Structural basis of the PE–PPE protein interaction in *Mycobacterium tuberculosis*

Received for publication, June 16, 2017, and in revised form, August 16, 2017Published, Papers in Press, August 23, 2017, DOI 10.1074/jbc.M117.802645


From the ©Centre for Protein Science and Crystallography, School of Life Sciences, the ©Department of Microbiology, and the ©School of Biomedical Sciences, Chinese University of Hong Kong, Hong Kong, China

Edited by Chris Whitfield

*Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis, has developed multiple strategies to adapt to the human host. The five type VII secretion systems, ESX-1–5, direct the export of many virulence-promoting protein effectors across the complex mycobacterial cell wall. One class of ESX substrates is the PE–PPE family of proteins, which is unique to mycobacteria and essential for infection, antigenic variation, and host–pathogen interactions. The genome of Mtb encodes 168 PE–PPE proteins. Many of them are thought to be secreted through ESX-5 secretion system and to function in pairs. However, understanding of the specific pairing of PE–PPE proteins and their structure–function relationship is limited by the challenging purification of many PE–PPE proteins, and our knowledge of the PE–PPE interactions therefore has been restricted to the PE25–PPE41 pair and its complex with the ESX-5 secretion system chaperone EspG5. Here, we report the crystal structure of a new PE–PPE pair, PE8–PPE15, in complex with EspG5. Our structure revealed that the EspG5-binding sites on PPE15 are relatively conserved among Mtb PPE proteins, suggesting that EspG5–PPE15 represents a more typical model for EspG5–PPE interactions than EspG5–PPE41. A structural comparison with the PE25–PPE41 complex disclosed conformational changes in the four-helix bundle structure and a unique binding mode in the PE8–PPE15 pair. Moreover, homology-modeling and mutagenesis studies further delineated the molecular determinants of the specific PE–PPE interactions. These findings help develop an atomic algorithm of ESX-5 substrate recognition and PE–PPE pairing.

Tuberculosis (TB), which is primarily caused by *Mycobacterium tuberculosis* (Mtb) infection, causes ~2 million deaths annually and therefore remains one of the most devastating diseases worldwide (1, 2). The recent emergence of multidrug-resistant TB and HIV co-infection has highlighted the urgent need for more effective new vaccines (3, 4). Therefore, it is critical to understand the virulent determinants and components of Mtb that are responsible for the host immune response and host–pathogen interactions during different stages of TB infection. The genomes of Mtb and other pathogenic mycobacteria have revealed the prominence of the pe and ppe gene families. For example, the Mtb H37Rv strain contains 99 pe genes and 69 ppe genes, thus highlighting the importance of this protein repertoire for mycobacterial survival and pathogenesis (5). Each PE or PPE protein contains a highly conserved N-terminal domain with a Pro-Glu or Pro-Pro-Glu motif, respectively. Most PE–PPE proteins also possess a variable C-terminal domain that contributes to structural and functional diversification within the protein family. Gene neighborhood and co-expression analyses suggest that PE and PPE proteins act in complexes (6, 7), and these interactions are well-exemplified by the PE25–PPE41 complex (8). Although the exact biological roles of most PE–PPE proteins remain unknown, some of them have been associated with antigenic variation (9–11), immune response modulation (12, 13), drug resistance (14–16), and Mtb virulence (17, 18).

PE–PPE proteins are commonly thought to be either secreted or presented on the cell surface, in line with their functional properties (19, 20). The secretion and surface translocation of PE–PPE proteins is associated with a unique, specialized set of type VII secretion systems (ESX-1 to ESX-5) (21, 22). Earlier studies have demonstrated that recognition of the ESX substrates by the cognate ESX machinery is mediated through a YXXXXD/E secretion signal motif and a WXG motif, which is present in many ESX substrates, including PE–PPE proteins, WxG100 family proteins, and some Esp proteins. However, this signal motif does not define the specificity of secretion system (23, 24). Recently, the crystal structure of the ESX-5–encoded chaperone EspG5 in complex with PE25–PPE41 was solved to reveal the molecular determinants of PPE secretion through specific binding with EspGs (25, 26). Current predictions suggest that ~95% of PPE proteins in Mtb interact with EspG5 and are secreted by the ESX-5 secretion system.

