obtained using the $EspB_{sm}$ full-length construct at 20 mg/ml in 0.1 M Bis-tris propane (pH 6.5), 200 mM $MgCl_2$, and PEG 3350, reaching a maximum size after 1 month. The $P3_221$ crystals were flash frozen directly, while the $P6_5$ crystals were cryoprotected in 20% glycerol prior to flash freezing. The data collected at the Canadian Light Source Canadian Macromolecular Crystallography Facility (CMCF) were integrated, scaled, and merged with Mosflm/Pointless (Bailey, 1994) and the structure was solved, built, and refined with the Phenix package (Adams et al., 2010) and Coot (Emsley and Cowtan, 2004). PHENIX Autosol located 13 selenium atoms in the substructure solution (figure of merit = 0.42 using data from 3.2 to 61.8 Å). The $P3_221$ structure served as a molecular replacement search model for solving the $P6_5$ native data set using PHASER (McCoy et al., 2007). The structures were analyzed and figures generated using UCSF Chimera (Pettersen et al., 2004).

### *M. smegmatis* Protein Complementation Assay

The coding sequence for $EspB_{sm}$ residues 70–100 was cloned into pUAB100, in-frame with the a C-terminal murine dihydrofolate reductase F[1,2] fragment and the coding sequence for $EccCb1_{sm}$ residues 341–617 was cloned into pUAB200, in-frame with a C-terminal murine dihydrofolate reductase F[3] fragment, and the assay was carried out as described (Singh et al., 2006). Plasmids were co-transformed into *M. smegmatis* $mc^{155}$ by electroporation and plated on 7H10 media containing 25 μg/ml kanamycin and 50 μg/ml hygromycin. Overnight cultures grown from glycerol stocks were adjusted to OD = 1.0, and 4 μl of cells were spotted on 7H11 plates containing 25 μg/ml kanamycin and 50 μg/ml hygromycin, in the absence or presence of 5 μg/ml trimethoprim.

### Multi-Angle Light Scattering

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

### Analytical Ultracentrifugation

Analytical ultracentrifugation experiments were carried out at 20°C using a Beckman XL-I analytical ultracentrifuge, an An-60 Ti rotor, and absorbance optics. $EspB_{tb\text{-}[1–460]}$ samples were loaded at 1.0 mg/ml and spun at 47,000 rpm in two-channel carbon-filled Epon centerpieces. The data were analyzed with Sedfit (Schuck, 2000).

### Negative Stain EM

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

To produce the EspB 3D reconstruction, a representative 2D class average corresponding to a side view of the EspB heptamer was rotationally extruded about the symmetry axis to form an initial 3D model using John Rubinstein's program "build_fspace_v2_00" (https://sites.google.com/site/rubinsteingroup/3-d-fourier-space). Single particle images belonging to class averages corresponding to both top and side views of the EspB complex (1,455 particles) were used to refine the initial model using EMAN2. The final resolution was determined using the FSC function using the 0.5 FSC criterion.

### Oligomeric Modeling

Starting with $EspB_{tb}$ homology based on the $EspB_{sm}$ structure (Song et al., 2013), we ran the Rosetta symmetric docking protocol (Soding, 2005) guided by density data to build a C7 symmetric system. In total, 25,000 docked models were generated. The 500 of lowest energy were selected and clustered, yielding four distinct clusters. The centroids of each cluster were then compared with crosslinking data. A full description of the modeling procedure is included in the Supplemental Materials and Methods.

### Crosslinking and Mass Spectrometry

Equimolar ratio of $^{14}N$- and $^{15}N$-metabolically labeled EspB were mixed to determine intra- or inter-protein origin of the identified crosslinks for both $EspB_{tb\text{-}[1–348]}$ and $EspB_{tb\text{-}[1–460]}$. Samples were crosslinked with 1 mM CBDPS-H8/D8, digested with trypsin or Proteinase K, and subjected to mass spectrometric analysis (Petrotchenko et al., 2014a). Data were analyzed using 14N15N DXMSMS Match program (Petrotchenko et al., 2014b). A full description of the crosslinking procedure is included in the Supplemental Materials and Methods.

## ACCESSION NUMBERS

The PDB accession numbers for the X-ray protein structures reported in this paper are 4WJ1 and 4WJ2. The PDB accession number for the EM density-fit model reported in this paper is 3J83. The EM databank accession number for the electron microscopy density map reported in this paper is EMD-6120.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, Supplemental Materials and Methods, and 3D molecular models and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.01.002.

## AUTHOR CONTRIBUTIONS

M.S. and N.C.J.S. designed the research. M.S. carried out cloning, protein preparation, crystallization. and X-ray crystallographic structure determination. D.S. and C.K.Y. acquired/processed EM data with some input from M.S. E.L. and M.S. carried out SEC-MALS experiments. M.S. and J.R.B. conducted the preliminary EM experiments. M.V. provided technical assistance with some cloning. D.G.C. and M.S. carried out the analytical ultracentrifugation experiments. F.D. carried out Rosetta modeling. K.A.T.M., E.V.P., and C.H.B. performed mass spectrometry/crosslinking experiments and analyzed data. M.S. and N.C.J.S. analyzed all data and wrote the paper.

