Supporting Information

SI Materials and Methods

Cloning, expression and purification of EspGs. Wild-type *M. tuberculosis* (*Mtb*) $EspG_3$ ($EspG_{3Mt}$), $EspG_{5Mt}$, and *M. smegmatis* (*Msmeg*) $EspG_3$ ($EspG_{3Ms}$) were amplified from strain Erdman (1) (Mtb) or strain mc^2-155 (*Msmeg*) genomic DNA and cloned into pDCE467 (a derivative of pET21a with modifications to the MCS, tags, and with an inserted loxP site to facilitate plasmid fusions for multi-protein expression (2)) using Gibson assembly (3). The resulting constructs had an N-terminal 6xHis tag followed by a TEV cleavage site. For binding experiments, an additional BirA tag was fused to the C-terminus, allowing for site-specific biotinylation and immobilization of the resulting proteins using streptavidin.

Plasmids for EspG expression were transformed into *E. coli* strains Rosetta 2 (EMD Millipore) for protein production. Cultures were grown shaking at 37° C to an optical density at 600 nm (OD_{600}) of approximately 0.8, then shifted to 15° C, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and incubated overnight (~16-24 h). Cells were harvested by centrifugation at 3500g for 20 minutes, then resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole pH 8.0, 1 mM dithiothreitol [DTT]) in 1/25 the culture volume (i.e., 40 mL / 1 L culture). Cell slurries were lysed by two passes through an EmulsiFlex C-3 cell disruptor (Avestin). The resulting lysates were cleared by centrifugation at 38,000g for 1 hr, and the supernatants were transferred to fresh tubes. A 1 mL bed volume of Ni-NTA agarose resin (Qiagen) was added to cleared lysates, and incubated for 30-60 min at 4C. After washing with ~100 mL of lysis buffer, bound protein was eluted from the resin using 50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole pH 8.0, 1 mM DTT, and 1-2 mM ETDA was added to the eluted fractions. Where specified, 6xHis tags were removed by digestion with TEV (~1:50 ratio of target:TEV, by mass) overnight at 4° C, followed by removal of TEV (6xHis tagged) and uncleaved target by incubation with Ni-NTA resin. After concentration to a final volume of 2 mL, proteins samples were further purified by gel filtration, using a Superdex 200 16/60 column

equilibrated in 10 mM Tris pH 8.0, 150mM NaCl, 1 mM DTT. Fractions containing monomeric EspGs were pooled and concentrated for crystallization trials.

Expression and purification of selenomethionine labeled $EspG_{3Ms}$. $EspG_{3Ms}$ protein derivatized with selenomethionine for phasing was produced in Rosetta 2 cells grown in M9 minimal media supplemented with selenomethionine while inhibiting methionine biosynthesis, essentially as previously described (4). Briefly, cultures were grown shaking at 37° C to an OD_{600} of approximately 0.6, then 100 mg/L lysine, 100 mg/L phenylalanine, 100 mg/L threonine, 50 mg/L isoleucine, 50 mg/L leucine, 50 mg/L valine, and 50 mg/L selenomethionine were added as solids. After 15 min, the culture was shifted to 15° C, induced with 1 mM IPTG, and incubated for ~24 hrs before harvesting and processing as described above for native proteins.

Cloning, expression and purification of PE25-PPE41 heterodimer. Wild-type *Mtb PE25* and *PPE41* genes were amplified from genomic DNA and cloned into pDCE467 and pDCE431 (similar to pDCE467, but with an R6Kγ origin of replication and a kanamycin resistance cassette), respectively, using Gibson assembly. The resulting plasmids were fused using Cre and transformed into Rosetta 2 cells for expression, selecting for both ampicillin and kanamycin resistance. PE25-PPE41 protein was expressed and purified essentially as described for EspGs, above, except that cultures were incubated at 37° C for ~4 hr during IPTG induction.

Cloning, expression and purification of $EspG_{5Mt}$-PE25-PPE41 ternary complex. Initial attempts to express and purify $EspG_{5Mt}$ resulted in a ~1:1 mixture of full-length and clipped/truncated protein, and both forms bound to PE25-PPE41. Similarly, co-expression of all three components yielded complex containing both full-length and clipped $EspG_{5Mt}$, and only low quality crystals were obtained. In an attempt to stabilize the $EspG_{5Mt}$-PE25-PPE41 ternary complex and possibly reduce

EspG clipping, we created a $EspG_{5Mt}$-PE25 fusion. As PE25 binds to PPE41 with very high affinity, this should recruit $EspG_{5Mt}$ to PPE41 via its association with PE25 and drive ternary complex formation. To this end, PE25 was fused to the C-terminus of $EspG_{5Mt}$ by overlap PCR, connected by a long flexible linker commonly used to join the V_H and V_L domains of an scFv (SSGGGSGGGSGGGS). The resulting cassette was cloned into pDCE467, fused to pDCE431-PPE41 using Cre, and transformed into Rosetta 2 cells. The $EspG_{5Mt}$-PE25-PPE41 complex was expressed and purified as described for EspGs, above. However, the apparent molecular weight of resulting protein on gel filtration was approximately double the expected size of ~76 kDa, suggesting that the predominant form of the chimeric protein is a dimer-of-trimers. In contrast, samples prepared by mixing purified PE25-PPE41 with $EspG_{5Mt}$ yielded a ternary complex of the expected size (~76 kDa), suggesting that the $EspG_{5Mt}$-PE25 fusion has resulted in a pair of domain-swapped complexes held together via the $EspG_{5Mt}$-PE25 linker. Such domain swapping is common in fusion proteins such as scFvs, but generally has little impact on structure or function. As the resulting protein was more homogenous with little apparent clipping in $EspG_{5Mt}$, we pooled fractions containing the domain swapped dimer and concentrated the protein for crystallization trials.

Structure determination. Purified $EspG_{3Mt}$ (28 mg/mL), $EspG_{3Ms}$ (25 mg/mL, selenomethionine derivatized), PE25-PPE41 heterodimer (10 mg/mL), and $EspG_{5Mt}$-PE25-PPE41 ternary complex (20 mg/mL) in 10 mM Tris 8.0, 150 mM NaCl . Crystallization trials using the JCSG core I-IV were performed using a Mosquito liquid handling system, using 100 nL protein + 100 nL reservoir solution per drop. The crystals used for data collection and structure determination were grown from the following conditions: $EspG_{3Mt}$ (0.2 M magnesium sulfate, 20% (w/v) polyethylene glycol [PEG] 3350), $EspG_{3Ms}$ (0.2 M lithium sulfate, 0.1 M phosphate-citrate pH 4.2, 10% (v/v) isopropanol), PE25-PPE41 heterodimer (0.1 M sodium acetate pH 4.5, 5% (w/v) PEG 1000, 50% (v/v) ethylene glycol), and $EspG_{5Mt}$-PE25-PPE41 (0.1 M HEPES pH 6.5, 5% (w/v) PEG 8000). Crystals

were flash cooled by plunging in liquid nitrogen after cryoprotection in the reservoir solution supplemented with the following: $EspG_{3Mt}$ (18% (v/v) ethylene glycol), $EspG_{3Ms}$ (30% (v/v) ethylene glycol), PE25-PPE41 heterodimer (no additional cryoprotectant), and $EspG_{5Mt}$-PE25-PPE41 (33% (v/v) ethylene glycol).

All diffraction data were collected at the Advanced Light Source (ALS) beamline 8.3.1. As no significant similarity to proteins of known structure could be identified, phase information for $EspG_{3Ms}$ was obtained from selenomethinone-derivatized crystals using multiwavelength anomalous dispersion (MAD) (5). A two wavelength MAD dataset was collected using the inverse beam approach with 1 degree wedges at a high energy remote wavelength (λ = 0.957 Å, data to 3.20 Å resolution) and a compromise wavelength halfway between the peak and inflection points (λ = 0.980 Å, data to 3.25 Å resolution; data collection strategy devised by J. Holton). The resulting diffraction data set was indexed in spacegroup $P3_221$ and integrated and scaled using the XDS package (6). Substructure solution, phasing, and initial automated model building were carried out using the SHELX C/D/E pipeline (7), which yielded a high quality, easily interpretable map to guide further model building and refinement (Fig. S1).