Phylogenetic analysis suggests that *pe*/*ppe* genes co-evolved with *esx* loci and underwent specific gene expansion (20, 27). For ESX-5, three duplicated gene clusters (ESX-5a, ESX-5b, and ESX-5c) are located distal to the ESX-5 region in the *Mtb* genome (28). Although little is known about the functions of...
these clusters, ESX-5a, which encodes two ESX proteins and PE8 and PPE15, is considered an accessory to the parental ESX-5 export apparatus and is responsible for the secretion of a subset of PE–PPE proteins (29). Up-to-date structural data of the PE–PPE protein complex are scarce, mainly because of difficulties associated with PE–PPE protein expression and purification (8). For example, individually expressed PE8s and PPEs are highly insoluble. The only available relevant crystallographic structure of a PE–PPE pair was initially published more than 10 years ago to describe PE25–PPE41 (8), however, the atomic details of other PE–PPE pairs are essential to a better understanding of the distinct protein repertoire required for Mtb infection and pathogenesis. Here, we report the molecular interaction of a novel PE–PPE pair, PE8–PPE15, which is located within ESX-5a and is phylogenetically distinct from PE25–PPE41. A structural comparison with EspG5–PPE41 reveals that the EspG5-binding interface on PPE15 is relatively conserved than EspG5–PPE41, suggesting that the EspG5–PPE15 structure could represent a typical model for EspG5–PPE interactions. Our structure also highlights the structural flexibility induced by the highly conserved prolines and glycines present in the four helix bundle of the PE–PPE complex. Using a homology model of three other PE–PPE pairs and mutagenesis analysis, we identify the molecular determinants of specific PE–PPE recognition.

Results and discussion

Production of the EspG5–PE81–99–PPE151–194 complex for crystallographic studies

Previous bioinformatics analyses predicted the interaction of PE8 (Rv1040c) with PPE15 (Rv1039c) (6, 7). Here, our team used yeast two-hybrid and pulldown assays to validate the direct interaction of these proteins. Because full-length PE8 and PPE15 are highly insoluble, truncated fragments were constructed to improve solubility and define the minimum binding regions (supplemental Fig. S1a). Our results showed that the N-terminal domains of PE8 (residues 1–99) and PPE15 (residues 1–194) are necessary for PE8–PPE15 complex formation. Subsequently, PE81–99 and PPE151–194 were co-expressed and co-purified with the intent to obtain a sufficient sample for structural analysis. However, the protein complex was prone to aggregation at high concentrations, and crystallization trials using this recombinant material failed to yield well-diffracting crystals. To improve the solubility and stability of the protein complex, we included EspG5, which has been reported as a specific chaperone for PE–PPE proteins (30), in the co-purification experiment. Because pe8 and ppe15 are located within the ESX-5a duplicated gene cluster, we hypothesized that EspG5 might interact with the PE8–PPE15 pair. We confirmed the binding of EspG5 to PE81–99–PPE151–194 using a pulldown assay (supplemental Fig. S1b) and further purified this ternary complex to a high level of homogeneity (Fig. 1a). The results of sedimentation velocity and static light scattering experiments yielded a molecular mass of 65–66 kDa for the EspG5–PE81–99–PPE151–194 complex, indicating that these three proteins exist in a stoichiometric ratio of 1:1:1 (Fig. 1b and supplemental Fig. S2). The frictional ratio of 1.96, obtained through a sedimentation velocity analysis, also revealed that the protein complex forms an elongated shape in solution (Fig. 1b).

