The structure of the endogenous ESX-3 secretion system

Nicole Poweleit
University of California, San Francisco

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Amino Acids, Peptides, and Proteins Commons, Bacteria Commons, Biophysics Commons, Microbiology Commons, Molecular Biology Commons, and the Structural Biology Commons

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution 4.0 License.
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMMS Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Title: The structure of the endogenous ESX-3 secretion system

Authors: Nicole Poweleit\textsuperscript{1,2}, Nadine Czudnochowski\textsuperscript{1,2}, Rachel Nakagawa\textsuperscript{1}, Donovan Trinidad\textsuperscript{1,2}, Kenan C. Murphy\textsuperscript{3}, Christopher Sassetti\textsuperscript{3}, Oren S. Rosenberg\textsuperscript{1,2,*}

Affiliations:
\textsuperscript{1}University of California, San Francisco, Department of Medicine, Division of Infectious Diseases, San Francisco, CA 94158, USA.
\textsuperscript{2}Chan-Zuckerberg Biohub, University of California, San Francisco, CA 94158, USA.
\textsuperscript{3}Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA, 01655, USA.

*Corresponding author. E-mail: oren.rosenberg@ucsf.edu

Abstract

The ESX (or Type VII) secretion systems are protein export systems in mycobacteria and many Gram-positive bacteria that mediate a broad range of functions including virulence, conjugation, and metabolic regulation. These systems translocate folded dimers of WXG100-superfamily protein substrates across the cytoplasmic membrane. We report the cryo-electron microscopy structure of an ESX-3 system, purified using an epitope tag inserted with recombineering into the chromosome of the model organism \textit{Mycobacterium smegmatis}. The structure reveals a stacked architecture that extends above and below the inner membrane of the bacterium. The ESX-3 protomer complex is assembled from a single copy of the EccB\textsubscript{3}, EccC\textsubscript{3}, and EccE\textsubscript{3} and two copies of the EccD\textsubscript{3} protein. In the structure, the protomers form a stable dimer that is consistent with assembly into a larger oligomer. The ESX-3 structure provides a framework for further study of these important bacterial transporters.
Introduction

Mycobacteria use a set of specialized secretion systems called ESX to transport proteins across their complex, diderm cell walls (Gröschel et al., 2016). Originally described as virulence factors in *Mycobacteria tuberculosis* (Guinn et al., 2004; Hsu et al., 2003; Lewis et al., 2003; Stanley et al., 2003), orthologs of ESX have since been discovered in most gram-positive bacteria (Bottai et al., 2017), and are more generally referred to as Type VII secretion systems (Bitter et al., 2009). In mycobacteria there are five paralogous ESX operons (ESX 1-5) each of which encodes an inner membrane translocon complex consisting of three conserved Ecc proteins: EccB, EccC, and EccD. A fourth protein, EccE is conserved in all ESX operons except the ancestral ESX-4 operon and is also considered a part of the ESX translocon complex as it copurifies with EccB, EccC, and EccD (Houben et al., 2012). All Type VII secretion systems translocate proteins in the WXG100-superfamily, which share a common two-helix hairpin structure and are found as homo- or heterodimers (Poulsen et al., 2014) and are mutually dependent for secretion with other substrates (Fortune et al., 2005). In contrast to the general secretory apparatus (Sec), ESX substrates have been shown to be secreted in their folded, dimeric state (Sysoeva et al., 2014).

Structural and functional information has been reported for truncated and isolated, soluble domains of the ESX translocon complexes and their homologs (Korotkova et al., 2015, 2014; Renshaw et al., 2005; Rosenberg et al., 2015; Strong et al., 2006; Wagner et al., 2016, 2014, 2013; Zhang et al., 2015; Zoltner et al., 2016). A low resolution, negative stain electron microscopy structure of ESX-5 shows a translocon complex assembled into a hexamer (Beckham et al., 2017). Structures of other proteins encoded in ESX operons including secreted substrates (Ilghari et al., 2011), substrate chaperons (Ekiert and Cox, 2014), and the protease MycP
(Solomonson et al., 2013) have been solved. Despite revealing important functional information about ESX, structures of overexpressed and isolated proteins are insufficient to understand the regulated secretion of fully folded substrates. We therefore undertook structural studies of an endogenously expressed ESX-3 complex from the model organism \textit{M. smegmatis} using cryo-electron microscopy (cryo-EM). During the preparation of this work for publication, a similar structure of the ESX-3 system expressed from a plasmid was published by another group. (Famelis et al., 2019)

The ESX-3 translocon complex is important for iron acquisition (Serafini et al., 2013; Siegrist et al., 2009), cell survival (Tinaztepe et al., 2016), and virulence in pathogenic mycobacteria (Tufariello et al., 2016), and its role in iron homeostasis is conserved in the model system, \textit{M. smegmatis}. (Siegrist et al., 2009) The ESX-3 translocon complex proteins are transcribed in a single operon (Li et al., 2017), and expression of the ESX-3 operon is dependent on the transcriptional regulator IdeR, which controls iron metabolism (Rodriguez et al., 2002) and is required for growth in the human pathogen \textit{M. tuberculosis} (Pandey and Rodriguez, 2014). The ESX-3 operon is 67\% identical between the non-pathogenic model organism \textit{M. smegmatis} and the pathogen \textit{M. tuberculosis} over the 4354 amino acids of the ESX-3 operon. This high degree of conservation and essential role in cell growth makes ESX-3 an important candidate for small molecule inhibition (Bottai et al., 2014), as blockade of ESX-3 will both inhibit virulence in \textit{M. tuberculosis} and kill a broad range of pathogenic mycobacteria.

\textbf{Results}

A major innovation made possible by the dramatic improvements in cryo-electron microscopy (Cheng, 2018) is the ability to examine challenging protein samples at atomic resolution, even when samples are only available at low concentrations. When coupled with
recent genetic manipulations that allow for facile insertion of chromosomal epitope and purification tags (Murphy et al., 2018), cryo-EM now holds the promise of routine structural characterization of many endogenously expressed protein complexes not previously tractable by structural techniques. We undertook the purification of the ESX-3 complex from the native host without the need for overexpression. To facilitate purification of the endogenous translocon complex, a cleavable GFP tag was inserted into the chromosome of *M. smegmatis* MC² 155 wild type and ΔideR (Dussurget et al., 1996) strains at the C-terminus of EccE₃ via the ORBIT method (Murphy et al., 2018) (Figure 1A, Figure 1-figure supplement 1). EccE₃ is the final gene in the 11 gene long ESX-3 operon making insertion at this site less likely to disrupt regulation and expression of the operon. Deletion of the gene for the iron acquisition regulator ideR greatly increases chromosomal expression of ESX-3 from negligible amounts of protein to a yield sufficient for purification and structure determination (Figure 1-figure supplement 2A). Components of ESX-3 were pulled down using an anti-GFP nanobody (Rothbauer et al., 2008) and the GFP tag was proteolytically cleaved. After size exclusion chromatography, the peak fractions were pooled and analyzed biochemically and by cryo-electron microscopy.

**Global Structure of the ESX-3 Dimer**

Four components of the ESX-3 translocon complex EccB₃, EccC₃, EccD₃ and EccE₃ were stably affinity-purified as a large molecular weight species of about 900 kDa (Figure 1B and 1C, Figure 1-figure supplement 2B). The sample was imaged by cryo-EM and reconstructed revealing a dimeric structure, which can be divided into four areas: the flexible periplasmic multimerization domain, the stable transmembrane region, the stable upper cytoplasmic region, and the flexible lower cytoplasmic motor domain. (Figure 1D, Table 1). While the peak fraction does not contain particles of a larger size consistent with higher order oligomers, thorough examination of the
void volume revealed a small number of particles in a higher oligomeric state (Figure 1- figure supplement 3A-E). The resolution of the ESX-3 dimer varies substantially in different parts of the electron microscopy map and this heterogeneity was partially resolved through data processing (Figure 1-figure supplement 4 and Figure 1-figure supplement 5). Initially, the entire ESX-3 dimer was reconstructed to 4.0 Å resolution (Figure 1-figure supplement 6A). Using symmetry expansion, and focused classification and refinement techniques, the resolutions of targeted regions of the ESX-3 complex were improved to 3.7 Å for the transmembrane region and upper cytoplasmic region (Figure 1-figure supplement 6B-D), 5.8 Å resolution for the periplasmic region (Figure 1-figure supplement 6E), and ~7 Å resolution for the lower cytoplasmic region (Figure 1-figure supplement 6F). The highest resolution maps for each region were combined and filtered to the threshold of the lowest resolution map to form an overall 10 Å combination map for the entire ESX-3 dimer.

