INTRODUCTION

Mycobacterial pathogens cause more than one million deaths each year (1). Key to their pathogenicity is the secretion of a wide range of virulence proteins via type VII secretion systems (T7SSs) (2). The ESX-5 T7SS is found almost exclusively in slow growing, pathogenic mycobacteria (3) and plays a key role in nutrient uptake and immune modulation during an infection (4–6). The importance of ESX-5 and other ESX secretion systems (ESX-1 to ESX-4) for the virulence of mycobacterial pathogens makes these membrane-spanning machineries a key target for the development of novel therapeutics (7). However, despite its importance, we still lack high-resolution structural information of a fully assembled core complex, which is critical for a mechanistic understanding of ESX-5 function and for informing drug development.

A previous low-resolution model of the ESX-5 complex revealed the presence of a hexameric core complex with dimensions suggesting that it spans the inner mycobacterial membrane (8). Unexpectedly, recent high-resolution structures of the ESX-3 complex from Mycobacterium smegmatis revealed a dimeric complex, instead of the expected hexameric pore assembly, and led to the suggestion that the dimer represented a building block of the full assembly. However, because of incomplete density for the membrane helices of EccC3, an accurate model of a hexameric pore could not be generated (9, 10). To address this question and reveal details of the ESX-5 structure, dynamics, and assembly, we used an integrated structural biology approach using single particle cryo–electron microscopy (cryo-EM), mass spectrometry (MS)–based cross-linking, and integrative modeling. Our data revealed a highly dynamic central ESX-5 pore embedded within a rigid membrane scaffold formed by six protomeric units, with a diameter ranging from 4 to 10 Å, suggesting that we captured ESX-5 in the partially open, secretion-competent state. Moreover, our structure showed an unanticipated break in symmetry between the membrane and cytoplasmic regions, with a sixfold symmetric arrangement, and the periplasmic region, which displayed a twofold symmetry. We propose that this conformational plasticity and the flexibility in both the periplasmic and cytoplastic regions are critical for protein substrate recognition, transport, and release.

RESULTS

Overall structural organization of the hexameric ESX-5 pore complex

To elucidate the structure of the T7SS, we purified the Mycobacterium xenopi ESX-5 complex for single-particle cryo-EM analysis (fig. S1). The initial 3.9-Å resolution map calculated without imposed symmetry revealed substantial differences in the interpretability of different regions (fig. S2). Consistent with the previous low-resolution model (8), the membrane and cytoplasmic regions of the complex both displayed sixfold symmetry, with the latter showing a larger extent of positional disorder toward the periphery of the complex. We therefore generated two new maps imposing C6 symmetry for modeling these regions: a global map with a 3.4-Å resolution and a locally refined map of the cytosolic part of the protomeric units at 3.0-Å resolution (Fig. 1A and table S1). This enabled the de novo building of the transmembrane (TM) and membrane-adjacent cytoplasmic domains to 88% completeness with a detectable sequence register (EccB5 18-73; EccC5 12-417; EccD5-1 23-502; EccD5-2 18-494; and EccE5 95-332) (Fig. 1D and fig. S4).

To model the periplasmic region of the ESX-5 complex, we generated a third map without imposed symmetry restraints with a resolution of 4.6 Å. In contrast to the sixfold symmetry of the membrane and cytoplasmic regions, the periplasmic region of the ESX-5 secretion complex displayed an approximate C2 symmetry. Since the resolution of this map was not sufficient for de novo model building, we used a homology model of the periplasmic EccB5 domain (11) and distance restraints obtained from cross-linking MS to build an integrative ensemble model. The resulting model was composed of all six copies of EccB5 (Figs. 1, B and C and 2B) at an...
We observed that one of the EccD5 molecules are colored to show their location in the segmented EM map in (A). Low-resolution density has been shown as a lighter shade for EccC5 and EccE5. (annotated them as EccD5-1 and EccD5-2, respectively. The EccD5 is proximal and the second one is distal to the central pore, and we views of the rigid core of the ESX-5 membrane complex.

Because of the observed sixfold symmetry of the TM and cytosolic segments of the ESX-5 complex, we first defined its overall arrangement as consisting of six identical protomeric subcomplexes. Each of these protomers comprises EccB5, EccC5, two copies of EccD5, and EccE5 (Fig. 2A and fig. S4). Within each protomeric unit, EccD5 was observed to assume an elliptical ring-shaped dimer composed of 22 (11 per EccD5 monomer) TM helices (TMHs) forming the core of the protomer. We observed that one of the EccD3 molecules is proximal and the second one is distal to the central pore, and we annotated them as EccD5-1 and EccD5-2, respectively. The EccD5 dimer was stabilized by hydrophobic interactions between TMH9 and TMH10 of EccD5-1 and the only EccB5 TMH establish a scaffold for the central pore axis (see below). In our model, contacts between TMH11 from EccD5-1 and the only EccB5 TMH establish a scaffold for the central pore. Although we observed that EccD5-2 TMH11 interacts with the TM domain of EccE5 in a similar manner, we have not included the TMHs of EccE5 in our high-resolution model due to weaker density, indicative of flexibility at the periphery (fig. S4).

