PE5-PPE4-EspG3 heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems

Zachary A. Williamson1, Catherine T. Chaton1, William A. Ciocca1*, Natalia Korotkova1, and Konstantin V. Korotkov1*

From the 1Department of Molecular & Cellular Biochemistry and the Center for Structural Biology, University of Kentucky, Lexington, KY 40536

Running Title: PE5-PPE4-EspG3 structure from ESX-3

#Present address: Department of Biological Sciences, Eastern Kentucky University, Richmond, KY 40475

*To whom correspondence should be addressed: Konstantin V. Korotkov: Department of Molecular & Cellular Biochemistry, University of Kentucky, Lexington, KY 40536; kkorotkov@uky.edu; Tel. (859) 323-5493; Fax (859) 257-2283.

Keywords: mycobacteria, tuberculosis, X-ray crystallography, protein complex, protein-protein interaction, type VII secretion system

ABSTRACT

*Mycobacterium tuberculosis* (*Mtb*) has evolved numerous type VII secretion (ESX) systems to secrete multiple factors important for both growth and virulence across their cell envelope. ESX-1, ESX-3, and ESX-5 systems have been shown to secrete a distinct set of substrates, including PE and PPE families of proteins, named for conserved Pro-Glu and Pro-Pro-Glu motifs in their N-termini. Proper secretion of the PE-PPE proteins requires the presence of EspG, with each system encoding its own unique copy. There is no cross-talk between any of the ESX systems and how each EspG is recognizing its subset of PE-PPE proteins is currently unknown. The only current structural characterization of PE-PPE-EspG heterotrimers is from the ESX-5 system. Here we present the crystal structure of the PE5mt-PPE4mt-EspG3mm heterotrimer from the ESX-3 system. Our heterotrimer reveals that EspG3mm interacts exclusively with PPE4mt in a similar manner to EspG5, shielding the hydrophobic tip of PPE4mt from solvent. The C-terminal helical domain of EspG3mm is dynamic, alternating between an ‘open’ and ‘closed’ form, and this movement is likely functionally relevant in the unloading of PE-PPE heterodimers at the secretion machinery. In contrast to the previously solved ESX-5 heterotrimers, the PE-PPE heterodimer of our ESX-3 heterotrimer is interacting with it’s chaperone at a drastically different angle, and presents different faces of the PPE protein to the chaperone. We conclude that the PPE-EspG interface from each ESX system has a unique shape complementarity that allows each EspG to discriminate amongst non-cognate PE-PPE pairs.

Tuberculosis is currently the deadliest infectious disease in the world, killing 1.5 million people in 2019 (1). The lack of an effective vaccine against the most prevalent pulmonary form of Tuberculosis, as well as the emergence of numerous multi-drug resistant strains of the causative agent, *Mycobacterium tuberculosis* (*Mtb*), highlights the growing need for more effective treatment options. Therefore, a more comprehensive understanding of the *Mtb* virulence machinery is needed to aid the development of new therapeutics.

*Mtb*, like all mycobacteria, contains a thick hydrophobic cell envelope that aids in protecting the mycobacterium from its environment. To overcome the limited permeability created by this envelope, mycobacteria have evolved specialized secretion systems to export proteins across their cell envelopes, the type VII secretion systems, also known as the ESX systems (2). Five different ESX systems encoded in the *Mtb* genome, and three are known to secrete proteins; ESX-1, ESX-3, and ESX-5 (3). Recently, the structures of the core complex of both ESX-3 (4,5) and ESX-5 (6) have been solved. The ESX-5 core complex has six-fold symmetry and sits on the inner membrane (6), while
the ESX-3 core complex was solved as a dimer that could be modeled onto the six-fold symmetry of the ESX-5 core complex (4,5). These systems are not functionally redundant, as their substrates are not re-routed to other ESX systems (7). The ESX systems secrete a variety of different substrates, each containing a general type VII secretion motif of YxxxD/E (8). A significant class of substrates being the PE and PPE proteins, named for conserved residues (Pro-Glu for PE and Pro-Pro-Glu for PPE) within their N-terminal domains (9,10). The N-terminal domains are about 110 (PE) or 180 (PPE) amino acids in length and interact together to form a PE-PPE heterodimer. A cytosolic chaperone, EspG, is required for proper folding and/or stability of the PE-PPE proteins and, ultimately their proper secretion (11,12). Each ESX system secretes a unique subset of PE-PPE heterodimers, and therefore each encodes an EspG that binds to only its corresponding heterodimers (11,12). The first structural insight into the EspG and PE-PPE interaction was revealed by analysis of the structure of the PE25-PPE41-EspG5 complex, a heterotrimer from ESX-5 (12,13). EspG5 interacts solely with PPE41 at the tip distal to the PE25 interaction and aids in preventing PE-PPE heterodimer aggregation in part by shielding a conserved hydrophobic tip on the PPE proteins, known as the hh motif (12). The additional structure of the ESX-5-related, PE8-PPE15-EspG3 heterotrimer, revealed similar interactions of the substrate PE-PPE dimer with the EspG3 chaperone (14). Despite high conservation among PPE proteins in the identified EspG3 binding region from PPE41, three residues vary depending on whether the PPE protein is secreted by ESX-1, ESX-3, or ESX-5 (12). Alteration of any or all of these positions in the ESX-5-dependent PPE41 did not disrupt PPE41-EspG3 binding (12). Based on this observation it has been suggested that structural elements outside of the EspG-binding region differentiate the ESX-5-specific PPE proteins from their ESX-1 and ESX-3 homologs to bind EspG3 (12).

This study was initiated to understand the how each EspG from the different ESX systems specifically recognizes its unique subset of cognate PE-PPE heterodimers. Here we present the structure of PE5-PPE4-EspG3 from ESX-3. This structure reveals a novel binding mode of PE-PPE proteins with the EspG chaperone and suggests the molecular mechanism by which the PE-PPE dimers are specifically targeted by cognate chaperones.

