



RESEARCH ARTICLE

Structural insights into substrate recognition by the type VII secretion system

Shuhui Wang^{1,2,7}, Kaixuan Zhou³, Xiaolin Yang^{1,2,7}, Bing Zhang^{1,2,7}, Yao Zhao^{1,2,7}, Yu Xiao^{1,2,7}, Xiuna Yang¹, Haitao Yang¹, Luke W. Guddat⁶, Jun Li¹✉, Zihe Rao^{1,3,4,5}✉

¹ Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

² CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

³ State Key Laboratory of Medicinal Chemical Biology, College of Life Sciences and College of Pharmacy, Nankai University, Tianjin 300353, China

⁴ Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China

⁵ National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

⁶ School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

⁷ University of Chinese Academy of Sciences, Beijing 100101, China

✉ Correspondence: lijun1@shanghaitech.edu.cn (J. Li), raozh@mail.tsinghua.edu.cn (Z. Rao)

Received April 10, 2019 Accepted July 2, 2019

ABSTRACT

Type VII secretion systems (T7SSs) are found in many disease related bacteria including *Mycobacterium tuberculosis* (*Mtb*). ESX-1 [early secreted antigen 6 kilodaltons (ESAT-6) system 1] is one of the five subtypes (ESX-1~5) of T7SSs in *Mtb*, where it delivers virulence factors into host macrophages during infection. However, little is known about the molecular details as to how this occurs. Here, we provide high-resolution crystal structures of the C-terminal ATPase₃ domains of EccC subunits from four different *Mtb* T7SS subtypes. These structures adopt a classic RecA-like α/β fold with a conserved Mg-ATP binding site. The structure of EccCb1 in complex with the C-terminal peptide of EsxB identifies the location of substrate recognition site and shows how the specific signaling module “LxxxMxF” for *Mtb* ESX-1 binds to this site resulting in a translation of the bulge loop. A comparison of all the ATPase₃ structures shows there are significant differences in the shape and composition of the signal recognition pockets across the family, suggesting that distinct signaling sequences of substrates are required to be specifically recognized by

different T7SSs. A hexameric model of the EccC-ATPase₃ is proposed and shows the recognition pocket is located near the central substrate translocation channel. The diameter of the channel is ~25-Å, with a size that would allow helix-bundle shaped substrate proteins to bind and pass through. Thus, our work provides new molecular insights into substrate recognition for *Mtb* T7SS subtypes and also a possible transportation mechanism for substrate and/or virulence factor secretion.

KEYWORDS type VII secretion system, *Mycobacterium tuberculosis*, ATPase, virulence factor, substrate recognition

INTRODUCTION

Mycobacterium tuberculosis (*Mtb*), the pathogen of human tuberculosis (TB), infects one-third of the world’s population resulting in more than one million deaths annually (World Health Organization 2018). Whilst there are several drugs available to treat TB, these are required to be administered over a long time-frames (up to years or longer), can be expensive and can have serious side-effects. Moreover, these current medications are losing their effectiveness due to the emergence of extensively and multi-drug resistant strains of *Mtb*. Thus, there is an urgent call for new anti-TB therapies to be developed.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13238-019-00671-z>) contains supplementary material, which is available to authorized users.

One potential drug target for *Mtb* is the type VII secretion system (T7SS), which secretes virulence factors during host infection (Stanley et al. 2003). However, very little is known about its molecular features to use as a target for rational structure-based drug design. *Mtb* T7SS was first identified by comparing the genome of *Mtb* H37Rv strain with those of the attenuated vaccine strains of *Mycobacterium bovis* Bacille Calmette-Guérin (BCG) and *Mycobacterium microti* (Pym et al. 2002). A 9.5 kb genomic segment, named RD1, was deleted from all vaccine strains (Mahairas et al. 1996; Behr et al. 1999). The RD1 region encodes nine proteins including two virulence factors, the 6 kDa early secretory antigenic target ESAT-6 (also called EsxA) (Sorensen et al. 1995; Brodin et al. 2004) and its dimerization partner, culture filtrate protein CFP-10 (also known as EsxB) (Berthet et al. 1998), which together form a newly identified and highly specialized secretion pathway, the ESX-1 secretion system [early secreted antigen 6 kDa (ESAT-6) system 1] (Stanley et al. 2003). In the genome of *Mtb*, five paralog *esx* gene clusters (ESX-1 to -5) have now been identified as T7SSs (Bitter et al. 2009). Each *esx* cluster encodes small secreted proteins including the Esx proteins themselves, a cytosolic ATPase EccA, and the core transmembrane components EccB, EccC, EccD, EccE and MycP (Bitter et al. 2009). The small secreted proteins are about 100 amino acids in length, containing the conserved WxG and Yxxx[D/E] motifs (Daleke et al. 2012).

The T7SSs play several important roles in *Mtb*. The ESX-1 secretion system is the best studied T7SS amongst the five ESXs. It secretes not only the major virulence factors EsxA and EsxB, but also ESX-1 secretion-associated proteins, EspA, EspB and EspC (Brodin et al. 2006; McLaughlin et al. 2007; Raghavan et al. 2008; Carlsson et al. 2009; Champion et al. 2009). Single-cell fluorescence resonance energy transfer (FRET) observed a pore-forming toxin formed by the ESX-1 secreted EsxA (Smith et al. 2008), finally allowing *Mtb* escape from the innate host immune responses (McLaughlin et al. 2007; van der Wel 2007; Wong and Jacobs 2011; Simeone et al. 2012). ESX-1 can also regulate host immunoreaction and contribute to granuloma formation and bacterial dissemination (Koo et al. 2008; Davis and Ramakrishnan 2009; Carlsson et al. 2010; Volkman et al. 2010; Stoop et al. 2011). Interestingly, ESX-1 is also found to be essential for conjugal DNA transfer in some mycobacteria (Flint et al. 2004; Gray et al. 2013; Derbyshire and Gray 2014). The mycobacterial ESX-3 secretion system has a function in metal homeostasis mediated by mycobactin (Siegrist et al. 2009; Serafini et al. 2013). The EsxG-EsxH complex, which is the substrate for ESX-3, is suggested to strongly induce interferon gamma secretion in T cells of mice infected with *Mtb* (Skjøt et al. 2002; Majlessi et al. 2003; Hervas-Stubbs et al. 2006). ESX-5 is the most recently evolved ESX system which is only present in slow-growing mycobacteria (Gey Van Pittius et al. 2001), and is the major pathway for secretion of PE family (Pro-Glu motif containing) and PPE family (Pro-Pro-Glu motif containing) proteins which are localized at the mycobacterial cell surface (Sampson 2011). *Mycobacterium marinum* *esx-5*

mutant showed hyper virulence in adult zebrafish, suggesting ESX-5 might have a role in downregulation of the host immune response (Weerdenburg et al. 2012). ESX-2 and ESX-4 may be non-essential systems and not host-oriented (Cole et al. 2001; Singh et al. 2015). Nevertheless, ESX-4 was reported to be crucial for conjugal recipient activity in *Mycobacterium smegmatis* (Gray et al. 2016).

Using negative staining electron microscopy, the core ESX-5 apparatus from *Mycobacterium xenopi*, which consists of EccB5, EccC5, EccD5 and EccE5, has been shown to form a 1.8 MDa membrane complex with six-fold symmetry (Beckham et al. 2017). EccC, the motor subunit of T7SS, is essential to assemble a stable membrane complex for the secretion process (Houben et al. 2012). It contains a two-pass transmembrane domain, an unknown function domain (DUF), and three Fts-SpoIIIE-like ATPase domains (ATPase₁, ATPase₂, and ATPase₃) (Pallen 2002). For ESX-1 in *Mtb*, EccC is split between the ATPase₁ and ATPase₂ domains, into two proteins, named EccCa1 and EccCb1 (Fig. 1A). Yeast two-hybrid experiment showed that the seven residues at the C-terminus of EsxB interact with EccCb1 as a signal sequence required for secretion (Champion et al. 2006). The flexible C-terminus of PE25, a substrate of ESX-5, has also been shown to be important for secretion (Daleke et al. 2012). In addition, a highly conserved motif (Yxxx[D/E]) has been identified in all known mycobacterial T7SS substrates, and is absolutely required for secretion (Daleke et al. 2012). The structure of the cytoplasmic portion of EccC in complex with the signal sequence of EsxB from *Thermomonospora curvata* and an experiment where the C-terminal sequences between *Tc*EsxB and *Mt*EsxB were swapped indicates that EccC specifically recognizes substrates from different species (Rosenberg et al. 2015). Thus, though there is some data regarding substrate recognition by T7SS, the molecular details about how this works have remained limited, especially in *Mtb*.

Here we have determined the crystal structures of the EccC-ATPase₃ domains of ESX-1, ESX-2, ESX-3 and ESX-5 from *Mtb* in a pre-activated state. Amongst them, we observed ATP binds in a similar mode at a highly conserved nucleotide-binding site. The structure of EccCb1 in complex with EsxB shows the precise interactions between the ATPase₃ domain and the specific signaling motif (LxxxMxF) at the C-terminus of EsxB after a translation of the bulge loop. Sequence and structural comparisons reveal that the substrate recognition pockets for the distinct *Mtb* T7SS subtypes differ significantly. These findings provide new insights into substrate recognition by T7SS, thus broadening our knowledge as to how T7SS secretes substrates and virulence factors.