In addition, we noted that EccD5 also acts as a main connector for the cytoplasmic regions of the ESX-5 protomer (Fig. 2A, cytoplasm inset, and fig. S4, A to D). The ubiquitin-like (Ubl) domains of EccD5 dimerize and interact with the cytoplasmic domain of EccE5 at the periphery of the ESX-5 complex. On the cytosolic side of the central pore, the EccD5 Ubl domain formed interactions with the EccC5 domain of unknown function (DUF), which follows the EccC5 stalk domain. The stalk domain is flanked by the long loop between TMH6 and TMH7 of EccD5-1 as well as the linker between EccD5-2 TMH1 and the EccD5-2 Ubl domain (fig. S4, B and D). Besides these interactions at the interface between the membrane and the cytoplasm (Fig. 2A, membrane cytoplasm inset), we observed several specific electrostatic interactions of the EccC5 stalk domain with the N-terminal EccB5 helix, situated parallel to the membrane. The density of our map was of sufficient quality to identify some of the electrostatic interactions unambiguously (fig. S4F). As for the membrane region, the EccD5 Ubl domain dimer seems to act as a central scaffold for the other ESX-5 cytosolic subunits (fig. S4D).

In our structure, the DUF domain was the most distal rigid component at the cytosolic face of the ESX-5 complex. We attributed the less interpretable density for the remaining C-terminal part of EccC5 to the increased flexibility of adenosine triphosphatase (ATPase) domains 1 to 3, as previously observed (8–10). To examine the conformational space of ATPase domains further, we generated the C1 map at low threshold values (fig. S6A). Integrative modeling estimated precision of 7 Å (Materials and Methods, fig. S5, and table S2). Taking the cytosolic, TM, and periplasmic segment models together, we generated a model of the overall architecture of the complete ESX-5 pore complex (Fig. 1).

**Architecture of the ESX-5 protomer**

Because of the observed sixfold symmetry of the TM and cytosolic segments of the ESX-5 complex, we first defined its overall arrangement as consisting of six identical protomeric subcomplexes. Each of these protomers comprises EccB5, EccC5, two copies of EccD5, and EccE5 (Fig. 2A and fig. S4). Within each protomeric unit, EccD5 was observed to assume an elliptical ring-shaped dimer composed of 22 (11 per EccD5 monomer) TM helices (TMHs) forming the core of the protomer. We observed that one of the EccD3 molecules is proximal and the second one is distal to the central pore, and we annotated them as EccD5-1 and EccD5-2, respectively. The EccD5 dimer was stabilized by hydrophobic interactions between TMH9 and TMH10 of EccD5-1 and the only EccB5 TMH establish a scaffold for the central pore axis (see below). In our model, contacts between TMH11 from EccD5-1 and the only EccB5 TMH establish a scaffold for the central pore. Although we observed that EccD5-2 TMH11 interacts with the TM domain of EccE5 in a similar manner, we have not included the TMHs of EccE5 in our high-resolution model due to weaker density, indicative of flexibility at the periphery (fig. S4).

In addition, we noted that EccD5 also acts as a main connector for the cytoplasmic regions of the ESX-5 protomer (Fig. 2A, cytoplasm inset, and fig. S4, A to D). The ubiquitin-like (Ubl) domains of EccD5 dimerize and interact with the cytoplasmic domain of EccE5 at the periphery of the ESX-5 complex. On the cytosolic side of the central pore, the EccD5 Ubl domain formed interactions with the EccC5 domain of unknown function (DUF), which follows the EccC5 stalk domain. The stalk domain is flanked by the long loop between TMH6 and TMH7 of EccD5-1 as well as the linker between EccD5-2 TMH1 and the EccD5-2 Ubl domain (fig. S4, B and D). Besides these interactions at the interface between the membrane and the cytoplasm (Fig. 2A, membrane cytoplasm inset), we observed several specific electrostatic interactions of the EccC5 stalk domain with the N-terminal EccB5 helix, situated parallel to the membrane. The density of our map was of sufficient quality to identify some of the electrostatic interactions unambiguously (fig. S4F). As for the membrane region, the EccD5 Ubl domain dimer seems to act as a central scaffold for the other ESX-5 cytosolic subunits (fig. S4D).