Diffraction data sets for $EspG_{3Mt}$ ($C222_1$, 2.85 Å resolution), the PE25-PPE41 heterodimer ($P222_1$, 1.95 Å resolution), and the $EspG_{5Mt}$-PE25-PPE41 ternary complex ($P6_122$, 2.45 Å resolution) were indexed, integrated, and scaled and merged using XDS (6). $EspG_{3Mt}$ and PE25-PPE41 were solved by molecular replacement using Phaser (8) with $EspG_{3Ms}$ and a previous PE25-PPE41 structure (PDB code: 2G38 (9)) as search models, respectively. For the $EspG_{5Mt}$-PE25-PPE41 ternary complex, we were unable to find $EspG_{5Mt}$ using either $EspG_{3Mt}$ or $EspG_{3Ms}$ as a search model. However, molecular replacement of a PE25-PPE41 dimer was readily achieved (~50% of the mass in the asymmetric unit), and after rigid body and restrained refinement, additional patchy density was observed for $EspG_{5Mt}$, most clearly defined for the α1-α2-α3 and α1'-α2'-α3' bundles (Fig. S2). Due to the asymmetry of the α1-α2-α3 and α1'-α2'-α3' bundles, EspG could only be fit into the map in one

orientation. A polyalanine model derived from $EspG_{3Ms}$ was manually fit into the map, and the curvature of the β sheet and position of the helices was manually adjusted in Coot (10), and regions of poor agreement were pruned from the model.

All structures were adjusted using Coot (10) and refined using Phenix (11), including simulated annealing. The final coordinates were validated using the JCSG Quality Control Server v3.0, which includes Molprobity (12) (publicly available at http://smb.slac.stanford.edu/jcsg/QC/). Data collection and refinement statistics for all structures are summarized in Tables S1 and S2.

Structural and bioinformatic analyses. Hydrogen bonds and van der Waals' contacts were calculated using HBPLUS (13) and CONTACSYM (14), respectively. Surface area buried upon binding was calculated with MS (15) and the PISA server (16). MacPyMol (DeLano Scientific) and UCSF Chimera (17) was used to render structure figures and for general manipulations. The DALI (18) and CATHEDRAL (19) servers were used to search for similar protein structures. Voronoia (20) was used to identify packing defects at the $EspG_{5Mt}$-PPE41 interface. PSI-BLAST (21) was used to identify all PPE proteins encode in the Mtb genome (Erdman strain) (1, 22). EspG sequences were identified using a PSI-BLAST (21) search of all mycobacterial proteins in the nr database, initiated with $EspG_{3Mt}$. After clustering into orthologous groups and removing identical sequences, we identified 35 $EspG_1$, 27 $EspG_2$, 66 $EspG_3$, and 17 $EspG_5$ unique sequences. Sequences were aligned using MUSCLE (23) and analyzed using custom scripts (available from the authors upon request). Sequence alignments were visualized with Jalview (24).

Expression and purification of BirA. *E. coli* biotin ligase (BirA enzyme) was expressed and purified in a manner similar to previous reports (25), but with an N-terminal His tag, essentially as previously described (26). Briefly, a pET21a derivative containing the *E. coli birA* gene (pDCE095) was transformed into BL21(DE3) cells, grown in shake flasks in low salt LB medium at 37°C to an

OD_{600} of ~0.7, then shifted to 23°C and induced with the addition of IPTG to a final concentration of 1mM. The culture was incubated at 23°C for ~16 hours after induction. The cells were lysed and homogenized by two passes through an EmulsiFlex C-3 cell disruptor, and BirA was purified by Ni-affinity (NiNTA resin, Qiagen), anion exchange (MonoQ column, GE Healthcare), and gel filtration. Purified BirA protein was concentrated down to 5mg/mL in 50mM Tris, pH 7.5, 200mM potassium chloride, 5% glycerol, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Biotinylation of EspGs for binding studies. After expression and NiNTA purification as described above, EspGs for binding studies were concentrated down to ~2-5mg/mL total protein. The EspGs were biotinylated by the addition of 25µg BirA enzyme/mg total protein, in a buffer of the following composition: 100mM Tris pH 8.0, 10mM ATP, 10mM MgOAc, 50µM biotin, with less than 50mM NaCl. The biotinylation reactions were incubated at overnight at 4 °C. Biotinylated EspGs were purified by size exclusion chromatography, and concentrated to ~5-20 mg/mL.

K_d determination and EspG binding specificity. K_d's were determined by bio-layer interferometry using an Octet Red instrument (ForteBio, Inc.). Biotinylated EspGs, at ~50 µg/mL in 1x kinetics buffer (1x PBS, pH 7.4, 0.01% BSA, and 0.002% Tween 20) were loaded onto streptavidin-coated biosensors and incubated with varying concentrations of PE25-PPE41 in solution. All binding data were collected at 30°C. The experiments comprised 5 steps: 1. Baseline acquisition (60 s); 2. HA loading onto sensor (600 s); 3. Second baseline acquisition (180-600 s); 4. Association of CR8020 for the measurement of k_{on} (180-600 s); and 5. Dissociation of CR8020 for the measurement of k_{off} (180-600 s). 7 concentrations of PE25-PPE41 were used, with the highest concentration being 50 nM. Baseline and dissociation steps were carried out in buffer only. The ratio of k_{on} to k_{off} determines the K_d reported here. To assess the specificity of EspG binding to PE25-PPE41 and the PPE chimeras, the same fixed concentration of the analyzed PPE was tested for binding to sensors loaded with X, Y, Z, or

an unloaded sensor (negative control). Little non-specific binding of the chimeras to the uncoated sensor was observed. All binding traces are reported in Fig. S11.

Yeast 2-hybrid protein-protein interaction assay. Genes encoding full-length $EccA_{5Mt}$, $EspG_{5Mt}$, PPE41, $MycP_{5Mt}$, ESAT-6, CFP-10, and the N-terminal 53 residues of $EccB_{5Mt}$ were cloned into bait or prey vectors (pJSC401 and pEG202), resulting in the fusion of either a LexA DNA binding domain or a B42 transactivation domain to the N-terminus of the protein of interest. pJSC401- and pEG202-derived plasmids were transformed into yeast strains W303A and EGY48/pSH18, respectively. After growth to saturation in selective media, matings were carried out on YPAD plates by spotting 2 uL of each culture on top of one another and incubating the plates at 30°C for 2 days. The matings were replica plated onto indicator plates (CSM -URA -HIS -TRP +Galactose +Raffinose +X-gal) and incubated at 30°C for 2 days. All data presented in Fig. S12 are derived from a single plate and single photograph, but were cropped to remove additional, unrelated expeirments.

References:

1. Miyoshi-Akiyama T, Matsumura K, Iwai H, Funatogawa K, Kirikae T (2012) Complete annotated genome sequence of Mycobacterium tuberculosis Erdman. *J Bacteriol* 194:2770.

2. Bieniossek C et al. (2009) Automated unrestricted multigene recombineering for multiprotein complex production. *Nat Methods* 6:447–450.

3. Gibson DG et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.

4. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229:105–124.

5. Guss JM et al. (1988) Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. *Science* 241:806–811.

6. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132.

7. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66:479–485.

8. McCoy AJ et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.

9. 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* 103:8060–8065.

10. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.

11. Afonine PV et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352–367.

12. Chen VB et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.

13. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793.

14. Sheriff S, Hendrickson WA, Smith JL (1987) Structure of myohemerythrin in the azidomet state at 1.7/1.3 A resolution. *J Mol Biol* 197:273–296.

15. Connolly ML (1983) Analytical molecular surface calculation. *J Appl Crystallogr* 16:548–558.

16. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797.

17. Pettersen EF et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.

18. Hasegawa H, Holm L (2009) Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* 19:341–348.

19. Redfern OC, Harrison A, Dallman T, Pearl FMG, Orengo CA (2007) CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. *PLoS Comput Biol* 3:e232.

20. Rother K, Hildebrand PW, Goede A, Gruening B, Preissner R (2009) Voronoia: analyzing packing in protein

structures. *Nucleic Acids Res* 37:D393–395.

21. Altschul SF et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

22. Cole ST et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393:537–544.

23. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.

24. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinforma Oxf Engl* 25:1189–1191.

25. Brown PH, Cronan JE, Grøtli M, Beckett D (2004) The biotin repressor: modulation of allostery by corepressor analogs. *J Mol Biol* 337:857–869.

26. Ekiert DC et al. (2011) A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* 333:843–850.

27. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336:1030–1033.

Fig. S1. Initial electron density map for EspG$_{3Ms}$ after phasing in SHELXE. Initial electron density map after phasing, contoured at 2σ (orange mesh). The backbone is easily traceable through most of the map, as illustrated by the fit to the final refined model (depicted as a blue Cα trace).