The ESX-3 dimer is comprised of ten total proteins, two copies each of EccB$_3$, EccC$_3$, and EccE$_3$ and four copies of EccD$_3$. Two pseudo-symmetric protomers referred to as i and ii, combine to form the ESX-3 dimer. Each protomer contains one copy of EccB$_3$, EccC$_3$, and EccE$_3$ and two conformationally distinct copies of EccD$_3$, referred to as EccD$_3$-bent and EccD$_3$-extended (Figure 1E) based on their highly asymmetric conformations. At 3.7 Å resolution, it was possible to build de novo atomic models for all observable amino acids in the transmembrane and upper cytoplasmic regions, except the two transmembrane helices of EccC$_3$ (Figure 1E and Figure 1-table supplement 1). The lower resolution regions of density, the EccC$_3$ transmembrane helices, EccC$_3$ ATPase 1, 2, and 3 domains and the EccB$_3$ periplasmic domain, were flexibly fit using homology models. Using this hybrid approach, a model of the entire ESX-3 dimer has been produced (Figure 1F).
EccD$_3$ forms a homodimer that encloses a large hydrophobic cavity

There are two copies of EccD$_3$ in each ESX-3 protomer (Figure 2A). The ubiquitin-like N-terminal domain of each EccD$_3$ molecule interacts with EccE$_3$ and EccC$_3$ in the cytoplasm, and a long linker joins the soluble domain of EccD$_3$ to 11 transmembrane helices. (Figure 2B and Figure 2- figure supplement 1A-F). The four EccD$_3$ molecules account for 44 of the total 54 transmembrane helices observed in the ESX-3 dimer. A distinct transmembrane cavity is formed by dimerization of the two copies of EccD$_3$ in each protomer with a cross-sectional diameter of ~20 x 30 Å without significant regions of constriction. Transmembrane helices 1, 9 and 10 interact across the cavity dimer interface in a tight bundle making passive lipid transport in and out of the cavity unlikely (Figure 2C). The inner surface of the periplasmic half of the cavity is composed primarily of hydrophobic residues and in our maps, eight extended densities consistent with hydrophobic lipid tails or detergent molecules line the periplasmic inner face of the cavity (Figure 2C). In contrast, the cytoplasmic face of the cavity has several polar residues and ordered hydrophobic densities are not visible.

In the cytoplasm below the membrane, a stable, upper cytoplasmic region is formed by interactions between the soluble domains of EccD$_3$-bent, EccD$_3$-extended, EccE$_3$, and EccC$_3$. The linker joining the cytoplasmic ubiquitin-like domain to transmembrane helix 1 (residues 100-127) of EccD$_3$ conserves a high sequence identity throughout evolution (Ashkenazy et al., 2016), yet it adopts two, distinct secondary structures resulting in the asymmetric placement of the cytoplasmic domains of EccD$_3$ (Figure 2D and Figure 2-figure supplement 1A-G). In EccD$_3$-bent, residues 100-127 are bent, folding into an α-helix and forming a nexus of stabilizing contacts with EccB$_3$ and EccC$_3$ (Figure 2-figure supplement 1H). In EccD$_3$-extended, residues 100-127 are
extended and fold into a shorter α-helix that interacts with EccE₃ and the cytoplasmic domain of EccD₃-bent (Figure 2-figure supplement 1I). This conformational flexibility suggests that if residues 100-127 were released from their associations with EccC₃ and EccE₃ they could rearrange into the alternative bent or extended conformation with little energetic barrier.

**EccC₃ and EccE₃ make extensive, stabilizing interactions with the asymmetric, cytoplasmic domains of EccD₃**

The next component of the stable upper cytoplasmic region is EccE₃. EccE₃ is positioned at the front of the ESX-3 dimer (Figure 3A), where the conserved transmembrane helix 1 of EccE₃ interacts with helix 11 of EccD₃-bent in the membrane. Helix 1 is followed by a second EccE₃ transmembrane helix, a linker helix, and then extends into the cytoplasm (Figure 3B and 3C and Figure 3-figure supplement 1A-D). The anti-parallel β-sheets of the cytoplasmic domain of EccE₃ have weak structural homology to glycosyl transferase proteins, however, the nucleotide binding pocket is absent in EccE₃, leaving it incapable of performing this function (Figure 3-figure supplement 1E and F, Figure 3-table supplement 1). EccE₃ does not have another obvious ligand or catalytic site. Two conserved helices in the cytoplasmic region of EccE₃ between amino acids 133 and 163 form extensive stabilizing interactions with both subunits of EccD₃ (Figure 3D, Figure 3-figure supplement 1G). These interactions hold the flexible linker of EccD₃-extended in the extended conformation and sterically hinder EccD₃-extended from assuming the bent conformation (Figure 3-figure supplement 1H). EccE₃ does not form direct protein-protein interactions with either EccB₃ or EccC₃ suggesting the contacts with EccD₃-extended and EccD₃-bent are extremely stable as EccE₃ was the tagged protein used to immunoprecipitate the ESX-3 dimer.
The final component of the stable upper cytoplasmic region is the domain of unknown function (DUF) of EccC3. EccC3 extends from the membrane into the upper and lower cytoplasmic regions (Figure 4A). Amino acids 1-33 and 94-403 of EccC3 were built de novo into the higher resolution region of the electron microscopy map revealing the structure of the DUF domain (Figure 4B, Figure 4-figure supplement 1A-C). The de novo model of the DUF has the typical Rossman fold of a nucleotide hydrolysis domain (Figure 4-figure supplement 1D and 1E) and its closest homolog by Dali search is the ATPase 1 domain of EccC of T. curvata. It is linked to the transmembrane domains by a long helical bundle making extensive contacts with the flexible linker region of EccD3-bent (Figure 4C). The DUF makes additional stabilizing contacts with the ubiquitin-like domains of EccD3-bent and EccD3-extended in the cytoplasm (Figure 4D).

When the transmembrane and upper cytoplasmic regions are compared between protomers, only the transmembrane helices of EccC3 and the N-terminal tail of EccB3 differ (Figure 4-figure supplement 2A and 2B), otherwise the protomers are superimposable. All four EccC3 transmembrane helices were modeled at 6 Å resolution through a combination of homology modeling and molecular dynamics. In protomer i, transmembrane helix 2 forms lipid mediated hydrophobic interactions with the transmembrane helix of EccB3 in protomer i, and transmembrane helix 1 interacts with transmembrane helix 2 of EccC3 in protomer ii. Transmembrane helix 1 of EccC3 in protomer ii is shifted relative to the protomer i conformation and does not directly interact with other proteins.

**The EccC3 motor domains are flexible and asymmetric across the dimer**

The motor domains containing the EccC3 ATPase 1, 2 and 3 hang below the DUF domain in the flexible lower cytoplasmic region. They were resolved at a lower resolution than the upper
cytoplasmic domain, but they are clearly asymmetric between protomers i and ii (Figure 4A).

Although the EccC3 ATPase 1 domains in both protomers are in a similar location relative to the DUF, the ATPase 2 and 3 domains do not super impose across protomers even at low resolution (Figure 4-figure supplement 2C) suggesting significant asymmetry between these domains. In protomer i, a homology model based on existing EccC structures fits well into the density; however in protomer ii, the interface between ATPase 1 and 2 is broken relative to the crystal structure with ATPase 2 and ATPase 3 rotated away from the crystal structure interface.

**EccB3 extends into the periplasm and stabilizes dimer formation**

The ESX-3 dimer is stabilized by cross-protomer interactions formed by the two EccB3 proteins. EccB3 begins in the cytoplasm with a flexible N-terminal tail leading into a linker helix, followed by a single pass transmembrane helix, and an extended periplasmic domain (Figure 5A and 5B, Figure 5-figure supplement 1A and B). The N-terminal tail of EccB3 from protomer i forms extensive cross-protomer contacts with EccB3, EccC3, EccD3-bent and EccD3-extended from protomer ii (Figure 5C, Figure 5-table supplement 1). The linker helix of EccB3 forms further protein-protein interactions with EccC3 and EccD3-bent. The transmembrane helix of EccB3 interacts with transmembrane helix 11 of EccD3-extended. Two hydrophobic tails consistent with a lipid or detergent molecules link the transmembrane helix of EccB3 to transmembrane helix 2 of EccC3 (Figure 5D). The two EccB3 periplasmic domains share a large interaction interface across the protomers further stabilizing dimerization. Homology models of two EccB3 proteins can be docked into the periplasmic domain (Figure 5-figure supplement 1C); however, this region is not resolved sufficiently to identify specific interactions. The majority of cross-protomer interactions involve EccB3, suggesting the periplasmic domain is essential for oligomerization.
A hexameric model of ESX-3

Previous reports have shown ESX-1 and ESX-5 form hexamers or higher order multimers (Beckham et al., 2017; Houben et al., 2012). We modeled a higher order oligomeric state of ESX-3 based on the low-resolution negative stain structure of ESX-5, which had C6 symmetry imposed during refinement (EMDB 3596). The ESX-3 translocon complex was modeled as a trimer of dimers by low pass filtering the density to 6 Å and docking into the ESX-5 negative stain map (Figure 6-figure supplement 1A). The model of the ESX-3 dimer transmembrane and upper cytoplasmic regions, including the EccC3 transmembrane helices modeled to 6 Å resolution, was fitted into the ESX-3 dimer map (Figure 6-figure supplement 1B). The angle between protomers in the trimer of dimers model alternates between 72° (the angle between protomers i and ii in the ESX-3 dimer map) and 48° and contains both experimentally observed conformations of EccC3. The complete model of the ESX-3 translocon complex was docked into the ESX-5 negative stain map in the same manner revealing major clashes between the low resolution periplasmic and motor domains in a hexameric form (Figure 6-figure supplement 1C). Accommodation of a hexameric complex would require extensive rearrangement of both EccC3 and EccB3.