In our structure, the DUF domain was the most distal rigid component at the cytosolic face of the ESX-5 complex. We attributed the less interpretable density for the remaining C-terminal part of EccC5 to the increased flexibility of adenosine triphosphatase (ATPase) domains 1 to 3, as previously observed (8–10). To examine the conformational space of ATPase domains further, we generated the C1 map at low threshold values (fig. S6A). Integrative modeling estimated precision of 7 Å ( Materials and Methods, fig. S5, and table S2). Taking the cytosolic, TM, and periplasmic segment models together, we generated a model of the overall architecture of the complete ESX-5 pore complex (Fig. 1).
EccB₅ forms an elongated interaction network around a central cleft in the periplasm

The EccB₅ model at the periplasmic face showed two distinct EccB₅ trimers arranged with an approximate twofold symmetry (EccB₅-A/EccB₅-B/EccB₅-C) used for further characterization below and EccB₅-A'/EccB₅-B'/EccB₅-C', highlighted in Fig. 2A, periplasm inset). The two trimers formed an elongated keel-shaped assembly with overall dimensions of 20 nm in length, 10 nm in width, and 8 nm in height divided by a central cleft. The assembly was formed by dimers of trimers with three distinct EccB₅ dimeric arrangements, denoted as "V-shape," "parallel," and "peripheral" (Fig. 2B). Of note, in our ESX-5 structure, the angle between the two V-branches in each of the two EccB₅ trimers was 113° (mean value; fig. S7), which is substantially different from the 85° angle of V-shaped EccB₃ conformation found in the ESX-3 dimer structure (9). However, the angles between the principal axis of the periplasmic domains of the protomers within each of the two EccB₅ trimers and the axis of the respective TM helix vary substantially (fig. S7A), demonstrating substantial flexibility of the hinge between the TM helix and the periplasmic domain within each EccB₅ subunit. Our data thus indicate that the transition of hexameric symmetry observed for the TM section of the ESX-5 pore complex and lower symmetry of the elongated periplasmic EccB₅ domain arrangement is crucial for the conformational dynamics and opening of the ESX-5 pore (see below).

Structural basis of hexameric ESX-5 assembly and central pore

Next, we examined our structural model to identify key interactions that support hexameric ESX-5 assembly. Within the TM segment, one of the main interactions required for hexamerization occurs between the TM domains of EccB₅ and EccC₅. We observed that the EccC₅ TMH1 of each protomer interacts with the EccB₅ TMH of the neighboring protomer via hydrophobic interactions and is further stabilized by the EccC₅ TMH2 (Fig. 2C). This domain swapping interaction was repeated in an anticlockwise fashion when viewed from the periplasm (Fig. 2D), thus generating an interlocking mechanism of next neighbor protomers within the TM section of the overall ESX-5 complex. The N terminus of EccB₅, which hooks into the loop between EccD₅-1 TMH10 and TMH11 of the neighboring protomer, mediates an additional interprotomer interaction at the cytoplasmic side (fig. S4F). Our model suggests that these EccB₅-EccD₅-1 interactions drive EccC₅ TMH1 domain swapping within the TM section of the ESX-5 complex. In turn, this may lead to changes in the orientation of the EccB₅ periplasmic domains, causing them to interlock and form a hexameric assembly in the periplasm. We propose that these steps are crucial in forming a secretion-competent complex with a central pore, required for substrate translocation.

A notable feature of our ESX-5 core complex is the central pore (Fig. 3). The TMHs of EccC₅ from each of the six protomers, most notably the TMH2 of one EccC₅ subunit that forms a domain swapping interaction with TMH1 of the neighboring subunit, contribute to the formation of this central pore (Fig. 3). In contrast to most other TMHs of the complex, the density of TMH2 is less well defined (Fig. 3C), which implies that this helix is more flexible and adopts a range of orientations, as indicated by our calculated ensemble model (Fig. 3D). A highly conserved proline P73 (Fig. 3, A to C, and fig. S8) triggers a notable kink in the middle of TMH2, which results in increased flexibility of its C-terminal part. As observed in other
transport proteins, proline residues positioned in related positions are key for their regulation and function (12, 13). Therefore, our structural findings suggest that P73 acts as a hinge point facilitating conformational flexibility of the pore required for substrate transport.

TMH2 also comprises a number of phenylalanine and methionine residues that are common to this helix across different ESX systems (Fig. 3A and fig. S8). In the ESX-5 structure, F66, F72, and F75 were observed to flank P73, orienting some of these aromatic side chains toward the inner surface of the pore and thus determining the pore diameter. To quantify this, we analyzed the diameter of the 100 highest-scoring ensemble models to sample the range of potential conformations of TMH2 (Fig. 3D). Our results showed that the narrowest constrictions of the central pore reduce the pore diameter to slightly below 1 nm. As this pore diameter is too narrow to secrete folded substrates with an estimated width of around 2.2 nm (14), we hypothesize that our model, in the absence of a bound substrate, represents a gated, secretion-competent state. Our structural findings indicate that even subtle changes in the interactions between the two EccC5 TMHs, the EccB5 TMH, and the innermost EccD5 TMH could induce conformational changes of the central ESX-5 pore, either toward a closed state or a more open state, to regulate substrate translocation. Comparing EccB angles between the hexameric ESX-5 pore complex and previously published dimeric ESX-3 subcomplexes (9, 10) demonstrate inherent flexibility within this arrangement, supporting our hypothesis (fig. S7, C and D).