RESULTS

Overall structure of EccC-ATPase₃

An initial attempt to obtain the crystal structure of EccC fragment from *Mtb* containing three or two ATPase domains was failed. Due to the significance of ATPase₃ domain in

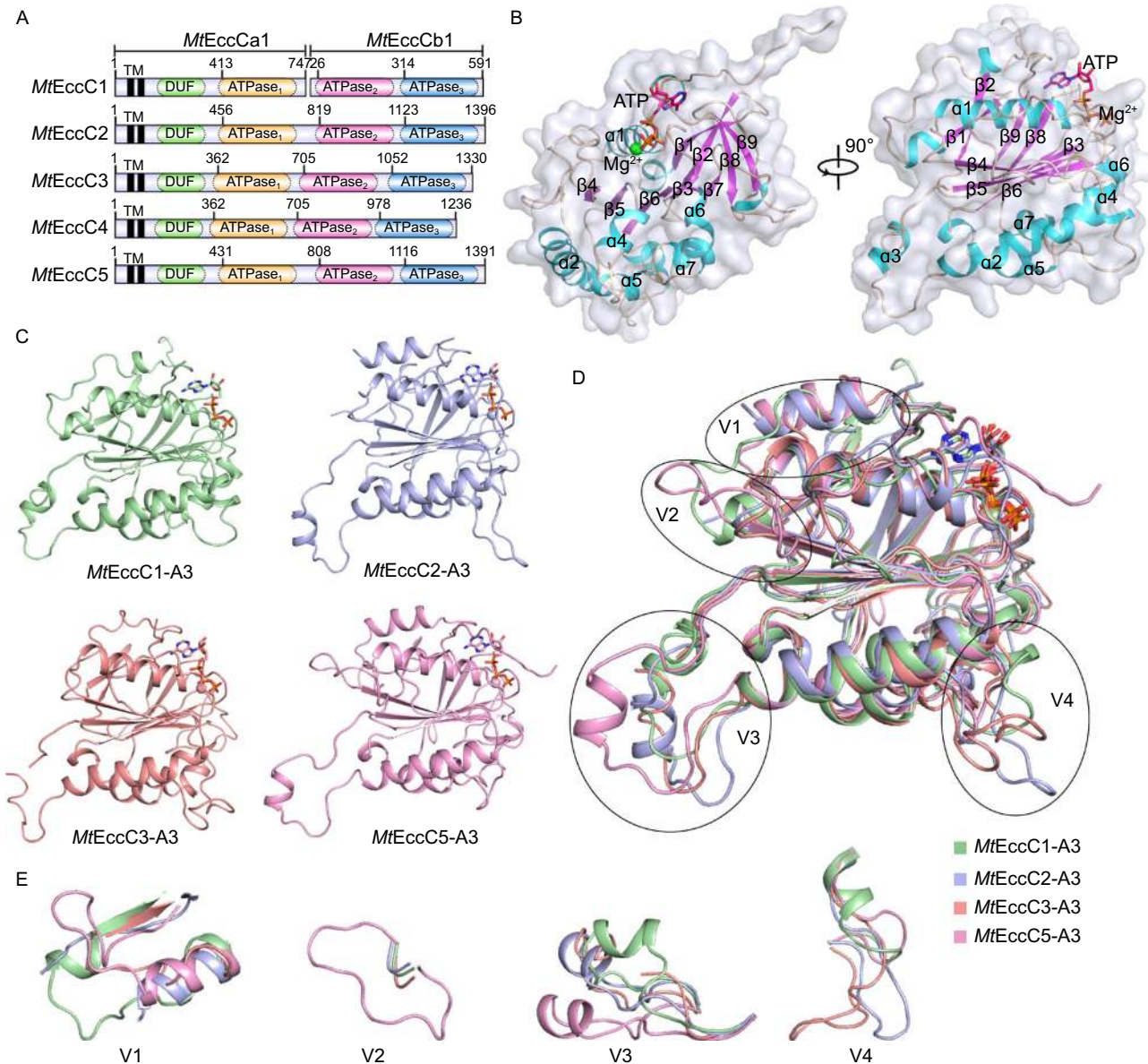


Figure 1. Overall structure of *MtEccC*-ATPase. (A) Domain arrangements of EccC proteins from *Mycobacterium tuberculosis*. (B) Cartoon and surface representation for *MtEccCb1*-ATPase₃. The helices and beta-sheets are colored cyan and magenta, respectively. ATP and the magnesium ion, are shown in stick representation and as a sphere, respectively. (C) Cartoon representation of *MtEccCb1*-ATPase₃ (C1A3, pale green), *MtEccC2*-ATPase₃ (C2A3, light blue), *MtEccC3*-ATPase₃ (C3A3, salmon) and *MtEccC5*-ATPase₃ (C5A3, pink). (D) Superposition of the four structures in (C). Structures of the four variable regions (V1–V4) are marked with circles. (E) Close-up views of the four variable regions in (D).

substrate interaction and recognition, we then tried to focus study on this region. Firstly, we solved the structure of the ATPase₃ domain of EccCb1 from *Mtb* at 2.10 Å resolution. The data collection and refinement statistics for this structure are given in Table S2. EccCb1-ATPase₃ crystallized with two molecules in the asymmetric unit. The overall structure adopts a classic RecA-like α/β fold, with a central five-stranded ($\beta 7$ - $\beta 3$ - $\beta 6$ - $\beta 5$ - $\beta 4$) parallel β -sheet, which is further extended by an additional anti-parallel sheet that includes

$\beta 2$ - $\beta 1$ - $\beta 9$ - $\beta 8$. The combined sheet wraps around helix $\alpha 1$ on one side while $\alpha 2$ and other four short helices ($\alpha 4$, $\alpha 5$, $\alpha 6$ and $\alpha 7$) cover the other side (Fig. 1B). The nucleotide-binding site is located at the N-terminal end of helix $\alpha 1$. A fragment containing $\alpha 3$ protrudes out of the main body of the structure from the end of $\alpha 2$. It then loops back to connect to $\beta 5$.

To explain structure-function differences amongst EccC-ATPase₃ domains, the crystal structures of *Mtb* ATPase₃ from ESX-2, -3, -5 were determined at 2.20 Å, 1.97 Å and

2.00 Å resolution, respectively (Fig. 1C and Table S2). Though the sequence identities are only 20%–30% amongst the four ATPase₃ domains, their atomic structures can be superimposed (Fig. 1D) and have *r.m.s.d.* values in the range 0.99–2.02 Å by comparing Ca atom pairs (Table S3). However, four structurally variable regions were observed (Figs. 1E and S1). One is at the region following the N-linker which connects to the ATPase₂ domain (variable region 1, V1). This region may be related to interactions with ATPase₂ domain and domain movement during substrate translocation. The second is the loop connecting α1 and β4 (V2), which is extremely long in EccC5-ATPase₃. V3 is located at the protruding segment between α2 and β5 and has different orientations in all of the structures. This region is possibly responsible for hexamer assembly (see DISCUSSION). V4 is the region located between β5 and α5 and has a variable length across the different ATPase₃ domains. Due to its gating position at the potential central channel of the hexamer model, we speculate that this region may be related to specific interactions with secreted substrates.

A structure-based Dali search (Holm and Laakso 2016) using the model of EccCb1-ATPase₃ showed a homologous EccC-ATPase₃ structure of T7SS from *Thermomonospora curvata* (PDB code: 4NH0) (Rosenberg et al. 2015) as the top-hit with an *r.m.s.d.* of 1.7 Å. The V1 and V4 regions can also be superimposed in the two structures. However, TcEccC-ATPase₃ has an additional β-strand instead of a bulge loop near β4. VirB4 (PDB code: 4AG6) (Walldén et al. 2012) and motor domain of FtsK (PDB code: 2IUT) (Massey et al. 2006) are also observed to have similar structures to EccCb1-ATPase₃. The former is the energetic ATPase in the Type IV secretion system, which implies that our ATPase domain structures have a similar energetic function during protein secretion. The latter is a DNA translocase that coordinates chromosome segregation and cell division in bacteria. A functional hexameric model for substrate translocation in that structure could be used as a reference model for EccC-ATPase₃ (see DISCUSSION).

The highly conserved nucleotide-binding site

It has been demonstrated that ATP binding to the ATPase₃ of EccC is essential for secretion (Rosenberg et al. 2015). In our structures, ATP and a magnesium ion bind to the nucleotide-binding site in a similar manner (Fig. S2A). This site is bordered by the clearly recognizable Walker A (GXXXXG[K/R][T/S]) and Walker B (hhhhDD with h = hydrophobic residue) motifs which include residues 376–383 and residues 472–477, respectively in MtEccCb1-ATPase₃ (Figs. 2A and S1). In addition, we found a motif, Motif1 (Dx [R/K]), that also participates in nucleotide binding. For EccCb1-ATPase₃, ATP-Mg²⁺ is inserted into a deep pocket (Fig. 2B). The adenosine ring of ATP is clamped in place by Arg327, Pro558, and Tyr576 and surrounded by Thr384, Ile385, Gln573, Ala574 and Pro575 (Fig. 2C). The ribose moiety of ATP is more solvent exposed but one of the ribose

hydroxyl groups does form a hydrogen bond with the side-chain of R327. The tri-phosphate group is held in the groove at the N-terminus of helix α1 mainly through main-chain interactions. The side chain of the conserved Lys382 in the Walker A motif points to the region between the β- and γ-phosphates implying that it may play an important role in ATP hydrolysis. Other positively charged residues, such as Lys379 in Walker A and Arg410 in Motif1, are also observed near the phosphate groups in the binding site and may also be involved in ATP tethering during hydrolysis.