Discussion

The ESX-3 structure presented here is purified without the addition of substrates or nucleotide. It is therefore likely to be in a conformation representing the end of the translocation cycle, awaiting either the direct binding of substrates, the binding of nucleotide or both, to reset a substrate-binding competent state. By fitting our dimer structure into a prior low resolution envelop we suggest a model of the oligomeric state of the complex, in close agreement with
Famelis et al. However, even allowing for major rearrangements in the more flexible regions of EccC$_3$ and EccB$_3$, a model built by the static trimerization of the experimentally determined dimer structure cannot itself explain the mechanism of action of ESX-3 secretion. The existence of an R-finger catalytic site for ATPase 1 of EccC$_3$ (Rosenberg et al., 2015) requires the R-finger of one protomer to insert into the active site of another protomer. Given the ~65 Å distance we observe between ATPase 1 domains, the completion and activation of the catalytic site of ATPase 1 by an R-finger will necessitate an extremely large rearrangement of the position of the ATPase domains. How might this rearrangement occur? We propose movements in the highly flexible EccD$_3$ linker lead to the release of EccE$_3$ and EccC$_3$ from their rigid positions, thus allowing for a rearrangement of the ATPase domains into an active conformation (Figure 6).

Once EccC$_3$ assembles into an oligomeric state, the substrate proteins will need to translocate through the inner membrane. We have considered two models for how pore formation and transit might occur: 1) though a pore created by the oligomerization of EccC$_3$ and EccB$_3$ or 2) through the large cavity created by the dimerization of EccD$_3$. In the first model (Figure 6A), the resting state of an ESX translocon complex is a hexamer, with disordered EccC ATPase domains free in the cytosol, stabilized by interactions with proteins not seen in the structures presented here (van Winden et al., 2016). It is possible the rare multi-dimer oligomeric state we see in the void volume, and also seen by Famelis et al., represents this state. In a hexamer model, the center of the multimer is formed by the transmembrane helices of EccC$_3$ and EccB$_3$, which create a cavity that could serve as a pore for translocation of substrates. These transmembrane helices are largely hydrophobic and do not contain obvious residues that would allow for the conductance of hydrated substrates. Thus the production of a protein transit channel would require either a large, conformational change in the transmembrane helices, likely
facilitated by movements in the cytoplasmic domains of EccC<sub>3</sub>, EccD<sub>3</sub> and EccE<sub>3</sub>, or a novel mechanism of action for transit through the central pore.

A hexameric pore created by EccC<sub>3</sub> agrees well with the documented mechanism of action for motor ATPases in the additional strand catalytic E (ADCE) division of P-loop NTPases (Erzberger and Berger, 2006), which includes EccC<sub>3</sub>. A hexameric pore also agrees with the proposed mechanism of action for other bacterial secretion systems, such as the Type IV secretion systems VirD4 coupling protein (Gomis-Rüth et al., 2001; Hormaeche et al., 2002), which is closely related evolutionarily to EccC<sub>3</sub> (Iyer et al., 2004). The hexamer model is thus firmly grounded in the motor ATPase and bacterial secretion systems literature, although the oligomeric state of VirD4 has recently been called into question (Redzej et al., 2017) and remains controversial (Llosa and Alkorta, 2017).

In a second, more speculative model, EccD<sub>3</sub> dimers form a channel for translocation of substrates (Figure 6B). The large cavity found in the EccD<sub>3</sub> dimer is striking and by structural homology, is unlike any other membrane protein in the Protein Data Bank. In our density maps, the EccD<sub>3</sub> dimer cavity appears capped on the periplasmic side by a dense layer of lipids. In contrast, on the cytoplasmic side the cavity does not exhibit bound lipids due to the polar residues lining the lower half of the cavity. The large cavity is of sufficient diameter to transit a folded EsxGH dimer, however given the strong hydrophobicity of the cavity the mechanism would not be mediated by water and would require a novel mechanism of secretion that has not been seen in other bacterial secretion systems. It is also possible that the cavity exists to transit a non-protein substrate such as a specific mycobacterial lipid. The ability to transport non-protein substrates could resolve some of the mysteries that remain about the relationships between ESX
systems, cell wall stability, lipid content, and nutrient acquisition (Barczak et al., 2017; Bosserman and Champion, 2017; Siegrist et al., 2014; Tufariello et al., 2016).

As each protomer contains an EccD3 cavity, the second model, proposing translocation through EccD3, does not require hexamerization. However, this model is also not incompatible with hexamerization, which would not block a substrate path through EccD3. Further, the role of the hexamer may not be to form a central channel for substrate transit. Rather, hexamerization could serve some other purpose. For example, it may tether functional dimers together, facilitate localization, or increase local concentration and allosteric control of enzymatic activity (Kuriyan and Eisenberg, 2007).

Although the ESX-3 structure presented here allows for mechanistic hypotheses about the transit of substrates across the inner-membrane, it does not provide sufficient information to allow for a structural model of transit across the outer, mycomembrane. The EccB periplasmic domain (Wagner et al., 2016) has been found to have similarity to the peptidoglycan binding phage protein PlyBC, which forms a ring inside the bacterial cell wall facilitating phage entrance into the cell. A hypothesis is that EccB is anchored to a larger outer membrane complex, but the purification conditions we have employed remove proteins required for its stabilization.

The description of ESX-3 presented here agrees in both protein composition and structure with the structure recently published by Famelis et al. Given the high sequence conservation among the ESX systems in mycobacteria and related actinobacteria, these structures likely represent an excellent framework for the structural modelling of the other ESX systems. Together, these structures provide a wealth of information about protein-protein interaction
interfaces and ESX complex architecture, which can be used to guide structure-based drug
design and to generate hypotheses for further mechanistic investigations.

**Materials and Methods**

**Strain construction**

*Mycobacteria smegmatis* MC²155 wild type and ΔIdeR cells were chromosomally tagged using
the ORBIT protocol (Figure S1). For wild type cells, the integrase and annealase expressing
plasmid was pKM444. For recombineering in the ΔIdeR strain, which already contained a
kanamycin resistance marker, we created a modified pKM444 plasmid where a zeocin
resistance cassette was inserted at the EcoIV restriction site. The tagging plasmid was pKM468
with a 3C protease cleavage site added before the eGFP tag via primer extension PCR. The
targeting oligo had the sequence: 5’

```
TGTGCATCACTGTTCTCCCGGCAACCACCTGCTGCACGTGAGCCAGGAGACTAC
CTAGGTGTTGTACACCACTGAGACCGCGCGTGGTTCGACCAGACCAACCGGCG
GATGACCCGCTTCTGCGCGTTCGACTGAACCCTTCACCGAGGCTCCG 3’.
```

*M. smegmatis* cells containing pKM444 were grown in an overnight liquid culture and induced
for annealase and integrase expression. Cells were prepared for electroporation and
electroporated with the targeting oligo and tagging plasmid. The transformed *M. smegmatis* were
plated on hygromycin containing 7H9 plates and incubated at 37°C for 3 days. Colonies were
verified for insertion of the tagging plasmid into the chromosome by PCR.