Analysis of the central pore of the ESX-5 system shows that this T7SS system displays intriguing similarities to other transport systems. The lining of secretion pores with bulky hydrophobic amino acids has been reported for other secretion machineries, in which the hydrophobic residues act to gate the pore, thus regulating secretion. For example, in A/B type toxins, such as the anthrax toxin from *Bacillus anthracis*, the phenylalanine rings or "Φ clamp" at the top of the channel restrict the passage of a folded lethal factor (15, 16). In another example, a highly conserved Met-Met-Met loop of the export apparatus complex of T3SS forms a molecular gasket, which constricts the channel to less than 10 Å to preclude secretion (17). These systems do not share any sequence similarity with the ESX systems, suggesting that a highly dynamic state of the pore is a fundamental principle of bacterial secretion machineries that evolved through convergent mechanisms of different TM protein scaffolds.
DISCUSSION
In conclusion, our cryo-EM structure of the hexameric *M. xenopi* ESX-5 pore complex reveals the architecture of the T7SS and the dynamic nature of the central secretion pore. According to our data, the considerably more flexible arrangements of the distal cytosolic ATPase domains of EccC5 and the dimer-of-trimers formed by the EccB5 periplasmic domains define the overall dynamics of the system, suggesting that major conformational changes occur during substrate recognition and substrate release into the periplasm.

Following the submission of our work to a preprint server, another structure of a related ESX-5 complex from *Mycobacterium tuberculosis* was made available (18). In the structure, the presence of an additional conserved subunit with protease activity (MycP5) plugs the periplasmic EccB5 subunit domains in a triangular arrangement. In our experiments, MycP5, despite being present in the expression plasmid, does not stably associate with the purified *M. xenopi* ESX-5 complex, suggesting a regulatory rather than structural role of this subunit. Although in both structures there is an unexpected symmetry break between the TM-cytosolic region and the periplasmic domains, the nature of the EccB5 intersubunit interactions and the level of structural dynamics are distinct. These differences seem to correlate with distinct pore conformations in both structures: closed in the presence of MycP5 and more open in the absence of MycP5 (Fig. 4). Comparison of the two structural models suggests that the periplasmic arrangement of the EccB5 subunits and the level of conformational flexibility induced in the absence of MycP5 could play a role in the opening of the pore. As this hypothesis is based on the comparison of the presently available structural data only, functional data are still required for its ultimate validation. Whether there is further opening of the central pore to accommodate ESX-5 substrates destined for secretion, beyond the conformation observed in our structure (Fig. 4), remains to be established and will likely require structural information obtained in the presence of a bound substrate.

Among the diverse repertoire of bacterial secretion machineries characterized to date, the overall architecture of the T7SS is distinct. On the other hand, as the subunit composition in different T7SS is highly conserved, we argue that our findings on the overall architecture and associated conformational dynamics of the ESX-5 secretion machinery are generally applicable to other T7SS members. Our data provide a key step toward a comprehensive mechanistic understanding of the T7SS secretion mechanism.

![Fig. 4. Scheme of conformational changes occurring in type VII secretion.](http://advances.sciencemag.org/)

**Fig. 4. Scheme of conformational changes occurring in type VII secretion.** (A) Side views of the ESX-5 overall conformations proceeding from an inactive to active complex. We propose that the closed state (i) based on the *M. tuberculosis* ESX-5 structure in the presence of the protease MycP5 (18) represents an inactive conformation. The second step toward active secretion (ii), based on the ESX-5 structure reported here, would represent a gated, secretion-competent conformation capable of initiating substrate translocation. In this conformation, the periplasmic EccB5 subunits are reorganized into a dimer-of-trimers arrangement, generating an elongated central cleft (B). The change in orientation of the EccB5 TMH reorients the EccC5 TMHs in the membrane to allow opening the central TM pore (C). In the closed conformation the same EccC5 TMHs are organized into four-helix bundles formed from two protomeric units without a visible pore. In all presently known structures, the cytosolic EccC5 domains adopt multiple conformations, demonstrating a high level of conformational plasticity requirements for ESX-5 substrate recognition and gating into the pore. We speculate that there could be an even more open conformation of the ESX-5 pore during active translocation of folded substrates, resulting from further reorientation of the EccB5 TMHs and the two pore-forming EccC5 TMHs (iii). EccB5 is shown in orange, pink, and red; EccC5 is shown in blue and light blue; substrate is shown in brown; and MycP5 is shown in yellow. For reasons of clarity, the remaining structural EccD5 and EccE5 components of the TM section are shown in gray. In (C), the EccC5 TMH1 is in light blue and TMH2 is in dark blue. The TMHs have been numbered according to the protomer they correspond to. The numbering of the closed conformation is based on Bundec et al. (18).
understanding of distinct steps in T7SS-targeted translocation across the thick multilayered mycobacterial membrane. As tuberculosis (TB) remains as the longest lasting pandemic in the history of human civilization, and the recent rise of multidrug-resistant TB strains, development of novel anti-TB therapeutics remains a top societal priority. As ESX-5 represents a major determinant for mycobacterial pathogenicity, our structural findings may facilitate future target-driven drug discovery against TB and other mycobacterial diseases.