The β- and γ-phosphates and Thr383, together with three water molecules are ligands for the Mg²⁺, and they form a regular octahedron around this ion (Fig. 2C). The six Mg-O distances are within the expected range of 2.0–2.2 Å. Interestingly, the two Asp residues (Asp476 and Asp477) in the Walker B motif, which usually contribute to stabilization of the Mg²⁺ ion, are too far away to interact directly with the Mg²⁺ ion in our model. Instead, they mediate stability indirectly through interacting with water molecules coordinated to the metal ion (Fig. 2C). At this point, EccC-ATPase₃ may represent a pre-activated state unfavorable for ATP hydrolysis, and this could explain why intact ATP is observed in our structure. Accordingly, little enzymatic activity (the reaction rate is very low) was observed when these ATPase₃ domains were assayed (Fig. S5). Even the EccC fragment containing two or three ATPase domains has low reaction rate, though the activity of three-domain group was relatively higher than the others. This pre-activated state is similar to the state of ATPase₃ domain of EccC from *Thermomonospora curvata* (Rosenberg et al. 2015). The conformation of our structure, together with previous reported ATPase₃ domain of TcEccC, is similar to the inhibited-state of F₁-ATPase where the inert ATP analog, AMP-PNP is bound (PDB code: 2CK3) (Bowler et al. 2006; Rosenberg et al. 2015). The homologous ATPase₂ and ATPase₃ domains of EssC from *Geobacillus thermodenitrificans* also have similar ATPase structures, and also bind ATP (Fig. S2B) (Zoltner et al. 2016). A ConSurf analysis (Ashkenazy et al. 2010; Celniker et al. 2013) of the ATPase₃ domains of 143 unique EccC proteins from different species and subtypes showed that nucleotide-binding site is highly conserved (Fig. S4A). Previous reported structure of ATPase₁ domain of TcEccC showed that the nucleotide binding residues were strikingly different from other two domains and there is only a sulfate ion in the active site when co-crystallization with high concentration of ATP (Rosenberg et al. 2015).

Substrate recognition by MtEccCb1-ATPase₃

To understand substrate recognition within ESX-1, interactions between MtEsxA/B and MtEccCb1-ATPase₃ were investigated. Size exclusion chromatography (SEC) showed that MtEsxA by itself is not able to bind to MtEccCb1-ATPase₃ (Fig. 3B), however, both the MtEsxAB heterodimer and MtEsxB homodimer do bind to it (Fig. 3A and 3C) with

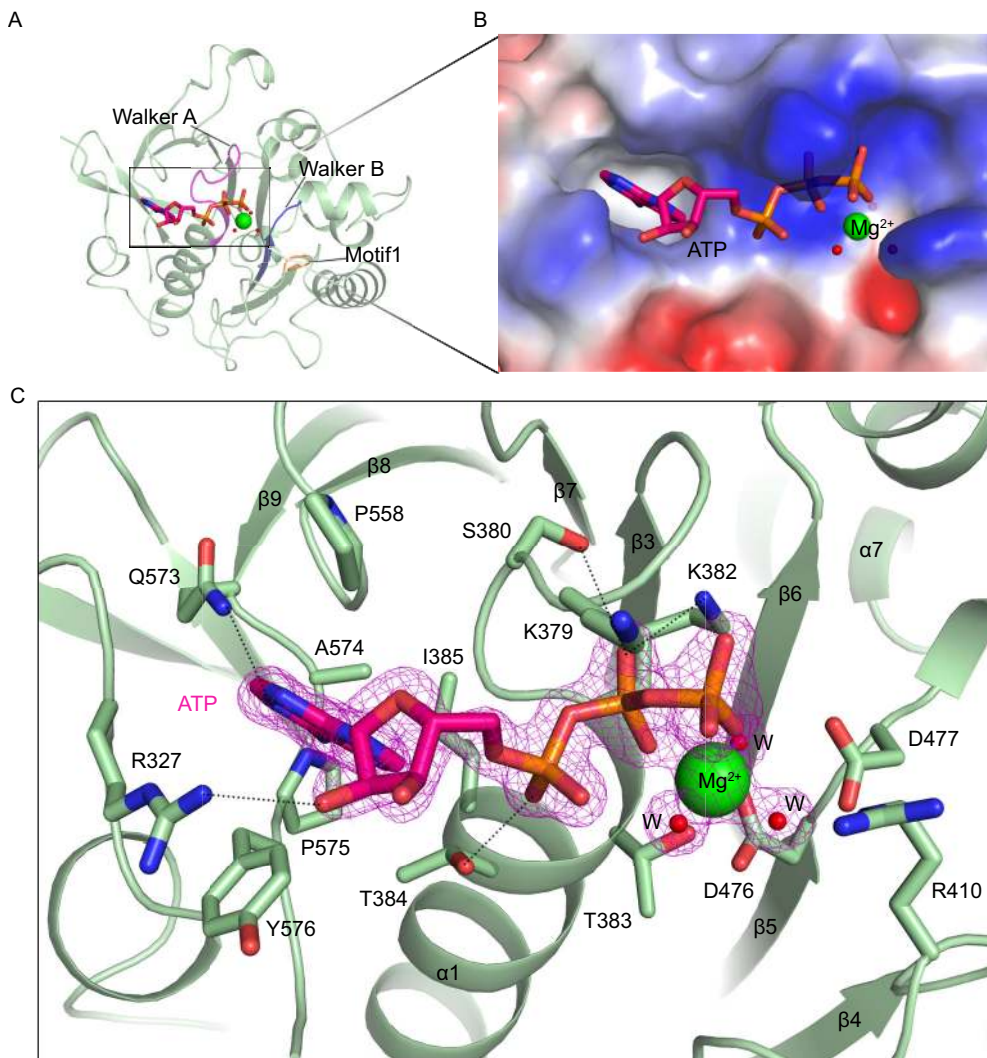
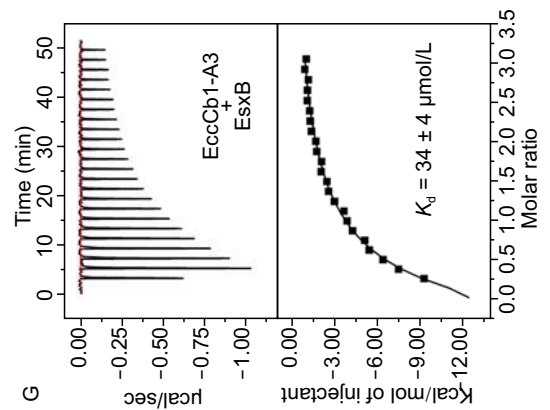
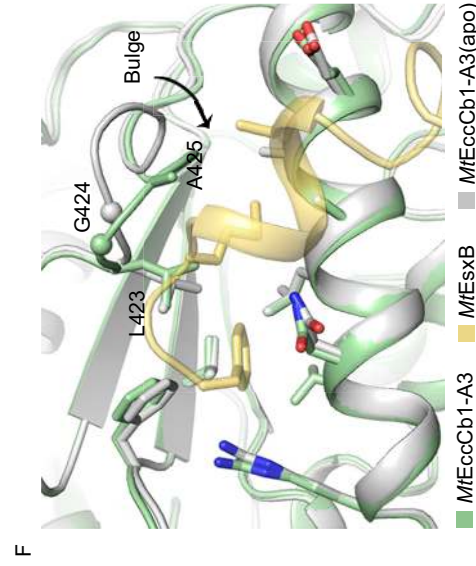
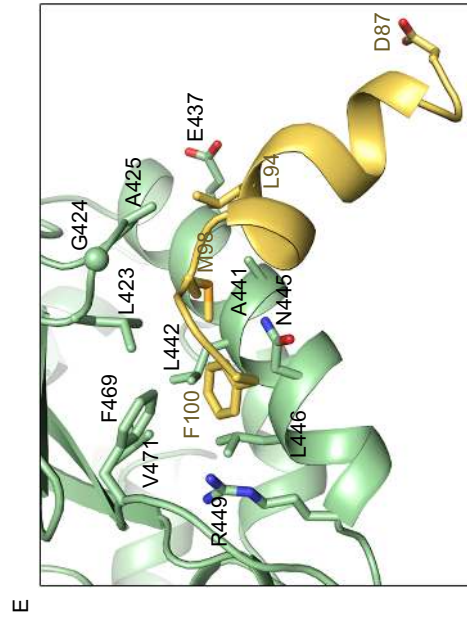
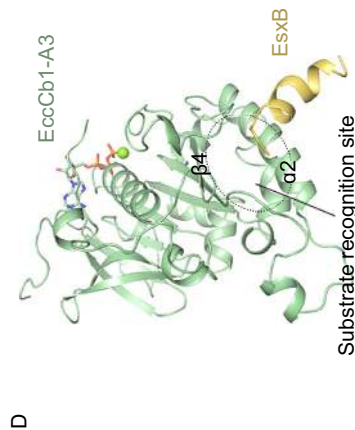
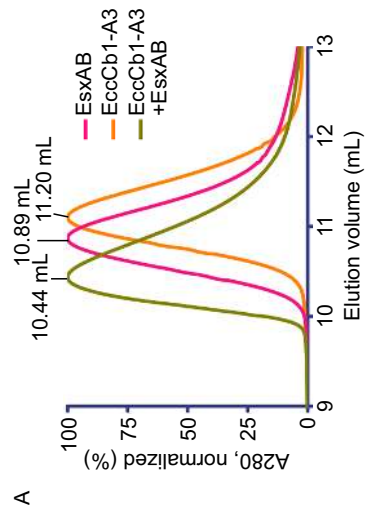
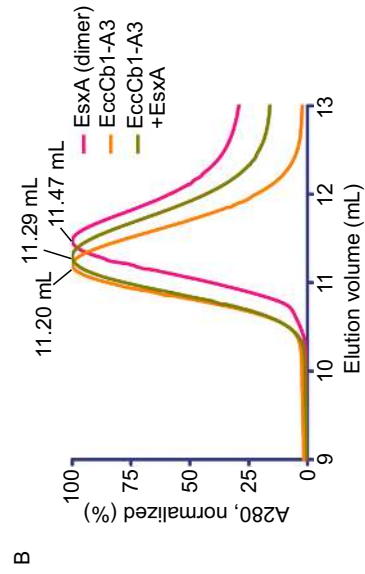
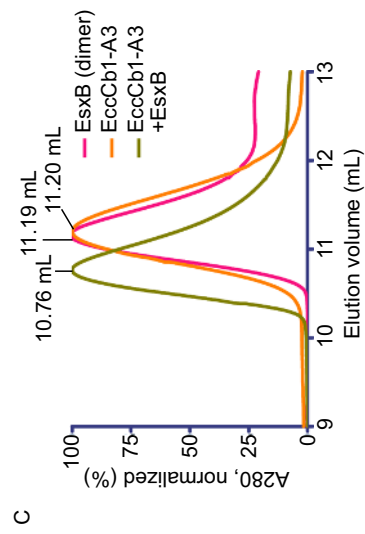