**Western Blotting**

100mL of EccE₃ tagged wild type and IdeR knock out cells were grown overnight to an OD₆₀₀ of
1.0-1.2. Cells were pelleted and resuspended in 1 mL of buffer (50 mM Tris-HCl pH 8.0, 150
mM NaCl, 1% DDM) and sonicated for 30 seconds. Cell lysates were run on a 4-20% SDS-PAGE gel (Genescript) and transferred to blotting paper using a BioRad Trans-blot turbo system. The blot was washed with PBS and blocked in a 5% milk/PBS-T solution for 1 hour. The blot was incubated with mouse anti-GFP monoclonal antibody (Roche) overnight. After rinsing with PBS-T, the blot was incubated with anti-mouse IgG HRP-conjugated antibody (R&D Systems) for 2 hours. After activation (Amersham) the blot was imaged on a BioRad ChemiDoc. The blot was stripped with stripping buffer (ThermoFisher Scientific) as per the manufactures instructions, and incubated overnight with rabbit anti-GroEL monoclonal antibody (Sigma-Aldrich). The blot was incubated with goat anti-rabbit IgG antibody HRP (GenScript) for 2 hours, activated (Amersham), and imaged on a BioRad ChemiDoc.

**Protein purification**

*M. smegmatis* was grown in 7H9 supplemented with 0.05% Tween 80 and 20 µg/mL kanamycin. After harvest, cells were washed three times with PBS and frozen in liquid nitrogen before lysis with a cryogenic grinder (SPEX SamplePrep). 24.9 g of powdered cell material was resuspended in 56.3 mL of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% DDM supplemented with protease inhibitor cocktail (SigmaFast) and 224 units Benzonase endonuclease. The suspension was stirred for 120 min at 4°C. After centrifugation for 30 min at 98,000 g, the supernatant was incubated with 1.4 mL of anti-GFP-nanobody resin for 110 min at 4°C. The resin was transferred to a column and washed sequentially with 28 ml of wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1% GDN), 14 mL of high salt wash buffer (50 mM Tris-HCl pH 8.0, 400 mM NaCl, and 0.1% GDN), and 14 mL of wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 0.1% GDN). To cleave off the purification tag, the resin was incubated o/n at 4°C with 70 units Pierce HRV 3C protease (Thermo Scientific Pierce) in 2.8 mL wash buffer supplemented with 0.2 mM
DTT. This resin was sedimented by gentle centrifugation (300 x g for 3 min), the supernatant collected, and the resin was subsequently washed with 1.4 mL wash buffer. The supernatant and wash fraction were combined and concentrated using an Amicon Ultra-4 centrifugal filter unit with a 100 kDa molecular weight cut-off. The sample was centrifuged for 6 min at 16,000 g before injection on a Superose 6 10/300 column equilibrated in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.021% GDN. Peak fractions were concentrated using a 0.5 ml centrifugal filter unit (Amicon, 100 kDa cut-off) to an A280 of 5.52 by Nanodrop reading.

**Blue-Native Polyacrylamide Gel Electrophoresis (BN-PAGE)**

BN-PAGE was carried out using the Invitrogen NativePAGE Novex Bis-Tris Gel system. Samples were prepared in a total volume of 10 µL using 0.25µl 5% G-250 sample additive. Electrophoresis was performed using light blue cathode buffer at a constant voltage of ~ 105 V for 3.5 h at 4°C. The gel was fixed and stained using the Pierce silver stain kit.

**Cryo electron microscopy – data acquisition**

Samples were frozen for cryo electron microscopy. Quantifoil R1.2/1.3, 400 mesh, cooper grids were glow discharged using a Solarus plasma cleaner (Gatan) with an H₂/O₂ mixture for 30 seconds. 2µL of sample were applied per grid and the grids were plunged into liquid ethane using a FEI Vitrobot Mark IV.

Initially, samples were screened, and test data sets were collected on a FEI Talos Arctica 200kV microscope equipped with a Gatan K2 Summit detector. For the initial screen of freezing conditions, 2,499 movies were collected at a magnification of 36,000 with a pixel size of 1.14, and a defocus range of -1.5 to -2.5 µm, an exposure time of 9 seconds, and a dose rate of 7 electrons/Å²/second. (Table S1). Data collection for the final structure presented in the main text.
was collected on a FEI Titan Krios at 300kV with a Gatan K2 Summit detector. Two imaging sessions were used. In the first imaging session, 2705 movies were collected at a magnification of 29,000 with a pixel size of 0.82, and a defocus of -0.4 to -1.2 µm, an exposure time of 10 seconds to collect 100 total frames, and a dose of 8 electrons/Å²/second. (Table S1). In the second imaging session, data was collected on the same microscope with the same detector, 4632 movies were collected at a magnification of 29,000 with a pixel size of 0.82, and a defocus range of -0.6 to -1.4 µm, an exposure time of 10 seconds to collect 80 total frames, and a dose of 6.7 electrons/Å²/second. Data used to analyze the void, plateau, and peak regions of the SEC profile were collected on a FEI Talos Arctica at 200kV with a Gatan K3 detector. All micrographs were collected at a magnification of 28,000 with a pixel size of 0.9, and a defocus range of -1.5 to -2.5 µm, an exposure time of 11.7 seconds to collect 117 total frames at a total dose of 56 electrons/Å²/second. For the void region, 1215 micrographs were collected.

**Cryo electron microscopy – data processing**

For all data, movies were motion corrected using MotionCor2 (Zheng et al., 2017) and CTF correction was performed using CTFFind4 (Rohou and Grigorieff, 2015). For the Arctica dataset, particles were picked using a gaussian blob in either RELION (Zivanov et al., 2018) or cisTEM (Grant et al., 2018) and initial 2D classification was performed to remove obvious artifactual particles. Initially, a shotgun approach was taken to generate several initial models using RELION, cisTEM, and cryosparc (Punjani et al., 2017). Once an initial model which contained realistic low-resolution features was generated, a user defined descent gradient was performed to improve the model with the goal of achieving accurate secondary structure features. First, all particles selected during 2D classification were refined in 3D against the randomly generated initial model. Second, a round of 3D classification with 4 classes and default RELION settings
was performed and the best class selected. Third, the best class was refined as a single class in 3D classification with increasing Tau2_Fudge and decreasing search angle size. The resulting EM density map had clear transmembrane helix densities and was used as the model for a new 3D reconstruction. This reconstruction was used to back project models for reference-based particle picking in RELION. Two rounds of 2D classification were performed and the best classes selected. One round of 3D classification was performed using the Tau2_Fudge value optimized during the previous run through (T=12) and the best class selected. A final 3D reconstruction of the Arctica data set yielded a map of about 4.7Å resolution (Figure S9). After motion correction and CTF determination, the final Titan Krios dataset was processed entirely using RELION. Particles were picked using a gaussian blob, and extracted as 4x binned particles. Two rounds of initial 2D classification were performed with T=3 on the binned particles and obvious artifactual particles were removed. The final reconstruction from the Arctica dataset was used as the initial model for a 3D reconstruction of the binned particles. 3D classification with 4 classes and the previously optimized Tau2_Fudge value, T=12, was performed on the binned particles. The two best class were selected and re-extracted without binning. A 3D reconstruction was performed. A mask was created for the high-resolution region of the reconstruction and 3D classification without image alignment was performed focused on this region. The best class was selected and the subsequent 4.0Å reconstruction is the consensus structure for the entire complex (Figure S3). Focused classification of each protomer, the periplasmic EccB region, and the ATPase 1,2, and 3 domains of EccC were performed. To perform focused classification, the center of mass of the region of interest was determined using chimera (Pettersen et al., 2004). Particles were recentered on this area and reextracted. Masks for the region of interest were generated and 3D classification without image alignment was
performed. The best class was selected and used for a focused 3D reconstruction without image
alignment of the region of interest. A reconstruction was generated and density outside of the
region of interest was subtracted. A final reconstruction of the masked and density subtracted
particles was then performed. This procedure improved the resolution of the protomer i to 3.75Å
and protomer ii 3.83Å, 5.8Å resolution for the EccB periplasmic domain, and ~7 Å resolution for
the EccC lower cytoplasmic region

To generate the symmetry expanded protomers based on non-point group symmetry (also
known as non-crystallographic symmetry or NCS), a transformation matrix between the two
protomers was calculated using chimera. Particles were then transformed and aligned using the
subparticles.py and star.py utilities in pyem (Asarnow et al., 2019) resulting in a particle stack
with twice as many particles as the input file, each focused on protomer i or protomer ii. Density
subtraction was performed to remove density outside of the symmetry expanded protomer, and
focused classification and refinement were performed as described above. This procedure
improved the resolution of the symmetry expanded protomer to 3.69Å resolution.