**MATERIALS AND METHODS**

**Molecular biology**

Polymerase chain reaction (PCR) was performed using Q5 DNA polymerase (New England Biolabs). For cloning, *Escherichia coli* DH5α was used. The eccD5 gene was amplified by PCR to include the N-terminal ubiquitin-like domain (residues 1 to 129) and inserted into the pMyNT vector using SíCE methods (19, 20), generating pMyNT-EccD5. The *M. xenopi* ESX-5 complex was expressed from pMV-ESX-5 vector in *M. smegmatis* as previously described (8).

**Protein expression and purification**

Expression vectors were transformed into *M. smegmatis* mc2155 *groEL1Δ* (21) and grown in Middlebrook 7H9 medium (BD Biosciences) supplemented with 0.2% (w/v) glucose (Carl Roth), 0.05% (v/v) Tween-80 (Carl Roth), and 0.2% (v/v) glycerol (Carl Roth) with appropriate antibiotics. For expression of the *M. xenopi* ESX-5 complex, cells were cultured to an optical density of 600 nm (OD600) of 1.5 and pelleted by centrifugation. EccD5 was expressed using an inducible promoter, cells were grown to OD600 of 1.0, induced with 1% acetic acid, cultured for further 24 hours at 37°C, and pelleted by centrifugation. The ESX-5 complex was purified as previously described in (8). The final sample was vitrified in 20 mM tris (pH 8) and 150 mM NaCl.

For the purification of EccD5, cells were resuspended in buffer A [20 mM tris (pH 8.0), 300 mM NaCl, and 20 mM imidazole] with EDTA-free protease inhibitors (Roche) and deoxyribonuclease (Sigma-Aldrich). Cells were lysed by high-pressure emulsification, and unbroken cells were removed by centrifugation at 4°C for 20 min (19,000 g).

**Cross-linking MS analysis**

Fifty micrograms of purified ESX-5 complex was cross-linked by addition of an iso-stoichiometric mixture of H12/D12 isotope-coded, di-succinimidyl-suberate (DSS) (Creative Molecules). The cross-linking reaction (final concentration of 1 mM) was incubated for 30 min at 37°C and quenched by addition of ammonium bicarbonate to a final concentration of 50 mM for 10 min at 37°C. Cross-linked proteins were denatured using urea and RapiGest (Waters) [final concentration of 4 M and 0.05% (w/v)], respectively. Samples were reduced using 10 mM dithiothreitol (30 min at 37°C), and cysteines were carbamidomethylated with 15 mM iodoacetamide (30 min in the dark). Protein digestion was performed using 1:100 (w/w) LysC (Wako Chemicals) for 4 hours at 37°C and then finalized with 1:50 (w/w) trypsin (Promega) overnight at 37°C after the urea concentration was diluted to 1.5 M. Samples were then acidified with 10% (v/v) trifluoroacetic acid and desalted using OASIS HLB μElution Plate (Waters). Cross-linked peptides were enriched using size exclusion chromatography (SEC) (22).

Collected SEC fractions were analyzed by LC-coupled tandem MS (MS/MS) using a nanoACQUITY Ultra Performance LC (UPLC) system (Waters) connected online to linear ion trap quadrupole (LTQ)–Orbitrap Velos Pro instrument (Thermo Fisher Scientific). Peptides were separated on a BEH300 C18 (75 mm by 250 mm by 1.7 mm) nanoACQUITY UPLC column (Waters) using a stepwise 60-min gradient between 3 and 85% (v/v) acetonitrile in 0.1% (v/v) formic acid. Data acquisition was performed using a top-20 strategy, where survey MS scans (mass/charge ratio range of 375 to 1600) were acquired in the Orbitrap (R = 30,000), and up to 20 of the most abundant ions per full scan were fragmented by collision-induced dissociation (normalized collision energy = 40, activation Q = 0.250) and analyzed in the LTQ. To focus the acquisition on larger cross-linked peptides, charge state 1, charge state 2, and unknown were rejected. Dynamic exclusion was enabled with a repeat count = 1, exclusion duration = 60 s, list size = 500, and mass window of ±15 parts per million (ppm). Ion target values were 1,000,000 (or 500-ms maximum fill time) for full scans and 10,000 (or 50-ms maximum fill time) for MS/MS scans. The sample was analyzed in technical duplicates.

