Structure of a PE–PPE–EspG complex from Mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion

Damian C. Ekiert and Jeffery S. Cox

Department of Microbiology and Immunology, University of California, San Francisco, CA 94143

Edited by Scott J. Hultgren, Washington University School of Medicine, St. Louis, MO, and approved September 8, 2014 (received for review May 19, 2014)

Nearly 10% of the coding capacity of the Mycobacterium tuberculosis (Mtb) genome is devoted to two highly expanded and enigmatic protein families called PE and PPE, some of which are important virulence/immunogenicity factors and are secreted during infection via a unique alternative secretory system termed “type VII.” How PE-PPE proteins function during infection and how they are translocated to the bacterial surface through the five distinct type VII secretion systems (ESAT-6 secretion system (ESX)) of M. tuberculosis is poorly understood. Here, we report the crystal structure of a PE-PPE heterodimer bound to ESX secretion-associated protein G (EspG), which adopts a novel fold. This PE-PPE-EspG complex, along with structures of two additional EspGs, suggests that EspG acts as an adaptor that recognizes specific PE-PPE protein complexes via extensive interactions with PPE domains, and delivers them to ESX machinery for secretion. Surprisingly, secretion of most PE-PPE proteins in M. tuberculosis is likely mediated by EspG from the ESX-5 system, underscoring the importance of ESX-5 in mycobacterial pathogenesis. Moreover, our results indicate that PE-PPE domains function as cis-acting targeting sequences that are read out by EspGs, revealing the molecular specificity for secretion through distinct ESX pathways.

Significance

Mycobacterium tuberculosis (Mtb) infects nearly a third of the global population, and understanding how Mtb establishes infection and evades host responses is key to development of improved therapies. Two mysterious protein families, called Pro-Glu motif–containing (PE) and Pro-Pro-Glu motif–containing (PPE) proteins, are highly expanded in Mtb and have been linked to virulence, but their function remains unknown. We have determined the crystal structure of a PE-PPE protein dimer bound to ESAT-6 secretion system (ESX) secretion-associated protein G (EspG), a component of the secretion system that translocates PE-PPE proteins to the bacterial cell surface. This structure reveals how each of the four EspGs in Mtb interacts with a different subset of the ~100 PE and ~70 PPE proteins, directing specific classes of PE-PPE “effector” proteins through separate secretory pathways.

Author contributions: D.C.E. designed research; D.C.E. performed research; D.C.E. and J.S.C. analyzed data; and D.C.E. and J.S.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freedly available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4W4L–4W4W).

1To whom correspondence may be addressed. Email: damian.ekiert@ucsf.edu or jefferycox.ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409345111/-/DCSupplemental.
ESX secretion-associated protein G (EspG) (22), another ESX-encoded protein of unknown structure and function. However, the molecular basis for this interaction and its functional importance in PPE secretion remain unclear.

Here, we report the crystal structures of three EspGs from *Mtb* and *Mycobacterium smegmatis* (*Msmg*), and a ternary complex between an *Mtb* PE-PPE pair and its cognate EspG. These structures define the key elements of the PPE–EspG interaction and, coupled with bioinformatics and biochemical interaction studies, suggest that EspGs function as adaptors to deliver PE–PPE complexes to their cognate ESX system for translocation across the plasma membrane. Moreover, our work shows that the vast majority of PE-PPE proteins in *Mtb* interact with EspG from the ESX-5 secretion system, which has been hypothesized to play a major role in PE-PPE protein secretion (9, 20, 23).

### Results and Discussion

**Structure of the EspG Encoded in the ESX-3 Gene Cluster of Mtb**

** Defines a Novel Protein Fold.** To better understand the role of EspG in PE-PPE protein secretion, we determined the crystal structure of the EspG paralog from the ESX-3 system (EspG3) of *Mtb* (EspG3<sub>Mtb</sub>) at a resolution of 2.85 Å (*SI Appendix*, Table S1). Because no structural homologs of EspG could be identified based upon the primary sequence, crystals of *Msmg* EspG3 (EspG3<sub>Msmg</sub>) derivatized with selenomethionine were phased by multiwavelength anomalous dispersion (MAD) (24) (*SI Appendix*, Fig. S1 and Table S2), and the resulting model was used to phase the EspG3<sub>Mtb</sub> dataset by molecular replacement. Here, we focus on the structure of EspG3<sub>Mtb</sub> and will return to discuss EspG3<sub>Msmg</sub> subsequently.