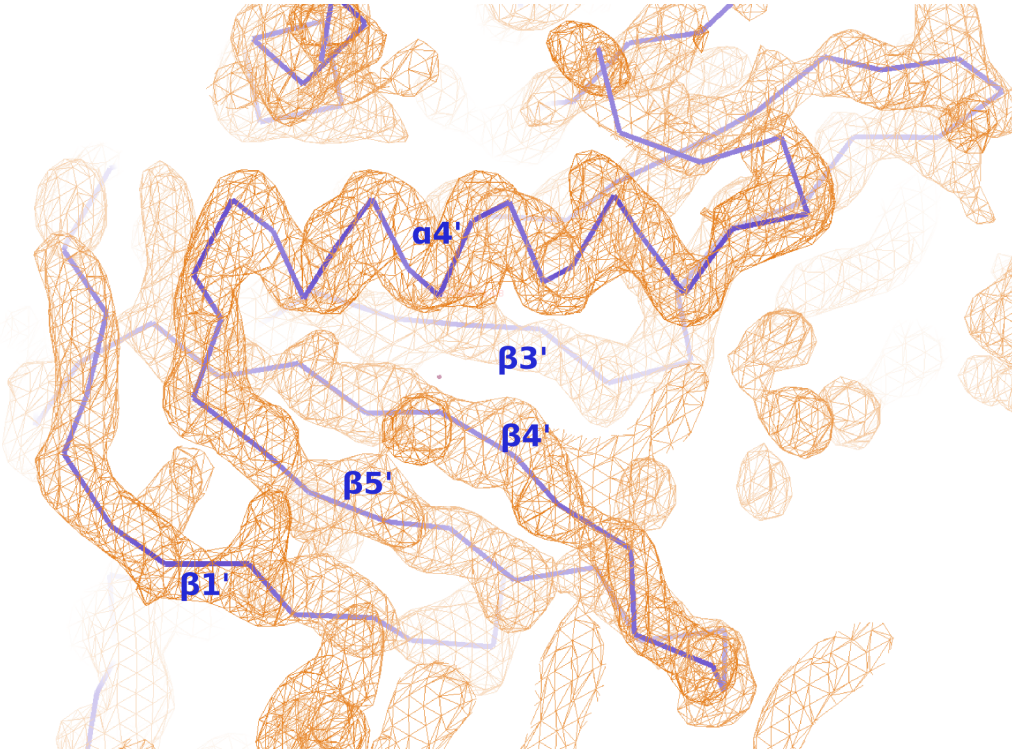

Fig. S2. Initial electron density of for $EspG_{5Mt}$ after molecular replacement of PE25-PPE41 in $EspG_{5Mt}$-PE25-PPE41 ternary complex. $2F_O$-F_C electron density map (orange mesh) for $EspG_{5Mt}$-PE25-PPE41 complex after molecular replacement and refinement of PE25-PPE41 only. Additional density was apparent for bound $EspG_{5Mt}$, allowing manual fitting of an EspG model (derived from $EspG_{3Ms}$). The final refined model is superimposed on the density as a blue Cα trace. a. strands from N-terminal half of β-sheet. b. Helices α1, α2, and α3 from N-terminal subdomain. c. Helices α1', α2', and α3' from C-terminal subdomain.

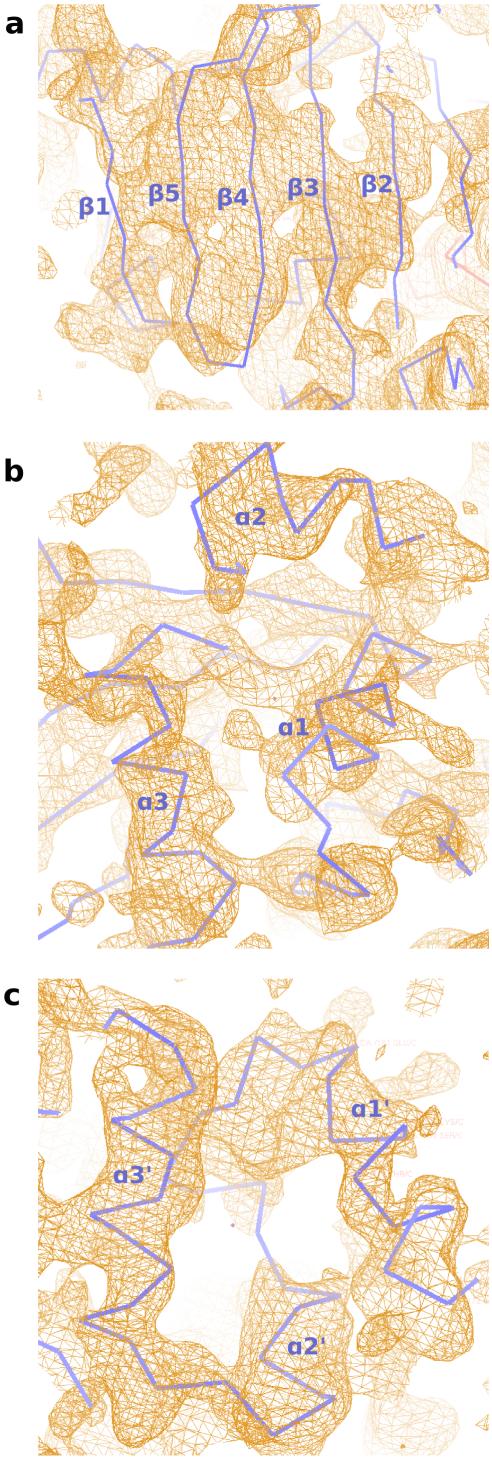

Fig. S3. Minor conformation changes occurring in PE25-PPE41 upon EspG binding. a. Ribbon representations of 2 unbound PE25-PPE41 conformations from 2G38 (blue and cyan) are superimposed on PE25-PPE41 from the $EspG_{5Mt}$-bound ternary complex (red). $EspG_{5Mt}$ binds at the bottom of the PPE as depicted here, but is omitted for clarity. b. Close-up view of helix-turn-helix EspG binding region at the tip of PPE41. Unbound PE25-PPE41 structures (blue and cyan) show modest flexibility in this region, and the EspG-bound conformation lies roughly between the two unbound states.