## ACKNOWLEDGMENTS

We thank the staff at the Canadian Light Source CMCF for assistance with data collection, and the Canadian Foundation of Innovation, the British Columbia Knowledge Development Fund for infrastructure funding. We thank Dr. Liam Worrall for advice regarding X-ray data collection and Dr. John Rubenstein for helpful suggestions related to EM data processing. The mDHFR plasmids were kindly provided by Dr. Yossef Av-Gay. M.S. is supported by a UBC Four-Year PhD Fellowship. N.C.J.S. thanks the Canadian Institute of Health Research and Howard Hughes Medical Institute International Scholar program for operating funding. N.C.J.S. is a Canada Research Chair Tier 1 in Antibiotic Discovery. C.K.Y. acknowledges operating support from the Natural Sciences and Engineering Research Council of Canada, salary awards from the Canadian Institutes for Health Research, Michael Smith Foundation for Health Research, an infrastructure grant from the Canadian Foundation for Innovation, and startup funds from the University of British Columbia. K.A.T.M., E.V.P., and C.H.B. were supported by a Genome Canada/Genome British Columbia Technology Development Grant.

Received: September 30, 2014
Revised: December 17, 2014
Accepted: December 23, 2014
Published: February 12, 2015

## REFERENCES

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

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

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

Arbing, M.A., Chan, S., Harris, L., Kuo, E., Zhou, T.T., Ahn, C.J., Nguyen, L., He, Q., Lu, J., Menchavez, P.T., et al. (2013). Heterologous expression of mycobacterial Esx complexes in *Escherichia coli* for structural studies is facilitated by the use of maltose binding protein fusions. PLoS One *8*, e81753.

Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Bergeron, J.R., Worrall, L.J., Sgourakis, N.G., DiMaio, F., Pfuetzner, R.A., Felise, H.B., Vuckovic, M., Yu, A.C., Miller, S.I., Baker, D., et al. (2013). A refined model of the prototypical *Salmonella* SPI-1 T3SS basal body reveals the molecular basis for its assembly. PLoS Pathog. *9*, e1003307.

Brennan, P.J., and Crick, D.C. (2007). The cell-wall core of *Mycobacterium tuberculosis* in the context of drug discovery. Curr. Top. Med. Chem. *7*, 475–488.

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

Callahan, B., Nguyen, K., Collins, A., Valdes, K., Caplow, M., Crossman, D.K., Steyn, A.J.C., Eisele, L., and Derbyshire, K.M. (2009). Conservation of structure and protein-protein interactions mediated by the secreted mycobacterial proteins EsxA, EsxB, and EspA. J. Bacteriol. *192*, 326–335.

Carlsson, F., Joshi, S.A., Rangell, L., and Brown, E.J. (2009). Polar localization of virulence-related Esx-1 secretion in mycobacteria. PLoS Pathog. *5*, e1000285.

Champion, P.A.D., and Cox, J.S. (2007). Protein secretion systems in mycobacteria. Cell. Microbiol. *9*, 1376–1384.

Champion, P.A.D., Stanley, S.A., Champion, M.M., Brown, E.J., and Cox, J.S. (2006). C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. Science *313*, 1632–1636.

Champion, P.A.D., Champion, M.M., Manzanillo, P., and Cox, J.S. (2009). ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol. Microbiol. *73*, 950–962.

Champion, M.M., Williams, E.A., Pinapati, R.S., and Champion, P.A. (2014). Correlation of phenotypic profiles using targeted proteomics identifies mycobacterial esx-1 substrates. J. Proteome Res. *13*, 5151–5164.

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

Chen, J.M., Zhang, M., Rybniker, J., Boy-Röttger, S., Dhar, N., Pojer, F., and Cole, S.T. (2013b). *Mycobacterium tuberculosis* EspB binds phospholipids and mediates EsxA-independent virulence. Mol. Microbiol. *89*, 1154–1166.

Chou, Y.T., and Gierasch, L.M. (2005). The conformation of a signal peptide bound by *Escherichia coli* preprotein translocase SecA. J. Biol. Chem. *280*, 32753–32760.

Coros, A., Callahan, B., Battaglioli, E., and Derbyshire, K.M. (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Mol. Microbiol. *69*, 794–808.

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

Daleke, M.H., van der Woude, A.D., Parret, A.H., Ummels, R., de Groot, A.M., Watson, D., Piersma, S.R., Jimenez, C.R., Luirink, J., Bitter, W., et al. (2012b). Specific chaperones for the type VII protein secretion pathway. J. Biol. Chem. *287*, 31939–31947.

de Jonge, M.I., Pehau-Arnaudet, G., Fretz, M.M., Romain, F., Bottai, D., Brodin, P., Honore, N., Marchal, G., Jiskoot, W., England, P., et al. (2007). ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J. Bacteriol. *189*, 6028–6034.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and Andre, I. (2011). Modeling Symmetric Macromolecular Structures in Rosetta3. PLoS One *6*, e20450.