EspG3<sub>Mtb</sub> adopts a novel, mixed \(\alpha\)-\(\beta\)-fold consisting of a 10-stranded \(\beta\)-sheet and eight \(\alpha\)-helices (Fig. L4). The large sheet is kinked in the middle of the 8-strand (\(\beta^3\)) by a nonclassical bulge due to a two-residue insertion, resulting in a slight bend of the sheet. The sheet is flanked on each end by a three-helix bundle (\(\alpha1\)-\(\alpha3\)); Fig. L4), and two additional helices (\(\alpha4\) and \(\alpha7\)) pack against the back face of the \(\beta\)-sheet. In contrast, a large portion of the opposite face of the sheet is solvent-exposed, forming the bottom of a shallow basin lined by nearby loops and helices. The C-terminal 18 residues, including several large hydrophobic residues, are largely disordered in the electron density map but may serve as an ideal docking site for EspG onto the larger ESX apparatus anchored in the plasma membrane.

Intriguingly, the overall EspG fold is approximately twofold symmetrical, with each half containing five strands and four helices (Fig. 1 B and C). The two halves of EspG have the same topology and connectivity, despite little detectable sequence similarity between the two halves. Thus, the extant EspG fold may have arisen from the duplication and fusion of an ancestral subdomain that has been heavily diversified over the course of evolution. Indeed, whereas a search of the Dali Database (25) was unable to identify proteins with a structure similar to the complete EspG, searches using either of the two subdomains identified several distantly related proteins with a similar core topology, although they were generally missing one or more helices and had additional elements not present in EspG (Fig. 1D).

Although the two halves of EspG bear striking resemblance to one another, there are two main regions of divergence between the two subdomains. First, helix \(\alpha2\) from the N-terminal subdomain is rotated \(\sim70^\circ\) from the position of helix \(\alpha2^3\) in the C-terminal subdomain (Fig. 1C, *Upper vs. Lower*). In addition, the short linker connects to a helix \(\alpha1^4\) in the N-terminal subdomain (\(\alpha1\)-\(\alpha2\)) and \(\alpha2^3\) in the C-terminal subdomain (Fig. 1C, *Upper vs. Lower*). No direct contacts are made between the two halves. Thus, the extant EspG fold may have arisen from the duplication and fusion of an ancestral subdomain that has been heavily diversified over the course of evolution. Indeed, whereas a search of the Dali Database (25) was unable to identify proteins with a structure similar to the complete EspG, searches using either of the two subdomains identified several distantly related proteins with a similar core topology, although they were generally missing one or more helices and had additional elements not present in EspG (Fig. 1D).

**EspG Encoded in the ESX-5 Gene Cluster of Mtb Interacts Exclusively with the PPE Subunit of PE25-PPE41.** EspG has been reported to interact with PPE proteins and is required for PE-PPE secretion (22). To understand how EspG recognizes PE-PPE complexes and mediates their secretion through the ESX, we determined the crystal structure of a ternary complex of the EspG encoded in the ESX-5 gene cluster of *Mtb* (EspG5<sub>Mtb</sub>) bound to the PE25-PPE41 dimer at a resolution of 2.45 Å (*SI Appendix*, Fig. S2 and Table S1). In addition, we determined a higher resolution structure of the unbound PE25-PPE41 dimer at a resolution of 1.95 Å (*SI Appendix*, Table S1). Our unbound PE25-PPE41 dimer is very similar to the previously reported structure (10) (rmsd of \(\sim4.3\), over 250 \(\alpha\)-helix atoms). The PE25-PPE41 dimer is entirely \(\alpha\)-helical and forms an elongated, cigar-shaped rod \(\sim110\) Å long and \(\sim20\) Å thick (Fig. 2C). The PPE subunit spans the full length of the molecule, whereas the smaller PE domain is more compact and binds only to one end of the PPE protein. EspG5<sub>Mtb</sub> binds to the opposite end of PPE41, distal from the PE25 binding site (Fig. 2A and B), resulting in only minor structural changes in PE25-PPE41 upon binding (*SI Appendix*, Fig. S3). No direct contacts are made between EspG5<sub>Mtb</sub> and PE25, with \(\sim10\) Å separates the closest pair of residues. It is noteworthy that the C termini of both the PE and PPE subunits also reside at the EspG-distal end of the PE25-PPE41 rod. Because many PE and PPE proteins encode large C-terminal domains of unknown function, binding of EspG far from their C termini may prevent steric clashes with these additional domains.