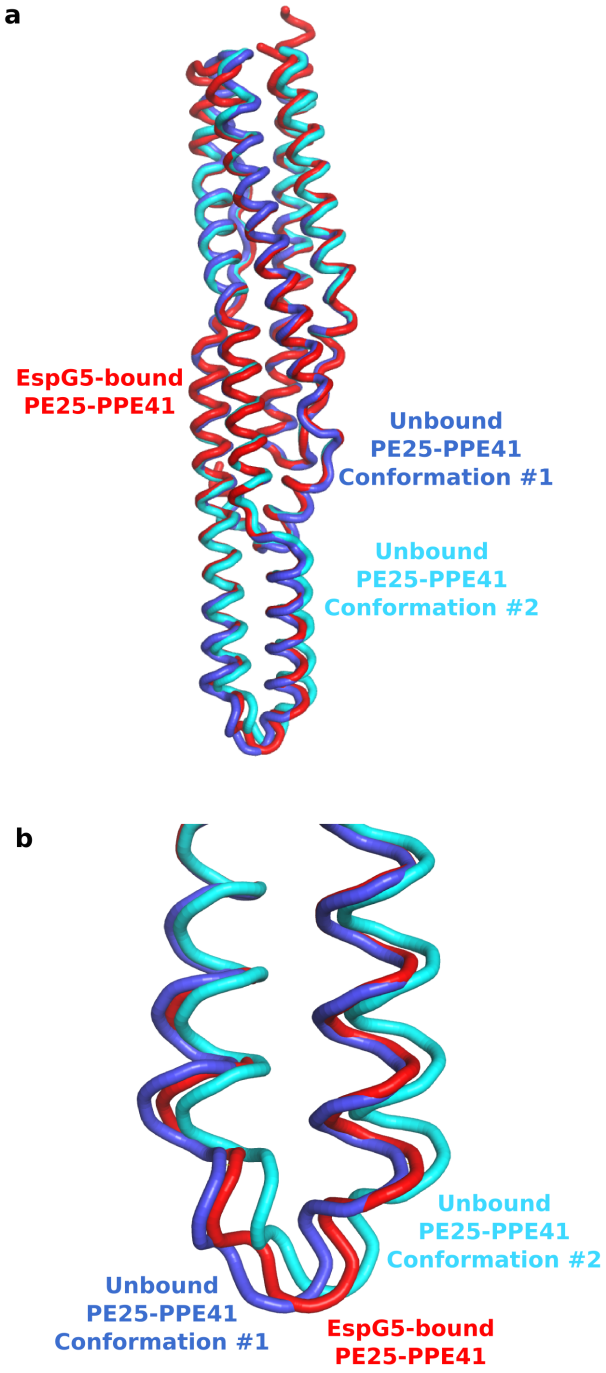

Fig. S4. Structure-based sequence alignment of EspG proteins from *Mtb*. $EspG_{1Mt}$, $EspG_{2Mt}$, $EspG_{3Mt}$, and $EspG_{5Mt}$ sequences were aligned with Muscle, then manually adjusted based upon the $EspG_{3Mt}$ and $EspG_{5Mt}$ crystal structures and $EspG_{1Mt}$ and $EspG_{2Mt}$ homology models. Approximate location of secondary structural elements are indicated below the alignment. Regions with good structural alignment between $EspG_{3Mt}$ and $EspG_{5Mt}$ are indicated with "*", while deviating or uncertain regions are indicated with "?". $EspG_{5Mt}$ residues that contact PPE41 are highlighted in yellow and in bold face.

```
* = Structurally aligned position
? = Structural variation or uncertainty
X = PPE contact residues on EspG5
```

```
                              ****************************????? ????????***********?? ?
EspG5       -----------MDQQSTRTDITVNVDGFWMLQALLDIRHVAPELRCRPYVST-DSNDWLNEHPGMAVMREQGIV-V
EspG2       -------------------MLTTTVDGLWVLQAVTGVEQTCPELGLRPLLPRLDTAERALRHPVAAELMAVGALDQ
EspG3       ------------MDATPNAVELTVDNAWFIAETIGAGTFPWVLAITMPYSD-AAQRGAFVDRQRDELTRMGLLSP
EspG1       MIMTGPSAAGRAGTADNVVGVEVTIDGMLVIADRLHLVDFPVTLGIRPNIPQ-EDLRDIVWEQVQRDLTAQGVLDL
                           -----                                ~~~~~~~~~~~~~~~~~~~~~~
                             β1 -------------                          α2
                                    α1


            ?????************************??????????????????????**********??????????
EspG5       NDAVNEQVAARMKVLAAPDLEVVALLSRGKLLYGVIDDENQPPGSRDIPDNEFRVVLARRG---------Q
EspG2       AGNADPMVREWLTVLLRRDLGLLVTIGVPGG-------------------EPTRAAICRFA---------T
EspG3       QGVINPAVADWIKVVCFPDRWLDLRYVGPASADGAC--------------ELLRGIVALRTGTGKTSNKTG
EspG1       HGEPQPTVAEMVETLGRPDRTLEGRWWRRDIGG-----------------VMVRFVVCRRG---------D
                ---------------                                 -----------          -
                     α3        ----------                          β3
                                  β2


            ******************??????????????????????????**********????????????????
EspG5       HWVSAVRVGNDITVDDVTVSD--------SASIAALVMDGLESIHHADPAAINAVNVPMEEMLEATKSWQESGF----
EspG2       WWVVLERHGNLVRLYPAGTASDEAGAGELVVGQVERLC------GVAEAAPLRPVTVDADELLHAVRDAGTLRSYLLS
EspG3       NGVVALRNAQLVTFTAMDIDD--------PRALVPILGVGL---AHRPPARFDEFSLPTRVGARADERLRS-------
EspG1       RHVIAARDGDMLVLQLVAPQVG-------LAGMVTAVL------GPAEPANVEPLTGVATELAE--------------
            ---------                    ----------------              ----------------
                β4    --------                 α4                -----      α1'
                        β5                                         β1'


            ??????????????????????????????????**********?????????????????
EspG5       NVFSGGDLRRMGI------SAATVAALGQALSDPA--AEVAVYARQYRD-----DAKGPSASVLSLKDGS
EspG2       QRLDVDQLQ----------MVTMAAD-------PTRSAHATLVALQAGVGPEKSARILVGDSTVAIVDTA
EspG3       GVPLGEVVDYLGIP-----ASARPVVESVFS--GPR-SYVEIVAGCNRD-----GRHTTTEVGLSIVDTS
EspG1       CTTASQLTQYGIAPASAR-VYAEIVG-------NPT-GWVEIVASQRHP-G---GTTTQTDAAAGVLDSK
            ~~~~~~~~~~                          -----------
                α2'             ~~----------~~        β2'          ------/\--------
                                     α3'                                 β3'


            *********???????*****************************?????????????????????
EspG5       GGRIALYQQARTAGSGEAWLAICPATPQLVQVGVKTVLDTLPYGE-WKTHSRV---------
EspG2       AGRICVESVTSGQ---RRYQVLSPGSRSDIGGAVQRLIRRLPAGDEWYSYRRVV--------
EspG3       AGRVLVSPSRAFDG--EWVSTFSPGTPFAIAVAIQTLTACLPDGQ-WFPGQRVSRDFSTQSS
EspG1       LGRLVSLPRRVGG---DLYGSFLPGTQQNLERALDGLLELLPAGA-WLDHTSDHAQASSRG-
            ---------~~              ---------------
                 β4'        ~~------       α4'
                               β5'
```

Fig. S5. Sequence and structural conservation/variation in the EspG "tongue" region. a. alignment of tongue region from *Mtb* EspG sequences. The $EspG_{2Mt}$ and $EspG_{5Mt}$ "tongues" are most similar to one another, and $EspG_1$ also shares some key sequence features. The $EspG_{3Mt}$ tongue appears to be the most divergent. b. Structural superposition of the "tongue" regions from $EspG_{3Mt}$ (blue), $EspG_{5Mt}$ (green), and $EspG_{3Ms}$ (pink). The conformation of the tongue is significantly different in $EspG_{5Mt}$, where it interacts extensively with PPE41.

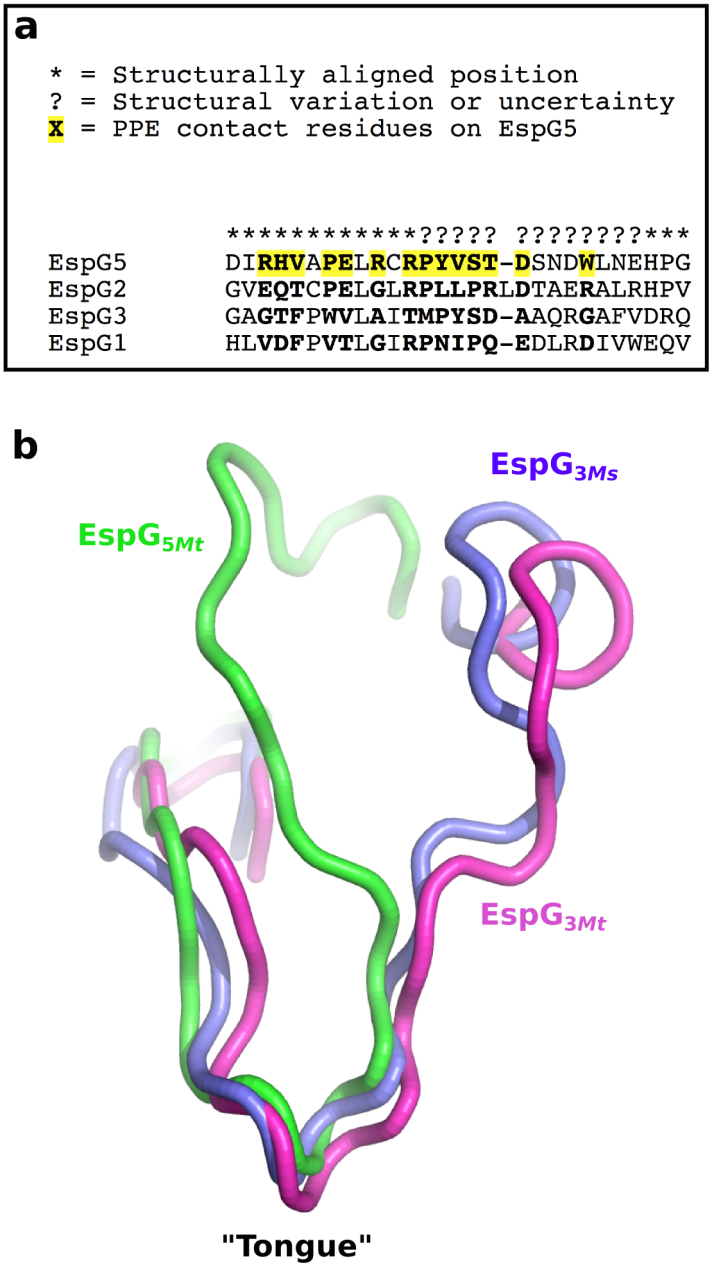

Fig. S6. Alignment of PPE proteins from Mtb. Alignment of the PPE domains from all PPE proteins identified in the Mtb Erdman genome using PSI-BLAST. The alignment was colored by sequence similarity (BLOSUM64 score) in Jalview. The PPE motif and EspG contact region are indicated. The sequences vary considerable across the domain, which is the best conserved part of each PPE protein. Even the PPE motif that gave rise to the family name is not strictly conserved.