**Atomic model building**

The cytoplasmic domain from the crystal structure of EccD1 (PDB 4KV2) was docked into the
cytoplasmic domains of the two EccD3 molecules and the sequence was mutated. The remaining
transmembrane domains of EccD3 and the residues 14-93 of EccB3 were built *de novo* in Coot
(Emsley et al., 2010) using baton building. The alpha helices of EccE3 and EccC3 were initially
modeled using the RaptorX (Källberg et al., 2012) homology server. The loops and strands of
EccE3 and EccC3 were built in Coot using baton building. All models were subsequently refined
individually, as a symmetry expanded protomer, left and right protomers, and as the full model
using phenix real space refine (Afonine et al., 2018), Coot, and the MDFF (Trabuco et al., 2009) server, Namdinator (Kidmose et al., 2018) (Table S2).

**Low resolution modeling**

The left and right protomer map, periplasmic focused refined map, and lower cytoplasmic focused refined map were all docked into the consensus map and added together using chimera. The combined map was filtered to 10Å resolution to match the lowest resolution component. Homology models for amino acids 94-516 of EccB3, the transmembrane helixes of EccC3, and 404-1268 of EccC3 were generated using RaptorX. These models were fit into the combined map density using the fit map to model utility in Chimera. The full model was refined using phenix.real_space_refine.

**Model interpretation and display**

Buried surface area between subunits was calculated by PISA(Krissinel, 2015). Atomic models for individual proteins were compared against the PDB using the DALI server (Holm and Laakso, 2016). Chimera and ChimeraX (Goddard et al., 2018) were used to display maps and models for figure creation. Consurf (Ashkenazy et al., 2016) was used to produce multisequence alignments and to color structural models by homology.

**Acknowledgments:** The ΔIdeR strain was provided by Gloria Marcela Rodriguez. Data was collected at the electron microscopy core facility at the University of California, San Francisco with the assistance of Alexander G. Myasnikov. We thank Daniel Asarnow for making his pyem software available on github ahead of publication. We thank Huong Kratochvil and William DeGrado for preliminary lipid modeling and helpful discussions. We thank Robert Stroud and the members of his laboratory for stimulating discussions.
Funding: Research reported here was supported by the NIH under award numbers 1RO1AI128214, 1U19AI135990-01, and P01AI095208. NP is funded in part by the NIH Host Pathogen Interaction training grant (5T32AI060537). Equipment in the UCSF EM facility used in this research was purchased with an NIH S10 grant (1S10OD021741).

Competing interests: Authors declare no competing interests.

Data and materials availability: The electron microscopy maps are deposited to EMDB, code 20820. The atomic models are deposited to the PDB, code 6UMM.

Table and figure legends

Figure 1. Overview of the ESX-3 tagging, purification, and structure
(A) The ESX-3 operon in M. smegmatis and the placement of the purification tag. Genomic deletion of ideR derepresses ESX-3 to boost expression for purification. (B) SDS-PAGE of purified ESX-3 shows four major bands corresponding to EccB₃, EccC₃, EccD₃, and EccE₃. (C) Blue native page of the purified ESX-3 complex shows a single large molecular weight band around 900 kDa. (D) Merged maps of all focused refinement maps (grey transparency) of the ESX-3 dimer filtered to 10 Å resolution. The transmembrane and upper cytoplasmic focused maps (3.7 Å) segmented by subunit showing one copy per protomer of EccB₃ (pink), EccC₃ (blue), EccE₃ (orange), EccD₃-bent (yellow), EccD₃-extended (green). (E) Atomic models of the transmembrane and upper cytoplasmic regions. (F) A combined map of the full complex filtered to 10 Å resolution (grey transparency) with full models for each protein, EccD₃-bent (yellow), EccD₃-extended (green), EccE₃ (orange), EccC₃ (blue), and EccB₃ (pink).
Table 1. Data collection and refinement statistics

Collection parameters for initial test data set and void peak analysis on the Talos Arctica and final collection on Titan Krios microscopes. Refinement details for initial model, consensus map, and focused refinements.

Figure 2. The structure of EccD

(A) EccD3-bent (yellow) and EccD3-extended (green) in the context of the overall ESX-3 dimer (grey transparency). (B) Atomic models of EccD3-bent and EccD3-extended (C) An unsharpened electron microscopy density map of the ESX-3 dimer shows extra densities consistent with lipid or detergent molecules (teal) on the periplasmic face of the EccD3 cavity. (D) EccD3-bent (yellow) and EccD3-extended (green) aligned based on the transmembrane regions shows two distinct conformations of the EccD3 cytoplasmic domains. Amino acids 100-127 of EccD3 adopt a bent (yellow) and an extended (green).

Figure 3. The structure and protein-protein interactions of EccE

(A) The placement of EccE3 in the overall ESX-3 dimer. (B) Atomic model of EccE3 (C) Transmembrane helix 1 of EccE3 interacts with transmembrane helix 11 of EccD3-bent (D) Two soluble helices of EccE3 interact with EccD3-extended and EccD3-bent.

Figure 4. The structure and protein-protein interactions of EccC

(A) The placement of EccC3 in the overall ESX-3 dimer. (B) Atomic model of the EccC3 DUF (C) The stalk helices of EccC3 interact with EccB3, EccD3-bent, and EccD3-extended (D) Interactions between EccC3 and the ubiquitin-like domains of EccD3-bent and EccD3-extended in the cytoplasm.
**Figure 5. The periplasmic multimerization domain**

(A) EccB₃ (pink) in the context of the overall ESX-3 dimer (grey transparency). EccB₃ has a single pass transmembrane domain which extends into a large periplasmic domain which was resolved at 5.8 Å resolution. (B) Atomic models of the EccB₃ cytoplasmic and transmembrane domains, amino acids 14-93 and 32-93. (C) The N-terminus of EccB₃ forms extensive cross-protomer contacts with EccC₃ (blue), EccD₃-bent (yellow), and EccD₃-extended (green). (D) An unsharpened map of the ESX-3 dimer reveals ordered densities consistent with lipids or detergent molecules mediating the interactions between the EccB₃ transmembrane helix (marked with a pink dot) and the EccC₃ transmembrane helices.

**Figure 6. Two models of the ESX-3 translocon complex**

ATPase activity entails, at a minimum, oligomerization of ATPase 1 to bring the R-finger (R) into proximity of the catalytic site, marked by the Walker A motif (WA). This requires at least 65 Å of movement from the position seen in the structure. (A) The first model of ESX substrate secretion involves trimerization of the ESX-3 dimer followed by multimerization of the EccC ATPase domains into a stack of one to four rings of ATPases (B) The second model shown through the function of a single protomer of the ESX-3 complex. Substrates are selected by interaction with ATPase 3 of EccC and transported via the upper cytoplasmic bridge to the EccD cavity for secretion.

**Figure 1-figure supplement 1. ORBIT tagging of the chromosomal copy of EccE₃**

(A) Cartoon of the ESX-3 operon present in the *M. smegmatis* MC²155 chromosome. (B) The targeting oligo used to insert the attp site into the chromosome. (C) The tagging plasmid
recombined into the chromosome. (D) Cartoon of the resulting change to the chromosome after the ORBIT protocol was performed.

**Figure 1-figure supplement 2. ESX-3 dimer purification optimization**

(A) Western blot of solubilized cell material with EccE3 tagged with GFP. Lane 1, wild type background. Lane 2, ΔIdER background. (B) Size exclusion profile for the ESX-3 dimer purification run on a Superose 6 column.

**Figure 1-figure supplement 3. Examination of the void volume**

(A) Size exclusion profile with void volume and peak fractions indicated. (B) SDS-PAGE of all fractions from the beginning of elution until the end of the peak fraction. (C) BN-PAGE of the void and peak fractions (D) Cryo electron micrograph of the void volume. Scale bar 50 nm (E) 2D class averages of the particles picked from the void volume, some classes reveal possible higher order oligomers. Scale bar 20 nm.

**Figure 1-figure supplement 4. Initial data collection and initial model generation**

An initial data set was collected on a Talos Arctica microscope. The final refinement from this data processing was used as the starting model for future data processing.

**Figure 1-figure supplement 5. Data processing workflow for final data collection**

Data was collected on a Titan Krios. Particles from the final reconstruction were subsequently used for focused classification and refinement.
Figure 1-figure supplement 6. Consensus and focused refinements

(A) Consensus refinement map and FSC curve for the ESX-3 complex. Focused refinement maps and FSC curves for (B) the left protomer of the ESX-3 complex. (C) the right protomer of the ESX-3 complex (D) a symmetry expanded protomer (E) the periplasmic domain of EccB (F) the lower ATPase domains of EccC. All maps are unsharpened and unmasked.

Figure 1-table supplement 1. Model refinement statistics

Final refinement validation for all amino acids which were modeled de novo.