Results

EspG forms a complex with PE5-PPE4 and binding is conserved across species

In order to understand the mechanism for the specificity of PE-PPE recognition by cognate chaperones a high-resolution structure of a heterotrimer produced by the ESX systems, other than ESX-5, was needed. Our efforts have focused on optimizing the ESX-3 PE-PPE-EspG heterotrimer for X-ray structural studies. Constructs of full-length PE5 (Rv0285), the conserved N-terminal PPE domain of PPE4 (Rv0286, residues 1-181), and EspG3 (Rv0289) from Mycobacterium tuberculosis (Fig. 1a) never formed high-resolution diffraction quality crystals, despite our best efforts. The difficulty could be due to some heterogeneity in the processing of EspG3mt within the Escherichia coli cell, as seen by the double band in Fig. 1b and Sup. Fig. 1a. Numerous variations of PE5-PPE4-EspG3 constructs were screened utilizing multiple mycobacterial species, different fusion approaches, and even mixing PE5-PPE4 dimers with EspG3 chaperones from different species (Sup. Table 1). This latter approach was inspired by the work done on the Plasmodium aldolase-thrombospondin-related anonymous protein complex (15), and in the end, produced the best crystals for further diffraction experiments. To ensure the mixed heterotrimers behaved the same in solution as the wild-type (WT) heterotrimer, size-exclusion chromatography with multi-angle light scattering (SEC-MALS) experiment was performed on both the WT PE5mt-PPE4mt-EspG3mt heterotrimer as well as the mixed PE5mt-PPE4mt-EspG3mm heterotrimer which contained the Mycobacterium marinum EspG3 (MMAR_0548) with 78% sequence identity to EspG3mt (Fig. 1b-c). Both heterotrimers form a 1:1:1 complex with experimental molecular weights of 56.2 kDa (Fig. 1b) for the full M. tuberculosis heterotrimer (theoretical heterotrimer molecular weight of 58.8 kDa) and 54.6 kDa (Fig. 1c) for the mixed heterotrimer with the M. marinum EspG3 (theoretical heterotrimer molecular weight of 58.1 kDa). Co-purification assays were run with both M. tuberculosis and M. marinum EspG3 with the M.
**PE5-PPE4-EspG3 structure from ESX-3**

*tuberculosis* PE4-PPE5, along with EspG3’s from *Mycolicibacterium smegmatis* (MSMEG_0622), *Mycolicibacterium hasiacum* (MHAS_04631), and *Mycobacterium kansasii* (MKAN_17015). Due to the His6-tag only being present on PE5mt, EspG3 co-purification required interaction with the PE5mt-PPE4mt heterodimer. Across all species that were tested, EspG3 co-purified with the PE5mt-PPE4mt heterodimer (Supplementary Fig. 1a-e). The binding of different EspG3s to the same PE-PPE heterodimer suggests a common protein-protein recognition mechanism within the ESX-3 family.

**Overall structure of PE5mt-PPE4mt-EspG3mm**

The PE5mt-PPE4mt-EspG3mm heterotrimer was able to form diffraction quality crystals, and two different crystal forms were observed that diffracted to 3.3 Å (I422) and 3.0 Å (P212121) (Table 1). Final refinement and data statistics are in Table 1. Overall there is little structural variation between the individual proteins across the copies present in the two crystal forms (Table 2).

For all structural analysis and comparisons, the first copy of the PE5mt-PPE4mt-EspG3mm heterotrimer from the higher resolution P212121 crystal form was used because it diffracted at a higher resolution and has the lowest B-factors from the non-crystallographic copies in the P212121 form. EspG3mm interacts solely with the tip of PPE4mt (Fig. 2), similar to EspG5 in the previously solved ESX-5 heterotrimers (12-14). However, the orientation of PE5mt-PPE4mt relative to EspG3mm is dramatically different than was observed for either ESX-5 heterotrimer, and the differences between them will be described in later sections. The YXXDE motif for ESX secretion of PE5mt is accessible for interactions with the rest of the ESX machinery, as it is located distal to the EspG3mt interaction (11). In both crystal forms, this secretion motif is disordered, similar to the motif in PE8mt from the PE8mt-PPE15mt-EspG5mt heterotrimer (14). The individual components of the PE5mt-PPE4mt-EspG3mm heterotrimer align well to the individual components of the previously reported ESX-5 heterotrimers, both PE25mt-PPE41mt-EspG3mt (4KXR and 4W4L) and PE8mt-PPE15mt-EspG3mt (5XFS), with only moderate variations (Table 3).

In a small-angle X-ray scattering (SAXS) experiment was done on the PE5-PPE4-EspG3 heterotrimer from *M. smegmatis*. Comparisons between this SAXS analysis and our crystal structure were performed to see if the solution-based characterization of the heterotrimer matched the X-ray-based characterization. We ran CRYSOX (17) on our crystal structure in comparison to the experimental scattering data from the *M. smegmatis* heterotrimer. The overall $\chi^2$ is 2.53, which is acceptable given that the heterotrimers are from different species with only 54.0-73.8% sequence identity across the different components (Sup. Fig. 2). The main differences are in the extreme high- and low-resolution areas, likely arising from differences in the primary structure between the two samples and from aggregation in the SAXS sample, respectively. Therefore, we are confident that the crystal structure is an appropriate model of the ESX-3 heterotrimer as it exists in solution.

**Interface between PPE4mt and EspG3mm**

The interface between EspG3mm and PPE4mt contains numerous hydrophobic interactions, multiple hydrogen bonds, and two salt bridges centered around E140PPE4mt (Fig. 2b-f). Overall the interface buries 3,121 Å$^2$ of solvent-accessible surface area, as calculated by the PISA server (18), and has the shape correlation $S_c$ value of 0.664 (19). The interface is comprised of 30 total residues from PPE4mt and 49 residues from EspG3mm (Fig. 3). The tip of PPE4mt containing the ends of $\alpha$4 and $\alpha$5 and the loop between them is inserted into a groove on EspG3 composed of its central $\beta$ sheet and C-terminal helical bundle. This bundle shields the hydrophobic tip of PPE4mt, including the hh motif of FF128, from solvent access. The tip of PPE4mt is interacting with EspG3 in such a way that the complex is unlikely to disengage at the ESX secretion machinery without structural rearrangement of the chaperone.