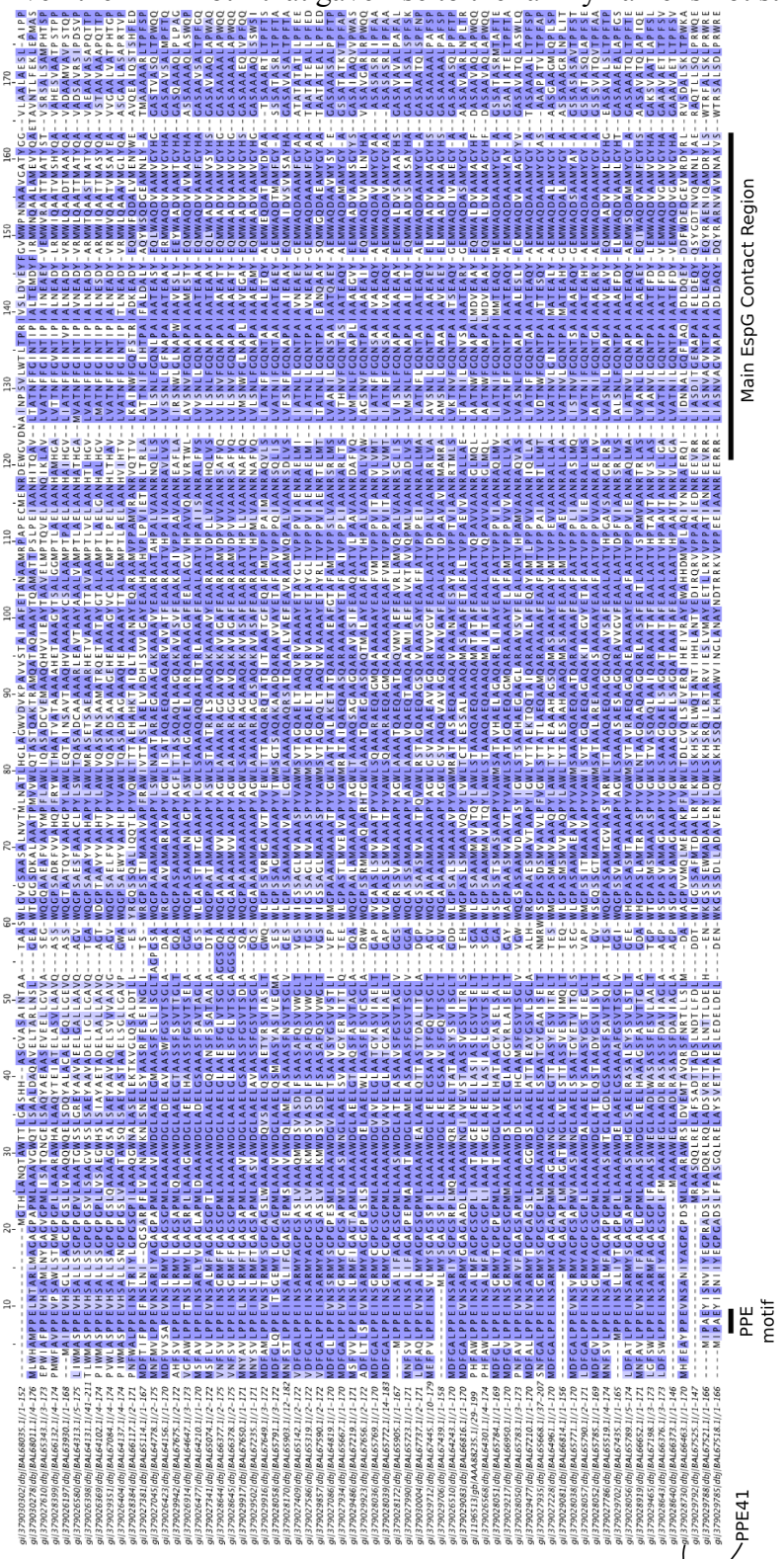

Fig. S7. Conserved residues in PPE helix-turn-helix motif suggest most PPEs adopt a similar conformation in the EspG binding region. a. Asn122 caps the α4 helix of most PPEs. b. Asn123 makes a loop-stabilizing hydrogen bond across the EspG-binding, helix-turn-helix motif at the tip of PPEs. In addition, the nearly invariant Gly126 at the tip of the helix-turn-helix motif adopts a positive phi angle, which is most accessible to Gly due to lack of a side chain.

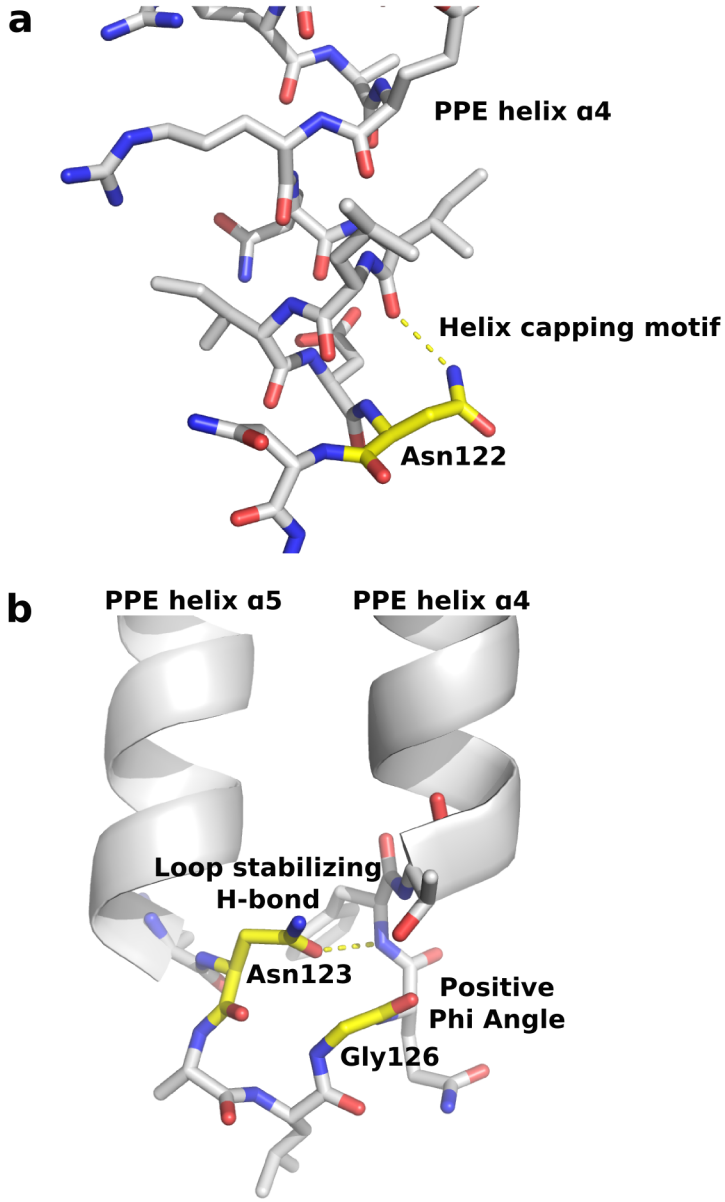

Fig. S8. Underpacking of the EspG$_{5Mt}$-PPE41 interface. Numerous packing defects at the EspG$_{5Mt}$-PPE41 interface (green spheres) leave additional space for larger side chains at several positions on PPE, allowing EspG$_{5Mt}$ to bind to a broader range of PPE protein sequences. Larger spheres and deeper green color indicate larger packing defects, while smaller spheres and whiter color indicate smaller defects. PE25 (peach), PPE41 (red), and EspG$_{5Mt}$ (purple) are shown in cartoon depiction.