The EspG-PPE interface is expansive, including 42 residues on EspG and 29 on PPE protein, with a total of 3,500 Å\(^2\) of solvent accessible surface buried upon binding (Fig. 2 C and D), which is nearly two- to threefold the buried surface area in a typical antibody/antigen interaction (27–32). Almost the entire EspG contact surface on PPE41 is derived from a single polypeptide segment (residues 116–155), which forms a helix-turn-helix motif at one end of the PE25-PPE41 rod (Fig. 2C). The tip of this helix-turn-helix is inserted into a shallow basin formed on one face of the EspG \(\beta\)-sheet (Fig. 2D), but the interface extends well outside this depression, with interactions running for eight helical turns along the \(\alpha5\) helix of PPE41. A small number of interactions are...
and three N-terminal PPE41 residues (blue). (Ekiert and Cox sequences lack a proline in and EspGtongue se-

(Fig. 3Ekiert and Cox 2013, Tables S1 and S2). The Pro46 kink reconfigures the adjacent, long α1-α2 tongue from an open V-shape in the EspGtongue to a more closed U-shape in EspGMt. The resulting EspGMt tongue cradles PPE41’s α5 helix over six helical turns (Fig. 2F), accounting for over 40% of the PPE–EspGt interaction. The vast majority of EspGt and EspGt sequences lack a proline in the vicinity of position 46, and consequently may form longer α2 helices with an α1-α2 loop structure that is more similar to the α1-α2 loop structure in EspGMs (Fig. 3D and SI Appendix, Table S3). In contrast, 100% of EspGt sequences analyzed have Pro46, and the α2 helix and α1-α2 loop of EspGt may consequently adopt a conformation similar to the conformation of EspGMs. Consistent with this prediction, the sequence of the α1-α2 loop tongue from EspGMs is similar to that of EspGMt, particularly at positions that contact PPE protein in the PE25-PPE41-EspGt crystal structure (SI Appendix, Fig. S5), whereas the EspGt tongue sequence is divergent. Taken together, variation in the PPE binding surface on EspG suggests that each EspG paralog may interact with distinct subsets of the PE–PPE repertoire.

EspG-Interacting Surface on PPE Proteins Is Conserved. In contrast to the variability of the EspGs, sequence analysis of the ~70 PPE proteins from Mtb Erdman reveal that the core of the EspG-interacting surface on PPE protein is surprisingly well-conserved, particularly in light of the high degree of sequence diversity among PPE proteins (mean pairwise identity of ~47%) (SI Appendix, Fig. S6). Twenty-nine PPE residues make at least one van der Waals contact also made between EspGt and three N-terminal PPE41 residues (His2, Glu4, and Pro7), although these glancing contacts are predicted to make little contribution to specificity or the overall binding energy. On EspGt, four major elements are involved in PPE recognition (Fig. 2E): (i) the convex surface of the β-sheet; (ii) helix α1 from the C-terminal EspG subdomain; (iii) an open, extended loop between strands β2 and β3; and (iv) the long, “tongue-shaped” α1-α2 loop. The α1-α2 loop alone accounts for ~40% of the interactions with PPE protein (Fig. 2F).