Figure 2. The nucleotide-binding site of *MtEcc*-ATPase₃. (A) The location of the nucleotide-binding site in *MtEccCb1*-ATPase₃. ATP and the magnesium ion, are shown as stick representations and as a sphere, respectively. The Walker A, Walker B and Motif 1 are colored magenta, blue and orange, respectively. (B) Close-up view of the nucleotide-binding pocket shown in electrostatic surface representation. ATP and the magnesium ion, are shown as stick representations and as a sphere, respectively. (C) Interactions between Mg²⁺, ATP and nearby amino acid residues. Green and red spheres represent Mg²⁺ and water molecules, respectively. The 2F_o-F_c density map (pink mesh) of ATP, Mg²⁺ and the coordinated water molecules is contoured at 1 σ . The dashed lines represent hydrogen bonds between ligands and the protein. W, water molecule.

binding affinity constants (K_d values) of 47 $\mu\text{mol/L}$ and 34 $\mu\text{mol/L}$, respectively, measured by the isothermal titration calorimetry (ITC) method (Figs. 3G, S6A and S6B). This is consistent with previous reported analysis from homologous proteins from *Thermomonospora curvata* (Rosenberg et al. 2015).

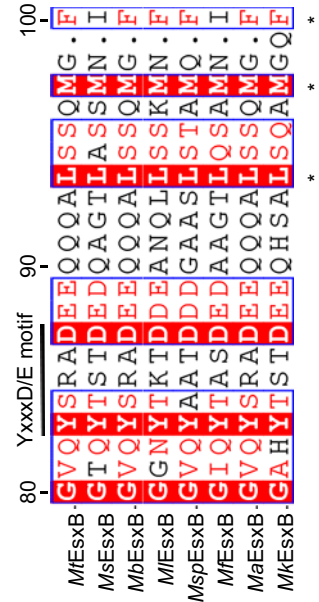
To determine how *MtEsxB* is recognized by *MtEccCb1*-ATPase₃, a structure of the complex was determined at 1.98 Å resolution. However, the main body of the helix bundle of *MtEsxB* was not observed in the structure. SDS-PAGE analysis of crystal sample showed that *MtEsxB* degraded severely (Fig. S7A). Only the C-terminal 14 residues of

MtEsxB was stable through strongly binding to *MtEccCb1*-ATPase₃ according to the high-resolution density. The first 11 residues form an α -helix while the last three residues form a short loop (Figs. 3D and S7B). PISA (Krissinel and Henrick 2007) analysis showed that in total, 414 Å² of surface area is buried at the interacting interface. Most importantly, it is the side chains of Met98 and Phe100 in *MtEsxB* that insert into the recognition pocket of ATPase₃, which is located in the gap between $\alpha 2$ and $\beta 4$. The phenyl ring of Phe100 is sandwiched by Asn445 and Phe469, with Leu442, Leu446, Arg449 and Val471 all located within van der Waals distance (Fig. 3E). Met98 is surrounded by Leu423, Ala425 and



H

EccCb1-A3	EsxB	K_D ($\mu\text{mol/L}$)
WT	1-100	34 \pm 4
WT	1-84	UD
WT	85-100	55 \pm 5
WT	85-100 (L94A)	UD
WT	85-100 (M98A)	UD
WT	85-100 (F100A)	UD
WT	85-100 (F100I)	UD



◀ **Figure 3. The C-terminal peptide of *MtEsxB* interacts with *MtEccCb1-ATPase₃*.** (A–C) Gel filtration analysis of interactions between Esx proteins and *MtEccCb1-ATPase₃* performed on a Superdex 75 column. The peak volumes are indicated on the top. (A) *MtEsxAB* binds to *MtEccCb1-ATPase₃* inducing a shift in elution volume. (B) *MtEsxA* alone does not bind to *MtEccCb1-ATPase₃*, thus there is no change in elution volume. (C) *MtEsxB* binds to *MtEccCb1-ATPase₃*, inducing a similar shift in elution volume as observed in (A). (D) Overall structure of *MtEccCb1-ATPase₃* (palegreen) complexed with *MtEsxB* (yelloworange). The substrate binding site is marked with a dashed circle. (E) Detailed interactions between *MtEsxB* and *MtEccCb1-ATPase₃*. Interacting residues and Asp87 in Yxxx[D/E] motif are shown as sticks. (F) *MtEccCb1-ATPase₃* in complex with *MtEsxB* is superimposed onto *MtEccCb1-ATPase₃* (apo). The bulge in the loop moves closer to *MtEsxB* to enhance substrate binding. (G) ITC assay shows the binding affinity of *MtEsxB* to *MtEccCb1-ATPase₃*. The data were representative of at least three repetitions. (H) The dissociation constant, K_d , is based on the ITC studies of *MtEsxB* and its truncations or peptides, to *MtEccCb1-ATPase₃*. The data were representative of at least three repetitions. WT, wild type; UD, data was undetermined. (I) Sequence alignment of EsxB from different *Mycobacterium* species including *Mycobacterium tuberculosis* (*Mt*), *Mycobacterium smegmatis* (*Ms*), *Mycobacterium bovis* (*Mb*), *Mycobacterium leprae* (*Ml*), *Mycobacterium sp.* (*Msp*), *Mycobacterium flavescens* (*Mf*), *Mycobacterium africanum* (*Ma*) and *Mycobacterium kyorinense* (*Mk*). The recognition residues are marked with stars. The Yxxx[D/E] motif is also labeled on top of sequences.

Leu442. Additionally, Leu94 is close to $\alpha 2$ and interacts with Glu437 and Ala441. This Leu94 appears to provide extra binding affinity for the complex though it is not inserted into the pocket. A previous report showed that Asp87 in the conserved Yxxx[D/E] motif of *MtEsxB* is required for secretion (Daleke et al. 2012), however this residue makes no interactions with *MtEccCb1-ATPase₃* in our structure (Fig. 3E). Possibly, an additional component, such as a chaperonin, might interact with this motif to contribute to secretion, or a second recognition by this motif is required in the later secretion steps. *MtEccCb1-ATPase₃* has a unique short bulge loop ⁴²³LGAGA⁴²⁷ near $\beta 4$, instead of a β -strand observed in other solved *MtEccC-ATPase₃* structures. Superposition of the *MtEsxB* bound and free *MtEccCb1-ATPase₃* structures showed that the bulge loop moves by ~ 3.5 Å to allow Gly424 and Ala425 to form interactions with *MtEsxB* (Fig. 3F). Meanwhile, the side chain of Leu423 also re-orientates to be closer to Met98 of *MtEsxB*. Thus, the shift of bulge loop could further strengthen substrate binding.

To further verify which residues on *MtEsxB* are crucial for substrate recognition, we made the *MtEsxB* truncation mutant lacking the 16-residues signal sequence and showed that it has no detectable binding affinity (Figs. 3H and S6C),

this is consistent with our structure that showed only C-terminal signaling sequence of *MtEsxB* was visible. We then synthesized individual peptides corresponding to the C-terminal 16-residues of *MtEsxB* with site-mutations and then measured their binding affinities with *MtEccCb1-ATPase₃*. The K_d value of the wild-type peptide is 55 $\mu\text{mol/L}$ (Figs. 3H and S6D), a value comparable to the *MtEsxB* protein. When the side chains of Met98 and Phe100 were individually mutated to alanine, binding is lost completely (Figs. 3H, S6F and S6G). Note that mutation of Leu94 to alanine also resulted in a loss of binding (Figs. 3H and S6E), confirming the role of Leu94. These results were also consistent with the two-hybrid analysis between the two proteins (Champion et al. 2006). Therefore, Leu94, Met98 and Phe100 of *MtEsxB* are all necessary for substrate recognition. A previous report showed that the G99A mutant abolished substrate binding (Champion et al. 2006; Rosenberg et al. 2015). By referring to our structural information, this could be explained by the loss of flexibility in main chain conformation caused by the addition of the side chain C β atom, which prevents Phe100 or Met98 from having an ideal orientation to bind to *MtEccCb1-ATPase₃*.