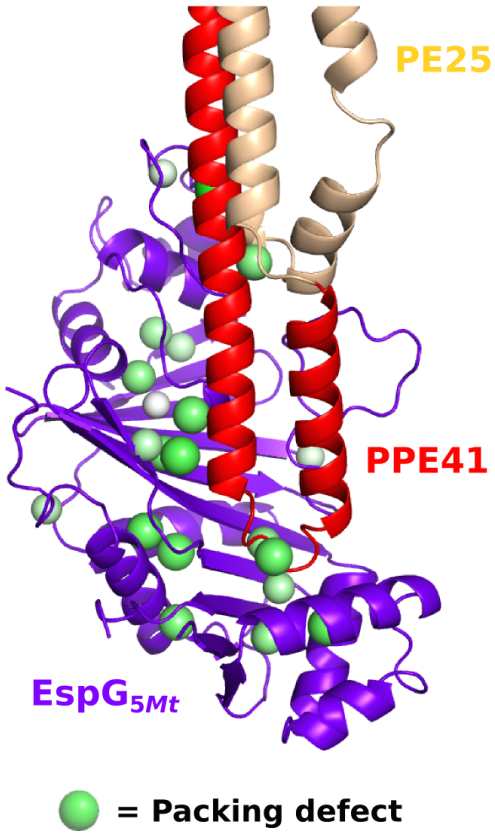

Fig. S9. Additional conserved EspG-interacting residues on PPE. a. Ile120 occupies a small pocket and makes a handful of hydrophobic contacts with Tyr91 and Val93 on $EspG_{5Mt}$, and most other PPEs have a small, apolar residue at this position. b. Thr129 is buried at the interface and makes a number of van der Waals contacts with $EspG_{5Mt}$, as well as a hydrogen bond to the PPE Gly126 backbone, which may help to stabilize the helix-turn-helix conformation at the PPE tip. Due to the tight packing around this residue, only small residues can be accommodated and ~97% of *Mtb* PPEs have a Thr, Ser, or Ala at this position, with a distinct preference for Thr/Ser to preserve the hydrogen bond to Gly126. c. While repacking of nearby side chains may allow larger residues, there is a strong bias towards Ala at PPE residue 133 (~86% of *Mtb* PPEs). d. Asp136 and Asp140 are buried at the PPE-EspG interface and may make salt bridges/electrostatic interactions with $EspG_{5Mt}$ residues Arg29/Arg104 and Arg112, respectively. A negatively charged Asp or Glu residue at position 136 and 140 is found in ~97% and ~42% of PPEs from *Mtb*, respectively. e. Gln127 makes 2 hydrogen bonds and several van der Waals contacts via the aliphatic chain residues, and ~71% of *Mtb* PPE proteins have Gln at his position, while most of the remaining sequences may retain the hydrophobic interaction with an Ile or Leu side chain.

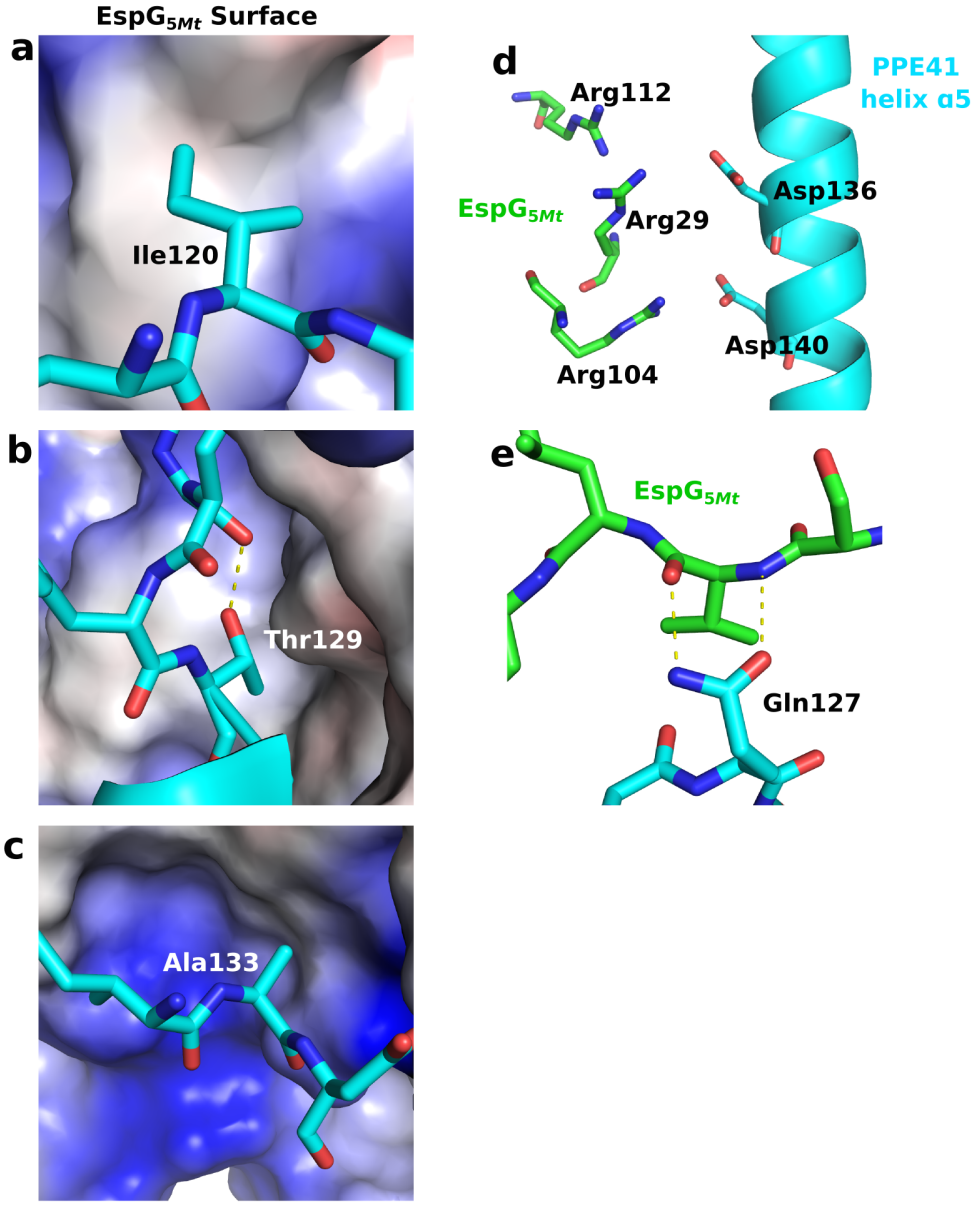

Fig. S10. Sequence alignment of PPE panel used for binding studies. a. Amino-acid sequences of the 8 representative PPE proteins selected binding experiments with EspGs. The selected PPE sequences ranged from 194 to 539 amino-acids in length, but were truncated to the region encoding the core PPE domain. Sequences are colored by BLOSUM64 conservation score and PPE41 positions that interact with $EspG_{5Mt}$ are indicated with *. b. Sequence alignment of chimeric PPE constructs used for binding experiments. The framework for all chimeras is derived from PPE41, with the major EspG binding region grafted from the PPEs of interest (from panel a). The graft region is indicated with a bar. Sequences are colored by BLOSUM64 conservation score and PPE41 positions that interact with $EspG_{5Mt}$ are indicated with *.