Overall structure of EspG5–PE81–99–PPE151–194

The purified EspG5–PE81–99–PPE151–194 ternary complex was readily amenable to crystallization trials and yielded crystals that diffracted to a 2.9 Å resolution (Table 1). The structure was determined via molecular replacement as implemented in the program suite phenoix mr_rosetta, using the EspG5–PE25–PPE41 structure (PDB code 4W4L) as the search model. The electron density map was clearly defined throughout the structure, except for residues 85–99 in PE8 and residues 174–194 in PPE15, in line with a secondary structure prediction by Phyre2 (31), suggesting that these regions are highly disordered. This disordered region of PE8 includes the YXXXD/E secretion motif. The final structure contained residues 7–299 of EspG5, residues 7–84 of PE8, and residues 1–173 of PPE15. The overall structure of EspG5–PE81–99–PPE151–194 is similar to the previously reported structure of EspG5–PE25–PPE41 (Fig. 1c) (25, 26), with RMSD of 0.485 Å (EspG5), 2.475 Å (PE), and 2.037 Å (PPE). EspG5 interacts exclusively with helices α4 and α5 of PPE15 at the opposite end of the PE81–99–PPE151–194 heterodimer, and no direct contact between EspG5 and PE81–99 was observed (Fig. 1, c and d). PE8 comprises two helices that interact with helices α1, α2, α3, and α5 of PPE15 to form a four-helix bundle. Structural comparisons between EspG5–PPE15 and EspG5–PPE41 and between PE8–PPE15 and PE25–PPE41 will be discussed in detail in later sections.

Recently, the atomic structure of a ESX-1 substrate EspB has been determined (32, 33). EspB adopts a PE–PPE-like fold, and superimposition of EspB (PDB code 4W1J) with PE8–PPE15 gives an RMSD of 2.114 Å (PE) and 1.764 Å (PPE) (supplemental Fig. S3a). Major structural differences lie on a short helix α1, extended α1–α2 loop and helix α2 in the PE domain of EspB, and a short α6–α7 loop for EspG binding in the PPE domain of EspB. We also compared the YXXXD/E secretion motif located in the C terminus of PE domain and the WXG motif in the helix-turn-helix region of PPE domain in PE8–PPE15, PE25–PPE41, and EspB (supplemental Fig. S3b). In the PE25–PPE41 and EspB structures, the YXXXD/E motif and the WXG motif are in close proximity, allowing van der Waals contact between Tyr78/Glu25 and Trp56/Phe41, and a hydrogen bond formation between Tyr78/EspB and Trp181/EspB. Interestingly, the electron density for the 87YXXXE motif in PE8 cannot be seen, and the side chain of Trp57 in the WXG motif of PPE15 is flipped away from the PE–PPE–binding interface. Although our current structure only contains the PE–PPE domain of PE8–PPE15, it is difficult to predict the orientation of the 87YXXXE motif, which is located in the linker region before the C-terminal 184 residues of the full-length PE8. It is likely that Tyr87 is in a flexible state, and its interaction with Trp57, if it exists, is distinct from that in PE25–PPE41 and EspB. However, the functional significance of these variations in ESX secretion needs further investigation.
EspG5–PPE15 provides a more typical model for EspG5–PPE interactions

A PDBsum (34) analysis of the crystal structure of EspG5–PE81–99–PPE151–194 showed that the contact surface between EspG5 and PPE15 measured 2654 Å, with an interface comprising 23 residues from EspG5 and 21 residues from PPE15. The interaction mainly involves helix H9251, the central β-sheet, the H9251–H9252 loop, and the H9252–H9253 loop of EspG5 and helices H9254 and H9255 of PPE15 (Fig. 2a). On PPE15, the main EspG5 contact regions are localized in helices H9254 and H9255, which contain residues 121–152. We further divided the binding interface of EspG5–PPE15 into three patches for comparison with the EspG5–PPE41 complex (Fig. 2, b and c). The first patch includes residues Val121, Asn124, Thr130, and Trp144 of PPE15, which interact with the β2–β3 loop in EspG5. The molecular interactions in this patch are mediated by a hydrogen bond formation between Asn124PPE15 and Tyr96EspG5 and by hydrophobic contacts of Val121PPE15 with Val98EspG5 and Trp144PPE15 with Arg109EspG5. An identified intramolecular hydrogen bond between Thr130 and Asn124 in PPE15 likely stabilizes the helix-turn-helix tip of PPE15 (supplemental Fig. S4a). Residues in this interaction patch are highly conserved among the Mtb PPE proteins, including PPE41.

The second interface patch is generated by the insertion of the helix-turn-helix tip of PPE15 into a hydrophobic pocket formed by the α1′-helix and central β-sheet of EspG5. Specifically, this patch comprises Val112PPE15, Leu126PPE15, Ile130PPE15, and Pro131PPE15 and Leu180PPE15, Leu216PPE15, Leu237PPE15, and Val241PPE15. This interaction patch is stabilized by hydrophobic contacts and hydrogen bonds between the residues in the two proteins, which are also highly conserved among the Mtb PPE proteins, including PPE41.