Figure 2-figure supplement 1. EccD3 map and model

Map to model fits for EccD3-bent for (A) transmembrane α-helix 1, amino acids 136-153, (B) transmembrane α-helix 1, amino acids 385-402, (C) soluble domain β-strands amino acids 10-15 and 90-96. Map to model fits for EccD3-extended for (D) transmembrane α-helix 1, amino acids 163-181 (E) transmembrane α-helix 11, amino acids 446-469 (F) soluble loop, amino acids 297-315. (G) Conservation mapping onto the residues of the EccD3 linker, amino acids 100-127. (H) Amino acids 100-127 from EccD3-bent (yellow) interact with EccB3 (pink) and EccC3 (blue). (I) Amino acids 100-127 of EccD3-extended (green), in a distinct conformation, interact with EccE3 (orange) and EccD3-bent (yellow)

Figure 3-figure supplement 1. EccE3 map and model

(A) Map and model comparison for transmembrane helix 1, amino acids 1 to 18. (B) Beta strand separation, amino acids amino acids 221 to 240. (C) a single beta strand, amino acids 117 to 126. (D) Conservation mapping onto TM1 of EccE3 and TM11 of EccD3bent (E) Comparison between the structure of EccE3 and the top DALI hit 5fxf-B. (F) Close up on the FADH (cyan) binding pocket of 5fxf-B which is not present in EccE3. (G) Conservation mapping onto the protein-
protein interaction soluble helix of EccE3, (H) Residues 133-163 of EccE3 interact substantially
with EccD3-bent and EccD3-extended, holding the flexible linker of EccD3-extended in the extended
conformation (H) EccE3 (orange) stabilizes the extended conformation of EccD3-extended (green)
and sterically hinders it from adopting the bent conformation (neon green).

**Figure 3-table supplement 1. Top Dali server hits**

Top 5 non-redundant hits from the Dali server compared against the full PDB for the *de novo-
built sections of EccB3, EccC3, EccD3, and EccE3.*

**Figure 4-figure supplement 1. EccC3 map and model**

- (A) Map and model of the EccC DUF domain, amino acids amino acids 97 to 130. (B) Two beta
  strands in the EccC DUF domain, amino acids 314 to 319 and 337 to 343. (C) Beta strand in the
  EccC DUF domain, amino acids 314 to 319. (D) Overlay of the model of the EccC DUF domain
  with the top DALI hit, 4NH0-A (white). (E) Close up of the ATP binding pocket of 4NH0-A.
  The ATPase fold remains intact in the DUF domain.

**Figure 4-figure supplement 2. Conformational differences between protomer i
and protomer ii**

- (A) Overlay of the focused refined maps of the transmembrane and upper cytoplasmic regions of
  protomer i (red) and protomer ii (blue). The two protomers are nearly identical except in the
  transmembrane domains of EccC3 and the N-terminal tail of EccB3. (B) Density subtraction of
  the two protomers in A highlights the regions of difference. Specifically, the transmembrane
  helixes of EccC3 adopt distinct conformations across protomers. (C) In protomer i, 4NH0, the
  crystal structure of EccC from *T. curvata*, fits well in the density. In protomer ii, ATPase 2 and 3
  fit well into the density but the position of ATPase 1 in the crystal structure, shown by the red
dotted circle, does not fit into the density. A density of about the right size for ATPase 1 is shown in yellow. Fitting of ATPase 1 in this density requires the disruption of the interface between ATPase 1 and ATPase 2.

**Figure 5-figure supplement 1. EccB₃ maps and models**

(A) Fit between map and model in the linker α-helix, amino acids 34 to 54. (B) Fit between map and model in the transmembrane helix, amino acids 65 to 85. (C) Fit between homology models of the soluble domain of EccB₃ and the periplasmic focused refinement map.

**Figure 5-table supplement 1. Buried surface area**

Buried surface area between subunits as measured by PISA.

**Figure 6-figure supplement 1. A hexameric model of the ESX-3 dimer**

(A) Fit of three copies of the transmembrane and upper cytoplasmic regions of the ESX-3 translocon complex into the ESX-5 hexameric structure (EMDB 3596). The ESX-3 translocon complex was filtered to 6 Å resolution to permit easy visualization of the EccC₃ transmembrane helices. (B) The model of the ESX-3 translocon complex for the full transmembrane and upper cytoplasmic domains (including homology models of the EccC₃ transmembrane helices accurate to 6 Å resolution) colored by protein. (C) The full model of the ESX-3 translocon complex including homology models for all three EccC₃ ATPase domains and the periplasmic domain of EccB₃ docked into the ESX-5 negative stain map. Red represents areas were atoms clash substantially.
References


**Data Collection**

<table>
<thead>
<tr>
<th>Collection</th>
<th>Initial Screening</th>
<th>Collection 1</th>
<th>Collection 2</th>
<th>Void peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microscope</td>
<td>Talos Arctica</td>
<td>Titan Krios</td>
<td>Titan Krios</td>
<td>Talos Arctica</td>
</tr>
<tr>
<td>Voltage (kV)</td>
<td>200</td>
<td>300</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>Detector</td>
<td>Gatan K2</td>
<td>Gatan K2</td>
<td>Gatan K2</td>
<td>Gatan K3</td>
</tr>
<tr>
<td>Pixel size (Å/pixel)</td>
<td>1.14</td>
<td>0.82</td>
<td>0.82</td>
<td>0.9</td>
</tr>
<tr>
<td>Exposure Time (s)</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>11.7</td>
</tr>
<tr>
<td>Electron dose (e-/Å²)</td>
<td>63</td>
<td>80</td>
<td>67</td>
<td>58</td>
</tr>
<tr>
<td>Defocus range (µm)</td>
<td>1.5-2.5</td>
<td>0.4-1.2</td>
<td>0.6-1.4</td>
<td>1.5-2.5</td>
</tr>
<tr>
<td>Number of micrographs</td>
<td>2499</td>
<td>2705</td>
<td>4632</td>
<td>1125</td>
</tr>
</tbody>
</table>

**Consensus Reconstruction**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Initial Screening</th>
<th>Collection 1 &amp; 2</th>
<th>Void peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td>Relion 2.1, cisTEM, and cryosparc</td>
<td>Relion 3.0</td>
<td>cisTEM</td>
</tr>
<tr>
<td># of particles, picked</td>
<td>240,000</td>
<td>778,149</td>
<td>259,333</td>
</tr>
<tr>
<td># of particles post, Class2D</td>
<td>138,000</td>
<td>554,901</td>
<td>640</td>
</tr>
<tr>
<td># of particles post, Class3D</td>
<td>46,830</td>
<td>362,438</td>
<td>NA</td>
</tr>
<tr>
<td># of particles post, skip align Class3D</td>
<td>NA</td>
<td>90,479</td>
<td>NA</td>
</tr>
<tr>
<td>Symmetry</td>
<td>C1</td>
<td>C1</td>
<td>NA</td>
</tr>
<tr>
<td>Map sharpening B-factor (Å²)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Final resolution (Å)</td>
<td>4.7 Å</td>
<td>4.0 Å</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Focused Refinements**

<table>
<thead>
<tr>
<th>Location of focus</th>
<th># of particles</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protomer i</td>
<td>76,967</td>
<td>3.8</td>
</tr>
<tr>
<td>Protomer ii</td>
<td>90,479</td>
<td>3.8</td>
</tr>
<tr>
<td>Symmetry expanded protomer</td>
<td>52,067</td>
<td>3.7</td>
</tr>
<tr>
<td>Periplasmic EccB</td>
<td>70,000</td>
<td>5.5</td>
</tr>
<tr>
<td>ATPase 1, 2, and 3</td>
<td>30,000</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 1 Data collection and refinement statistics**

Collection parameters for initial test data set and void peak analysis on the Talos Arctica and final collection on Titan Krios microscopes. Refinement details for initial model, consensus map, and focused refinements.
<table>
<thead>
<tr>
<th>Model Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total amino acids</td>
<td>3,156</td>
</tr>
<tr>
<td>All-atom clashscore</td>
<td>13.49</td>
</tr>
<tr>
<td>Ramachandran outliers</td>
<td>0.00%</td>
</tr>
<tr>
<td>Ramachandran allowed</td>
<td>5.39%</td>
</tr>
<tr>
<td>Ramachandran favored</td>
<td>94.61%</td>
</tr>
<tr>
<td>Rotamer Outliers</td>
<td>0.00%</td>
</tr>
<tr>
<td>C\text{\textgreek{b}} Deviations</td>
<td>0%</td>
</tr>
<tr>
<td>Cis-proline</td>
<td>0%</td>
</tr>
<tr>
<td>Cis-general</td>
<td>0.00%</td>
</tr>
<tr>
<td>Twisted proline</td>
<td>0%</td>
</tr>
<tr>
<td>Twisted general</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