PPE Binding Surface on EspG Differs Between Paralogs. The Mtb genome encodes four EspG paralogs, each sharing less than ~25% pairwise sequence identity (SI Appendix, Fig. S4), and nearly 70 highly variable PPE proteins. The crystal structure of the PE–PPE–EspG complex defines the interacting surfaces between EspG and PPE proteins, allowing for an analysis of how sequence and structural variation in these regions may be involved in determining the specificity of EspG–PPE interactions. Comparison of EspGMs and EspGt provides clues into how specificity may be mediated. Additionally, we solved a crystal structure of EspGMs at a resolution of 2.80 Å (Fig. 3A and SI Appendix, Tables S1 and S2). Although EspGMs and EspGt are orthologs, they are only ~60% identical at the sequence level, and consequently may provide additional insight into the structurally variable and/or conserved regions of the PPE-binding surface on EspG.

EspGt adopts the same overall fold as EspGt and EspG, (Fig. 3B; EspGt vs. EspGt: rmsd of ~2.8 Å over 231 Ca atoms; EspGt vs. EspG: rmsd of ~1.6 Å over 239 Ca atoms).

However, structural deviations in two key regions significantly alter the overall shape of the PPE binding surface and may have important implications for PPE binding specificity. First, a number of residues that interact with PPE protein reside in loops of variable length and structure (SI Appendix, Fig. S4), such as the β2-β3 loop, which ranges in length from approximately seven residues in EspGMs and homology models of EspGt to 12 residues in EspGt and 26 residues in EspG (Fig. 3C). Second, the α2 helix and the α1-α2 loop differ considerably between EspGt and EspG (Fig. 3D). In EspGt and also EspG, α2 is approximately five turns long and would clash with a PPE protein bound in the orientation observed in the EspGMt–PPE–EspGt complex. In contrast, the α2 helix in EspGt is broken after ~2.5 turns by a Pro at position 46 (Pro46), which is strictly conserved among mycobacterial EspGt (Fig. 3D and SI Appendix, Table S3). The Pro46 kink reconfigures the adjacent, long α1-α2 tongue from an open V-shape in the EspGt structures to a more closed U-shape in EspGt. The resulting EspGt tongue cradles PPE41’s α5 helix over six helical turns (Fig. 2F), accounting for over 40% of the PPE–EspGt interface.
PPE proteins carrying an Asn residue at this position. (B) EspG interacts with a single EspG. Based upon the structure of the EspG<sub>SM</sub>-PE25-PPE41 complex and conservation of the PPE contact surface between PPE41 and most other PPE proteins, we predict that most PPE proteins in Mtb will interact with EspGs, consistent with the immunological results of Sayes et al. (7). However, the degree of sequence variation on both sides of the interface makes it difficult to rule out other models, such as each EspG interacting with numerous PPE proteins and facilitating PPE secretion from each ESX cluster. To explore the specificity of the PPE–EspG interaction further, we selected a panel of eight diverse PPE proteins from the Mtb genome for additional study (SI Appendix, Fig. S10A). This panel included six proteins encoded outside of ESX clusters (PPE14, PPE17, PPE20, PPE36, PPE41, and PPE60) and two proteins encoded within the ESX-1 and ESX-2 loci (PPE68<sup>ESX-1</sup> and PPE69<sup>ESX-2</sup>). Despite moderate variation in the EspG-binding surface of these PPE proteins, all but one were predicted to bind EspGs. The remaining sequence, PPE69<sup>ESX-2</sup>, is a clear outlier among all of the PPE sequences in Mtb, with several polymorphisms predicted to disrupt binding to EspGs because it (i) lacks a helix-capping residue at position 122, (ii) has a Gly126Trp position compared with other PPE proteins: Ala124 instead of Val, Leu, Phe, or Trp; Leu125 instead of Phe; and Gly147 instead of Ala or Val. Several other positions make key interactions or may be otherwise conserved due to constraints imposed by the binding interface (details are provided in SI Appendix, Fig. S9). Thus, the EspG-interacting surface on PPE41 is moderately conserved, and most of the observed variation is predicted to have only a modest impact on binding to EspG<sub>SM</sub>.