To assign the fragment ion spectra, raw files were converted to centroid mzXML format using a raw converter and then searched using xQuest (23) against a FASTA database containing the sequences of the cross-linked proteins. Posterior probabilities were calculated using X!Prophet (23), and results were filtered using the following parameters: false discovery rate = 0.05, minimum Ascore = 0.95, MS1 tolerance window of –4 to +7 ppm, and identity (Id) score > 36.

**X-ray crystallography and data processing**

EccD5 crystallized in initial conditions from the Morpheus screen (Molecular Dimensions) containing 0.06 M MgCl2, 0.03 M CaCl2, 0.1 M tris(bicine (pH 8.5), 10% OEG 20k, 20% poly(ethylene glycol) monomethyl ether 550. Diffraction data were collected at EMBL beamline P13 at the PETRA III storage ring (DESY, Hamburg, Germany). The data were processed with XDS (24) and merged with AIMLESS (24, 25), and the relevant statistics are shown in table S1. We used the EccD5 model from *M. tuberculosis* [Protein Data Bank (PDB) ID: 4KV2 (26)] as a molecular replacement candidate (45% sequence identity to *M. xenopi* EccD5). After the successful placement of the model using Phaser (27), manual building was performed in Coot (28). The model was refined using REFMAC5 (29).

**Cryo-EM sample preparation and data acquisition**

For cryo-EM, 3.6 μl of the ESX-5 void peak fraction was applied on freshly glow-discharged Quantifoil R2/1 Cu 200 mesh grids with 2-nm continuous carbon. The sample was blotted for 2 s and vitrified in a liquid propane/ethane mix using a Vitrobot Mark IV at 10°C and 100% humidity. The grid was screened at the cryo-EM facility at the Centre for Structural Systems Biology (Hamburg, Germany), and high-resolution cryo-EM data were collected on a Titan Krios operated at 300 kV (Thermo Fisher Scientific FEI) equipped with a K3 direct detection camera (Gatan) and a Bio-Quantum K3 energy filter (Gatan) operated by SerialEM (30) at the EMBL Cryo-Electron Microscopy Service Platform (Heidelberg, Germany). A total of 27,873 movies with 40 frames were recorded in counting mode, with a total dose of 49.34 e/Å2 and a pixel size of 0.645 Å. The underfocus range was set to 0.7 to 1.7 μm, with a step size of 0.1 μm.
Atomic model building and refinement

As there are no reliable, high-resolution structures of any of the ESX-5 components or their homologs available in the PDB (33), we built de novo a model of the TM and nearby cytoplasmic regions of the complex (EccB5 18-73; EccC5 12-417; EccD5-1 23-502; EccD5-2 18-494; and EccE5 95-352). An initial model was traced into a masked, focused refinement map using ARP/wARP cryo-EM module with default parameters (34). Next, domains for which we solved the high-resolution crystal structure (EccD5, residues 17 to 107; fig. S3, A to C) were fitted into the focused refinement map as rigid bodies using a Jiggle Fit tool from Coot (28) (fig. S6). The resulting model was completed manually using Coot in regions with local resolution allowing for unambiguous de novo model tracing. The interpretation of poorly resolved map regions was aided by alternative blurring and sharpening of the map in Coot. We used an iterative approach, where each manual model building step was followed with sequence assignment using findMySequence program (35), which allowed for an identification and correction of tracing errors (insertions and deletions). Loops that were resolved in the density but difficult to trace manually were built using the RosettaES density-guided enumerative-sampling algorithm from the Rosetta suite (36). The complete protomer model built into a focused refinement map was expanded to a complex using symmetry operations directly from the C6-symmetrized map using phenix.find_ncs_from_density (37) and completed manually in Coot. Apart from solving minor symmetry conflicts, we traced the model fragments that were resolved in the symmetrized map only. These included two TMHs of EccC5 (TMH1 and TMH2; residues 37 to 94). First, TMH1 was built de novo in Coot into the better resolved density and assigned to the sequence using findMySequence program (35), which allowed for an unambiguous determination of the helix direction and sequence register. Subsequently, the second TM helix (TMH2) was built using the RosettaES density-guided enumerative-sampling algorithm followed with refinement with C6 symmetry. We also added to the model the most distant to the central pore EccD5-2 helix (TMH11) based on a model of the corresponding helix in EccD5-1. Geometry of the models was improved in ChimeraX (38) using ISOLDE (39) tool. Last, models of the protomer and full complex were refined against corresponding maps using phenix.real_space_refine (40), with nonbonded restraints weight increased to 200. For the complex, additional restraints to the initial model coordinates and strict rotamer matching were used.