Ekiert, D.C., and Cox, J.S. (2014). Structure of a PE-PPE-EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. Proc. Natl. Acad. Sci. USA *111*, 14758–14763.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fortune, S.M., Jaeger, A., Sarracino, D.A., Chase, M.R., Sassetti, C.M., Sherman, D.R., Bloom, B.R., and Rubin, E.J. (2005). Mutually dependent secretion of proteins required for mycobacterial virulence. Proc. Natl. Acad. Sci. USA *102*, 10676–10681.

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

Gao, L.Y., Guo, S., McLaughlin, B., Morisaki, H., Engel, J.N., and Brown, E.J. (2004). A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. Mol. Microbiol. *53*, 1677–1693.

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

Houben, E.N., Korotkov, K.V., and Bitter, W. (2013). Take five—Type VII secretion systems of mycobacteria. Biochim. Biophys. Acta *1843*, 1707–1716.

Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., et al. (2003). The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc. Natl. Acad. Sci. USA *100*, 12420–12425.

Huang, D., and Bao, L. (2014). Mycobacterium tuberculosis EspB protein suppresses interferon-gamma-induced autophagy in murine macrophages. Journal of microbiology, immunology, and infection. Published online November 21, 2014. http://dx.doi.org/10.1016/j.jmii.2014.11.008.

Ilghari, D., Lightbody, K.L., Veverka, V., Waters, L.C., Muskett, F.W., Renshaw, P.S., and Carr, M.D. (2011). Solution structure of the *Mycobacterium tuberculosis* EsxG.EsxH complex: functional implications and comparisons with other *M. tuberculosis* Esx family complexes. J. Biol. Chem. *286*, 29993–30002.

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

Korotkova, N., Freire, D., Phan, T.H., Ummels, R., Creekmore, C.C., Evans, T.J., Wilmanns, M., Bitter, W., Parret, A.H.A., Houben, E.N.G., et al. (2014). Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25–PPE41 dimer. Mol. Microbiol. *94*, 367–382.

Larrson, A.M. (2009). Protein Crystallization, Second Edition. (International University Line).

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

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

McLaughlin, B., Chon, J.S., MacGurn, J.A., Carlsson, F., Cheng, T.L., Cox, J.S., and Brown, E.J. (2007). A mycobacterium ESX-1-Secreted virulence factor with unique requirements for export. PLoS Pathog. *3*, 1051–1061.

Millington, K.A., Fortune, S.M., Low, J., Garces, A., Hingley-Wilson, S.M., Wickremasinghe, M., Kon, O.M., and Lalvani, A. (2011). Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. Proc. Natl. Acad. Sci. USA *108*, 5730–5735.

Niederweis, M., Danilchanka, O., Huff, J., Hoffmann, C., and Engelhardt, H. (2009). Mycobacterial outer membranes: in search of proteins. Trends Microbiol. *18*, 109–116.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative staining and image classification—powerful tools in modern electron microscopy. Biol. Proced. Online *6*, 23–34.

Ohol, Y.M., Goetz, D.H., Chan, K., Shiloh, M.U., Craik, C.S., and Cox, J.S. (2010). *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell Host Microbe *7*, 210–220.

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

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

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

Petrotchenko, E.V., Serpa, J.J., Makepeace, K.A., Brodie, N.I., and Borchers, C.H. (2014b). (14)N(15)N DXMSMS Match program for the automated analysis of LC/ESI-MS/MS crosslinking data from experiments using (15)N metabolically labeled proteins. J. Proteomics *109*, 104–110.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Poulsen, C., Panjikar, S., Holton, S.J., Wilmanns, M., and Song, Y.-H. (2014). WXG100 protein superfamily consists of three subfamilies and exhibits an $\alpha$-helical C-terminal conserved residue pattern. PLoS One *9*, e89313.

Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norman, J., et al. (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J. *24*, 2491–2498.

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

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jimenez, C.R., Daffe, M., et al. (2010). Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. PLoS Pathog. *6*, e1000794.

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

Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., and Enninga, J. (2012). Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. PLoS Pathog. *8*, e1002507.

Singh, A., Mai, D., Kumar, A., and Steyn, A.J.C. (2006). Dissecting virulence pathways of *Mycobacterium tuberculosis* through protein-protein association. Proc. Natl. Acad. Sci. USA *103*, 11346–11351.

Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J.J., Liu, J., McDonald, K.L., Szyk, A., LaRonde-LeBlanc, N., and Gao, L.Y. (2008). Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. Infect. Immun. *76*, 5478–5487.

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

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

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

Strong, M., Sawaya, M.R., Wang, S.S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA *103*, 8060–8065.

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

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

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

Watson, R.O., Manzanillo, P.S., and Cox, J.S. (2012). Extracellular *M. tuberculosis* DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. Cell *150*, 803–815.

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

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