**Mutations cause disruptions in the PPE4-EspG3 interface**

In order to probe the interface of the crystal structure and test the importance of interacting residues, we made several mutations on both PPE4 and EspG3 sides of the interface and opted to use the cognate PE5mt-PPE4mt-EspG3mt heterotrimer to test our mutations. The PISA output (18) of the interface was analyzed along with sequence alignments of the current known ESX-3 PPE proteins (Sup. Fig. 3) and alignments of the EspG3
used in this study (Sup. Fig. 4) to select which residues in the interface would be mutated.

PPE4mt is well conserved along the interface among ESX-3-specific PPE proteins (Sup. Fig. 3), and we targeted strictly conserved residues in the interface. We selected N127 and N132 because they contain buried hydrogen bonds, F128 and F129 because they are the hh motif and contribute a large amount of solvation energy to the interface according to PISA (18), and E140 because it is part of the salt bridges in the interface. We ran co-purification pull-down assays with mutated PPE4mt and EspG3mt (Table 4). As described earlier, EspG3mt is only co-purified with the PE5-PPE4 heterodimer if it forms a complex. The introduction of charges into the buried hydrogen bonds with N127D and N132E was unable to break the PPE4mt-EspG3mt interaction, and neither was the charge reversal of E140R, as all three mutations co-purify with EspG3mt (Sup. Fig. 5a). This suggests that disruption of any of these single positions is not sufficient to abolish PPE4mt-EspG3mt interaction. Conversely, the introduction of charged residues into the hh motif with F128R or F129E did disrupt the interface and prevented EspG3mt from being co-purified (Sup. Fig. 5a), as it interrupts with the hydrophobic environment deep within the EspG3mt binding pocket.

The interface of EspG3mt is also well conserved amongst the various EspG3s tested in this study (Sup. Fig. 4), and again, we targeted strictly conserved residues. We selected R208 and E212 because they contain buried hydrogen bonds, R87 and R102 because they form the salt bridge within the interface, and S231 because it sits at the top of the groove of EspG3 and could sterically block entrance into the pocket. Neither single mutation of the salt bridge, R87E or R102E, was able to prevent co-purification of EspG3mt (Sup. Fig. 5b). Also, the introduction of a charged residue with R208E was unable to prevent the interaction (Sup. Fig. 5b). In contrast, E212R was sufficient to prevent co-purification, as well as S231Y (Sup. Fig. 5b), as both prevent the hydrophobic tip of PPE4mt from interacting with the binding pocket of EspG3mt either by charge repulsion or steric hindrance. Thus, our mutations on both PPE4mt and EspG3mt highlight the importance of the hydrophobic environment deep within the PPE4mt-EspG3mt interface.

Structure of EspG3 in and out of heterotrimer complex

Our structure is the first of EspG3 solved in complex with a cognate PE-PPE dimer, and thus we wanted to compare it to the previously solved unbound EspG3 structures. In total, there are six available EspG3 structures, four of EspG3ms (PDB codes: 4L4W, 4RCL, 5SXL, and 4W4J (13,16)), one EspG3mt (4W4J (13)), and one EspG3mm (5DLB (16)). These six structures can be classified into two different forms, an ‘open’ and a ‘closed’ form. The differentiation between these two forms is the orientation of the C-terminal helical bundle relative to the core β-sheet. The EspG3mm structure (5DLB) is representative of the ‘open’ form, and one of the EspG3ms structures (4RCL) is representative of the ‘closed’ form. Analysis of EspG3ms as it exists in the PE5mt-PPE4mt-EspG3mm heterotrimer was done relative to these two representative structures. The overall alignment of the representative structures to the bound EspG3mm was good with r.m.s.d. of 2.1 Å and 1.9 Å for the ‘open’ and ‘closed’ forms, respectively (Fig. 4a). Inspection of these alignments show the majority of differences to be within the arrangement of the C-terminal helical bundles, with the bound form of EspG3mm being in close to the orientation found in the ‘closed’ form (Fig. 4b-c). The bound EspG3mm cannot be any closer to the ‘closed’ form orientation because the C-terminal helical bundle is making contacts with PPE4mt. We hypothesized that this C-terminal helical bundle is dynamic and closes on cognate PPE proteins upon interaction. A comparison between the bound EspG3mm structure and the ‘open’ EspG3mm was performed with the DynDom server to test this hypothesis (20). DynDom identified a moving domain within the structures that was located in the C-terminal helical bundle (Fig. 4d). DynDom’s analysis also performed a whole structure alignment that agreed with the previous Dali alignment in Fig 4a-b. DynDom performed alignments between the fixed domains (residues 11-168 and 189-279) and the moving domains (residues 168-188), which resulted in much better alignments with r.m.s.d. of 1.76 Å and 0.86 Å, respectively. Therefore, the moving domain, the C-terminal helical bundle, is essentially structurally identical between PPE4mt-bound EspG3mm and the ‘open’ EspG3mm, and its rotation of 30.2° and translation of 0.8 Å is moderately perturbing the fixed domain. Since the moving
domain is making extensive contact with PPE$_{4_{mt}}$ and PPE$_{4_{mt}}$ would sterically clash with the current orientation of the C-terminal helical bundle, the movement from the ‘closed’ to the ‘open’ orientation could be significant in releasing the secreted PE-PPE dimers from the chaperone at the secretion machinery.

Comparison of ESX-3 and ESX-5 PE-PPE-EspG heterotrimers

A vastly different binding mode is observed when comparing the ESX-3-specific PE$_5_{mt}$-PPE$_4_{mt}$-EspG$_{3_{mm}}$ heterotrimer to the previously published ESX-5-specific heterotrimers. As mentioned earlier, there is good agreement when comparing individual components of the ESX-3-specific heterotrimer to the available ESX-5-specific heterotrimers (Table 3). The difference between the two sets of heterotrimers became apparent when they were aligned via EspG (Fig. 5a-b, Sup. Fig. 6). Our results focused on comparisons with the PE$_{25_{mt}}$-PPE$_{41_{mt}}$-EspG$_{5_{mt}}$ (4KXR) heterotrimer, but the same differences were present with the PE$_{8_{mt}}$-PPE$_{15_{mt}}$-EspG$_{5_{mt}}$ (5XFS) heterotrimer. The interaction angle of the different PE-PPE heterodimer with EspG is drastically different between the two heterotrimers, with 30° angle difference (Fig. 5b). Another difference lies within the hh motif loops of PPE$_{25_{mt}}$ (α4-α5 loop) and PPE$_{4_{mt}}$ (α5-α6 loop) (Fig. 5c). In PPE$_{25_{mt}}$, this loop is seven residues long and undertakes a compact conformation that is not altered during EspG$_{5_{mt}}$ binding (12). In contrast, in PPE$_{4_{mt}}$, this loop is nine residues long and has an extended conformation, and this difference was rapidly apparent when PPE$_{25_{mt}}$ and PPE$_{4_{mt}}$ were aligned (Fig. 5c).