**Figure 1-table supplement 1. Model Refinement Statistics.** Final refinement validation for all amino acids which were modeled de novo
<table>
<thead>
<tr>
<th>Hit Rank</th>
<th>PDB Chain ID</th>
<th>Z score</th>
<th>RMSD</th>
<th>lali</th>
<th>nres</th>
<th>%id</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EccB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6kre-b</td>
<td>2.7</td>
<td>5.8</td>
<td>59</td>
<td>444</td>
<td>7</td>
<td>T-complex protein 1 subunit alpha</td>
</tr>
<tr>
<td>2</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>EccC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4nh0-A</td>
<td>14.8</td>
<td>3.3</td>
<td>207</td>
<td>861</td>
<td>16</td>
<td>Cell division FtsK/SpoIIE (EccC)</td>
</tr>
<tr>
<td>2</td>
<td>2iuu-A</td>
<td>14.5</td>
<td>4.9</td>
<td>198</td>
<td>408</td>
<td>16</td>
<td>DNA translocase FtsK</td>
</tr>
<tr>
<td>3</td>
<td>4nla-A</td>
<td>13.9</td>
<td>3.4</td>
<td>205</td>
<td>555</td>
<td>18</td>
<td>Cell division FtsK/SpoIIE</td>
</tr>
<tr>
<td>4</td>
<td>4lya-A</td>
<td>12.8</td>
<td>3.4</td>
<td>198</td>
<td>521</td>
<td>14</td>
<td>EssC</td>
</tr>
<tr>
<td>5</td>
<td>5fvo-A</td>
<td>11.9</td>
<td>3.2</td>
<td>182</td>
<td>508</td>
<td>13</td>
<td>EssC</td>
</tr>
<tr>
<td></td>
<td>EccD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4kv2-A</td>
<td>9.9</td>
<td>1.6</td>
<td>70</td>
<td>90</td>
<td>30</td>
<td>ESX-1 Secretion system protein EccD1</td>
</tr>
<tr>
<td>2</td>
<td>3qky-A</td>
<td>9.3</td>
<td>10.3</td>
<td>138</td>
<td>251</td>
<td>7</td>
<td>Outer membrane assembly lipoprotein (BamD)</td>
</tr>
<tr>
<td>3</td>
<td>6e9v-J</td>
<td>9.2</td>
<td>7.2</td>
<td>156</td>
<td>211</td>
<td>10</td>
<td>Dhf79 filament</td>
</tr>
<tr>
<td>4</td>
<td>3j99-G</td>
<td>9.2</td>
<td>8.3</td>
<td>150</td>
<td>286</td>
<td>7</td>
<td>Vesicle-fusing ATPase</td>
</tr>
<tr>
<td>5</td>
<td>5wfu-A</td>
<td>9.1</td>
<td>5.8</td>
<td>119</td>
<td>234</td>
<td>3</td>
<td>14-3-3 protein beta/alpha</td>
</tr>
<tr>
<td></td>
<td>EccE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5fxf-B</td>
<td>7.5</td>
<td>7.1</td>
<td>184</td>
<td>525</td>
<td>6</td>
<td>Eugenol oxidase</td>
</tr>
<tr>
<td>2</td>
<td>5l6g-A</td>
<td>7.1</td>
<td>6.6</td>
<td>168</td>
<td>473</td>
<td>9</td>
<td>FAD linked oxidase-like protein</td>
</tr>
<tr>
<td>3</td>
<td>2axr-A</td>
<td>7.0</td>
<td>7.0</td>
<td>166</td>
<td>479</td>
<td>10</td>
<td>Glucooligosaccharide oxidase</td>
</tr>
<tr>
<td>4</td>
<td>6f73-B</td>
<td>7.0</td>
<td>7.9</td>
<td>164</td>
<td>548</td>
<td>12</td>
<td>MTVA0615</td>
</tr>
<tr>
<td>5</td>
<td>4p52-A</td>
<td>6.6</td>
<td>11.8</td>
<td>98</td>
<td>315</td>
<td>8</td>
<td>Homoserine kinase</td>
</tr>
</tbody>
</table>

Figure 3-table supplement 1. Top Dali server hits. Top 5 non-redundant hits from the Dali server compared against the full PDB for the de novo built sections of EccB3, EccC3, EccD3, and EccE3.
### Interaction interfaces

<table>
<thead>
<tr>
<th>Monomer 1</th>
<th>Monomer 2</th>
<th>Buried surface area (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer i</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>3154</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>3116.3</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>EccB - protomer i</td>
<td>1370.2</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>EccB - protomer ii</td>
<td>1295.6</td>
</tr>
<tr>
<td>EccE - protomer i</td>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer i</td>
<td>1157.2</td>
</tr>
<tr>
<td>EccE - protomer ii</td>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>1122.4</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>EccC - protomer i</td>
<td>881.9</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>EccC - protomer ii</td>
<td>825.1</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer i</td>
<td>EccC - protomer i</td>
<td>779.9</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>EccC - protomer ii</td>
<td>746.6</td>
</tr>
<tr>
<td>EccB - protomer i</td>
<td>EccC - protomer i</td>
<td>635.7</td>
</tr>
<tr>
<td>EccE - protomer i</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>608.5</td>
</tr>
<tr>
<td>EccE - protomer ii</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>600.1</td>
</tr>
<tr>
<td>EccB - protomer i</td>
<td>EccC - protomer ii</td>
<td>576.1</td>
</tr>
<tr>
<td>EccB - protomer ii</td>
<td>EccC - protomer ii</td>
<td>502.3</td>
</tr>
<tr>
<td>EccB - protomer i</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>482.1</td>
</tr>
<tr>
<td>EccB - protomer ii</td>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>297.2</td>
</tr>
<tr>
<td>EccB - protomer i</td>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>288.6</td>
</tr>
<tr>
<td>EccB - protomer ii</td>
<td>EccB - protomer ii</td>
<td>255.6</td>
</tr>
<tr>
<td>EccC - protomer i</td>
<td>EccC - protomer ii</td>
<td>213.6</td>
</tr>
<tr>
<td>EccB - protomer i</td>
<td>EccD&lt;sub&gt;bent&lt;/sub&gt; - protomer ii</td>
<td>168.9</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>EccB - protomer ii</td>
<td>56.6</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer ii</td>
<td>49.9</td>
</tr>
<tr>
<td>EccD&lt;sub&gt;extended&lt;/sub&gt; - protomer i</td>
<td>EccC - protomer ii</td>
<td>40.1</td>
</tr>
</tbody>
</table>

### Cross-protomer interaction interfaces

| EccB - protomer i | EccB, EccC, EccD - protomer ii | 1482.7 |
| EccC - protomer i  | EccC - protomer ii | 213.6   |
| EccD<sub>extended</sub> - protomer i | EccB, EccC, EccD<sub>extended</sub> - protomer ii | 146.6  |

**Figure 5-table supplement 1. Buried surface area.** Buried surface area between subunits as measured by PISA.
Figure 1. Overview of the ESX-3 tagging, purification, and structure. (A) The ESX-3 operon in *M. smegmatis* and the placement of the purification tag. Genomic deletion of ideR derepresses ESX-3 to boost expression for purification. (B) SDS-PAGE of purified ESX-3 shows four major bands corresponding to EccB3, EccC3, EccD3, and EccE3. (C) Blue native page of the purified ESX-3 complex shows a single large molecular weight band around 900 kDa. (D) Merged maps of all focused refinement maps (grey transparency) of the ESX-3 dimer filtered to 10 Å resolution. The transmembrane and upper cytoplasmic focused maps (3.7 Å) segmented by subunit showing one copy per protomer of EccB3 (pink), EccC3 (blue), EccE3 (orange), EccD3-bent (yellow), EccD3-extended (green). (E) Atomic models of the transmembrane and upper cytoplasmic regions. (F) A combined map of the full complex filtered to 10 Å resolution (grey transparency) with full models for each protein, EccD3-bent (yellow), EccD3-extended (green), EccE3 (orange), EccC3 (blue), and EccB3 (pink).
**Figure 1-figure supplement** 1. ORBIT tagging of the chromosomal copy of EccE3. (A) Cartoon of the ESX-3 operon present in the M. smegmatis MC2155 chromosome. (B) The targeting oligo used to insert the attP site into the chromosome. (C) The tagging plasmid recombined into the chromosome. (D) Cartoon of the resulting change to the chromosome after the ORBIT protocol was performed.
Figure 1-figure supplement 2. ESX-3 dimer purification optimization. (A) Western blot of solubilized cell material with EccE3 tagged with GFP. Lane 1, wild type background. Lane 2, ΔIdeR background. (B) Size exclusion profile for the ESX-3 dimer purification run on a Superose 6 column.
Figure 1-figure supplement 3. Examination of the void volume. (A) Size exclusion profile with void volume and peak fractions indicated. The ~100 is a contaminant that is eliminated computationally. (B) SDS-PAGE of all fractions from the beginning of elution until the end of the peak fraction. (C) BN-PAGE of the void and peak fractions (D) Cryo electron micrograph of the void volume. Scale bar 50 nm (E) 2D class averages of the particles picked from the void volume, some classes reveal possible higher order oligomers. Scale bar 20 nm.
Figure 1 - supplement 4. Initial data collection and initial model generation. An initial data set was collected on a Talos Arctica microscope. The final refinement from this data processing was used as the starting model for future data processing.
Picked ~778,000 particles bin4 and 2x class2D in RELION