Figure 1. Overview of the EspG5–PE81–99–PPE151–194 protein complex. a, elution profile of the EspG5–PE81–99–PPE151–194 complex from size-exclusion chromatography using Superdex 200. Peak fractions as indicated were analyzed by SDS-PAGE analysis. b, a sedimentation velocity ultracentrifugation analysis of the purified EspG5–PE81–99–PPE151–194 complex determined the following: molecular size of 65.0 kDa, frictional ratio of 1.9, suggested ratio of 1:1:1, and elongated shape. The calculated molecular masses of EspG5, PE81–99, and PPE151–194 are 35.0, 10.0, and 20.0 kDa, respectively. c, the crystal structure of the M. tuberculosis EspG5–PE81–99–PPE151–194 complex is depicted as a cartoon in two views with 180° rotation. EspG5 (warm pink) binds exclusively with PPE151–194 (cyan), whereas PE81–99 (yellow) interacts with PPE151–194 to form a four-helix bundle. d, the contact surfaces between EspG5 and PE81–99, and between PPE151–194 and PE81–99. The molecular surfaces of EspG5 and PPE151–194 are colored according to the electrostatic potential. PPE151–194 (left, cyan) and PE81–99 (right, yellow) are depicted in cartoon mode.
(supplemental Fig. S4, b and c). Although the majority of PPE proteins adopt hydrophobic residues at residues equivalent to 125, 126, and 131 in PPE15, PPE41 contains a glutamine residue in the position equivalent to residue 128 in PPE15. Gln127 in PPE15, PPE41 contains a glutamine residue between the PPE complexes. Interestingly, a comparison of the interactions dues, as well as four salt bridges: Asp42

interface patches 1 and 2 appear to be common among EspG5–loop of EspG5. Interestingly, a comparison of the interactions between the main chain atoms of Val241 in EspG5, suggesting a relatively interface area, the binding affinity of EspG5–PPE41 is likely miner herein represents a typical model for EspG5–PPE interaction. The presence of numerous highly conserved proline, glycine, and alanine residues in helices α1–α2 of PE and in helices α2–α3 of PPE proteins suggest that these helices may display different degrees of helical bending required for specific PE–PPE pair formation.

As in PE25–PPE41, PE8–PPE15 complex formation is mediated by both electrostatic and hydrophobic interactions. Both complexes contain a hydrogen bond (Ser48 in the α2 helix of PE8 interacts with Tyr154 in the α5 helix of PPE15), and the interior of the four-helix bundle is lined with multiple hydrophobic contacts. However, distinct salt bridges and hydrogen bonds are found at the upper and lower areas of the PE8–PPE15 complex. We identified four sites in PE8–PPE15 interactions and further validated their importance using mutagenesis and pulldown assays (Fig. 4, a and b). These four sites include a salt bridge (Glu46

Comparison of PE8–PPE15 with PE25–PPE41 reveals structural plasticity and a unique binding mode

Although PE8–PPE15 and PE25–PPE41 exhibit very similar folding characteristics, pronounced bending was observed in the four-helix bundle distal from the EspG5-binding area (Fig. 3a and supplemental Fig. S5). Specifically, the helical pairs α1 and α2 in PE8 and α1 and α3 in PPE15 are tilted by ~26–29 and 20–23°, respectively, leading to dramatic shifts in the helical directions (Fig. 3b). In these four tilted helices, the kinks start at similar longitudinal positions and are facilitated by either a proline (Pro35

Recognition specificity of PE–PPE proteins

Table 1

<table>
<thead>
<tr>
<th>Data collection and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>EspG5–PE81–99–PPE151–194 (PDB code 5XFN)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell dimensions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>β (°)</td>
</tr>
<tr>
<td>γ (°)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>CC1/2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)*</td>
</tr>
<tr>
<td>No. reflections</td>
</tr>
<tr>
<td>Rmerge/Reject (%)</td>
</tr>
<tr>
<td>No. of atoms</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>B-factors</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>RMSD</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
</tr>
<tr>
<td>Favored</td>
</tr>
<tr>
<td>Allowed</td>
</tr>
<tr>
<td>Disallowed</td>
</tr>
</tbody>
</table>

*The values given in parentheses are for the highest-resolution shell.