This loop conformation also made each PPE protein incompatible with the other’s binding mode. When looking at the PPE alignment in the context of the ESX-3 heterotrimer, the α4-α5 loop of PPE$_{25_{mt}}$ does not align over the central groove of EspG$_{3_{mm}}$ and instead sterically clashes the central β sheet of the chaperone (Fig. 5d). The tip of PPE$_{41_{mt}}$ would have to undergo a drastically new tip confirmation in order to bind in the opening of EspG$_{3_{mm}}$. In the context of the ESX-5 heterotrimer, the α5-α6 loop does not align with the central groove of the chaperone, and instead, PPE$_{4_{mt}}$’s hh motif sterically clashes with the C-terminal helical bundle of EspG$_{5_{mt}}$ (Fig. 5e). Also, none of the salt bridges between PPE$_{41_{mt}}$ and EspG$_{5_{mt}}$ are conserved in PPE$_{4_{mt}}$. Specifically, D134$^{PPE4_{mt}}$-$K235^{EspG5_{mt}}$, D140$^{PPE4_{mt}}$-$R109^{EspG5_{mt}}$, and D144$^{PPE4_{mt}}$-$R27^{EspG5_{mt}}$; that are all replaced with hydrophobic residues in PPE$_{4_{mt}}$: either T137$^{PPE4_{mt}}$, V144$^{PPE4_{mt}}$, and L147$^{PPE4_{mt}}$, respectively.

Discussion

In this work, we present the first structure of the PE$_5_{mt}$-PPE$_{4_{mt}}$-EspG$_{3_{mm}}$ heterotrimer, which is from the ESX-3 system. Our structure is a mixed heterotrimer, and we presented evidence that EspG$_3$ from numerous mycobacterial species can bind the PE$_5_{mt}$-PPE$_{4_{mt}}$ heterodimer. Conservation of the EspG$_3$s used in this study ranged from 57-83% identity, yet an enrichment in conservation is observed within PPE4-interacting residues (Sup. Fig. 4). The ability of EspG$_3$ from numerous mycobacterial species to bind PE$_5_{mt}$-PPE$_{4_{mt}}$ suggests that the recognition mechanism is conserved within ESX systems across species. Overall the PE$_5_{mt}$-PPE$_{4_{mt}}$ interaction is similar to the previously reported PE-PPE-EspG heterotrimers (12-14) in that PPE$_{4_{mt}}$’s tip is solely interacting with EspG$_{3_{mm}}$ and the general secretion motif of YxxxD/E, on PE$_{5_{mt}}$, is at the distal end of the PE$_5_{mt}$-PPE$_{4_{mt}}$ heterodimer. In all copies of PE$_{5_{mt}}$, this motif is unstructured as it is in the PE8-PPE15-EspG$_5$ heterotrimer (14), and similarly, W63$^{PPE4_{mt}}$ is pointed away from this secretion motif. This arrangement is distinct from the PE25-PPE41-EspG$_5$ heterotrimers (12,13) and EspB, an ESX-1 substrate that has a similar structural fold to the PE-PPE heterodimers (21,22). PE8$_{mt}$ contains an expanded C-terminal domain, and since the secretion motif is located in the linker between the C-terminal domain and the PE domain the orientation of the secretion motif was unclear (14). PE$_{5_{mt}}$ does not have an expanded C-terminal domain and is just the conserved PE domain, yet its secretion motif is still unstructured in our heterotrimer. Therefore, it is still unclear the exact significance of the structural variations in the ESX secretion motif, and further work is still needed.

Our structure is the first of EspG$_3$ bound to a cognate PE-PPE heterodimer. In comparisons of the various published EspG$_3$ structures, we identified two different forms that relate to the orientation of the C-terminal helical bundle, an
‘open’ and a ‘closed’ form. EspG3nm, when bound to the PE5mt-PPE4mt heterodimer, is in a conformation slightly different than the ‘closed’ form due to interactions with the tip of PPE4mt. We also found that the C-terminal helical bundle is a dynamic domain and shifts between the ‘open’ and ‘closed’ forms via a hinge movement (Fig. 4d). The functional significance of this domain movement could be two-fold. First, the plasticity of the C-terminal helical bundle could allow EspG3 to accommodate any variation in the ESX-3-specific PPE tips. While the tips of ESX-3-specific PPE proteins is mostly conserved (Sup. Fig. 3), there is some variations at the end of α5 that could alter the tertiary structure and thus slightly alter the interactions with the PE-PPE heterodimers at the ESX-3 secretion machinery. It is unlikely that PPE4mt could be removed from its interactions with EspG3nm without either movement of the C-terminal helical bundle or steric clashes with the C-terminal helical bundle. Movement of this helical bundle and release of PPE4mt would likely require energy input, and a candidate to provide that energy is EccA. EccA is an ATPase (2) and interacts with both EspG and PPE proteins in yeast two-hybrid experiments (13,23). Recent structures of the ESX machinery from both ESX-3 (4) and ESX-5 (6) suggests overall six-fold symmetry of the core ESX machinery within the inner membrane, and EccA could not only be acting to provide the energy required to uncouple the PE-PPE heterodimers from their EspG chaperone but also to provide a platform for interaction with the core secretion machinery as EccA is likely hexameric when functional.