Sequence alignment analysis of EsxB from several *Mycobacterium* species showed that the Leu94 and Met98 are highly conserved, while the C-terminal residue be either Phe or Ile (Fig. 3I). The F100I mutant of *MtEsxB* peptide was also synthesized but it did not bind to *MtEccCb1-ATPase₃* (Figs. 3H and S6H), suggesting that ESX-1 systems from different *Mycobacterium* species recognize substrates specifically, using either the LxxxMxF or LxxxMxI pattern. Taking this data and analysis together, we have provided detailed structural information on the interactions between *MtEccCb1-ATPase₃* and *MtEsxB*, which gives a clear explanation for the substrate recognition pattern at the C-terminus of EsxB for *Mtb* ESX-1.

Structural comparison for substrate-specific recognition

Since the structure of the *Thermomonospora curvata* EccCb + EsxB (*TcEccCb*+*TcEsxB*) complex (Rosenberg et al. 2015) has been determined, it can be compared with our structure of *MtEccCb1-ATPase₃*+*MtEsxB* complex. In common is that both substrates use their C-terminal sequences for signaling. However, they differ in the composition of the residues that are present. The *TcEsxB* C-terminal sequence is “⁹⁸VQALLNG¹⁰⁴” while in *MtEsxB* it is “⁹⁴LSSQMGF¹⁰⁰”. Structural superposition showed that the two EsxB proteins bind to the same position on ATPase domains, however the binding pockets are different and accordingly the bound peptides adopt different conformations (Fig. 4A–D). In terms of secondary structure, the C-terminus of *TcEsxB* forms a helix to participate in binding while *MtEsxB* also has a helix but its C-terminal end finishes with a short loop. Moreover, the two EsxB substrates bind in opposite orientations.

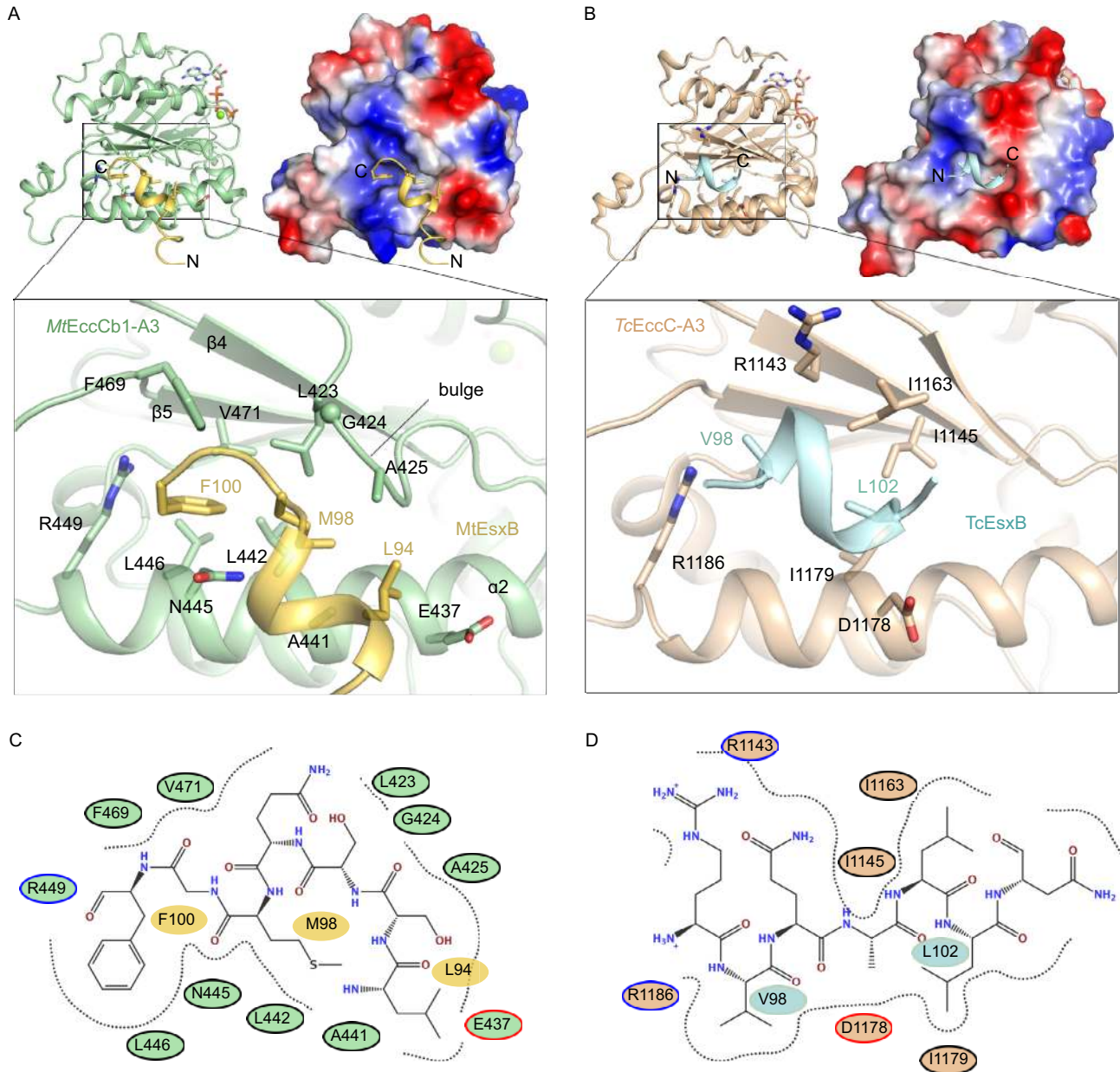


Figure 4. Substrate-specific recognition of EccCb1-ATPase₃ from *Mycobacterium tuberculosis* and EccC-ATPase₃ from *Thermomonospora curvata*. (A) (Upper left) A cartoon image of *MtEccCb1*-ATPase₃ (pale green) in complex with *MtEsxB* (yellow). (Upper right) The same complex but shown as an electrostatic surface. (Bottom image) A zoom-in view of the interactions at binding pocket. N/C, the N-/C-terminal end of peptide. (B) (Upper left) A cartoon image of *TcEccC*-ATPase₃ (wheat) in complex with the C-terminal region of *TcEsxB* (pale cyan) (PDB code: 4N1A). (Upper right) The same complex but *TcEccC*-ATPase₃ is shown as an electrostatic surface. (Bottom image) A zoom-in view of the interactions at binding pocket. (C) Schematic diagram of the interactions between *MtEccCb1*-ATPase₃ and *MtEsxB* at the recognition pocket. (D) Schematic diagram of the interactions between *TcEccC*-ATPase₃ and *TcEsxB* at the recognition pocket.

Detailed analysis of the *TcEccCb*+*TcEsxB* structure showed that Val98 and Leu102 of *TcEsxB* both insert deep into the recognition pocket at positions similar to Phe100 and Met98 in *MtEsxB*. This despite the fact that the valine and leucine side-chains are comparatively smaller than

phenylalanine and methionine. Another difference between the binding modes is that there is no equivalent to Leu94 of *MtEsxB* in *TcEsxB*. Comparison of recognition pockets in the two ATPase₃ structures showed that there is an additional β-strand in *TcEccCb*-ATPase₃ instead of the bulge loop near

$\beta 4$ in *MtEccCb1*-ATPase₃ (Fig. 4A and 4B). The binding of Phe100 of *MtEsxB* is through clamping between Asn445 and Phe469 in *MtEccCb1*-ATPase₃, however, the corresponding residues are both replaced by alanine in *TcEccCb*-ATPase₃ thus clamping cannot occur in *TcEsxB*. Instead, van der Waals' forces by residues surrounding Val98 are the main contributors to binding. In terms of buried surface areas, 371 Å² is buried in the interface between *TcEccCb*-ATPase₃ and *TcEsxB*, slightly smaller than the 414 Å² in the *MtEccCb1*-ATPase₃+*MtEsxB* complex. This is in reasonable agreement with the fact that the binding affinities are comparable, ~15 μmol/L for the *TcEccCb*+*TcEsxB* (Rosenberg et al. 2015) and 34 μmol/L for the *MtEccCb1*-ATPase₃+*MtEsxB* complex, despite the fact that these affinities were measured by different techniques (fluorescence vs ITC). These distinct aspects of substrate binding suggest that T7SSs from different species could use substrate-specific recognition patterns for secretion.