Fig. 4. Most PPE proteins bind EspG<sub>SM</sub> with high affinity. (A) Schematic of chimeric PPE proteins used in binding experiments, with the grafted region highlighted in red. (B) PE25-PPE41 binds to EspG<sub>SM</sub> with high affinity by bio-layer interferometry (BLI). The raw data are plotted in blue, whereas the best global fits are shown in red. The uppermost curve corresponds to PE25-PPE41 binding at 500 nM, with progressively lower responses at 250, 125, 62.5, 31.3, 15.6, and 7.81 nM. (C) Summary of qualitative BLI binding experiments for PPE proteins and EspGs, with (+) or without (−) binding (also SI Appendix, Fig. S11).

with EspG, but many of these residues make just a few interactions per residue, appear somewhat flexible, and are loosely packed at the interface. Consequently, many of these residues probably make only small individual contributions to the binding energy, although in aggregate, they may have a modest effect on binding.

Three contact residues are well-conserved across Mtb PPE proteins, despite making relatively minor interactions with EspG. However, these residues play critical roles in stabilizing the helix-turn–helix motif of the PPE tip, suggesting that the EspG-binding surface of most PPE proteins adopts a similar conformation. First, the Asn122 side chain caps the C-terminal end of PPE helix-4, and ~91% of Mtb PPE proteins have a helix-capping residue at this position (SI Appendix, Fig. S7A). Second, the Asn123 side chain reaches across the helix-turn–helix motif and forms a hydrogen bond with the Phe128 backbone, stabilizing the conformation of the connecting loop (SI Appendix, Fig. S7B). Asn123 is highly conserved, with ~97% of Mtb PPE proteins carrying an Asn residue at this position. Third, residue Gly126 lies at the tip of the helix-turn–helix motif and adopts a positive phi angle (SI Appendix, Fig. S7B). This region of the Ramachandran plot gives strong preference to Gly, and Gly126 is almost universally conserved (64 of 65 Mtb PPE sequences). The strong conservation of these three residues suggests that the overall conformation of the helix-turn–helix motif is similar in most Mtb PPE proteins, allowing us to model the interaction between various PPE-EspG pairs and make predictions about affinity and specificity, which we subsequently tested.

After excluding residues that make no side-chain interactions (i.e., with little sequence constraint at that position) or relatively low-quality contacts, 10 PPE residues remain that likely have the greatest impact on binding affinity and specificity (Fig. 2C). Many of these residues make buried hydrophobic interactions with EspG, such as Ala124, Leu125, Trp143, and Gly147. Trp143, in particular, is well-packed at the interface, making a total of ~22 van der Waals contacts and a cation–π interaction with Arg104 from EspG. In contrast, Ala124, Leu125, and Gly147 are each “underpacked” (SI Appendix, Fig. S8), potentially leaving space for the accommodation of larger hydrophobic side chains at these positions. Indeed, PPE41 has a smaller residue at each of these positions compared with other PPE proteins: Ala124 instead of Val, Leu, Phe, or Trp; Leu125 instead of Phe; and Gly147 instead of Ala or Val. Several other positions make key interactions or may be otherwise conserved due to constraints imposed by the binding interface (details are provided in SI Appendix, Fig. S9). Thus, the EspG-interacting surface on PPE41 is moderately conserved, and most of the observed variation is predicted to have only a modest impact on binding to EspG<sub>SM</sub>.