7,337 micrographs

MotionCor2, ctffind4

~554,000 particles

3D Classification in RELION

~362,000 particles

4.2Å resolution

Re-extract particles without binning

3D Reconstruction

90,479 particles

4.0Å resolution

Focused classification and refinement

3D Reconstruction

Final model from arctica collection

Figure 1-figure supplement 5. Data processing workflow for final data collection. Data was collected on a Titan Krios. Particles from the final reconstruction were subsequently used for focused classification and refinement.
Figure 1-figure supplement 6. Consensus and focused refinements. (A) Consensus refinement map and FSC curve for the ESX-3 complex. Focused refinement maps and FSC curves for (B) the left protomer of the ESX-3 complex. (C) the right protomer of the ESX-3 complex (D) a symmetry expanded protomer (E) the periplasmic domain of EccB (F) the lower ATPase domains of EccC. All maps are unsharpened and unmasked.
Figure 2. The structure of EccD3

(A) EccD3-bent (yellow) and EccD3-extended (green) in the context of the overall ESX-3 dimer (grey transparency). (B) Atomic models of EccD3-bent and EccD3-extended (C) An unsharpened electron microscopy density map of the ESX-3 dimer shows extra densities consistent with lipid or detergent molecules (teal) on the periplasmic face of the EccD3 cavity. (D) EccD3-bent (yellow) and EccD3-extended (green) aligned based on the transmembrane regions shows two distinct conformations of the EccD3 cytoplasmic domains. Amino acids 100-127 of EccD3 adopt a bent (yellow) and an extended (green) conformation.
**Figure 2-figure supplement 1. EccD3 map and model.** Map to model fits for EccD3-bent for (A) transmembrane α-helix 1, amino acids 136-153, (B) transmembrane α-helix 1, amino acids 385-402, (C) soluble domain β-strands amino acids 10-15 and 90-96. Map to model fits for EccD3-extended for (D) transmembrane α-helix 1, amino acids 163-181 (E) transmembrane α-helix 11, amino acids 446-469 (F) soluble loop, amino acids 297-315. (G) Conservation mapping onto the residues of the EccD3 linker, amino acids 100-127. (H) Amino acids 100-127 from EccD3-bentt (yellow) interact with EccB3 (pink) and EccC3 (blue). (I) Amino acids 100-127 of EccD3-extended (green), in a distinct conformation, interact with EccE3 (orange) and EccD3-bent (yellow).
Figure 3. The structure and protein-protein interactions of EccE3. (A) The placement of EccE3 in the overall ESX-3 dimer. (B) Atomic model of EccE3 (C) Transmembrane helix 1 of EccE3 interacts with transmembrane helix 11 of EccD3-bent (D) Two soluble helices of EccE3 interact with EccD3-extended and EccD3-bent.
Figure 3-figure supplement 1. EccE3 map and model. (A) Map and model comparison for transmembrane helix 1, amino acids 1 to 18. (B) Beta strand separation, amino acids amino acids 221 to 240. (C) a single beta strand, amino acids 117 to 126. (D) Conservation mapping onto TM1 of EccE3 and TM11 of EccD3bent (E) Comparison between the structure of EccE3 and the top DALI hit 5fxf-B. (F) Close up on the FADH (cyan) binding pocket of 5fxf-B which is not present in EccE3. (G) Conservation mapping onto the protein-protein interaction soluble helix of EccE3, (H) Residues 133-163 of EccE3 interact substantially with EccD3-bent and EccD3-extended, holding the flexible linker of EccD3-extended in the extended conformation (H) EccE3 (orange) stabilizes the extended conformation of the EccD3-extended linker (green) and sterically hinders it from adopting the bent conformation (neon green).
Figure 4. The structure and protein-protein interactions of EccC3. (A) The placement of EccC3 in the overall ESX-3 dimer. (B) Atomic model of the EccC3 DUF (C) The stalk helices of EccC3 interact with EccB3, EccD3-bent, and EccD3-extended (D) Interactions between EccC3 and the ubiquitin-like domains of EccD3-bent and EccD3-extended in the cytoplasm.
Figure 4-figure supplement 1. EccC3 map and model

(A) Map and model of the EccC DUF domain, amino acids 97 to 130. (B) Two beta strands in the EccC DUF domain, amino acids 314 to 319 and 337 to 343. (C) Beta strand in the EccC DUF domain, amino acids 314 to 319. (D) Overlay of the model of the EccC DUF domain with the top DALI hit, 4NH0-A (white). (E) Close up of the ATP binding pocket of 4NH0-A. The ATPase fold remains intact in the DUF domain.
Figure 4-figure supplement 2. Conformational differences between protomer i and protomer ii

(A) Overlay of the focused refined maps of the transmembrane and upper cytoplasmic regions of protomer i (red) and protomer ii (blue). The two protomers are nearly identical except in the transmembrane domains of EccC3 and the N-terminal tail of EccB3.  
(B) Density subtraction of the two protomers in A highlights the regions of difference. Specifically, the transmembrane helixes of EccC3 adopt distinct conformations across protomers. 
(C) In protomer i, 4NH0, the crystal structure of EccC from T. curvata, fits well in the density. In protomer ii, ATPase 2 and 3 fit well into the density but the position of ATPase 1 in the crystal structure, shown by the red dotted circle, does not fit into density. A density of about the right size for ATPase 1 is shown in yellow. Fitting of ATPase 1 in this yellow region requires the disruption of the interface between ATPase 1 and ATPase 2.
Figure 5. The periplasmic multimerization domain. (A) EccB3 (pink) in the context of the overall ESX-3 dimer (grey transparency). EccB3 has a single pass transmembrane domain which extends into a large periplasmic domain which was resolved at 5.8Å resolution. (B) Atomic models of the EccB3 cytoplasmic and transmembrane domains, amino acids 14-93 and 32-93. (C) The N-terminus of EccB3 forms extensive cross-protomer contacts with EccC3 (blue), EccD3-bent (yellow), and EccD3-extended (green). (D) An unsharpened map of the ESX-3 dimer reveals ordered densities consistent with lipids or detergent molecules mediating the interactions between the EccB3 transmembrane helix (marked with a pink dot) and the EccC3 transmembrane helices.
Figure 5-figure supplement 1. EccB3 maps and models (A) Fit between map and model in the linker α-helix, amino acids 34 to 54. (B) Fit between map and model in the transmembrane helix, amino acids 65 to 85. (C) Fit between homology models of the soluble domain of EccB3 and the periplasmic focused refinement map.
Figure 6. Two models of the ESX-3 translocon complex. ATPase activity entails, at a minimum, oligomerization of ATPase 1 to bring the R-finger (R) into proximity of the catalytic site, marked by the Walker A motif (WA). This requires at least 65 Å of movement from the position seen in the structure. (A) The first model of ESX substrate secretion involves trimerization of the ESX-3 dimer followed by multimerization of the EccC ATPase domains into a stack of one to four rings of ATPases (B) The second model shown through the function of a single protomer of the ESX-3 complex. Substrates are selected by interaction with ATPase 3 of EccC and transported via the upper cytoplasmic bridge to the EccD cavity for secretion.
**Figure 6-figure supplement 1. A hexameric model of the ESX-3 dimer.** (A) Fit of three copies of the transmembrane and upper cytoplasmic regions of the ESX-3 translocon complex into the ESX-5 hexameric structure (EMDB 3596). The ESX-3 translocon complex was filtered to 6 Å resolution to permit easy visualization of the EccC3 transmembrane helices. (B) The model of the ESX-3 translocon complex for the full transmembrane and upper cytoplasmic domains (including homology models of the EccC3 transmembrane helices accurate to 6 Å resolution) colored by protein. (C) The full model of the ESX-3 translocon complex including homology models for all three EccC3 ATPase domains and the periplasmic domain of EccB3 docked into the ESX-5 negative stain map. Red represents areas were atoms clash substantially.