Comparison of recognition pockets of EccC-ATPase domains

A previous report showed that though EsxA and EsxB are absent in supernatants from Δ EsxB and Δ EccD1 deletion *Mtb* strains, the other Esx proteins which may have a signaling module different to "LxxxMxF" are still secreted at the expected levels (Champion et al. 2006). Thus, ESX-2-5 could recognize their own signaling pattern for their substrate Esx proteins even in the absence of EsxA and EsxB. Here, our structural data showed that the recognition pocket in each EccC-ATPase₃ is different. The signal binding pocket of *MtEccC3*-ATPase₃ is much wider than *MtEccCb1*-ATPase₃ (Fig. 5A and 5C); the pocket of *MtEccC5*-ATPase₃ is much flatter (Fig. 5D) while *MtEccC2*-ATPase₃ has a shallow recognition pocket (Fig. 5B). The different shapes and compositions of signal recognition pockets (Figs. 5 and S3) further imply that different T7SS subtypes could specifically recognize individual substrates. In addition, conservation analysis of 143 EccC-ATPase₃ from different species showed that the signal recognition pocket is hypervariable (Fig. S4B). Interestingly, though ATPase₁ and ATPase₂ of *TcEccC* and *GtEssC* are homologous with ATPase₃ of EccC in sequence and structure, their recognition pocket region can interact with the linker peptide connecting two ATPase domains, and auto-inhibit the ATPase activity (Rosenberg et al. 2015; Zoltner et al. 2016). Thus, the functional roles of the recognition pocket of T7SS ATPase domains are in accord with the sequence and structure variation.

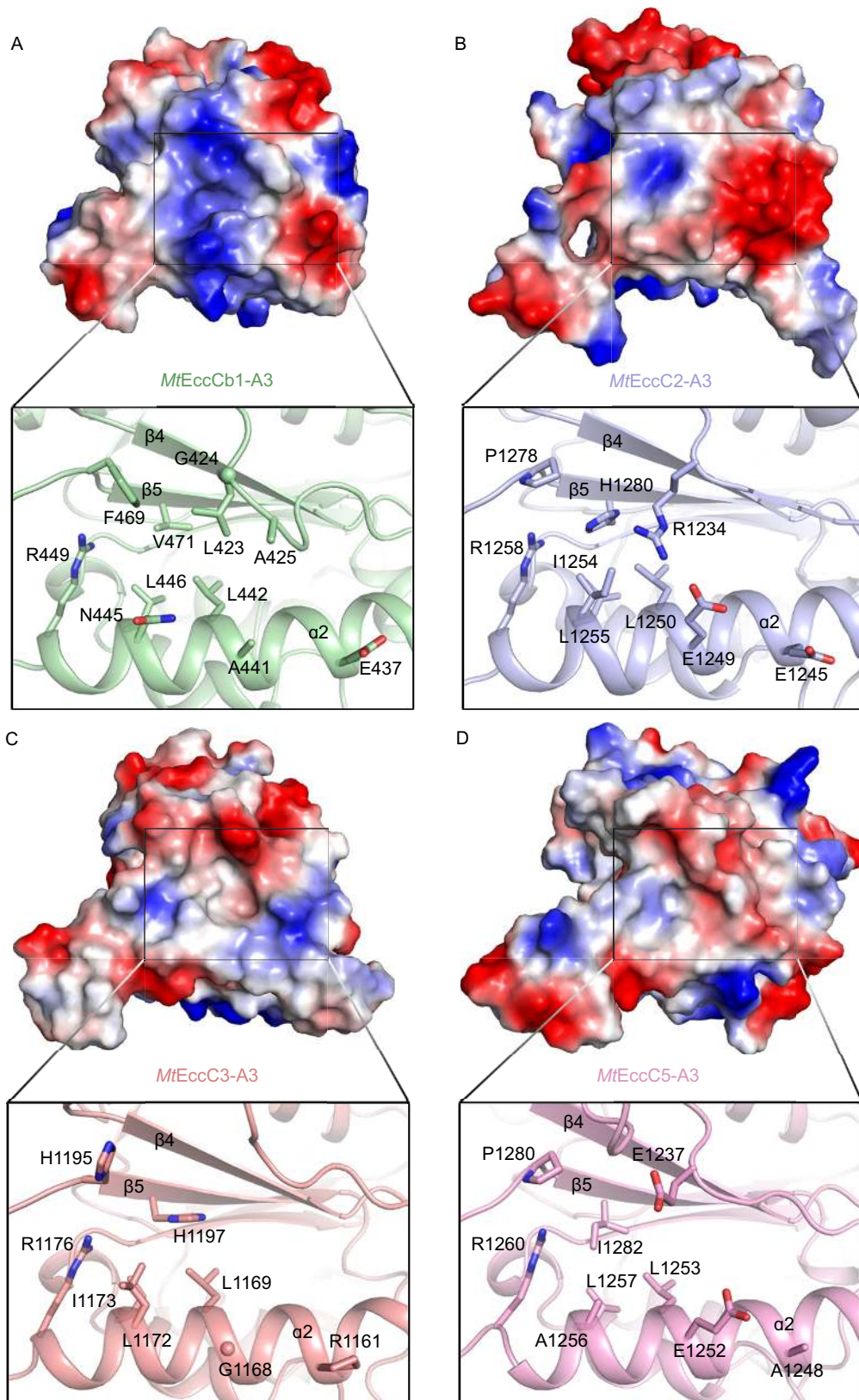
DISCUSSION

Type VII secretion systems are responsible for secreting virulence factors and other effectors (Unnikrishnan et al. 2017). EccC proteins, which typically contain three ATPase domains, have been demonstrated to be required for

secretion of substrates by maintaining a stable membrane complex and by acting as a motor to translocate substrates across membrane (Stanley et al. 2003; Houben et al. 2012). In this study, we have provided high-resolution crystal structures of ATPase₃ domains of EccC proteins from *Mtb*, all in complex with ATP. Like the property of *TcEccC* and *GtEssC* (Rosenberg et al. 2015; Zoltner et al. 2016), we proposed that these ATPase₃ domains are in a state of pre-activation and other components or conformational changes may be needed for their activation.

Negative staining electron microscopy showed the ESX-5 apparatus (including the core subunit EccC) from *Mycobacterium xenopi* forms a 1.8 MDa transmembrane complex with six-fold symmetry (Beckham et al. 2017). NanoLC-MS/MS analysis of the large T7SS membrane bound complex showed that it contains six copies of EccC (Houben et al. 2012). The EccC is a putative ATPase that belongs to the well-conserved FtsK/SpoIIIE family, of which different members have also been shown to form hexamers (Massey et al. 2006). Thus, a hexameric model for EccCb1-ATPase₃ can be proposed based on the hexamer structure of FtsK ATPase domain (Massey et al. 2006) through superposition. Indeed, this is similar to what we observed in the packing of our EccC5-ATPase₃ crystal structure which has six-fold screw axis symmetry (Fig. S8A and S8B). In our hexameric model (Fig. S9A), the ATP binding site is located at the interface between two neighboring ATPase domains. Conformational changes and ATP hydrolysis for the hexamer are required when T7SS is translocating substrate. Similar to FtsK, the protruding segment between $\alpha 2$ and $\beta 5$ interacts with the loop between $\beta 4$ and $\alpha 2$ of the neighboring subunit. Such an interaction mode can also be observed in the helical packing of EccC5-ATPase₃ (Fig. S8C).

The hexameric model of EccCb1-ATPase₃ forms a central channel with an inner diameter of about 25 Å (Fig. S9A), matching the size of helix bundle of substrates, for example, EsxAB dimer (Fig. S9B) (Renshaw et al.). The substrate recognition pocket is located at the bottom side, and close to the central pore. The inner surface of the channel is composed of the loop in the V4 region, $\alpha 6$ and the loop between $\beta 7$ and $\beta 8$. Note that the loop in the V4 region is at the gating position of the channel and thus may be responsible for the initial entry of substrate. This loop varies in length and residue composition for different ESX systems allowing them to bind their own substrates. Variation of the loop at the gating position and diverse residues inside the channel may also contribute to the substrate selection and translocation for different subtypes of T7SS. The hexameric model could also be proposed for ATPase₁ and ATPase₂ domains. We could expect that EccC may use three layers of hexamer structure to drive helix-bundle substrate translocation across membrane (Fig. S9C), this was similar with the proposed hexameric EssC-C model (Zoltner et al. 2016). Considering the flexibility of ATPase domains of EccC5 in the previous reported ESX-5 model, conformational changes are needed to induce hexamer formation.



◀ **Figure 5. Structural comparison of the signal recognition pocket of EccC-ATPase₃ domains.** *MtEccCb1-ATPase₃* (A), *MtEccC2-ATPase₃* (B), *MtEccC3-ATPase₃* (C) and *MtEccC5-ATPase₃* (D) are drawn in electrostatic surface representation. Inserts provide a detailed view of the signal recognition pocket. Comparable to *MtEccCb1-ATPase₃*, residues that are likely involved in substrate binding are shown as stick models.

Loss of the secretion function of T7SS coincides with a reduction in bacterial replication, weakened inflammation, decreasing granuloma formation and an increase in the host's survival (Hsu et al. 2003; Gao et al. 2004; Guinn et al. 2004; Volkman et al. 2004; Brodin et al. 2006). Considering the fact that the RD1 region, which encodes components of T7SS including EccCb1, is the main distinction between vaccine strain BCG and virulent strains (Mahairas et al. 1996; Behr et al. 1999), the ATPase₃ domain of EccC, especially *MtEccCb1*, is therefore a valuable target for developing new anti-TB therapies or vaccines. Based on this study, small molecules or peptides could be designed to bind at the substrate recognition site to block virulence factor secretion. Given the success of secretion of engineered yeast ubiquitin fused to the "LSSQMGF" peptide fragment by ESX-1 (Champion et al. 2006), next generation of vaccine strains could be engineered to secrete sets of immunodominant factors without causing disease, by using the signaling sequence.