Previous studies showed that each EspG only recognize PE-PPE heterodimers from their cognate systems (11,12). Despite the structures of two different PE-PPE-EspG heterotrimers from ESX-5 (12-14), it was still unclear how EspG5 was differentiating from cognate and non-cognate PE-PPE heterodimers. Our structure represents the first PE-PPE-EspG heterotrimer from ESX-3 and allows for direct comparisons between the ESX-3 and ESX-5 heterotrimers. Our structure reveals that PE5mt-PPE4mt interacts with EspG3nm at a different angle of interaction than what was shown for either ESX-5 heterotrimer. This difference in interaction angle presents a different face of PPE4mt to EspG3nm. We hypothesize that this is a conserved feature of the ESX-3 PE-EspG interaction, as both characterized ESX-5 PE-PPE heterodimers (12-14) display the same face to EspG; despite 33% sequence identity between PPPE41 and PPE15. Therefore, we hypothesize that each ESX system has a unique shape complementarity between its subset of PPE proteins and their cognate EspG chaperone, and these unique shapes are likely not compatible for interaction with non-cognate chaperones. Our structure is also the first of an ESX-3-specific PE-PPE heterotrimer. PE5mt-PPE4mt shares the same global conformation as the previously solved PE-PPE heterodimers, yet it differs drastically in PPE4mt in the loop between α5-α6, which contains the hh motif. This longer, more extended loop interacts deeper in the cleft of EspG3nm and is subsequently more shielded from solvent. It is possible that the longer, extended loop conformation is a feature of ESX-3 PPE proteins and could play an essential role in EspG3 recognition.

In conclusion, we presented the first structure of a PE-PPE-EspG heterotrimer from the ESX-3 system. This structure allowed us to compare the interactions of EspG3 and a cognate PPE protein to the previously described EspG5-PPE interactions. We hypothesize that shape complementarity is a key feature of distinguishing cognate and non-cognate PPE proteins from the EspG chaperones.

Experimental Procedures

Bacterial strains and growth conditions

The Escherichia coli Rosetta2(DE3) strains grown in Luria-Bertani (LB) medium or on LB agar at 37 °C. When needed, antibiotics were included at the following concentrations: chloramphenicol at 10 μg/ml, streptomycin at 50 μg/ml, and kanamycin at 50 μg/ml.

Expression and purification of PE5-PPE4-EspG3 heterotrimers

Optimized DNA sequences based on the amino acids of full-length PE5 and PPE4 residues 1-180 from M. tuberculosis were obtained from Invitrogen and put into a pRSF-NT vector (24) using NcoI and HindIII restriction sites, which contains an N-terminal His6 tag on PE5 that is
cleavable by TEV protease. EspG3mt, EspG3mm expression plasmid was constructed as described previously (12). Mutations in PPE4mt, EspG3mm, and EspG3mm were introduced with Gibson assembly mutagenesis (SGI-DNA).

Co-expression of all heterotrimers was performed as described previously (12). Briefly, E. coli strains containing the appropriate PE5mt-PPE4mt and EspG3 plasmids were induced with 0.5 mM IPTG when they reached an OD at 600 nm of 0.5-0.8 and then continued to shake at 16 °C for 20 h. Cells were harvested by centrifugation. Cells were then resuspended in lysis buffer (300 mM NaCl, 20 mM Tris pH 8.0, and 10 mM imidazole) and 1:100 Halt protease inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA). Cells were lysed using an EmulsiFlex-C5 homogenizer (Avestin, Ottawa, ON, Canada). The soluble lysate was purified over a Ni-NTA column (G-Biosciences, St. Louis, MO). Eluted protein was dialyzed against lysis buffer without imidazole and incubated with 1:20 mg of TEV protease at 4 °C for 20 h before being re-applied to the Ni-NTA column. Flow-through and washes were pooled and concentrated for size-exclusion chromatography over a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences, Marlborough, MA) that was equilibrated in buffer A (100 mM NaCl and 20 mM HEPES pH 7.5).

**Crystallization, data collection, and structure solution**

Purified protein was concentrated to 4.195 mg/ml. Initial screening was done using the MCSG Crystallization Suite (Anatrace, Maumee, OH). This initial screening produced the P21_21_2 crystals that were grown in 200 mM NH4 tartrate and 20% PEG 3350. Optimization around three others hits from the initial crystal screening containing NaCl as the precipitant and various buffers ranging from a pH 5.5 to 8.0 produced the I422 crystals, which were grown in 2.0 M NaCl and 100 mM bis-tris, pH 6.5. Crystals were transferred to cryoprotectant solution, which contained the crystallization solution supplemented with either 20% (P21_21_2) or 25% (I422) glycerol and then flash-cooled in liquid N2. Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Data were processed using XDS and XSCALE (25). Molecular replacement using Phaser (26) was used to solve the structure of both crystal forms. First, the PE25mt-PPE41mt dimer (PDB: 4KXR (12)) and EspG3mm (PDB: 5DLB (16)) were used as search models for the I422 dataset. Later, an early model of the I422 structure was used as a search model for the P21_21_2 dataset. DENSITY MODIFICATION AND ORIGINAL MODEL. The starting model for both forms was then iteratively rebuilt and refined using Coot and phenix.refine (27,28). The final structure for both crystal forms was refined in phenix.refine, with the P21_21_2 form using non-crystallographic symmetry restraints (TLS?). All data collection and refinement statistics are listed in Table 1. The final model was assessed using Coot and the MolProbity server (29) for quality.

**Size-exclusion chromatography multi-angle light scattering (SEC-MALS)**

Proteins were expressed and purified as described and then passed over an AKTA pure with an inline Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences), miniDAWN TREOS, and Optilab T-reX (Wyatt Technologies, Santa Barbara, CA). The system was equilibrated and run in buffer A. Samples were loaded at a volume of 500 μL at a concentration of 2-4 mg/ml and the system was run at 0.5 mL/min. Analysis of light scattering data was done using Astra (Wyatt Technologies). Molecular weight determination was done by analyzing peaks at one-half their maximum. Graphics were prepared using Prism (Graphpad Software, La Jolla, CA).

**Sequence analysis**

Sequence analysis was performed using the EMBL-EBL analysis tools, specifically the Clustal Omega program (30). Rendering of sequence analysis was done with the ESPript server (31).