Fig. 5. Putative protein binding site spans the PPE-EspG interface, suggesting a model for EccA-mediated dissociation of the PE–PPE-EspG complex. (A) PPE41 (white surface) bound to EspG<sub>SM</sub> (purple). Sequence analysis of the PPE surface surrounding the conserved Pro-Pro-Glu motif reveals a patch of conserved residues (orange) that are part of a larger hydrophobic surface (cyan), just adjacent to the j2-j3 loop. (B) Possible model for EccA-mediated dissociation of PE–PPE protein from EspG. Binding of the EccA ATPase across the PPE-EspG interface, perhaps coupled with ATP hydrolysis, may facilitate the release of PE-PPE protein for secretion through ESX and recycling of EspG.
substitution that is too bulky to be accommodated at the interface and unable to adopt a positive phi angle, and (iii) has a 2-a insertion in the vicinity of the 122-126 loop.

Assaying PPE-EspG binding is complicated by the fact that PPE proteins are predicted to interact with specific PE partners and function as obligate heterodimers (33); yet, aside from PE25-PPE41, very few pairs of interacting PE and PPE proteins have been identified (10, 34). Indeed, recombinant expression of PPE proteins alone typically yields nonnative, aggregated, or insoluble material (10). To overcome these technical challenges, we constructed chimeric PPE41s by replacing the key EspG-interacting region of PPE41 with the corresponding sequence from another PPE protein of interest (Fig. 4A). The resulting chimeras are referred to as “PPE14-like,” “PPE17-like,” etc. (SI Appendix, Fig. S10B). The chimeric PPE proteins were coexpressed with PE25 to produce soluble PE-PPE heterodimers. Using this approach, we produced sufficient quantities of seven of the eight PPE proteins selected for the panel, and only the PPE36 chimera failed to express well. Binding between EspG and PPE proteins was assayed by biolayer interferometry. Binding of the PE25-PPE41 heterodimer to EspG5M was robust, yielding a dissociation constant \((K_d)\) of 1.3 nM (Fig. 4B). In contrast, binding of PE25-PPE41 to EspG5M was undetectable (Fig. 4C and SI Appendix, Fig. S11).

Remarkably, five of the six soluble PPE chimeras also bound specifically to EspG5M with high affinity, including the PPE14-like, PPE17-like, PPE20-like, PPE60-like, and PPE68-like (Fig. 4C and SI Appendix, Fig. S11). In contrast, the remaining PPE69-like chimera failed to bind EspG5M, consistent with its unusual sequence features in the EspG binding region. Instead, the PPE69-like chimera bound specifically to EspG3M (Fig. 4C and SI Appendix, Fig. S11). Taken together, our structural, bioinformatic, and biochemical data support a model wherein EspGs specifically recognize PPE domains and the vast majority of PPE proteins in \(Mtb\) interact with EspG5M, whereas a small number of more specialized proteins interact with the remaining EspGs.

### Model for EccA-Mediated Dissociation of the EspG–PE–PPE Complex.

In contrast to PE and PPE proteins, which are translated through the ESX, EspG is not secreted (22). In light of the nanomolar \(K_d\) between PPE protein and EspG, this observation raises questions about how the PE-PPE dimer is released from the complex for secretion. The length and conformation of the \(\beta3\)-loop is arguably the most distinctive feature of EspG\(_{S5M}\) compared with the shorter loops in the EspG3 structures (Fig. 3C). However, despite the large size and striking conformation of the \(\beta3\)-loop, it makes only a moderate number of contacts with PPE protein and much of the loop remains solvent-exposed. Intriguingly, this loop lies in close proximity to the highly conserved Pro-Pro-Glu motif (Fig. 5 A–C) that gave rise to the “PPE” protein family name (2), but it makes only glancing contacts with Pro7 and is unlikely to explain its broad conservation. The PPE surface surrounding the Pro-Pro-Glu motif is well-conserved (Fig. 5 A–C), encompassing \(~500\) \(\text{Å}^2\) of solvent accessible surface area and including residues 6, 10, 109, 113, 139, and 143. This conserved patch is part of a larger, generally hydrophobic surface (Fig. 5 A–C), consistent with the hypothesis that this region may be important for protein–protein interactions. Thus, along with the adjacent \(\beta3\)-loop from EspG, the conserved PPE surface is poised to act as a docking site for other ESX components, such as ESX conserved component A (EccA).