MATERIALS AND METHODS

Cloning and expression

The DNA coding sequences of *MtEccCb1-ATPase₃* (residues 315-591), *MtEccC2-ATPase₃* (residues 1,127-1,396), *MtEccC3-ATPase₃* (residues 1,060-1,330), *MtEccC5-ATPase₃* (residues 1,125-1,391), as well as *MtEsxB* alone and *MtEsxB-EsxA* loci in tandem were PCR amplified from *Mycobacterium tuberculosis* H37Rv genomic DNA using pairwise primers listed in Table S1. The PCR product of *MtEccCb1-ATPase₃* was cloned into the pET-M3C plasmid, modified from pET-32-M3C (Novagen), fused with N-terminal 6xHis tag and a Rhinovirus 3C protease cleavage site, while constructs of *MtEccC3-ATPase₃* and *MtEccC5-ATPase₃* were made in the pET-22b vector (Novagen) fused with C-terminal 6xHis tag. The sequence of *MtEsxB* alone, *MtEsxB-EsxA* tandem sequence and *MtEccC2-ATPase₃* each was cloned into pET-32-M3C (Novagen) vector which expresses a thioredoxin (Trx)-His tag followed by a Rhinovirus 3C protease cleavage site fused at the N-terminus of the protein. Site-directed mutagenesis was performed using the TaKaRa MutanBEST Kit. The mutants were introduced by the PCR method using the *EsxB* expression plasmid as a template, with pairs of primers encoding the mutations at the sites of substitution. All constructs were verified by sequencing and then transformed into *Escherichia coli* BL21 (DE3) strain for expression. The bacteria were cultured in Luria-Bertani media supplemented with 100 µg/mL

ampicillin at 37 °C to an OD₆₀₀ of 0.6. Protein expression was induced by the addition of 0.2 mmol/L IPTG for 20 h at 16 °C. Cells were harvested after centrifugation at 4,000 rpm for 30 min and frozen at -80 °C.

Protein purification

All ATPase₃ domain proteins and substrate Esx proteins were purified by a similar method. Briefly, cells were thawed and resuspended in buffer A (20 mmol/L Hepes pH 7.0, 150 mmol/L NaCl, 5% (w/v) glycerol, 1 mmol/L MgCl₂, 5 mmol/L ATP). The resuspended cells were then lysed by passing through a French Press at 800 bar after adding 1 mmol/L PMSF. Cell debris was then removed by centrifugation at 18,000 rpm for 30 min at 4 °C. The supernatant was applied to Ni-NTA agarose beads (GE Healthcare) for 2 h at 4 °C. The beads were rinsed with buffer A containing 30 mmol/L imidazole. For *MtEccCb1-ATPase₃*, *MtEccC2-ATPase₃* and Esx proteins, the N-terminal tag was cleaved by 3C protease and then eluted. For *MtEccC3-ATPase₃* and *MtEccC5-ATPase₃*, the recombinant protein was eluted from the beads with buffer A containing 300 mmol/L imidazole. Then the sample was concentrated and purified using a 5mL Hitrap Q HP (GE life science) column followed by size exclusion chromatography (SEC) using a Superdex 75 HR 10/30 (GE life science) column. The peak fractions were pooled and concentrated to approximately 10 mg/mL using a 10 kDa cut-off spin concentrator (Millipore). The separately purified *MtEccCb1-ATPase₃* and *MtEsxB* were mixed in a 1:1 molar ratio, incubated and purified again by gel filtration. The fractions containing complex were pooled and concentrated for crystallization.

Crystallization

Crystallization trials were performed by hanging-drop vapor diffusion method at 16 °C. The protein solution, diluted to 8-10 mg/mL, was mixed in a 1:1 (v/v) ratio with the reservoir solution. Crystals of *MtEccCb1-ATPase₃*, *MtEccC2-ATPase₃*, *MtEccC3-ATPase₃*, *MtEccC5-ATPase₃* and *MtEccCb1-ATPase₃+MtEsxB* were grown from condition I [0.49 mol/L sodium phosphate monobasic monohydrate, 0.91 mol/L potassium phosphate dibasic, pH 6.9], condition II [0.1 M BIS-TRIS pH 6.5, 45% (v/v) Polypropylene glycol P 400], conditions III [0.1 mol/L Hepes/sodium hydroxide pH 7.5, 0.2 mol/L sodium chloride, 20% (w/v) PEG3000], condition IV [0.2 mol/L sodium dihydrogen phosphate monohydrate pH 4.5, 20% (w/v) PEG3350] and condition V [0.1 mol/L MES pH 6.5, 0.2 mol/L sodium chloride, 25% (w/v) PEG3350], respectively. After optimization for each condition, crystals were harvested using glycerol as cryo-protectant, flash-cooled and stored in liquid nitrogen for data collection.

Data collection and structure determination

X-ray data were collected on beamlines BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility (SSRF) and beamline BL41XU at SPring-8. Data sets were processed, merged and scaled using HKL2000 (Otwinowski and Minor 1997). The initial phases for *MtEccCb1-ATPase₃*, *MtEccC2-ATPase₃*, *MtEccC3-ATPase₃* and *MtEccC5-ATPase₃* were solved by the molecular replacement method using PHASER (McCoy et al. 2007) using the structure of the ATPase₃ domain of EccC from *Thermomonospora curvata* (PDB

code: 4NH0) (Rosenberg et al. 2015) as the search template. Model building was performed automatically using phenix.autobuild. Manual building in COOT (Emsley and Cowtan 2004) and refinement in PHENIX (Adams et al. 2010) were carried out iteratively for several rounds to obtain the final models. The structure of *MtEccCb1-ATPase₃* in complex with *MtEsxB* was then solved by molecular replacement using model of *MtEccCb1-ATPase₃* as search template. Data collection and structure refinement statistics are summarized in Table S2.

Isothermal titration calorimetry (ITC)

ITC was performed at 20 °C with a MicroCal iTC200 instrument (GE Healthcare). Proteins and peptides were prepared in ITC buffer containing 20 mmol/L Hepes (pH 7.0), 150 mmol/L NaCl, 1 mmol/L MgCl₂, 1 mmol/L ATP and 5% (w/v) glycerol. The concentration of *MtEccCb1-ATPase₃*, *Esx* proteins and peptides were 50 μmol/L, 1 mmol/L and 1.2 mmol/L, respectively. Control experiments were performed under same experimental conditions except that the sample in the syringe was replaced with the ITC buffer. This allowed the calculation of heat of dilution for the protein. The acquired ITC data were analyzed by the Origin 7.0 (GE Healthcare) program using the "One Set of Binding Sites" fitting model.

ATPase activity assay

ATPase activities were measured using the ATPase/GTPase Activity Assay Kit (MAK-113, Sigma-Aldrich) according to the manufacturer's instruction. The purified protein was diluted to 10 μmol/L with Assay Buffer (20 mmol/L Hepes pH 7.0, 150 mmol/L NaCl, 5% (w/v) glycerol, 5 mmol/L ATP, 5 mmol/L MgCl₂). 20 μL of the reaction mixture containing the diluted protein was incubated for 1 h at 20 °C. Then, 100 μL of malachite green reagent was added into each reaction well and incubated for 10 min. After that, the absorbance at 620 nm were measured, proportional to the enzyme activity present.

Data availability

Coordinates and structure factors for *MtEccCb1-ATPase₃*, *MtEccC2-ATPase₃*, *MtEccC3-ATPase₃*, *MtEccC5-ATPase₃* and *MtEccCb1-ATPase₃+MtEsxB* complex have been deposited in the Protein Data Bank, under accession codes 6JD4, 6JD5, 6J17, 6J18 and 6J19. All other data are available from the authors upon request.

ACKNOWLEDGEMENTS

We are extremely grateful to the National Centre for Protein Science Shanghai (Protein Expression and Purification System) for their instrumental support and technical assistance. We thank the staff from beamlines BL18U and BL19U1 at Shanghai Synchrotron Radiation Facility (Shanghai, China) and beamline BL41XU at SPring-8 (Hyogo, Japan) for assistance during data collection. We also thank Prof. Rongguang Zhang and Dr. Jinwei Zhu for providing laboratory resources. This work was supported by Grants from the National Key Research and Development Program of China (Grant No. 2017YFC0840300), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB08020200), and the National Natural Science Foundation of China (Grant Nos. 81520108019, 31500607).

ABBREVIATIONS

BCG, Bacille Calmette-Guérin; DUF, unknown function domain; ESX-1, early secreted antigen 6 kilodaltons (ESAT-6) system 1; FRET, fluorescence resonance energy transfer; *Gt*, *Geobacillus thermodenitrificans*; ITC, isothermal titration calorimetry; *Mtb/Mt*, *Mycobacterium tuberculosis*; PDB, Protein Data Bank; SEC, size exclusion chromatography; SSRF, Shanghai Synchrotron Radiation Facility; T7SS, type VII secretion system; TB, tuberculosis; *Tc*, *Thermomonospora curvata*; Trx, thioredoxin; V1, variable region 1.

AUTHOR CONTRIBUTIONS

Z.R. initiated and supervised the project. Z.R., J.L., S.W. designed experiments. S.W. carried out cloning, purified the proteins, grew and optimized the crystals. S.W., B.Z., Y.Z., J.L. collected the diffraction data. S.W., J.L. solved the structures. S.W., K.Z. performed ITC experiments. J.L., S.W., Y.X., H.Y., Xiaolin Y., Xiuna Y., L. W.G., Z.R. analyzed and discussed the results. J.L., S.W., Xiaolin Y., L.W.G., Z.R. prepared and wrote manuscript with the help of all the authors.