**Structural analysis**

All structural figures were generated using PyMol (http://www.pymol.org) and Chimera/ChimeraX (chimera ref). Electrostatic surface potentials were calculated using the APBS Electostatics plugin in PyMol (32). Structural alignments were performed with the Dali server (33).

**SAXS Data Comparison and Ab Initio Model Reconstruction**
PE5\textsubscript{ms}-PPE4\textsubscript{ms}-EspG\textsubscript{3ms} heterotrimer SAXS data (SASDDX2 (16)) was compared to a single copy of the mixed PE5\textsubscript{mt}-PPE4\textsubscript{mt}-EspG\textsubscript{3mm} heterotrimer structure (PDB code: 6UUJ) using CRYSOL (17). Ab initio reconstruction of the envelope was completed using GASBOR (34). Monomeric symmetry was used as a constraint for GASBOR. Twenty ab initio models were generated and averaged using the DAMAVER software package (35). DAMSEL rejected only one model.

Data availability
Coordinates and structure factors were deposited in the Protein Data Bank with accession codes 6UUJ ($P2_12_12_1$) and 6VHR ($I422$). All other data generated during this study are included in the article and supplementary information files.

Acknowledgments
The work in this study was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health grants P20GM103486 and P30GM110787, and by the National Institute of Allergy and Infectious Diseases grant R01AI119022 to K.V.K. W.A.C. was supported by the National Science Foundation Research Experiences for Undergraduates (REU) grant 1358627. Use of SER-CAT is supported by its member institutions (see www.ser-cat.org/members.html), and equipment grants (S10 RR25528 and S10 RR028976) from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, contract W-31-109-Eng-38. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Conflicts of Interest
The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Organization, W. H. (2019) Tuberculosis Fact Sheet 2019. World Health Organization, Geneva, Switzerland
2. Bitter, W., Houben, E. N., Bottai, D., Brodin, P., Brown, E. J., Cox, J. S., Derbyshire, K., Fortune, S. M., Gao, L. Y., Liu, J., Gey van Pittius, N. C., Pym, A. S., Rubin, E. J., Sherman, D. R., Cole, S. T., and Brosch, R. (2009) Systematic genetic nomenclature for type VII secretion systems. PLoS Pathog 5, e1000507
3. Groschel, M. I., Sayes, F., Simeone, R., Majlessi, L., and Brosch, R. (2016) ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14, 677-691
4. Famelis, N., Rivera-Calzada, A., Degliesposti, G., Wingender, M., Mietrach, N., Skehel, J. M., Fernandez-Leiro, R., Bottcher, B., Schlosser, A., Llorca, O., and Geibel, S. (2019) Architecture of the mycobacterial type VII secretion system. Nature 576, 321-325
5. Poweleit, N., Czudnochowski, N., Nakagawa, R., Trinidad, D., Murphy, K. C., Sassetti, C. M., and Rosenberg, O. S. (2019) The structure of the endogenous ESX-3 secretion system. Elife 8
6. Beckham, K. S., Ciccarelli, L., Bunduc, C. M., Mertens, H. D., Ummels, R., Lugmayr, W., Mayr, J., Rettel, M., Savitski, M. M., Svergun, D. I., Bitter, W., Wilmanns, M., Marlovits, T. C., Parret, A. H., and Houben, E. N. (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. Nat Microbiol 2, 17047
7. Abdallah, A. M., Gey van Pittius, N. C., Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., and Bitter, W. (2007) Type VII secretion--mycobacteria show the way. Nat Rev Microbiol 5, 883-891
8. Daleke, M. H., Ummels, R., Bawono, P., Heringa, J., Vandenbroucke-Grauls, C. M., Luirink, J., and Bitter, W. (2012) General secretion signal for the mycobacterial type VII secretion pathway. Proc Natl Acad Sci U S A 109, 11342-11347
9. Ates, L. S. (2019) New insights into the mycobacterial PE and PPE proteins provide a framework for future research. Mol Microbiol
10. Delogu, G., Brennan, M. J., and Manganelli, R. (2017) PE and PPE Genes: A Tale of Conservation and Diversity. *Adv Exp Med Biol* **1019**, 191-207

11. Daleke, M. H., van der Woude, A. D., Parret, A. H., Ummels, R., de Groot, A. M., Watson, D., Piersma, S. R., Jimenez, C. R., Lurink, J., Bitter, W., and Houben, E. N. (2012) Specific chaperones for the type VII protein secretion pathway. *J Biol Chem* **287**, 31939-31947

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

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

14. Chen, X., Cheng, H. F., Zhou, J., Chan, C. Y., Lau, K. F., Tsui, S. K., and Au, S. W. (2017) Structural basis of the PE-PPE protein interaction in Mycobacterium tuberculosis. *J Biol Chem* **292**, 16880-16890

15. Bosch, J., Buscaglia, C. A., Krumm, B., Ingason, B. P., Lucas, R., Roach, C., Cardozo, T., Nussenzwieg, V., and Hol, W. G. (2007) Aldolase provides an unusual binding site for thrombospondin-related anonymous protein in the invasion machinery of the malaria parasite. *Proc Natl Acad Sci U S A* **104**, 7015-7020

16. Tuukkanen, A. T., Freire, D., Chan, S., Arbing, M. A., Reed, R. W., Evans, T. J., Zenkevicuute, G., Kim, J., Kahng, S., Sawaya, M. R., Chaton, C. T., Wilmanns, M., Eisenberg, D., Parret, A. H. A., and Korotkov, K. V. (2019) Structural Variability of EspG Chaperones from Mycobacterial ESX-1, ESX-3, and ESX-5 Type VII Secretion Systems. *J Mol Biol* **431**, 289-307

17. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOl— a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768-773

18. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797

19. Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-950

20. Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144-154

21. Korotkova, N., Piton, J., Wagner, J. M., Boy-Rottger, S., Japaridze, A., Evans, T. J., Cole, S. T., Pojer, F., and Korotkov, K. V. (2015) Structure of EspB, a secreted substrate of the ESX-1 secretion system of Mycobacterium tuberculosis. *J Struct Biol* **191**, 236-244

22. Solomonson, M., Setiaputra, D., Makepeace, K. A. T., Lameignere, E., Petrochkeno, E. V., Conrady, D. G., Bergeron, J. R., Vuckovic, M., DiMaio, F., Borchers, C. H., Yip, C. K., and Strynadka, N. C. J. (2015) Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. *Structure* **23**, 571-583