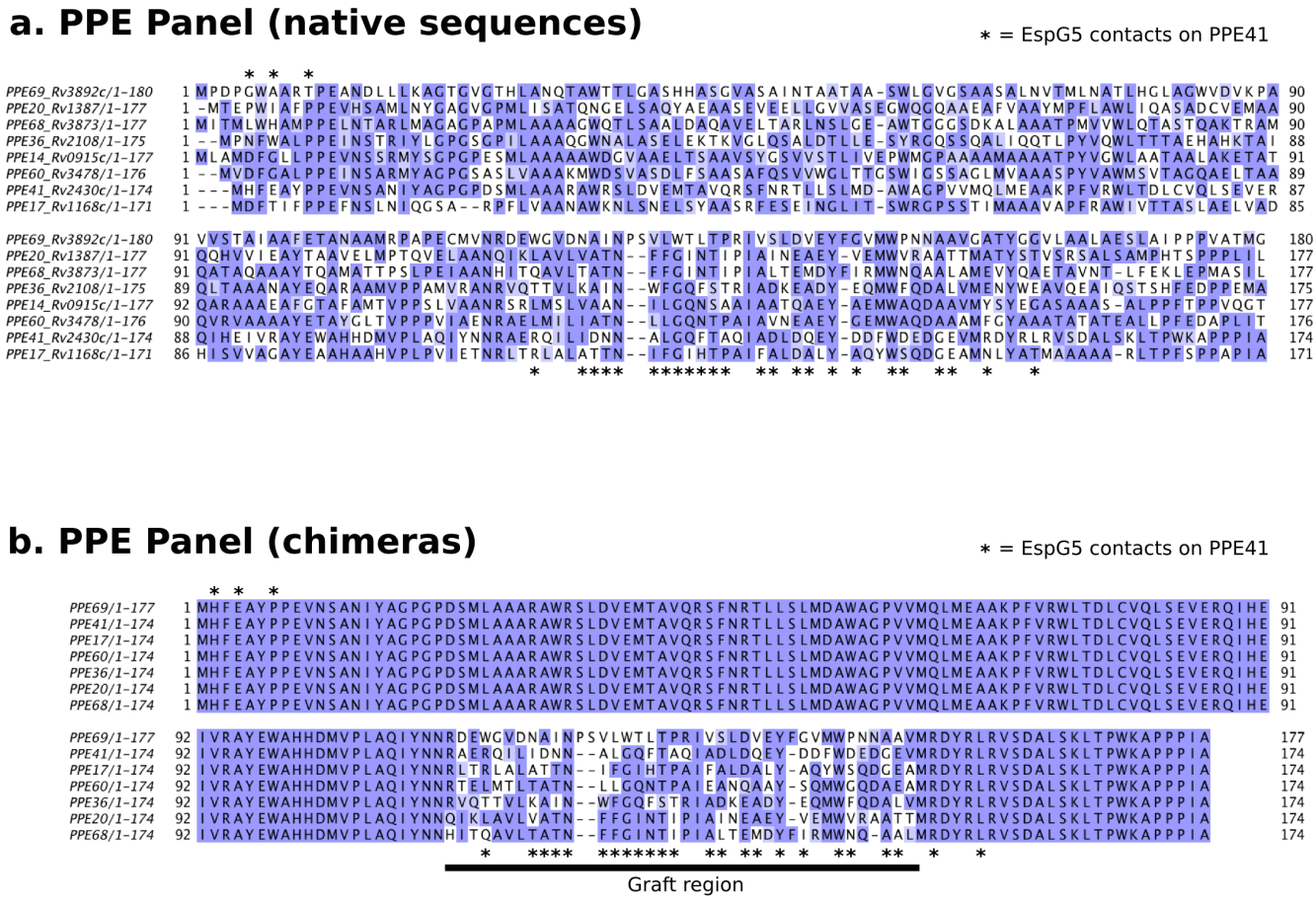

Fig. S11. Biolayer interferometry traces for binding of PPE chimeras to EspGs. Raw binding data for PPE41 and PPE chimeras (as indicated on graphs) to $EspG_{3Mt}$ (red traces), $EspG_{5Mt}$ (light blue traces), and $EspG_{3Ms}$ (green traces). As a negative control, binding of the chimeras to an identical empty sensor (no EspG protein loaded) was monitored (dark blue traces).

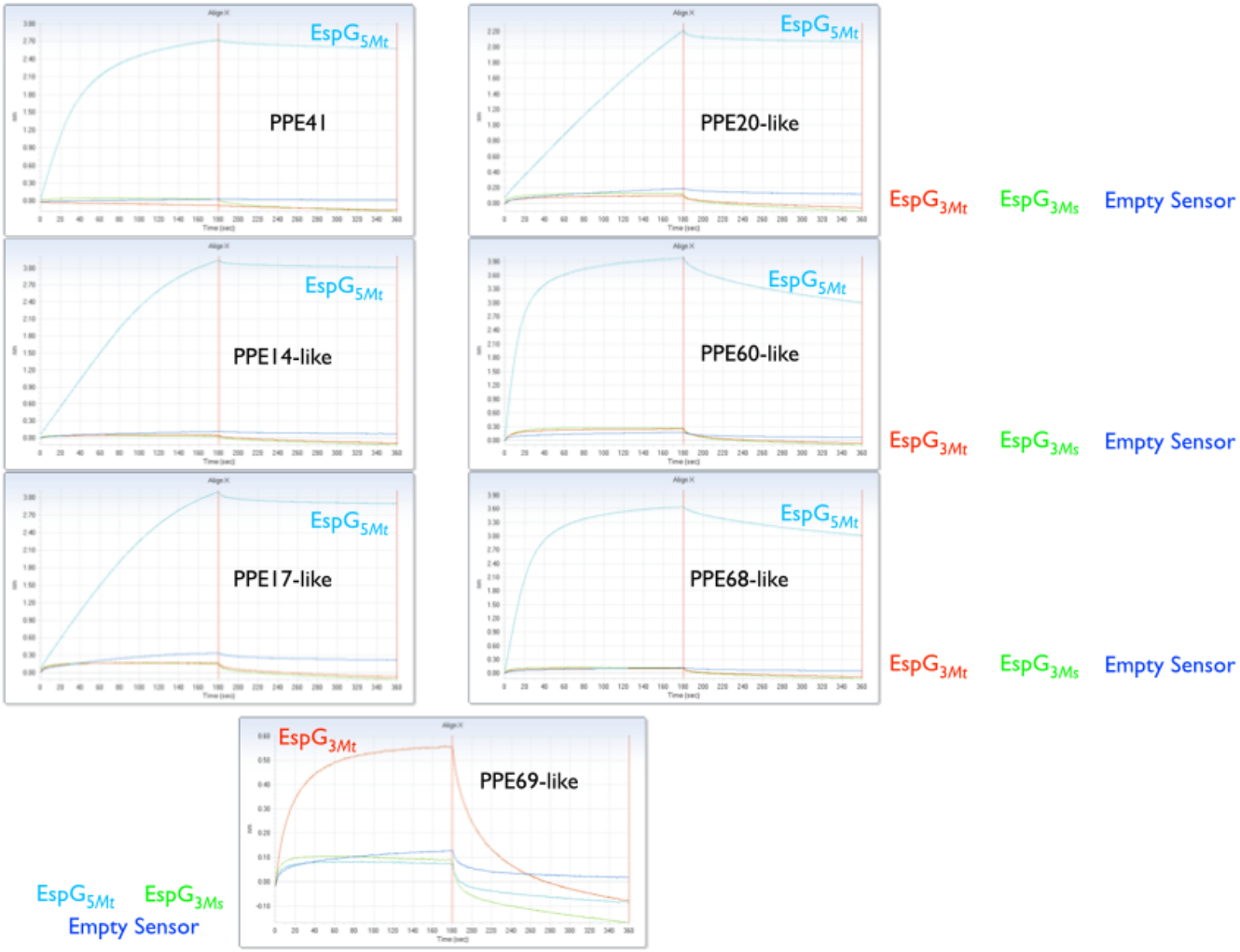

Fig. S12. EccA and EspG interact in a yeast 2-hybrid assay. $EccA_{5Mt}$ fused to a transcriptional activation domain (AD) and $EspG_{5Mt}$ fused to a DNA-binding domain (DBD) interact in a yeast 2-hybrid assay, leading to beta-galactosidase production and blue growth on indicator plates containing X-gal. In addition, known interactions between $EspG_{5Mt}$ and PPE41; ESAT-6 and CFP-10; and the weak homodimerization of ESAT-6 are detected. In contrast, no interaction is observed between $EccA_{5Mt}$ or $EspG_{5Mt}$ and negative control proteins ($MycP_{5Mt}$ and the N-terminal domain of $EccB_{5Mt}$, respectively).