*The values given in parentheses are for the highest-resolution shell.

EspG5 to PE25–PPE41 using microscale thermophoresis (Fig. 2d). The calculated dissociation constants of EspG5/PE81–99–PPE151–194 is 132 nM, whereas that of EspG5/PE25–PPE41 is 51 nM, indicating that EspG5–PPE15 or most EspG5–PPE proteins have a slightly weaker binding affinity than EspG5–PPE41.
Recognition specificity of PE–PPE proteins

PE8–PPE15 interaction was totally impaired in the PPE15 quintuple mutant (Fig. 4b). These results indicate that the conserved hydrogen bond (Ser48\textsubscript{PE8}–Tyr154\textsubscript{PPE15}) is critical for minimal binding of PE and PPE proteins but strong and specific PE–PPE complex formation involves multiple binding sites along the helix bundles. To confirm that these PPE15 mutants were properly folded, their expression and solubility were examined by immunoblotting (supplemental Fig. S6). All PPE15 mutants exhibited solubility similar to that of the wild-type protein. When we analyzed the PE25–PPE41 structure and sequence alignment, the equivalent residues at these four sites were found to be mainly non-polar (Fig. 4c). The PE25–PPE41 complex binding interface contains one salt bridge (Arg\textsuperscript{24}_{PE25}:Glu\textsuperscript{37}_{PPE41}) and one hydrogen bond (Glu\textsuperscript{17}_{PE25}:Thr\textsuperscript{48}_{PPE41}). These two interactions are not seen in the PE8–PPE15 complex, which contains the equivalent residues of Ala\textsuperscript{24}_{PER} and Glu\textsuperscript{37}_{PPE15} and Gln\textsuperscript{17}_{PER} and Val\textsuperscript{48}_{PPE15} (supplemental Fig. S7). These findings suggest that apart from the conserved hydrogen bond (Ser48\textsubscript{PE8}–Tyr154\textsubscript{PPE15}) and hydrophobic contacts buried in the helix bundle, the two PE and PPE complexes have adopted unique sets of complementary residues that are essential for binding affinity and specificity.

PE–PPE interaction requires specific set of complementary residues and helical bending

We further extended our understanding to other PE–PPE complexes according to our obtained PE8–PPE15 structure. A homology detection by HHpred (35) identified 8 PE proteins and 29 PPE proteins in Mtb that share more than 45% sequence identities with PE8 and PPE15, respectively (supplemental Table S1). Of these, we selected three PE–PPE pairs that were previously predicted by a bioinformatics analysis (6): PE27–
PPE43, PE13–PPE18, and PE32–PPE65, and examined their interactions using yeast two-hybrid assays (Fig. 5a). These three PE–PPE pairs and PE8–PPE15 are classified in the same phylogenetic sublineage IV and are believed to have co-evolved and co-expanded (supplemental Fig. S8). PE8–PPE15, PE13–PPE18, and PE32–PPE65 also constitute the three ESX-5 duplicated clusters (ESX-5a, ESX-5b, and ESX-5c) (28) (Fig. 5b), whereas PE27–PPE43 is associated with the ESX-5 secretion system (27). Homology models of these three PE–PPE pairs were generated using Modeller (36), and their binding interfaces were analyzed by PDBsum (34) and compared with the five interacting sites identified in PE8–PPE15 (Figs. 4 and 5c).