EccA is an ATPase that may be involved in the EspG–PPE interaction because (i) EccA proteins are only encoded in ESX gene clusters that also encode PE and PPE proteins, (ii) EccA mutants accumulate PPE-bound EspG (22), and (iii) EccA interacts with both PPE protein (35) and EspG (SI Appendix, Fig. S12) in yeast two-hybrid assays. Because the \(\beta3\)-loop is one of the most variable regions of EspG, this loop may potentially allow for specific recruitment of the cognate EccA paralog (e.g., recruitment of EccA\(_{5M}\) by EspG\(_{5M}\), etc.). Binding across the interface near the conserved site on PE protein and the \(\beta3\)-loop on EspG might allow EccA to associate the EspG–PE–PPE complex, passing the PE-PPE proteins off to the rest of the ESX system for secretion and recycling EspG to recruit additional PE-PPE proteins from the cytoplasm (Fig. 5D). Because EccA is an active ATPase (36), ATP hydrolysis is likely involved in the dissociation of the EspG–PE–PPE complex, although future studies will be required to understand this process better, as well as the relative contributions of the ATPase and tetratricopeptide repeat (TPR) (37) domains.

### Model of PE-PPE Secretion by ESX1–ESX5 in Mtba

Of the five ESX clusters in the \(Mtb\) genome, ESX-1, ESX-2, ESX-3, and ESX-5 encode EspG, PE, and PPE paralogs. In addition, numerous other PE and PPE genes are scattered throughout the genome. Because each cluster is thought to encode a complete, independent ESX system, it is generally assumed that PE, PPE, and EspG proteins from a given cluster will interact specifically with each other, and only associate with other components from the same ESX cluster (Fig. 6). However, our results suggest that some cross-talk may occur between clusters. For example, PE69 is encoded within the ESX-1 locus in \(Mtb\) and Mycobacterium marinum, and it was previously shown to interact with EspG, from the same cluster (22). However, the PPE68-like chimera also binds to EspG5M with high affinity (Figs. 4C and 6). Similarly, PE69 is encoded within the ESX-2 cluster and presumably interacts with EspG3M; however, in addition, our PPE69-like chimera binds EspG3M (Fig. 4C). Although binding of the ESX-3–encoded PPE protein to EspG5M has not been directly demonstrated, the Msme ESX-3 orthologs form a stable complex on gel filtration (SI Appendix, Fig. S13), and the contact surfaces on PPE protein and EspG3 are highly conserved between Msme and \(Mtb\).

Taken together, our data support a model consisting of two classes of PE–EspG interactions (Fig. 6). First, the PE, PPE, and EspG proteins from each ESX cluster interact with one another, accounting for a relatively small segment of the overall PPE repertoire. This intracluster association probably resembles the situation in nonpathogenic mycobacteria, such as Msme,
where the ESX-1 and ESX-3 clusters each encode a single PE, PPE, and EspG protein and no other PE or PPE genes are found elsewhere in the genome. In some cases, EspG and PPE proteins may cross-react with components from another cluster, although it remains unclear if this cross-reactivity is functionally relevant or merely arises from incomplete orthogonalization of the different ESX clusters. Second, the non–ESX-encoded PPE proteins interact with EspGβ, which includes the vast majority of the PPE sequences (Fig. 6). The dramatic expansion and diversification of the PE and PPE protein families, including the distribution of most of these genes throughout the genome and outside ESX clusters, appears to coincide with the acquisition of ESX-5 in pathogenic mycobacteria, which is consistent with our finding that most PPE proteins interact with EspGβ and are likely secreted in an ESX-5–dependent manner (23).

Although the interaction with EspGβ sheds light on how PE and PPE proteins interface with ESXs, their molecular functions remain to be defined. One intriguing possibility is that the PE and PPE domains serve as targeting modules, directing any cargo domains attached to their C termini to the ESX for translocation and anchoring to the cell surface (38), a process that is important for modulating interactions with the host during infection (39). In this model, EspGβ would play a role akin to the signal recognition and anchoring to the cell surface (38), a process that is important for modulating interactions with the host during infection (39).

To address this model, EspGβ would play a role akin to the signal recognition and anchoring to the cell surface (38), a process that is important for modulating interactions with the host during infection (39).

**Materials and Methods**

Recombiant proteins were expressed in *Escherichia coli* and purified by standard procedures. All diffraction data were collected at the Advanced Light Source beamline 8.3.1 and solved by MAD or molecular replacement. EspG and PE-PPE binding studies and Kd were assayed using an Octet RED instrument (Forte Bio). Further detailed information is provided in SI Appendix, SI Materials and Methods.

**ACKNOWLEDGMENTS.** We thank G. Bhabha (University of California, San Francisco) for discussions and for figure preparation; A. Murzin (Medical Research Council) for access to the Octet instrument. This work was supported by National Institutes of Health Grant R01 AI0801727 (to J.S.C.). D.C.E. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (Grant DRG-2140-12). The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract DE-AC02-05CH11231.

1. Udvardia ZS, Amale RA, Aljani KK, Rodrigues C (2012) Totally drug-resistant tuberculosis in India. Clin Infect Dis 54(6):579–581.
2. Cole ST, et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685):537–544.
3. Sander CA, Sl, et al. (2001) Expression, characterization, and subcellular localization of the Mycobacterium tuberculosis PPE gene RvH1x7. Tuberculosis (Edinb) 81(6):305–317.
4. Cascioferro A, et al. (2007) PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. Mol Microbiol 66(6):1536–1547.
5. Kruh NA, Troutt J, Izzo A, Prreni J, Dobos KM (2010) Portrait of a pathogen: The Mycobacterium tuberculosis proteome in vivo. PLoS ONE 5(11):e13938.
6. Banu S, et al. (2002) Are the PE–PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Mol Microbiol 44(1):9–19.
7. Sayes F, et al. (2012) Strong immunogenicity and cross-reactivity of Mycobacterium tuberculosis ESX-5 type VII secretion: Encoded PE-PPE proteins predict vaccine potential. Cell Host Microbe 11(4):352–363.
8. McDevie OE, et al. (2012) Comparative analysis of Mycobacterium tuberculosis pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. PLoS ONE 7(6):e30593.
9. Gey Van Pittius NC, et al. (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene region. BMC Evol Biol 6:76.
10. Strong M, et al. (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc Natl Acad Sci USA 103(21):7658–7663.
11. Abdalla AM, et al. (2006) A specific secretion system mediates PE41 transport in pathogenic mycobacteria. Mol Microbiol 62(3):667–679.
12. 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(22):13001–13006.
13. Baptista C, Barreto HC, São-José C (2013) High levels of DegU-P activate an Esx-6-like secretion system in Bacillus subtilis. PLoS ONE 8(7):e67840.
14. Syssova TA, Zapedia-Rivera MA, Huppert LA, Burton BM (2014) Dimer recognition and secretion by the ESX secretion system in Bacillus subtilis. Proc Natl Acad Sci USA 111(21):7653–7658.
15. Garufi G, Butler E, Misliakas D (2008) ESAT-6-like protein secretion in Bacillus anthracis. J Bacteriol 190(2):7004–7017.
16. Zahner M, Fyfe PK, Hunter T, Hunter WIN (2013) Characterization of Staphylococcus aureus EspB, an integral membrane component of the Type VII secretion system: Atomic resolution crystal structure of the cytoplasmic segment. Biochem J 449(2):469–479.
17. Akre San Roman S, et al. (2010) A heterodimer of EspA and EspB is involved in spoligotyping and is secreted by a type VII secretion system in *Streptomyces coelicolor*. Microbiology 156(Pt 6):1719–1729.
18. Houben EN, Korotkov KV, Bitter W (2014) Take five–Type VII secretion systems of *Mycobacterium tuberculosis*. Biometals 27(4):859–872. Epub 2014 Aug 19. doi: 10.1007/s10530-014-0952-9.
19. Maharias GG, Sabo PI, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol 178(12):3742–3750.