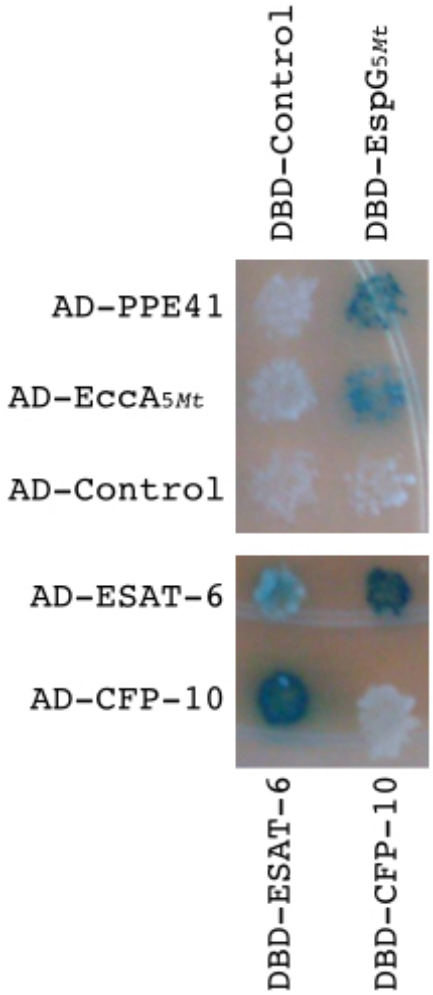

Fig. S13. Msmeg EspG$_3$-PE$_3$-PPE$_3$ form a stable complex on gel filtration. After co-expression of 6xHis-tagged PPE$_{3Ms}$, FLAG-tagged PE$_{3Ms}$, and Strep-tagged EspG$_{3Ms}$ in E. coli, 6xHis-tagged PPE$_{3Ms}$ was purified by immobilized metal affinity chromatography and the eluted protein was separated over a Superdex 200 16/60 gel filtration column. The complex eluted at the expected molecular weight for a 1:1:1 complex, and Western blotting against the epitope tags confirmed that all three components were present in the peak fraction.

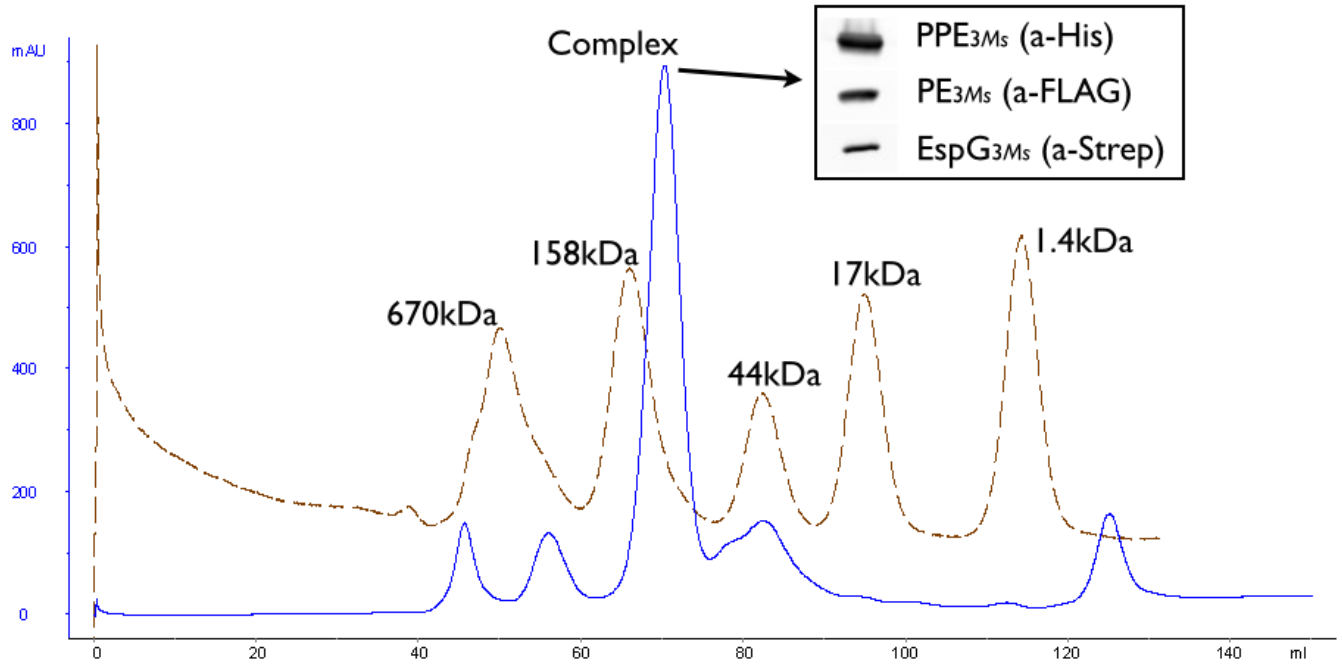

Table S1. Data collection and refinement statistics for crystal structures.

	EspG$_{3Mt}$	**EspG$_{3Ms}$**	**EspG$_{5Mt}$-PE25-PPE41**	**PE25-PPE41**
Data collection				
Space group	$C222_1$	$P3_221$	$P6_122$	$P222_1$
Cell dimensions				
a, *b*, *c* (Å)	75.0, 96.3, 94.5	95.1, 95.1, 171.4	138.6, 138.6, 169.7	41.0, 49.5, 284.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Wavelength (Å)	1.116	0.957	1.116	1.116
Resolution (Å)	50-2.85 (2.92-2.85)[1]	50-2.80 (2.87-2.80)[1]	50-2.45 (2.51-2.45)[1]	50-1.95 (2.00-1.95)[1]
Observations	58,890	497,430	3,124,112	301,689
Unique reflections	8,298	22,760	35,773	43,451
Redundancy	7.1 (7.4)	21.9 (22.5)	87.3 (89.3)	6.9 (7.1)
Completeness (%)	99.9 (100.0)	100.0 (100.0)	99.3 (98.9)	99.7 (99.9)
CC* [2]	1.00 (0.85)	1.00 (0.79)	1.00 (0.81)	1.00 (0.91)
CC1/2 [2]	1.00 (0.60)	1.00 (0.45)	1.00 (0.50)	1.00 (0.70)
$I / \sigma I$	10.8 (1.1)	19.8 (1.2)	26.1 (1.2)	12.1 (1.1)
R_{sym}	0.18 (1.56)	0.15 (3.15)	0.32 (8.70)	0.09 (1.94)
R_{pim}	0.07 (0.68)	0.03 (0.71)	0.03 (0.91)	0.04 (0.81)
Refinement				
Resolution (Å)	50-2.85	50-2.80	50-2.45	50-1.95
Reflections (work)	7,450	20,449	34,267	41,710
Reflections (free)	828	2,271	1,460	1,714
R_{work} (%) / R_{free} (%)	24.0 / 28.6	20.6 / 24.6	21.5 / 23.5	22.0 / 25.8
No. atoms				
Protein	2,015	4,029	4,262	4,070
Ligand/ion	5	10	0	0
Water	11	0	82	101
B-factors				
Protein	77.6	103.1	84.8	66.4
Ligand/ion	70.5	114.7	na	na
Water	59.7	na	62.3	56.2
R.m.s. deviations				
Bond lengths (Å)	0.002	0.004	0.004	0.007
Bond angles (°)	0.66	0.82	0.73	0.87
Ramachandran				
Favored	96.2	99.1	96.5	98.8
Outliers	0.4	0.0	0.4	0.0
PDB Code [3]	4W4I	4W4J	4W4L	4W4K

[1] Values in parentheses are for highest-resolution shell.

[2] See Karplus and Diederichs (27).

Table S2. Data collection and phasing statistics for $EspG_{3Ms}$ crystal structure.

	$EspG_{3Ms}$	
Data collection		
Space group	$P3_221$	
Cell dimensions		
a, *b*, *c* (Å)	95.2, 95.2, 171.5	
α, β, γ (°)	90.0, 90.0, 120.0	
	Peak	Remote
Wavelength (Å)	0.980	0.957
Resolution (Å)	50-3.25 (3.33-3.25) [1]	50-3.20 (3.28-3.20) [1]
Observations	320,163	670,399
Unique reflections	27,597	28,814
Redundancy	11.6 (11.7)	23.3 (23.5)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
CC* [2]	1.00 (0.89)	1.00 (0.91)
CC1/2 [2]	1.00 (0.48)	1.00 (0.53)
$I / \sigma I$	15.0 (1.3)	15.0 (1.3)
R_{sym}	0.16 (2.34)	0.26 (3.25)
R_{pim}	0.04 (0.54)	0.04 (0.52)
FOM	0.57	

[1] Values in parentheses are for highest-resolution shell.

[2] See Karplus and Diederichs (27).

Table S3. Conservation of Pro46 across EspG paralogs.

	% Pro46	% Pro near 46	Pro46 Present?	
			Mtb	*Msmeg*
All $EspG_1$	0	3	No	No
All $EspG_2$	100	100	Yes	na*
All $EspG_3$	2	11	No	No
All $EspG_5$	100	100	Yes	na*

* No ESX-2 or ESX-5 clusters in *Msmeg*.