Lipid analysis

To determine whether unassigned densities in the EM map may correspond to lipids copurified with the complex, we extracted lipids from purified ESX-5 samples using methanol-chloroform extraction. Cold methanol (200 μl) was added to 50 μg of ESX-5 complex in 20 mM tris (pH 8) and 150 mM NaCl and vortexed thoroughly. The sample was kept on ice, and 500 μl of cold chloroform and 200 μl of water were added and incubated for 10 min before centrifugation for 500 rpm for 5 min at 4°C. The phase-separated chloroform layer (300 μl) was removed and dried using a SpeedVac for 20 min at room temperature. Dried samples were resuspended in 100 μl of methanol before LC-MS analysis, performed as previously described (41).
Integrative modeling

The models of the hexameric assembly of the periplasmic domains of EccB5 (amino acids 74 to 490) and cytoplasmic ATPase domains of EccC5 (amino acids 431 to 1390) were built using an integrative modeling protocol similar to what was previously used by us (43–45). The modeling procedure described in more detail below is implemented as a custom software based on Integrative Modeling Platform (46) version 2.13 and Python Modeling Interface (47) further described in (48). All additional code and input files necessary to reproduce the steps will be released on Zenodo repository upon publication.

The TM region of the ESX-5 structure built de novo as above and a homology model of the monomeric EccB5 periplasmic domain and EccC5 ATPase domains were used as input for modeling. The homology models of EccB5 and EccC5 were built using Modeller (47) based on the crystal structure of EccB1 of M. tuberculosis (PDB ID: 3X3M (11)) and EccC of Thermomonospora curvata (PDB ID: 4NH0 (49)) and using the sequence alignment obtained from the HHpred server (50). The nonsymmetrized (C1) EM map and available cross-links were used as modeling restraints. Owing to the low resolution (<10 Å) of the periplasmic and cytoplasmic regions, the high-frequency noise in the EM map was removed using a Gaussian filter with a SD of 3 Å for EccB5 and 5 Å for EccC5. In addition, to limit the conformational space, the fitting was performed using only a segment of the EM map not yet occupied by the TM region of the ESX-5 structure. The models were additionally restrained using high-confidence cross-links above an xQuest (23) ld score of 36. At this threshold, two and six cross-links could be mapped to the EccB5 and EccC5 sequences, respectively, and used for modeling (fig. S1D and table S4).

As the first step of the modeling, large libraries of alternative fits to the EM map of the monomeric EccB5 and EccC5 structures were generated using the FitMap tool of the UCSF Chimera (32). The fitting was performed using 100,000 random initial placements, cross-correlation about the mean as the fitting score [Chimera’s “cam” score (32), equivalent to Pearson correlation coefficient], and the requirement of at least 80% of the input structure being covered by the EM map envelope defined at a permissive density threshold. This resulted in 9268 unique alternative fits for EccB5 and 5068 fits for EccC5 after clustering.

Second, the resulting alternative fits of the monomeric structures and the TM region of the ESX-5 structure were built de novo as above and used as input for the simultaneous fitting of six copies of EccB5 using the EM map and cross-link restraints and likewise for six copies of EccC5. The fitting was performed through simulated annealing Monte Carlo optimization that generates alternative configurations of the fits precalculated as above. The optimization was performed independently 4000 times with 12,000,000 Monte Carlo steps for each run for EccB5 and 2500 times with 12,000,000 Monte Carlo steps for each run for EccC5. The sampling exhaustiveness was assessed by ensuring that (i) the score converges in individual runs, (ii) no new better scoring models appear with extra runs, and (iii) the score distributions in two random samples of the models are statistically similar (fig. S6, A to C). The scoring function for the optimization was a sum of the EM fit restraint represented as the P values of the precalculated domain fits [calculated as described in (43–45)], cross-linking restraints, clash score, connectivity distance between neighboring domains, a term preventing overlap of the protein mass with the TM region, and a two- or sixfold symmetry restraint for EccB5 and EccC5, respectively. During the optimization, the structures were simultaneously represented at two resolutions—in Ca-atom representation and a coarse-grained representation—in which each 10-residue stretch was converted into one bead. The 10-residue bead representation was used for all restraints to increase computational efficiency except for the domain connectivity and cross-link restraints, for which the Ca-only representation was used for reasons of accuracy.