23. Teutschbein, J., Schumann, G., Mollmann, U., Grabley, S., Cole, S. T., and Munder, T. (2009) A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of Mycobacterium tuberculosis. *Microbiological research* **164**, 253-259

24. Korb, K. V., Debra, J. R., and Hol, W. G. J. (2013) A dodecameric ring-like structure of the N0 domain of the type II secretin from enterotoxigenic *Escherichia coli*. *J Struct Biol* **183**, 354-362

25. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132

26. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
PE5-PPE4-EspG3 structure from ESX-3

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

28. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221

29. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315

30. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641

31. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324

32. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128

33. Holm, L. (2019) Benchmarking Fold Detection by DaliLite v.5. *Bioinformatics* 34. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953

35. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864

36. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci* **29**, 128-140
Table 1. Data collection and refinement statistics.

| Data Collection          | PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> (PDB ID 6UUJ) | PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> (PDB ID 6VHR) |
|-------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| Wavelength (Å)          | 1.000                                                         | 1.000                                                         |
| Space group             | P<sub>2</sub><sub>1</sub>,<sub>2</sub><sub>1</sub>                                                             | I<sub>4</sub>2<sub>2</sub>                                                                 |
| Cell Dimensions:        |                                                              |                                                               |
| a, b, c (Å)             | 72.26, 158.63, 209.31                                         | 219.14, 219.14, 104.44                                         |
| α, β, γ (°)             | 90, 90, 90                                                   | 90, 90, 90                                                   |
| Resolution (Å)          | 39.51 – 3.00 (3.08 – 3.00)<sup>a</sup>                       | 35.73 – 3.30 (3.39 – 3.30)                                    |
| R<sub>sym</sub>         | 0.131 (1.56)                                                 | 0.087 (2.18)                                                 |
| R<sub>pim</sub>         | 0.070 (0.848)                                                | 0.029 (0.524)                                                |
| CC<sub>1/2</sub><sup>b</sup> | 0.998 (0.590)                                           | 0.999 (0.597)                                               |
| I/σ                     | 9.45 (1.14)                                                  | 15.70 (1.31)                                                 |
| Completeness (%)        | 99.1 (99.5)                                                  | 99.8 (100)                                                   |
| Multiplicity            | 4.2 (4.4)                                                    | 10.3 (9.5)                                                   |

| Refinement              |                                                              |                                                               |
| Resolution (Å)          | 39.51 – 3.00                                                 | 35.73 – 3.30                                                  |
| No. reflections (total/free) | 48463/2462                                      | 19335/928                                                    |
| R<sub>work</sub>/R<sub>free</sub> | 0.266/0.303                                      | 0.248/0.266                                                   |
| Number of atoms:        |                                                              |                                                               |
| Protein                 | 14535                                                       | 3643                                                         |
| Ligand/ion              | 0                                                          | 0                                                           |
| Water                   | 4                                                          | 0                                                           |
| B-factors:              |                                                              |                                                               |
| Protein                 | 101.6                                                       | 173.2                                                        |
| Water                   | 70.6                                                        |                                                              |
| All atoms               | 101.6                                                       | 173.2                                                        |
| Wilson B                | 87.9                                                        | 147.3                                                        |
| R.m.s. deviations:      |                                                              |                                                               |
| Bond lengths (Å)        | 0.002                                                       | 0.002                                                        |
| Bond angles (°)         | 0.53                                                        | 0.503                                                        |
| Ramachandran distribution<sup>c</sup> (%) |                                                              |                                                               |
| Favored                 | 96.42                                                       | 91.34                                                        |
| Allowed                 | 3.58                                                        | 8.04                                                         |
| Outliers                | 0                                                          | 0.62                                                         |
| Rotamer outliers<sup>c</sup> (%) | 0.28                                                      | 0                                                            |
| Clashscore<sup>d</sup>  | 7.02                                                        | 6.94                                                         |
| MolProbity Score<sup>e</sup> | 1.62                                                      | 1.89                                                         |

<sup>a</sup>Values in parentheses are for the highest resolution shell.
<sup>b</sup>CC<sub>1/2</sub> correlation coefficient is defined in (Karplus and Diederichs) and was calculated with XSCALE (Kabsch).
<sup>c</sup>Calculated with the MolProbity server (http://molprobity.biochem.duke.edu) (Chen or Williams).
<sup>d</sup>Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.
<sup>e</sup>MolProbity Score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution (Chen or Williams).
Table 2. Structural Variations in copies of PE5<sub>mt</sub>-PPE4<sub>mt</sub>-EspG<sub>3mm</sub> structure in r.m.s.d. (Å).

|         | PE5<sub>mt</sub> | PPE4<sub>mt</sub> | EspG<sub>3mm</sub> |
|---------|-----------------|-----------------|-------------------|
| **Aligned to 6UUJ copy 1** |                 |                 |                   |
| 6UUJ copy 2 | 0.2             | 0.3             | 0.4               |
| 6UUJ copy 3 | 0.4             | 0.2             | 0.4               |
| 6UUJ copy 4 | 0.3             | 0.2             | 0.4               |
| 6VHR     | 0.4             | 0.5             | 0.6               |
| **Aligned to 6UUJ copy 2** |                 |                 |                   |
| 6UUJ copy 1 | 0.2             | 0.3             | 0.4               |
| 6UUJ copy 3 | 0.3             | 0.3             | 0.3               |
| 6UUJ copy 4 | 0.3             | 0.3             | 0.4               |
| 6VHR     | 0.4             | 0.5             | 0.6               |
| **Aligned to 6UUJ copy 3** |                 |                 |                   |
| 6UUJ copy 1 | 0.4             | 0.2             | 0.4               |
| 6UUJ copy 2 | 0.3             | 0.3             | 0.3               |
| 6UUJ copy 4 | 0.4             | 0.3             | 0.4               |
| 6VHR     | 0.5             | 0.5             | 0.6               |
| **Aligned to 6UUJ copy 4** |                 |                 |                   |
| 6UUJ copy 1 | 0.3             | 0.2             | 0.4               |
| 6UUJ copy 2 | 0.3             | 0.3             | 0.4               |
| 6UUJ copy 3 | 0.3             | 0.3             | 0.4               |
| 6VHR     | 0.4             | 0.5             | 0.7               |
Table 3. Structural variations between individual components of ESX-3 heterotrimer and the previously published ESX-5 heterotrimers in r.m.s.d. (Å)