The conserved hydrogen bond observed in Ser48
$^{PE8}$–Tyr154
$^{PPE15}$ was also found in these three PE–PPE pairs. However, for the other four PE–PPE binding sites, variations were noted. For site 1, the salt bridge between Glu46
$^{PE8}$ and Arg14
$^{PPE15}$ was conserved in PE27–PPE43 and PE13–PPE18. However, this site was replaced by hydrophobic contacts in PE32 (Leu46) and PPE65 (Leu15). The alignment of all PE–PPE proteins in
$^{Mtb}$ revealed that 60% of PE–PPE complexes proteins contain the equivalent residues Glu and Arg, suggesting that most complexes adopt a salt bridge to maintain contact between the α2 helix of PE and the α1 helix of PPE. At site 2, hydrogen bonding between Gln51
$^{PE8}$ and Ser93
$^{PPE15}$ was only conserved in PE27–PPE43. In PE13–PPE18 and PE32–PPE65, however, Thr68
$^{PE13}$ can form a hydrogen bond with Thr171
$^{PPE18}$ (Fig. 5c). Interestingly, the residues at sites 3 and 4 were more variable. The hydrogen bond network in PE32–PPE65 is mediated through Gln73
$^{PE32}$ with Tyr46
$^{PPE65}$ and Gln73
$^{PPE65}$. Although PE13–PPE18 and PE27–PPE43 lack hydrogen bonds at sites 3 and 4, helical packing in the lower parts of the helix bundles is facilitated respectively by an Arg68
$^{PE13}$::Glu171
$^{PPE18}$ salt bridge and a Lys17
$^{PE27}$::Gln83
$^{PPE43}$ hydrogen bond. An additional hydrogen bond (His58
$^{PE32}$–Gln83
$^{PPE65}$) appears in the middle of the helix bundle in PE32–PPE65. Taken together, although only some of the interactions are highly conserved, PE–PPE proteins adopt specific sets of complementary residues for complex formation.

To test whether the specific interacting residues identified in PE8–PPE15 are the major determinants of binding specificity, we created various PE25 mutants, including PE25 A51Q, PE25 A51Q/L46E, PE25 A51Q/L46E/A70Q, and PE25 A51Q/L46E/A70Q/L73H. We hypothesized that the substitution of these residues in PE25 with their equivalents from PE8 would allow an interaction with PPE15 (Fig. 4a). Results from a GST pull-down assay revealed that none of these mutants could interact with PPE15 (Fig. 5d). We therefore considered that multiple sites along the interface are required for stabilization of the whole PE–PPE complex, which would explain why the single-, double-, and triple-amino acid PE25 mutant failed to interact with PPE15. However, we expected that PE25 A51Q/L46E/A70Q/L73H, which contained all complementary residues (including the conserved Ser48 in PE25) for the PPE15 interaction, would bind to PPE15. Although we did not find any electrostatic repulsion in the structural model, other determinants might contribute to the PE–PPE binding specificity and may have been responsible for the failure of PE25 A51Q/L46E/A70Q/L73H to pull down PPE15. As described in Fig. 3, PE8–
PPE15 and PE25–PPE41 exhibit various degrees of helical bending. The crystal structure of the PE8–PPE15 complex shows that Gln70 and His73 are positioned close to the kink of helix H92512 in PE8. Therefore, although the mutant PE25 A51Q/L46E/A70Q/L73H contains residues equivalent to those in PE8, residues 70 and 73 in the PE25 mutant are distal from Tyr72 and Tyr45 in PPE15 and result in no interaction. Likely, the binding specificity of the PE8–PPE15 complex is defined by the specific set of complementary residues in the binding interface, as well as the conformation of helices in the bundle.

PE–PPE family members contribute a sophisticated protein repertoire to mycobacteria and are strongly associated with the pathogenesis and virulence of these organisms. However, this set of proteins is poorly understood, particularly regarding the formation and functions of PE and PPE pairs. The first crystal structure of the PE25–PPE41 complex, which was published more than 10 years ago, highlighted a conserved hydrophobic interface within the PE–PPE complex. Here, the crystal structure of a new PE–PPE pair, PE8–PPE15, in complex with EspG5 has elucidated the molecular basis underlying the binding specificities of PE–PPE pairs. In conjunction with our biochemical analysis, we propose a model for PE–PPE recognition (Fig. 6). Extensive hydrophobic contacts along the heterodimeric interface comprise the basic criterion for PE–PPE complex formation. The hydrogen bond observed between a highly conserved Ser48 on PE and Tyr154 on PPE15 is not indicated.
Helical conformation stability is essential for EspG5 binding, the critical determinants of PE–PPE binding specificity depend on the coupling of multiple complementary residues positioned along the helix bundle, as well as the helical conformations of PE and PPE. Helical bending will determine whether these complementary residues are brought together for salt bridge and hydrogen bond formation and consequent PE–PPE interaction. On the other hand, analysis of the molecular surfaces of PE81–99–PPE151–194 and PE25–PPE41 revealed differences in the electrostatic surfaces (supplemental Fig. S9). PE81–99–PPE151–194 displays a relatively more hydrophobic and negatively charged surface, whereas PE25–PPE41 contains positively charged patches on each face of the four-helix bundle. It appears that the structural and functional properties of each PE–PPE protein is shaped by its unique electrostatic surface and extent of helical bending. However, the importance of the C-terminal domains of PE–PPE proteins cannot be excluded. Currently, PE–PPE structures are available for complexes in sublineage III (PE25–PPE41) and sublineage IV (PE8–PPE15 in this study). Therefore, structural solutions of other PE–PPE complexes, particularly those of the most recently evolved sublineage V, will provide a more comprehensive understanding of the evolution of this distinct protein family.

Experimental procedures

Plasmid construction

Full-length and truncated versions of PE8 and PPE15, and EspG5 were amplified from M. tuberculosis strain H37Rv genomic DNA (ATCC). PE8 and PE81–99 were cloned into...
expression vector pGEX-6p-1 (GE Healthcare) via BamHI and SalI sites. EspG5 and PPE151–194 were cloned into vector pAC28 via NdeI and EcoRI and NdeI and BamHI sites, respectively (37). All mutations were introduced by using the QuikChange site-directed mutagenesis kit (Stratagene Corp., La Jolla, CA). All plasmid constructs obtained were confirmed by a DNA sequencing service (BGI) and then subjected to protein expression in *Escherichia coli* strain. 

**Protein expression and purification**

The recombinant PE81–99–PPE151–194 were co-expressed in *E. coli* strain BL21 (DE3), whereas the pAC-EspG5 was expressed individually. Transformed bacteria were grown at 37 °C to an OD600 of 0.4–0.6. Protein expression was then induced by 0.4 mM isopropyl-D-thiogalactopyranoside at 20 °C for 16–20 h. Cells expressing PE81–99–PPE151–194 and EspG5 were harvested and co-lysed with sonication in buffer of 20 mM HEPES, pH 7.5, 300 mM NaCl, 5 mM DTT, 5% glycerol. Lysate was cleared by centrifugation at 48,384 g for 1 h, and the lysate containing EspG5–PE81–99–PPE151–194 were mixed with Ni-NTA agarose (Macherey Nagel). After cleavage of the fusion tag from the protein with PreScission protease overnight at 4 °C, the proteins were eluted in lysis buffer supplemented with 50 mM l-arginine and further purified using Mono Q 5/50 GL ion exchange column (GE Healthcare) and Superdex 200 (GE Healthcare) size-exclusion column. Purified protein complex containing EspG5–PE81–99–PPE151–194 were pooled and concentrated in buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl for crystallization trials.

**Pulldown assay**

For PE8–PPE15 interaction studies, GST-tagged PE81–99 and His-tagged PPE15 or PPE15 mutants were co-expressed and lysed in buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl, 5 mM DTT, and 5% glycerol. Clear lysate was mixed with glutathione agarose 4B beads (Macherey Nagel) and incubated for 2 h, followed by washing with lysis buffer for 8 times. Input material and the beads were boiled with SDS loading dye and analyzed by SDS-PAGE. For Western blotting detection, PPE15 or PPE15 mutants were probed with primary anti-His antibody (1:5000) (GE Healthcare). Same procedures were applied for PE25-PPE15 interaction analysis, but lysate containing co-expressed His-tagged PPE15 and GST-tagged PE25 or PE25 mutants were used. For nickel pull down, GST-tagged PE8 or -PE81–99 and His-tagged PPE15 or PPE151–194 were co-expressed and lysed in buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl, 20 mM imidazole, and 5% glycerol. Clear lysate was mixed with Ni-NTA agarose (Macherey Nagel) and incubated for 1 h, followed by washing with lysis buffer eight times. Pulldown products were analyzed by SDS-PAGE. All experiments were performed in triplicate.

**Size-exclusion chromatography/static light scattering (SEC/SLS)**

The purified protein complex EspG5–PE81–99–PPE151–194 was injected into Superdex 200 analytical (GE Healthcare) size-exclusion column pre-equilibrated with buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl. The experiments were performed at a predetermined temperature of 25 °C. Eluted protein from gel filtration was directed into a miniDawn light scattering detector and an Optilab DSP refractometer (Wyatt Technologies). The data were analyzed using the software ASTRA.

**Analytical ultracentrifugation**

Analytical ultracentrifugation experiments were performed using a Beckman proteomeLab XL-I analytical ultracentrifuge. The sample at a concentration of 1.6 AU absorbance at A280 was spun using at rotor 60-Ti at a speed of 42,000 rpm at 16 °C for 12 h. The data were collected at 280 nm in a continuous mode with a scan range from 6.05 to 7.20 cm. The data were processed according to continuous sedimentation coefficient distribution model using Sedfit (38) to determine the sedimentation coefficients.

**Crystallization and structure determination**

The EspG5–PE81–99–PPE151–194 crystals were grown by using the sitting drop vapor diffusion method. Crystals were obtained from optimized conditions containing 200 mM NaCl, 100 mM Tris, pH 8.5, 25% (w/v) PEG3350 after incubation at 16 °C for 4 days. For data collection, crystals were transferred to cryo protectant with 20% glycerol and immediately frozen under liquid nitrogen. X-ray diffraction data were collected at 100 K at Beamline 13B1 of the National Synchrotron Radiation Research Center in Taiwan. A 2.9 Å complete data set was processed by the iMosflm (39). EspG5–PE81–99–PPE151–194 complex crystal belongs to the space group of P2₁2₁2₁ with unit cell dimensions a = 54.74 Å, b = 69.96 Å, and c = 203.55 Å, and there is one complex per asymmetric unit. Phase determination was solved by molecular replacement using *phenix.*mr_rosetta* (40, 41). Subsequent iterative refinement with the *phenix.*refine* and manual model inspection and rebuilding with Coot (42) resulted in final Rwork/Rfree values of 21.33%/26.24%. A summary of X-ray data collection and model refinement statistics is shown in Table 1. The molecular graphics images were pro-
duced with PyMOL. The protein coordinates were submitted to Protein Data Bank with PDB code 5XF5.

**Sequence analysis of PE and PPE proteins**

All the sequence alignments were generated using Clustal Omega (43) and rendered by the ESPript server (44).

**Microscale thermophoresis**

EspG5, PE81–99–PPE151–194 and PE25–PPE41 were purified using affinity chromatography followed by gel filtration chromatography with final buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl. The microscale thermophoresis experiments were conducted using the Monolith NT.115 instrument (NanoTemper Technologies). In brief, 20 μM PE81–99–PPE151–194 or PE25–PPE41 was fluorescently labeled using the NanoTemper protein labeling kit RED-NHS (Amine Reactive). Labeled PE81–99–PPE151–194 or PE25–PPE41 were diluted to 0.4 μM and then mixed with 0.3–20 μM EspG5 in a final buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl. The reactions were set at 25 °C with 40% microscale thermophoresis power for 5-s/30-s laser off/on/off times. The microscale thermophoresis data were analyzed to obtain the dissociation constant $K_d$ by using the software MO.Affinity Analysis.

**Yeast two-hybrid screen**

The yeast two-hybrid screen was performed twice according to the Matchmaker™ Gold yeast two-hybrid manual (Clontech). Briefly, recombinant pGBK7-T7 DNA-BD and pGADT7 DNA-AD plasmids were used to co-transform into Saccharomyces cerevisiae. Positive transformants having bait–prey interaction were selected on selective SD/−Trp/(DDO) agar plates, the colonies on the DDO plates were patched onto SD/−His/(−Trp/X/α-Gal)/−Ade/−Leu/(−Trp/X/α-Gal)/−His/(−Leu/(−Trp/X/α-Gal) agar plates (Clontech). The plate was incubated at 30 °C for 3 days. Those pairs detected both on double and quadruple selection plates were identified as potential interaction pairs. PE25–PPE41 was used as a positive control.

**Comparative modeling of other PE–PPE complexes**

The structures of PE27–PPE43, PE13–PPE18, and PE32–PPE65 complexes were predicted by homology modeling using our solved crystal structure of EspG5–PE81–99–PPE151–194 as the template by program MODELLER9.18 (36). Sequences of individual PE and PPE proteins were aligned to PE8 and PPE15 by the program Clustal Omega (43). The sequence alignment was edited interactively using the program Chimera.

**References**

4. Kaufmann, S. H. (2016) How can we improve the existing vaccine for tuberculosis to combat the growing number of multi-resistant strains?
Recognition specificity of PE–PPE proteins