Last, top-scoring models from the previous step were subjected to a refinement coupled to an analysis of exhaustiveness of conformational sampling and estimation of model precision using a procedure proposed by Viswanath et al. (51). To this end, the models from the first modeling stage (simultaneous fitting based on the alternative fits) were split into two random subsets. The top 30 models from each subset were refined using a Monte Carlo simulated annealing optimization in which the structures were moved in the EM map with small rotational and translational increments. The scoring function consisted of cross-correlation to the EM map, domain connectivity restraint, clash score, a term preventing overlap of the protein mass with the TM region, and a two- or sixfold symmetry restraint as above. For EccC5, the monomeric homology model was split into two rigid bodies at the boundary between the ATPase domains 1 and 2, and an elastic network restraint was applied to enable limited flexibility between these domains. Each of the 30 models was refined with 200 independent runs with 260,000 steps. The top-scoring models from each of the two runs were selected, leading to two independent samples of refined models (about 1000 models in each sample). For EccC5, the multiple fitting step converged to a single top-scoring model; thus, only a single model was selected for the refinement, and the samples were generated by splitting the resulting refined models. The scores of the two samples were compared to each other to ensure convergence (fig. S5, D and E). The highest sampling precision at which sampling was exhaustive was determined on the basis of the root mean square deviation (RMSD) comparisons between all models and clustering at incremental RMSD thresholds using the statistical tests provided by Viswanath et al. (51) (fig. S5F). The two samples were then clustered at the resulting precision level (fig. S6G), and for each cluster, the model precision, defined as the average RMSD distance to cluster centroid, was calculated. The top 10 scoring models from all refined models were taken as the final ensemble model of the ESX-5 with the EccB5 (fig. S5H). All the top 10 models satisfied both EccB5 cross-link restraints (with a distance threshold of 30 Å; fig. S5I). The models will be deposited in the PDB-dev database upon publication.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/26/eabg9923/DC1

REFERENCES AND NOTES

Acknowledgments: We thank F. Weis and W. Hagen (EMBL) for access to the KRIO microscope of the EMBL Structural and Computational Biology Unit, Heidelberg. We also acknowledge access to the EM facility at the Centre for Structure Systems Biology, Hamburg, for sample preparation and screening. X-ray crystallography data were collected at beamline P13 operated by the EMBL Hamburg Unit at the PETRA III storage ring (DESY, Hamburg, Germany). We acknowledge the EMBL Metabolomics Core Facility and the technical support provided by the SPC facility at EMBL Hamburg. A. Parret, L. Ciccarelli, G. Pompidor, and V. Rantos are thanked for their contributions during the different stages of the project. Funding: This work was supported by the Joachim-Herz-Stiftung (Hamburg) via...
the Infectophysics consortium (to M.W.) and an add-on fellowship (to K.S.H.B.) and by the Federal Ministry of Education and Research of Germany (FKZ 031 L0100). **Author contributions:** M.W. supervised and supported the project. K.S.H.B produced samples of the *M. xenopi* ESX-5 complex suitable for cross-linking MS and structural analysis, including biophysical characterization and x-ray crystallography. S.A.M. performed high-resolution cryo-EM experiments. C.R. processed and interpreted the data, with support by E.M. K.S.H.B. and M.R. conducted cross-linking MS experiments and supervised by M.M.S. G.C. conducted and developed approaches for rapid structural model building and interpretation. J.K. and D.S.Z. performed integrative modeling. K.S.H.B., C.R., G.C., J.K., and M.W. wrote the paper.

**Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The structure of EccD129 and the ESX-5 complex have been deposited in RCSB Protein Data Bank with the respective PDB IDs 7B9F and 7B7J. EM maps have been deposited to the Electron Microscopy Data Bank (EMDB) under the following IDs, C6 full map (EMD ID -12105), C1 map (EMD-12674), and the C6 local refined map (EMD-12103). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 8 February 2021
Accepted 10 May 2021
Published 25 June 2021
10.1126/sciadv.abg9923

Structure of the mycobacterial ESX-5 type VII secretion system pore complex
Katherine S. H. Beckham, Christina Ritter, Grzegorz Chojnowski, Daniel S. Ziemianowicz, Edukondalu Mullapudi, Mandy Rettel, Mikhail M. Savitski, Simon A. Mortensen, Jan Kosinski and Matthias Wilmanns

Sci Adv 7 (26), eabg9923.
DOI: 10.1126/sciadv.abg9923

ARTICLE TOOLS http://advances.sciencemag.org/content/7/26/eabg9923
SUPPLEMENTARY MATERIALS http://advances.sciencemag.org/content/suppl/2021/06/21/7.26.eabg9923.DC1
REFERENCES This article cites 45 articles, 5 of which you can access for free http://advances.sciencemag.org/content/7/26/eabg9923#BIBL
PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title Science Advances is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).