COMPLIANCE WITH ETHICS GUIDELINES

Shuhui Wang, Kaixuan Zhou, Xiaolin Yang, Bing Zhang, Yao Zhao, Yu Xiao, Xiuna Yang, Haitao Yang, Luke W. Guddat, Jun Li and Zihe Rao declare that they have no conflict of interest. This article does not contain any studies with human or animal subjects performed by the any of the authors.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213–221
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38: W529–W533
- Beckham KS, Ciccarelli L, Bunduc CM, Mertens HDT, Ummels R, Lugmayr W, Mayr J, Rettel M, Savitski MM, Svergun DI et al (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. *Nat Microbiol* 2:17047
- Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284:1520–1523

- Berthet FX, Rasmussen PB, Rosenkrands I, Andersen P, Gicquel B (1998) A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology* 144(Pt 11):3195–3203
- Bitter W, Houben EN, Luirink J, Appelmek BJ (2009) Type VII secretion in mycobacteria: classification in line with cell envelope structure. *Trends Microbiol* 17:337–338
- Bowler MW, Montgomery MG, Leslie AG, Walker JE (2006) How azide inhibits ATP hydrolysis by the F-ATPases. *Proc Natl Acad Sci USA* 103:8646–8649
- Brodin P, Rosenkrands I, Andersen P, Cole ST, Brosch R (2004) ESAT-6 proteins: protective antigens and virulence factors? *Trends Microbiol* 12:500–508
- Brodin P, Majlessi L, Marsollier L, de Jonge MI, Bottai D, Demangel C, Hinds J, Neyrolles O, Butcher PD, Leclerc C et al (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect Immun* 74:88–98
- Carlsson F, Joshi SA, Rangell L, Brown EJ (2009) Polar localization of virulence-related Esx-1 secretion in mycobacteria. *PLoS Pathog* 5:e1000285
- Carlsson F, Kim J, Dumitru C, Barck KH, Carano RAD, Sun M, Diehl L, Brown EJ (2010) Host-detrimental role of Esx-1-mediated inflammasome activation in mycobacterial infection. *PLoS Pathog* 6:e1000895
- Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem* 53:199–206
- Champion PA, Stanley SA, Champion MM, Brown EJ, Cox JS (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* 313:1632–1636
- Champion PA, Champion MM, Manzanillo P, Cox JS (2009) ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. *Mol Microbiol* 73:950–962
- Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D et al (2001) Massive gene decay in the leprosy bacillus. *Nature* 409:1007–1011
- Daleke MH, Ummels R, Bawono P, Heringa J, Vandenbroucke-Grauls CMJE, Luirink J, Bitter W (2012) General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci USA* 109:11342–11347
- Davis JM, Ramakrishnan L (2009) The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* 136:37–49
- Derbyshire KM, Gray TA (2014) Distributive conjugal transfer: new insights into horizontal gene transfer and genetic exchange in mycobacteria. *Microbiol Spectr* 2(1):MGM2-0022-2013
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132
- Flint JL, Kowalski JC, Karnati PK, Derbyshire KM (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci USA* 101:12598–12603
- Gao L-Y, Guo S, McLaughlin B, Morisaki H, Engel JN, Brown EJ (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* 53:1677–1693
- Gey Van Pittius NC, Gamielien J, Hide W, Brown GD, Siezen RJ, Beyers AD (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* 2:RESEARCH0044
- Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM (2013) Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. *PLoS Biol* 11:e1001602
- Gray TA, Clark RR, Boucher N, Lapierre P, Smith C, Derbyshire KM (2016) Intercellular communication and conjugation are mediated by ESX secretion systems in mycobacteria. *Science* 354:347–350
- Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinson DM, Smith S, Sherman DR (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 51:359–370
- Hervas-Stubbs S, Majlessi L, Simsova M, Morova J, Rojas MJ, Nouzé C, Brodin P, Sebo P, Leclerc C (2006) High frequency of CD4⁺ T cells specific for the TB10.4 protein correlates with protection against *Mycobacterium tuberculosis* infection. *Infect Immun* 74:3396–3407
- Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44:W351–W355
- Houben EN, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jiménez CR, Ottenhoff TH, Luirink J, Bitter W (2012) Composition of the type VII secretion system membrane complex. *Mol Microbiol* 86:472–484
- Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M et al (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci USA* 100:12420–12425
- Koo IC, Wang C, Raghavan S, Morisaki JH, Cox JS, Brown EJ (2008) ESX-1-dependent cytolysis in lysosome secretion and inflammasome activation during mycobacterial infection. *Cell Microbiol* 10:1866–1878
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797
- Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 178:1274–1282
- Majlessi L, Rojas MJ, Brodin P, Leclerc C (2003) CD8⁺-T-cell responses of *Mycobacterium*-infected mice to a newly identified major histocompatibility complex class I-restricted epitope shared by proteins of the ESAT-6 family. *Infect Immun* 71:7173–7177
- Massey TH, Mercogliano CP, Yates J, Sherratt DJ, Lowe J (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell* 23:457–469
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674

- McLaughlin B, Chon JS, MacGurn JA, Carlsson F, Cheng TL, Cox JS, Brown EJ (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* 3:e105
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326
- Pallen MJ (2002) The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? *Trends Microbiol* 10:209–212
- Pym AS, Brodin P, Brosch R, Huerre M, Cole ST (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* 46:709–717
- Raghavan S, Manzanillo P, Chan K, Dovey C, Cox JS (2008) Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature* 454:717–721
- Renshaw PS, Lightbody KL, Veverka V, Muskett FW, Kelly G, Frenkiel TA, Gordon SV, Hewinson RG, Burke B, Norman J et al (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* 24:2491–2498
- Rosenberg OS, Dovala D, Li X, Connolly L, Bendebury A, Finer-Moore J, Holton J, Cheng Y, Stroud RM, Cox JS (2015) Substrates control multimerization and activation of the multidomain ATPase motor of type VII secretion. *Cell* 161:501–512
- Sampson SL (2011) Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin Dev Immunol* 2011:497203
- Serafini A, Pisu D, Palu G, Rodriguez GM, Manganelli R (2013) The ESX-3 secretion system is necessary for iron and zinc homeostasis in *Mycobacterium tuberculosis*. *PLoS ONE* 8:e78351
- Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng T-Y, Siddiqi N, Fortune SM, Moody DB, Rubin EJ (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc Natl Acad Sci USA* 106:18792–18797
- Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, Enninga J (2012) Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS Pathog* 8:e1002507
- Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, Nieselt K, Krause J, Vera-Cabrera L, Cole ST (2015) Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc Natl Acad Sci USA* 112:4459–4464
- Skjöt RLV, Brock I, Arend SM, Munk ME, Theisen M, Ottenhoff TH, Andersen P (2002) Epitope mapping of the immunodominant antigen TB10.4 and the two homologous proteins TB10.3 and TB12.9, which constitute a subfamily of the esat-6 gene family. *Infect Immun* 70:5446–5453
- Smith J, Manoranjan J, Pan M, Bohsali A, Xu J, Liu J, McDonald KL, Szyk A, LaRonde-LeBlanc N, Gao LY (2008) Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in *Mycobacterium marinum* escape from the vacuole. *Infect Immun* 76:5478–5487
- Sorensen AL, Nagai S, Houen G, Andersen P, Andersen AB (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* 63:1710–1717
- Stanley SA, Raghavan S, Hwang WW, Cox JS (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci USA* 100:13001–13006
- Stoop EJ, Schipper T, Rosendahl Huber SK, Nezhinsky AE, Verbeek FJ, Gurcha SS, Besra GS, Vandenbroucke-Grauls CM, Bitter W, van der Sar AM (2011) Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. *Dis Model Mech* 4:526–536
- Unnikrishnan M, Constantinidou C, Palmer T, Pallen MJ (2017) The enigmatic Esx proteins: looking beyond mycobacteria. *Trends Microbiol* 25:192–204
- van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ (2007) M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 129:1287–1298
- Volkman HE, Clay H, Beery D, Chang JCW, Sherman DR, Ramakrishnan L (2004) Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. *PLoS Biol* 2:e367
- Volkman HE, Pozos TC, Zheng J, Davis JM, Rawls JF, Ramakrishnan L (2010) Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. *Science* 327:466–469
- Walldén K, Williams R, Yan J, Lian PW, Wang L, Thalassinou K, Orlova EV, Waksman G (2012) Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proc Natl Acad Sci USA* 109:11348–11353
- Weerdenburg EM, Abdallah AM, Mitra S, de Punder K, van der Wel NN, Bird S, Appelmek BJ, Bitter W, van der Sar AM (2012) ESX-5-deficient *Mycobacterium marinum* is hypervirulent in adult zebrafish. *Cell Microbiol* 14:728–739
- Wong KW, Jacobs WR Jr (2011) Critical role for NLRP3 in necrotic death triggered by *Mycobacterium tuberculosis*. *Cell Microbiol* 13:1371–1384
- World Health Organization (2018) Global tuberculosis report 2018
- Zoltner M, Ng WMAV, Money JJ, Fyfe PK, Kneuper H, Palmer T, Hunter WN (2016) EssC: domain structures inform on the elusive translocation channel in the Type VII secretion system. *Biochem J* 473(13):1941–1952