|       | PE$_{5_{mt}}$-PPE$_{41_{mt}}$-Esp$_{G_{5_{mt}}}$ (4KXR) | PE$_{25_{mt}}$-PPE$_{41_{mt}}$-Esp$_{G_{5_{mt}}}$ (4W4L) | PE$_{8_{mt}}$-PPE$_{15_{mt}}$-Esp$_{G_{5_{mt}}}$ (5XFS) |
|-------|--------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------|
| PE$_{5_{mt}}$ | 2.3                                                   | 2.3                                                   | 1.4                                                   |
| PPE$_{4_{mt}}$ | 3.3                                                   | 3.4                                                   | 2.5                                                   |
| Esp$_{G_{3_{mm}}}$ | 2.4                                                   | 2.7                                                   | 2.3                                                   |
Table 4. Summary of analysis of PE5mt-PPE4mt-EspG3mt interactions *in vitro*.

| PPE4 mutations |  |
|----------------|---|
| N127D          | + |
| F128R          | - |
| F129E          | - |
| N132E          | + |
| E140R          | + |

| EspG3 mutations |  |
|-----------------|---|
| R87E            | + |
| R102E           | + |
| R208E           | + |
| E212R           | - |
| S231Y           | - |
Figure 1. Solution characterization of the PE5-PPE4-EspG3 heterotrimer. A, schematic showing design and molecular weights for constructs used in this study. PE5 from *M. tuberculosis* (*Rv0285*) contains an N-terminal His6-tag that is connected to the gene via a TEV protease cleavable linker. PPE4 (*Rv0286*) was truncated after its N-terminal PPE domain. Full-length copies of both *M. tuberculosis* (*Rv0289*) and *M. marinum* (*MMAR_0548*) EspG3 were used. B and C, elution profile of PE5\textsubscript{mt}-PPE4\textsubscript{mt}-EspG3\textsubscript{mt} (B) and PE5\textsubscript{mt}-PPE4\textsubscript{mt}-EspG3\textsubscript{mm} (C), with the right Y-axis showing the MALS-measured molecular weight. Inset shows an SDS-PAGE image of the major peak fraction.
Figure 2. Crystal structure of the PE5$^{\text{mt}}$-PPE4$^{\text{mt}}$-EspG3$^{\text{mm}}$ heterotrimer and selected interactions in PPE4$^{\text{mt}}$-EspG3$^{\text{mm}}$ interface. A. Crystal structure of the PE5$^{\text{mt}}$-PPE4$^{\text{mt}}$-EspG3$^{\text{mm}}$ heterotrimer in a cartoon representation with two views related by a rotation of 180°. EspG3 interacts exclusively with the tip PPE4, distal to PE5. B-F. Interacting residues are shown with main chain and side chain in stick form with electron density map (2Fo-Fc shown at 1.0 σ) covering side chains and hydrogen bonding in grey dashed lines.
Figure 3. Interface between $\text{PPE}_4\text{mt}$ and $\text{EspG}_3\text{mm}$. A, Surface representation of $\text{PPE}_4\text{mt}$ and $\text{EspG}_3\text{mm}$ shown in an ‘open book’ view. Interacting residues are colored in blue ($\text{PPE}_4\text{mt}$) and orange ($\text{EspG}_3\text{mm}$), with mutated residues highlighted in light orange ($\text{PPE}_4\text{mt}$) and light blue ($\text{EspG}_3\text{mm}$). Another view of $\text{PPE}_4\text{mt}$, related by a $180^\circ$ rotation, is also shown. B, The same orientations as panel A but with the surface colored according to surface potential as calculated by APBS (32). The interacting residues are circled within a black line.
Figure 4. EspG3 exists in multiple structural forms. A, An ‘open’ (5DLB, EspG3mm, sand) and ‘closed’ (4RCL, EspG3ms, maroon) conformation of EspG3 were aligned to EspG3mm (6UUJ, orange) as it is bound to PPE4mt. Overall the different conformations align well to the bound conformation of EspG3mm with r.m.s.d. of 2.1 Å (‘open’) and 1.9 Å (‘closed’). B and C, Closeups highlighting the different orientations of the α5 helices in the ‘open’ (B) and ‘closed’ (C) EspG3 structures as compared to EspG3mm bound to PPE4mt. D, Movement regions defined in EspG3 as it moves from the ‘open’ conformation to the bound conformation in two different views related by a 90° rotation. The rotation axis for the moving domain is shown in gray. Each conformation maintains the same coloring as in panel A, with the hinge between the moving and fixed domains colored blue (‘open’) and light blue (bound).
Figure 5. PE5<sub>mt</sub>-PPE4<sub>mt</sub> interacts with EspG<sub>3mm</sub> chaperone in a unique mode compared to ESX-5 PE-PPE dimers. A, Structural alignment of the ESX-3 and ESX-5 heterotrimers via the EspG chaperones (36) reveals a difference in the angle of interaction between the PE-PPE heterodimers with their respective chaperone. B, Top view of alignment from panel A. C, Superposition of PPE4<sub>1mt</sub> and PPE4<sub>mt</sub> highlights difference in hh loop conformations between ESX-3 (PPE4<sub>mt</sub>) and ESX-5 (PPE4<sub>1mt</sub>). D-E, Superposition of PPE alignment from panel C in context of EspG<sub>3mm</sub> interaction (D) and EspG<sub>5mt</sub> interaction (E) shows the incompatibility of each PPE protein with non-cognate chaperone binding.
PE5-PPE4-EspG$_3$ heterotrimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems

Zachary A Williamson, Catherine T Chaton, William A Ciocca, Natalia Korotkova and Konstantin V Korotkov

J. Biol. Chem. published online July 16, 2020

Access the most updated version of this article at doi: 10.1074/jbc.RA120.012698

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts