Architecture of a lipid transport system for the bacterial outer membrane

Damian C. Ekiert[1,2,*,§], Gira Bhabha[2,§], Garrett Greenan[2], Sergey Ovchinnikov[3], Jeffery S. Cox[1,‡], Ronald D. Vale[2,4]

[1] Department of Microbiology and Immunology, The University of California, San Francisco, 600 16th St, San Francisco, CA 94158, USA.

[2] Department of Cellular and Molecular Pharmacology, The University of California, San Francisco, 600 16th St, San Francisco, CA 94158, USA.

[3] Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.

[4] Howard Hughes Medical Institute, The University of California, San Francisco, 600 16th St, San Francisco, CA 94158, USA.

[§] These authors contributed equally to this work.

[‡] Present Address: Department of Molecular and Cell Biology, The University of California, Berkeley, 388 Li Ka Shing Center, Berkeley, CA 94720, USA.

[*] To whom correspondence should be addressed. Email: damian.ekiert@ucsf.edu

1

One Sentence Summary:
Structural characterization of a widely-conserved transport system suggests a new model for lipid trafficking in double-membraned bacteria and organelles.

Abstract:
How phospholipids are trafficked between the bacterial inner and outer membranes through the intervening hydrophilic space of the periplasm is not known. Here we report that members of the mammalian cell entry (MCE) protein family, which were previously implicated in outer membrane function, form hexameric rings with a central hydrophobic channel that can mediate phospholipid transport. The *E. coli* MCE protein MlaD forms part of a large multi-protein complex in the inner membrane that includes an unconventional ABC transporter. MlaC, a soluble lipid-binding protein, interacts with MlaD and an outer membrane protein complex, suggesting that it ferries lipids between the two membranes. In contrast, the syringe-like architecture of a second *E. coli* MCE protein, PqiB, creates a continuous channel of sufficient length to potentially span the entire periplasmic space. This work reveals a system of highly conserved protein-based channels that translocate lipids between the inner and outer membranes of bacteria and some eukaryotic organelles.

The bacterial outer membrane (OM) serves a critical role as an environmental barrier, restricting the traffic of small molecules such as antibiotics into the cell (1, 2). While double-membranes are generally regarded as a hallmark of gram-negative species, such as *E. coli*, several gram-positive species are recognized as having a *bone fide* OM, including the important human pathogen *M. tuberculosis* (3). Mutations that weaken or disrupt OM integrity reduce virulence in pathogenic bacterial species (4–7) and increase their susceptibility to antibiotics and detergents (8–11), suggesting that targeting pathways important for OM biogenesis may be a fruitful approach for development of new therapies for bacterial disease. The OM of most gram-negative bacteria is asymmetric, with an outer leaflet rich in lipopolysaccharide (LPS) and an inner leaflet composed primarily of phospholipids (12–14). Decades of research have uncovered a detailed, but still evolving, picture of the pathways and mechanisms involved in the biosynthesis, transport, and insertion of LPS into the OM, as well as the insertion of OM porins and lipoproteins (1, 15, 16). Yet, it remains largely unknown how nascent phospholipids are transported between the inner membrane (IM) and OM.

Proteins of the MCE superfamily [originally thought to mediate Mammalian Cell Entry in *M. tuberculosis* (17)] are defined by the presence of one or more conserved MCE domains at the sequence level, but have no similarity to proteins of known structure or function. MCE proteins are ubiquitous among double-membraned bacteria (18) (Fig. 1A) and are also found in eukaryotic chloroplasts (19), which are also surrounded by a double membrane. In contrast, MCEs are absent in bacteria bounded by a single membrane. MCE proteins are important virulence factors in TB and other bacterial pathogens (20–27), and have been suggested to play a role in the transport of lipids (19, 28–32), cholesterol and steroids (20, 33, 34), and other hydrophobic molecules (35). Thus, MCE proteins are likely to be critically important for OM function, but their precise roles and mechanism of action remains unclear.
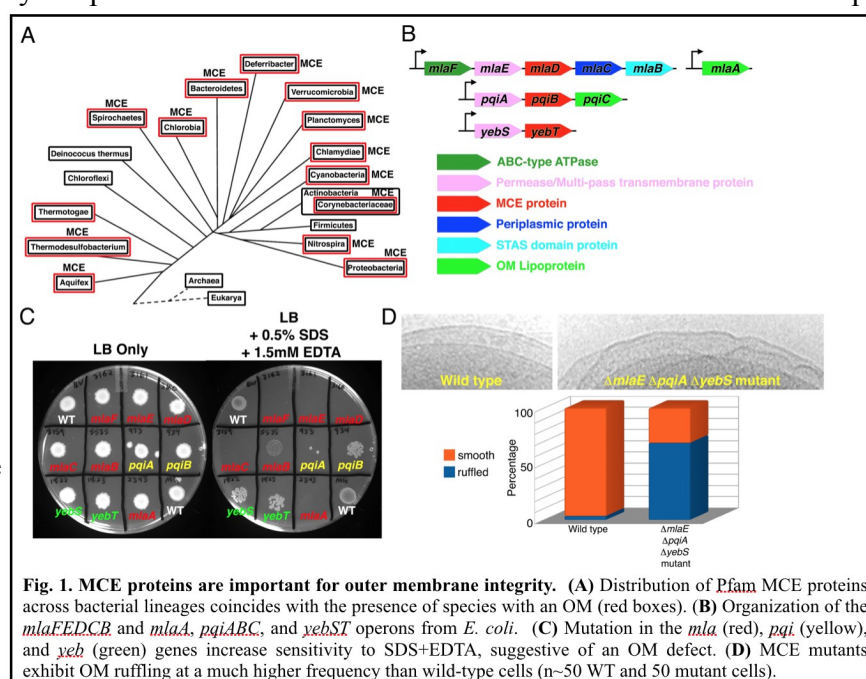

Fig. 1. MCE proteins are important for outer membrane integrity. (A) Distribution of Pfam MCE proteins across bacterial lineages coincides with the presence of species with an OM (red boxes). (B) Organization of the *mlaFEDCB* and *mlaA*, *pqiABC*, and *yebST* operons from *E. coli*. (C) Mutation in the *mla* (red), *pqi* (yellow), and *yeb* (green) genes increase sensitivity to SDS+EDTA, suggestive of an OM defect. (D) MCE mutants exhibit OM ruffling at a much higher frequency than wild-type cells (n~50 WT and 50 mutant cells).

The *E. coli* reference genome (strain MG1655) contains 3 MCE genes (Fig. 1B): *mlaD* (maintenance of OM lipid asymmetry D), *pqiB* (paraquat inducible B), and *yebT* (gene of unknown function; also known as MAM7). These MCE genes are part of three different operons, each of which also encode a multi-pass transmembrane protein (MlaE, PqiA, and YebS)(Fig. 1B). The *mla* operon in addition encodes an ABC-type (ATP binding cassette) ATPase (MlaF), a predicted periplasmic protein MlaC, and a STAS domain containing protein, MlaB. The presence of an ATPase and a multi-pass transmembrane protein in some MCE operons has led to the hypothesis that MCE systems may encode an ABC transporter (18, 25, 26), although the putative permease/transmembrane subunits have no detectable homology to any known ABC transporter. The *mlaFEDCB* genes also are thought to function together with an OM lipoprotein, MlaA, encoded elsewhere in the genome (Fig. 1B)(31). Mutations in the *mla* genes increase the phospholipid content in the outer leaflet of the OM, suggesting a role in maintaining OM asymmetry (31). However, how the proteins encoded by the *mla* operon interact and function is not understood.

Mutations in MCE operons result in OM defects. Mutations in the *mla* genes result in increased sensitivity to a combination of SDS and EDTA (SDS+EDTA), suggestive of a mild OM defect, which has been attributed to a loss of OM asymmetry (*31*). To assess whether other MCE proteins are similarly involved in OM maintenance or biogenesis, we constructed *E. coli* strains bearing single deletions of the MCE-encoding genes *pqiB* and *yebT*, and their associated putative permease genes, *pqiA* and *yebS*. These mutations in the *pqi* and *yeb* operons increased sensitivity of bacteria to 0.5% SDS + 1.25 mM EDTA (Fig. 1C). We also constructed a triple mutant lacking the putative permease subunit from each operon (Δ*mlaE* Δ*pqiA* Δ*yebS*). The triple permease mutant is viable and shows no discernible growth defect relative to wild-type in liquid LB medium (fig. S1). Thus, the Mla, Pqi, and Yeb systems are non-essential for *E. coli* growth in culture despite being critical for resistance to environmental stresses, including those likely encountered by pathogens during infection (*20–27*). The Δ*mlaE* Δ*pqiA* Δ*yebS* mutant exhibited SDS+EDTA sensitivity similar to the Δ*mlaE* single mutant (*36*). But in contrast to any of the single mutants, the triple mutant is more susceptible to outer membrane stressors, as this strain is uniquely sensitive to SDS alone and, unexpectedly, is also sensitive to 100 mM CaCl$_2$ (fig. S2) (*37*). Interestingly, exposure of Δ*mlaE* Δ*pqiA* Δ*yebS* mutant cells to the DNA stain DAPI resulted in strong nucleoid staining, whereas staining of wild-type cells was much weaker and largely confined to the cell surface (fig. S3), indicating that the function of the OM as a barrier to exclude hydrophobic molecules is severely compromised in the mutants. This is consistent with previous work suggesting that *E. coli* MCE mutants are more susceptible to hydrophobic antibiotics (*11*). Staining of live cells with propidium iodide, a membrane-impermeable DNA dye which stains cells with a disrupted plasma membrane, confirmed that only a small population of PI positive cells was present in the Δ*mlaE* Δ*pqiA* Δ*yebS* mutant cultures, suggesting that the IM remains intact (fig. S3). Overall, these results are consistent with a more severe OM defect in the triple mutant, suggesting that the Mla, Pqi, and Yeb systems act in concert to maintain the integrity of the OM barrier.
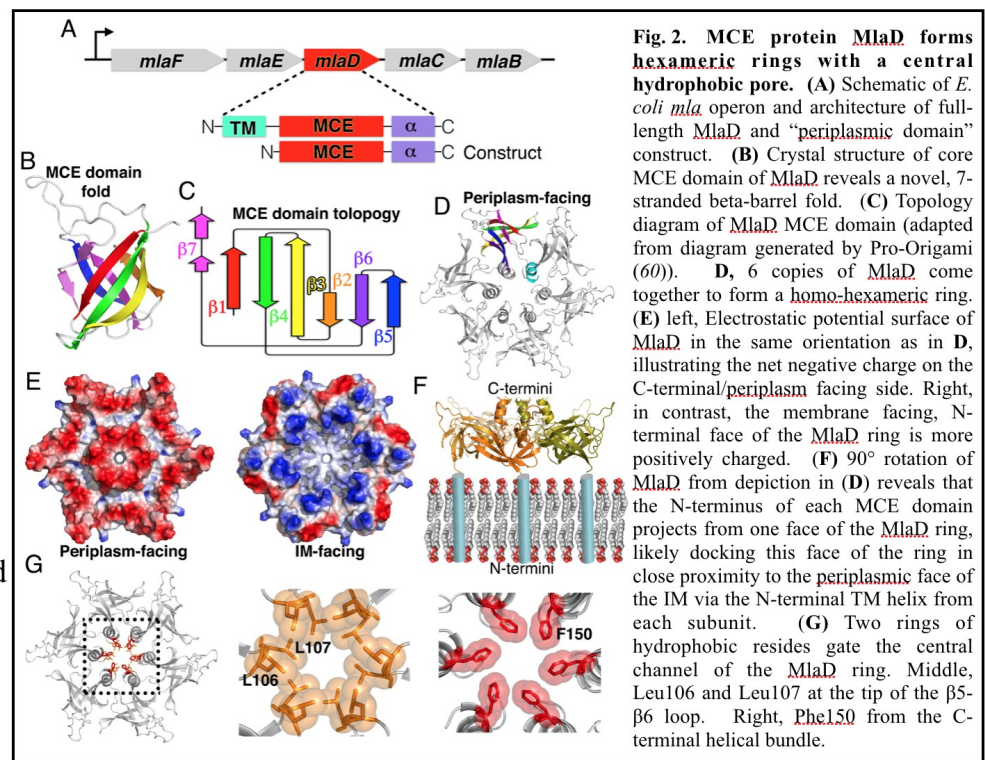
To better understand the nature of the OM defect in these mutants, we examined intact Δ*mlaE* Δ*pqiA* Δ*yebS* triple mutant *E. coli* cells by transmission electron microscopy. Strikingly, most triple mutant Δ*mlaE* Δ*pqiA* Δ*yebS* cells exhibited cell surface ruffling, suggestive of gross disorganization of the OM and periplasm (Fig. 1D), whereas the IM and OM of wild-type cells were smooth with a well-defined periplasm in between (Fig. 1D). This is consistent with results from *S. coelicolor* in which mutations in the single MCE gene cluster led to global changes in the cell surface (*38*). Thus, MCE proteins are important for maintaining the morphology of the OM and the uniform spacing between the IM and OM.

MCE protein MlaD forms a hexameric ring. As MCE proteins have no detectable similarity to proteins of known structure or function, we set out to obtain structural information using a combination of X-ray crystallography and electron cryo-microscopy (cryo-EM). MlaD is predicted to be anchored to the IM by a single N-terminal transmembrane helix, with its MCE domain residing in the periplasm (Fig. 2A). We determined the crystal structure of the complete periplasmic domain of *E. coli* MlaD (residues 32-183) at 2.85 Å resolution and its core MCE domain at 2.15 Å resolution (tables S1 and S2; fig. S4). The MCE core domain adopts a novel, 7-stranded β-barrel fold, encompassing just over 100 residues (Fig. 2, B and C). Most of the connecting loops are fairly short, with the exception of the β6-β7 loop, which projects prominently from one end of the barrel (Fig. 2B). A search of the Protein Databank for structural homologs using DALI (*39*) did not identify any other proteins that share this fold, although we identified two classes of similar domains that differ in the arrangement of one or more β-strands (fig. S5). The six copies of MlaD are assembled into a homo-hexameric ring with the C-terminal α-helical domain extending from one face of the ring (Fig. 2D). The interface between subunits consists of two main parts. First, the core MCE domains interact in a "head-to-tail" fashion, mediated primarily by the connecting loops at each end of the β-barrel fold, accounting for more than half (~900 Å^2) of the total interface (1,660 Å^2). Second, the C-terminal helical regions from

each protomer come together to form a 6-helix bundle at the center of the ring (760 Å^2 interface). Truncation of the C-terminal α-helical region led to the formation of heptameric instead of hexameric rings (fig. S6) suggesting that interactions between the C-terminal α-helices may play a role in determining the overall oligomeric state and packing of the MCE subunits. It is interesting to note that in mycobacterial species such as *M. tuberculosis*, MCE operons usually encode six distinct MCE proteins (fig. S7) (*18*). In light of the observed homo-hexameric assembly of MlaD, it seems likely that mycobacterial MCE proteins will form hetero-hexamers consisting of a single copy of each of the MCE proteins from an operon.

The N-terminus of each MCE domain is connected by a short peptide linker to an N-terminal transmembrane helix, which places the N-terminal face of the ring in close proximity to the negatively charged phospholipid head groups (Fig. 2E). Consistent with this orientation, the N-terminal, membrane-facing surface of the ring is considerably more basic than the C-terminal, periplasm-facing side (Fig. 2E). The C-terminal helix from each subunit associates to form a hollow barrel very similar in structure to computationally designed 6-helix coiled-coils (*40, 41*), leaving a solvent accessible channel running through the center of the ring (Fig. 2G).

Fig. 2. MCE protein MlaD forms hexameric rings with a central hydrophobic pore. (A) Schematic of *E. coli mla* operon and architecture of full-length MlaD and "periplasmic domain" construct. (B) Crystal structure of core MCE domain of MlaD reveals a novel, 7-stranded beta-barrel fold. (C) Topology diagram of MlaD MCE domain (adapted from diagram generated by Pro-Origami (*60*)). (D), 6 copies of MlaD come together to form a homo-hexameric ring. (E) left, Electrostatic potential surface of MlaD in the same orientation as in D, illustrating the net negative charge on the C-terminal/periplasm facing side. Right, in contrast, the membrane facing, N-terminal face of the MlaD ring is more positively charged. (F) 90° rotation of MlaD from depiction in (D) reveals that the N-terminus of each MCE domain projects from one face of the MlaD ring, likely docking this face of the ring in close proximity to the periplasmic face of the IM via the N-terminal TM helix from each subunit. (G) Two rings of hydrophobic resides gate the central channel of the MlaD ring. Middle, Leu106 and Leu107 at the tip of the β5-β6 loop. Right, Phe150 from the C-terminal helical bundle.
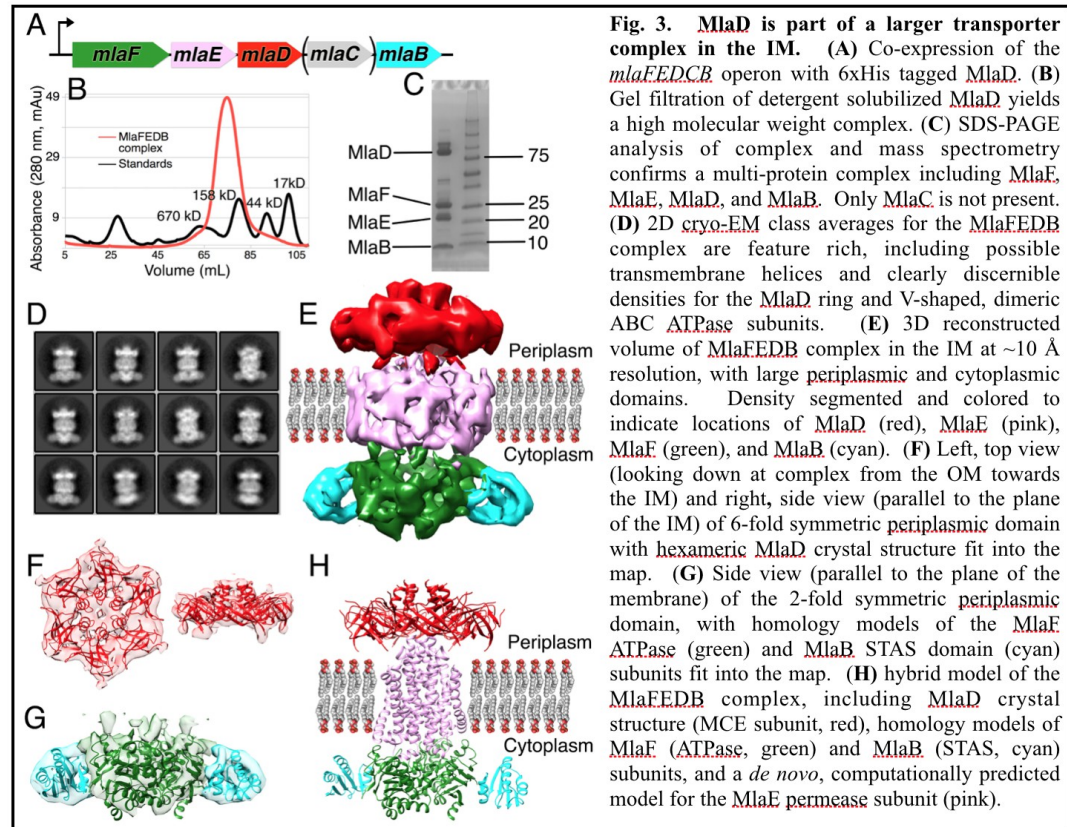
This channel is lined with hydrophobic residues and has two notable constriction points: the first is formed by Leu106 and Leu107 at the tip of the β5-6 loop and second by Phe150 in the C-terminal helix (Fig. 2G). These narrowings, in particular the constriction formed by the Phe150 ring, are reminiscent of the Phe427 Φ-clamp in the central pore of anthrax protective antigen (PA), where they play a key role in the transport of protein toxins through the PA translocon (*42*). At the narrowest point created by the Leu106/Leu107 loops, the channel diameter is ~6 Å, large enough to allow the passage of small hydrophobic molecules like phospholipids or sterols. The weaker electron density in this region of the crystal structure suggests that these loops are at least somewhat flexible and therefore may allow somewhat larger molecules to pass through. Based upon its proposed orientation (Fig. 2F), the MlaD ring is poised to function as a channel to facilitate trafficking of small hydrophobic molecules to or from the inner membrane.

Architecture of the MlaFEDB transporter complex. An MCE protein implicated in lipid transport in plants, TRIGALACTOSYLDIACYLGLYCEROL 2 (TGD2), forms part of a high-molecular weight complex in the chloroplast inner membrane (*19, 30, 43*). To assess whether MlaD forms a complex with any of the other Mla proteins, we over-expressed the full *mlaFEDCB* operon in *E. coli* (Fig. 3A), solubilized the membrane fraction with detergent, and determined whether any other Mla proteins co-purify with His-tagged MlaD. After affinity purification and gel filtration (Fig. 3B), SDS-PAGE analysis of the purified MlaD sample revealed the presence of three additional proteins

5

forming a stable complex with MlaD (Fig. 3C). Excision of these bands and protein identification by mass spectrometry confirmed that these proteins were MlaF (an ABC-type ATPase), MlaE (a multi-pass transmembrane protein and likely permease), and MlaB (STAS domain containing protein); the co-expressed MlaC protein, however, was absent.

To determine its architecture, we analyzed the purified MlaFEDB complex by single particle cryo-EM. The 2D class averages were feature-rich, and the locations of individual components within the complex were clearly identifiable (Fig. 3D). Some inherent flexibility in the complex is apparent in these 2D class averages, suggesting that the relative orientation of the components

Fig. 3. **MlaD is part of a larger transporter complex in the IM.** (A) Co-expression of the *mlaFEDCB* operon with 6xHis tagged MlaD. (B) Gel filtration of detergent solubilized MlaD yields a high molecular weight complex. (C) SDS-PAGE analysis of complex and mass spectrometry confirms a multi-protein complex including MlaF, MlaE, MlaD, and MlaB. Only MlaC is not present. (D) 2D cryo-EM class averages for the MlaFEDB complex are feature rich, including possible transmembrane helices and clearly discernible densities for the MlaD ring and V-shaped, dimeric ABC ATPase subunits. (E) 3D reconstructed volume of MlaFEDB complex in the IM at ~10 Å resolution, with large periplasmic and cytoplasmic domains. Density segmented and colored to indicate locations of MlaD (red), MlaE (pink), MlaF (green), and MlaB (cyan). (F) Left, top view (looking down at complex from the OM towards the IM) and right, side view (parallel to the plane of the IM) of 6-fold symmetric periplasmic domain with hexameric MlaD crystal structure fit into the map. (G) Side view (parallel to the plane of the membrane) of the 2-fold symmetric periplasmic domain, with homology models of the MlaF ATPase (green) and MlaB STAS domain (cyan) subunits fit into the map. (H) hybrid model of the MlaFEDB complex, including MlaD crystal structure (MCE subunit, red), homology models of MlaF (ATPase, green) and MlaB (STAS, cyan) subunits, and a *de novo*, computationally predicted model for the MlaE permease subunit (pink).

is not rigidly fixed. This also limits the final resolution of our density maps. 3D classification and refinement produced a ~10 Å resolution density map (Fig. 3E and fig. S8). The six-fold symmetry of the MlaD ring is readily apparent at one end of the map (Fig. 3F). The hexameric, MlaD crystal structure fits the density of this region very well (Fig. 3F), with the C-terminal helices forming a knob at the apex. Additional density for the N-terminal transmembrane helices extend down towards the membrane. The opposite end of the map is approximately two-fold symmetric, consistent with the dimeric state of known ABC transporters. Homology models of MlaF and MlaB could be generated with reasonable confidence (*44*), and were used to fit into the density. Two copies of the MlaF ATPase can be fit into the map (Fig. 3G), leaving two small densities on either side. These two densities are approximately the size and shape of a STAS domain, allowing for the docking of two copies of an MlaB homology model into the map; one domain is associated with each of the ABC subunits (Fig. 3G). The remaining central density most likely corresponds to the transmembrane subunit, MlaE, along with six additional transmembrane helices from the MlaD subunits. Based upon the volume of the transmembrane region and the apparent 2-fold symmetry of the map, we infer that there are most likely two copies of the MlaE subunit, although the modest resolution of our cryo-EM reconstruction does not allow us to trace the MlaE subunit *de novo*. The apparent symmetry mismatch between the hexameric MlaD and the dimeric permease assemblies implies that the MlaD-MlaE interface cannot be identical for each of the six MlaD subunits.

Recently, a model for the MlaE structure was generated from a powerful new approach to protein structure prediction using residue contact restraints based upon co-evolving residues (*45*). This *de novo* model for MlaE is dimeric and closely resembles the shape of the permease subunit observed in our 2D class averages (Fig. 3H). This predicted MlaE structure is unique among ABC transporter

permeases, and fits within the density of the transmembrane region. Additional density surrounds the MlaE model, which may be derived from the six MlaD transmembrane segments associating with the sides of MlaE within the transmembrane region (*46*). Indeed, several residues in the MlaD transmembrane helices show signatures of co-evolution with residues in the membrane spanning portion of MlaE, suggesting a physical interaction between these regions (fig. S9A). Several additional residues in the MlaF ATPase appear to be co-evolving with residues in MlaE. These residue lie at the end of MlaE predicted to be oriented towards the cytoplasm, and on the surface of MlaF typically used by other ABC ATPases to interact with the transporter permease subunit (fig. S9B), supporting the placement and orientation of the Mla subunits in our cryo-EM map.

MlaC is a lipid-binding protein. Many ABC transporters from double-membraned bacteria encode small substrate binding proteins that ferry substrates across the periplasm. Although MlaC has no detectable similarity to known substrate binding proteins, it is encoded in the *mla* operon (Fig. 4A), targeted to the periplasm, and may play a role in shuttling substrates to and from the MlaFEDB transporter complex. To better understand how MlaC might be involved in substrate recognition, we determined the crystal structure of MlaC at ~1.50 Å resolution. MlaC adopts a mixed α/β-fold, consisting of a highly twisted, 4-stranded β-sheet and a bundle of 7 α-helices (Fig. 4B). Two divergent homologs of MlaC were previously deposited by the Northeast Structural Genomics Consortium (NESG) without any associated publications (2QGU and 4FCZ; 27% and 22% sequence identity and Cα rmsd of 2.5 Å and 3.0 Å with *E. coli* MlaC, respectively). Outside of this small group of three MlaC-like proteins, a structure-based comparison to other proteins in the PDB using DALI revealed more distant similarities to the nuclear transport factor 2 (NTF2) protein superfamily and to the parental, Cystatin-like fold (fig. S10).
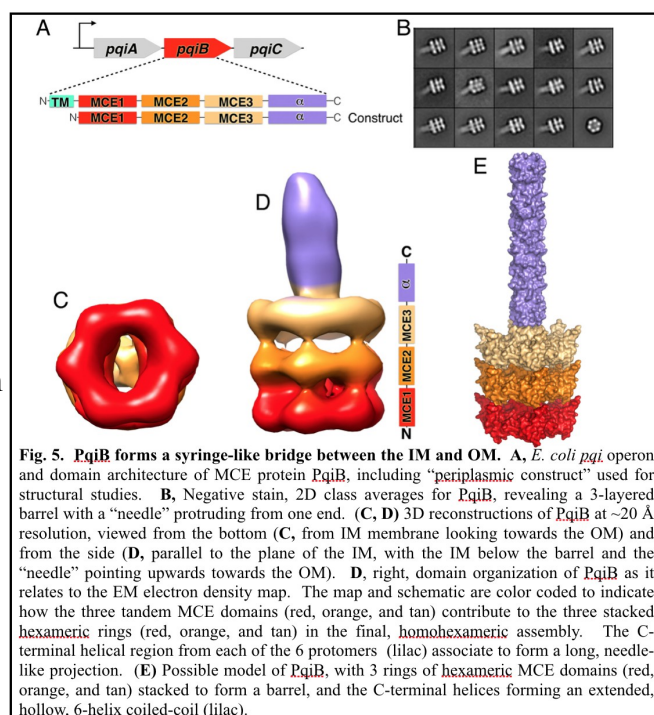
Closer examination of our MlaC electron density maps revealed the presence of a large ligand bound within a deep hydrophobic pocket in the protein core (Fig. 4, C and D). Endogenous ligands are observed in a surprisingly large number of structures (*47*), and it appears that a high-affinity ligand co-purified with MlaC and remained bound in our crystals. This roughly "horse-shoe"-shaped density closely resembled two fatty acid chains esterified to a phosphoglycerol head group, which we have modeled as phosphatidic acid (Fig. 4, B and C). The two acyl chains of the lipid insert deep into the protein

Fig. 4. **MlaC is a periplasmic protein that shuttles lipids across the periplasm.** (A) MlaC is encoded in the *mla* operon and is the only component not stably present in the MlaFEDB complex. (B) Crystal structure of MlaC adopts a mixed α/β fold with 4 β-strands (light blue) and 7 α-helices (dark blue). MlaC-bound lipid shown as spheres. (C) Cut away view of MlaC surface, revealing deep hydrophobic pocket in MlaC containing the two acyl chain "tails" of a bound lipid. (D) Unbiased 2Fo-Fc electron density map for bound lipid (prior to refinement with lipid ligand), contoured at 1 sigma. (E) In contrast to the extensive interactions between MlaC and the lipid tails, no significant interactions are made between MlaC and the lipid head group. MlaC interacts with the IM, MlaFEDB complex (F) and MlaC also interacts with the OM MlaA-OmpF complex (G) by biolayer interferometry, suggesting MlaC may ferry lipid substrates between the IM and OM components of the Mla transport system.

core where they are completely protected from solvent (Fig. 4C), with little remaining space to accommodate larger lipids with more than two acyl chains. In contrast, the head group is almost
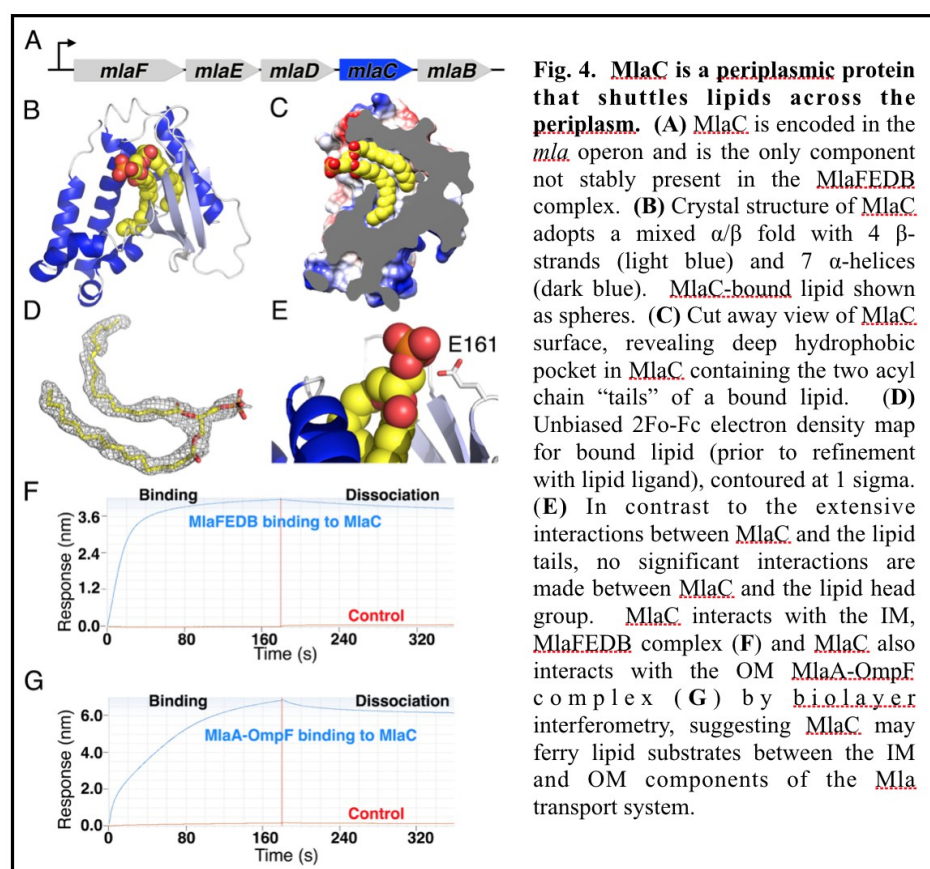
entirely solvent exposed and makes only minor contacts with the protein (Fig. 4E). Due to the limited volume of the hydrophobic cavity and the lack of head group interactions, we predict that MlaC is capable of binding to a wide range of diacyl lipids with little head group specificity. The NESG crystal structure of an MlaC-like protein from *R. solanacearum* (2QGU) was also crystallized with a ligand of unknown origin bound in its hydrophobic pocket (fig. S11), which was modeled as phosphatidylethanolamine. It has been previously noted that MlaC is a predicted homolog of 2QGU (*31*), and our results show that MlaC is indeed a lipid-binding protein. In another MlaC-like structure in the PDB (4FCZ), no ligands were explicitly modeled, but our re-refinement of the deposited coordinates and structure factors revealed additional electron density in the hydrophobic pocket of this protein (fig. S11 and table S3), suggestive of a tetra-acyl, cardiolipin-like lipid. The observation of lipid ligands in MlaC, 2QGU, and 4FCZ support the notion that MlaC and its close homologs are lipid-binding proteins.

MlaC binds IM and OM protein complexes. Since MlaC is a soluble lipid-binding protein predicted to localize to the periplasm, we hypothesized that MlaC may shuttle lipid substrates across the periplasm between the IM MlaFEDB transporter and an OM complex, where lipids would be inserted into or removed from the OM. Lipoprotein MlaA (which is not part of the *mla* operon in *E. coli*) has emerged as candidate component for this OM complex, as it has been implicated in Mla function (*31*) and forms a complex with the OM porin proteins, OmpC and OmpF (*48*). Thus, we set out to detemine whether MlaC could interact directly with purified MlaFEDB and MlaA-OmpF protein using biolayer interferometry. We over-expressed and purified MlaA from *E. coli*, which co-purified as a stable complex with OmpF (*49*). We observed robust and specific binding of both MlaFEDB (Fig. 4F) and MlaA-OmpF assemblies (Fig. 4G) to MlaC, as well as a direct interaction between MlaC and MlaD (fig. S12). These results demonstrate that MlaC is capable of binding to both the IM and OM complexes, supporting a model where MlaC plays a central role in transferring a lipid substrate from one membrane to another. This has some interesting parallels the Lol lipoprotein transport pathway (*50*), in which the LolA protein plays a similar role to MlaC; it sequesters lipid modifications during the transport of nascent lipoproteins across the periplasm.
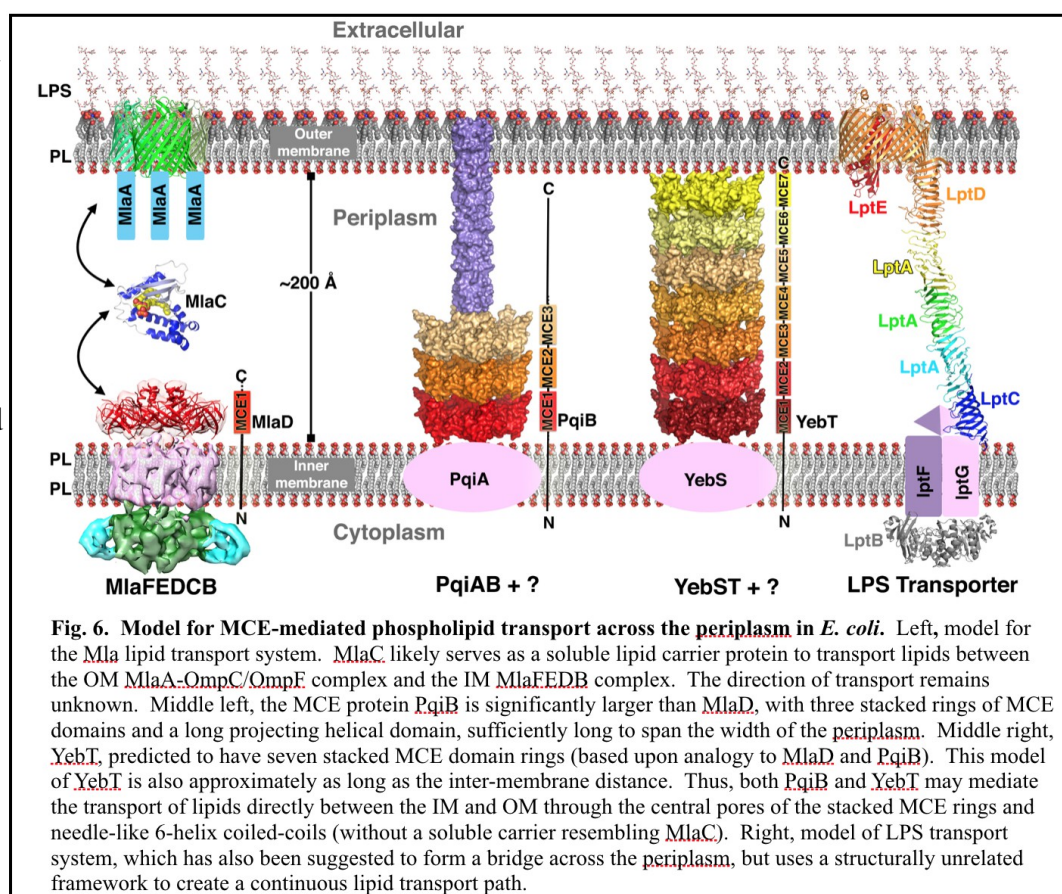
PqiB forms a large assembly of stacked MCE domains. No other homologs of MlaC are encoded in the *E. coli* genome, raising the question of how the Pqi and Yeb systems might move substrates across the periplasm. To gain insight into this question, we purified the periplasmic domain of *E. coli* PqiB and obtained a negative-stain EM reconstruction at ~20 Å resolution (Fig. 5, A and B). PqiB's three tandem MCE domains come together to form a large, cylindrical barrel, ~90 Å in diameter and ~100 Å long, with the N-terminal, middle, and C-terminal MCE domains self-associating to form three stacked hexameric rings (Fig. 5B-D). A long, narrow projection emanates from the top of the barrel, resulting in an overall assembly reminiscent of a needle and syringe (Fig. 5D). Recombinant PqiB proteins containing only the three MCE domains form barrels without needles, showing that the C-terminal regions from each subunit come together to form this needle-like structure (fig. S13). Secondary structure prediction suggests that this C-terminal region is almost entirely α-helical. To create such a long needle, most of the ~100

Fig. 5. PqiB forms a syringe-like bridge between the IM and OM. A, *E. coli pqi* operon and domain architecture of MCE protein PqiB, including "periplasmic construct" used for structural studies. **B,** Negative stain, 2D class averages for PqiB, revealing a 3-layered barrel with a "needle" protruding from one end. **(C, D)** 3D reconstructions of PqiB at ~20 Å resolution, viewed from the bottom **(C,** from IM membrane looking towards the OM) and from the side **(D,** parallel to the plane of the IM, with the IM below the barrel and the "needle" pointing upwards towards the OM). **D,** right, domain organization of PqiB as it relates to the EM electron density map. The map and schematic are color coded to indicate how the three tandem MCE domains (red, orange, and tan) contribute to the three stacked hexameric rings (red, orange, and tan) in the final, homohexameric assembly. The C-terminal helical region from each of the 6 protomers (lilac) associate to form a long, needle-like projection. **(E)** Possible model of PqiB, with 3 rings of hexameric MCE domains (red, orange, and tan) stacked to form a barrel, and the C-terminal helices forming an extended, hollow, 6-helix coiled-coil (lilac).

residues in the C-terminal region would have to take the form of a linear α-helix, which potentially could yield a very long 6-helix coiled-coil structure, up to 125-150 Å long and ~30 Å in diameter with a hollow core (Fig. 5E). This type of structure is essentially an elongated version of the 6-helix coiled-coil at the C-terminus of our MlaD crystal structure. Measurements of PqiB in our negative stain electron micrographs suggest a total length of ~225-250 Å from the bottom of the barrel to the tip of the needle, indicating that PqiB is sufficiently long to directly span the distance between the IM and OM (*E. coli* periplasmic width ~200-225 Å) (*3*, *51*). Based upon this model for PqiB, we propose that the seven MCE domains of *E. coli* YebT will form a stack of seven hexameric rings, resulting in an elongated barrel that may also be capable of spanning the width of the periplasm.

Model for lipid transport by MCE systems. These structural data allow us to generate a model for how MCE transport systems traffic lipids across the periplasm (Fig. 6). As the direction of transport remains to be determined, we present this model in the context of lipid export to the outer membrane. However, these systems could equally drive lipid import towards the inner membrane, with the steps outlined occurring in the reverse order. In the case of the Mla system, the MlaFEDB complex uses energy from ATP hydrolysis to extract lipids from the IM, translocates these lipids through the hydrophobic pore of MlaD, and then loads them into MlaC. Lipid-bound MlaC then traverses the periplasm, delivering the lipid to the MlaA-OmpC/OmpF complex for insertion into the OM. In contrast, in the Pqi system, the syringe-like MCE protein PqiB is poised to bridge the gap between the IM and OM directly, forming a continuous lipid translocation pathway that transverses the entire periplasmic space without the need for a soluble lipid carrier protein (Fig. 6). The other major lipid component of the OM, LPS, is transported across a periplasm-spanning bridge comprised of the LptA, LptC, and LptD proteins (Fig. 6) (*52–58*). However, in contrast to the open groove utilized by the LPS transport system, we propose that MCE transporters translocate lipids through a closed channel formed by the pores of multiple stacked MCE rings (e.g., YebT), or a combination of rings and a hollow, needle-like projection formed by long C-terminal helices (e.g., PqiB, and the MCE proteins from *M. tuberculosis* and chloroplasts; fig. S14). Thus, evolution appears to have developed at least three unique strategies for lipid transport between the IM and OM of double-membraned bacteria and chloroplasts: 1) the use of a soluble lipid carrier protein such as MlaC; 2)

Fig. 6. Model for MCE-mediated phospholipid transport across the periplasm in *E. coli*. Left, model for the Mla lipid transport system. MlaC likely serves as a soluble lipid carrier protein to transport lipids between the OM MlaA-OmpC/OmpF complex and the IM MlaFEDB complex. The direction of transport remains unknown. Middle left, the MCE protein PqiB is significantly larger than MlaD, with three stacked rings of MCE domains and a long projecting helical domain, sufficiently long to span the width of the periplasm. Middle right, YebT, predicted to have seven stacked MCE domain rings (based upon analogy to MlaD and PqiB). This model of YebT is also approximately as long as the inter-membrane distance. Thus, both PqiB and YebT may mediate the transport of lipids directly between the IM and OM through the central pores of the stacked MCE rings and needle-like 6-helix coiled-coils (without a soluble carrier resembling MlaC). Right, model of LPS transport system, which has also been suggested to form a bridge across the periplasm, but uses a structurally unrelated framework to create a continuous lipid transport path.

a long hydrophobic groove, as in the LPS transporter; and 3) elongated lipid translocation tunnels as in PqiB and YebT.

Many important questions relating to the function of MCE transporters remain to be explored. Our structures do not reveal the direction of lipid transport. In the case of LPS transport, the IM LptBFG complex uses ATP hydrolysis to extract LPS molecules from the IM and transport them through the LptACD bridge to the OM. Perhaps the MlaFEDB complex uses ATP hydrolysis in a similar manner to drive phospholipid export, although genetic approaches have suggested that these proteins may be involved in retrograde transport of mislocalized phospholipids from the OM to the IM (*31*). It is also unclear why *E. coli* and many other organisms encode multiple MCE systems. While they may serve partially redundant roles, their very distinct structures suggest they may have distinct substrate specificities, transport lipids in different directions, or might be subject to unique regulatory control. Intriguingly, despite being almost universally conserved from double-membraned bacteria to chloroplasts, MCE homologs are conspicuously missing in the mitochondria, and it is unclear if they have been functionally replaced during domestication of this organelle. Future studies will begin to unravel these and other questions to further understand the assembly and maintenance of the outer membranes of bacteria and organelles.

References and Notes:

1. H. Nikaido, Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).

2. T. Nakae, Outer membrane of Salmonella. Isolation of protein complex that produces transmembrane channels. *J. Biol. Chem.* **251**, 2176–2178 (1976).

3. C. Hoffmann, A. Leis, M. Niederweis, J. M. Plitzko, H. Engelhardt, Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc. Natl. Acad. Sci.* **105**, 3963–3967 (2008).

4. J. E. Somerville, L. Cassiano, R. P. Darveau, Escherichia coli msbB Gene as a Virulence Factor and a Therapeutic Target. *Infect. Immun.* **67**, 6583–6590 (1999).

5. J. S. Cox, B. Chen, M. McNeil, W. R. Jacobs, Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature.* **402**, 79–83 (1999).

6. Q. Kong *et al.*, Phosphate Groups of Lipid A Are Essential for Salmonella enterica Serovar Typhimurium Virulence and Affect Innate and Adaptive Immunity. *Infect. Immun.* **80**, 3215–3224 (2012).

7. X. Wang, A. A. Ribeiro, Z. Guan, S. N. Abraham, C. R. H. Raetz, Attenuated virulence of a Francisella mutant lacking the lipid A 4′-phosphatase. *Proc. Natl. Acad. Sci.* **104**, 4136–4141 (2007).

8. C. A. Schnaitman, J. D. Klena, Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655–682 (1993).

9. I. M. Helander *et al.*, rfaP mutants of Salmonella typhimurium. *Eur. J. Biochem.* **185**, 541–546 (1989).

10. A. M. Kropinski, L. Chan, F. H. Milazzo, Susceptibility of lipopolysaccharide-defective mutants of Pseudomonas aeruginosa strain PAO to dyes, detergents, and antibiotics. *Antimicrob. Agents Chemother.* **13**, 494–499 (1978).

11. R. J. Nichols *et al.*, Phenotypic Landscape of a Bacterial Cell. *Cell.* **144**, 143–156 (2011).

12. J. Smit, Y. Kamio, H. Nikaido, Outer membrane of Salmonella typhimurium: chemical analysis and freeze-fracture studies with lipopolysaccharide mutants. *J. Bacteriol.* **124**, 942–958 (1975).

13. Y. Kamio, H. Nikaido, Outer membrane of Salmonella typhimurium: accessibility of phospholipid head groups to phospholipase C and cyanogen bromide activated dextran in the external medium. *Biochemistry (Mosc.).* **15**, 2561–2570 (1976).

14. P. F. Mühlradt, J. R. Golecki, Asymmetrical Distribution and Artifactual Reorientation of Lipopolysaccharide in the Outer Membrane Bilayer of Salmonella typhimurium. *Eur. J. Biochem.* **51**, 343–352 (1975).

15. N. Ruiz, D. Kahne, T. J. Silhavy, Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nat. Rev. Microbiol.* **7**, 677–683 (2009).

16. J. M. May, D. J. Sherman, B. W. Simpson, N. Ruiz, D. Kahne, Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer membrane. *Phil Trans R Soc B.* **370**, 20150027 (2015).

17. S. Arruda, G. Bomfim, R. Knights, T. Huima-Byron, L. W. Riley, Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. *Science.* **261**, 1454–1457 (1993).

18. N. Casali, L. W. Riley, A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics.* **8**, 60 (2007).

19. K. Awai, C. Xu, B. Tamot, C. Benning, A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. *Proc. Natl. Acad. Sci.* **103**, 10817–10822 (2006).

20. A. K. Pandey, C. M. Sassetti, Mycobacterial persistence requires the utilization of host cholesterol. *Proc. Natl. Acad. Sci.* **105**, 4376–4380 (2008).

21. A. Gioffré *et al.*, Mutation in mce operons attenuates Mycobacterium tuberculosis virulence. *Microbes Infect.* **7**, 325–334 (2005).

22. R. H. Senaratne *et al.*, Mycobacterium tuberculosis strains disrupted in mce3 and mce4 operons are attenuated in mice. *J. Med. Microbiol.* **57**, 164–170 (2008).

23. J. Rengarajan, B. R. Bloom, E. J. Rubin, Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8327–8332 (2005).

24. S. M. Joshi *et al.*, Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc. Natl. Acad. Sci.* **103**, 11760–11765 (2006).

25. L. Zhang *et al.*, The mammalian cell entry (Mce) protein of pathogenic Leptospira species is responsible for RGD motif-dependent infection of cells and animals. *Mol. Microbiol.* **83**, 1006–1023 (2012).

26. C. D. Carpenter *et al.*, The Vps/VacJ ABC Transporter Is Required for Intercellular Spread of Shigella flexneri. *Infect. Immun.* **82**, 660–669 (2014).

27. S. Nakamura *et al.*, Molecular basis of increased serum resistance among pulmonary isolates of non-typeable Haemophilus influenzae. *PLoS Pathog.* **7**, e1001247 (2011).

28. C. Xu, J. Fan, W. Riekhof, J. E. Froehlich, C. Benning, A permease- like protein involved in ER to thylakoid lipid transfer in Arabidopsis. *EMBO J.* **22**, 2370–2379 (2003).

29. C. Xu, J. Fan, A. J. Cornish, C. Benning, Lipid Trafficking between the Endoplasmic Reticulum and the Plastid in Arabidopsis Requires the Extraplastidic TGD4 Protein. *Plant Cell Online*. **20**, 2190–2204 (2008).

30. R. Roston, J. Gao, C. Xu, C. Benning, Arabidopsis chloroplast lipid transport protein TGD2 disrupts membranes and is part of a large complex. *Plant J.* **66**, 759–769 (2011).

31. J. C. Malinverni, T. J. Silhavy, An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. *Proc. Natl. Acad. Sci.* **106**, 8009–8014 (2009).

32. H. A. Sutterlin *et al.*, Disruption of lipid homeostasis in the Gram-negative cell envelope activates a novel cell death pathway. *Proc. Natl. Acad. Sci.* **113**, E1565–E1574 (2016).

33. W. W. Mohn *et al.*, The Actinobacterial mce4 Locus Encodes a Steroid Transporter. *J. Biol. Chem.* **283**, 35368–35374 (2008).

34. L. I. Klepp *et al.*, Impact of the deletion of the six mce operons in Mycobacterium smegmatis. *Microbes Infect.* **14**, 590–599 (2012).

35. K. Kim, S. Lee, K. Lee, D. Lim, Isolation and Characterization of Toluene-Sensitive Mutants from the Toluene-Resistant Bacterium Pseudomonas putida GM73. *J. Bacteriol.* **180**, 3692–3696 (1998).

36. Zero colonies of mutant strains grew at the lowest dilution plated in this experiment, corresponding to at least a 3-log reduction in CFUs relative to the WT.

37. As divalent cations are thought to be important for the stabilization of interactions between LPS molecules in the OM, and the ratio of LPS to divalents may be important to maintain the appropriate degree of membrane fluidity/rigidity.

It is possible that high concentrations of CaCl2 lead to saturation of all available divalent binding sites, leading to dissociation of LPS-LPS interactions instead of cross-linking. This effect may be exacerbated in the mutant due to an already weakened outer membrane.

38. L. C. Clark *et al.*, Mammalian cell entry genes in Streptomyces may provide clues to the evolution of bacterial virulence. *Sci. Rep.* **3** (2013), doi:10.1038/srep01109.

39. L. Holm, P. Rosenström, Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).

40. N. R. Zaccai *et al.*, A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* **7**, 935–941 (2011).

41. A. R. Thomson *et al.*, Computational design of water-soluble α-helical barrels. *Science*. **346**, 485–488 (2014).

42. J. Jiang, B. L. Pentelute, R. J. Collier, Z. H. Zhou, Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature*. **521**, 545–549 (2015).

43. R. L. Roston, J. Gao, M. W. Murcha, J. Whelan, C. Benning, TGD1, -2, and -3 Proteins Involved in Lipid Trafficking Form ATP-binding Cassette (ABC) Transporter with Multiple Substrate-binding Proteins. *J. Biol. Chem.* **287**, 21406–21415 (2012).

44. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

45. S. Ovchinnikov *et al.*, Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife*. **4**, e09248 (2015).

46. In addition some of this density may be due to the amphipol use to stablize the transmembrane region in the absence of detergent.

47. A. Kumar *et al.*, Ligands in PSI structures. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1309–1316 (2010).

48. Z.-S. Chong, W.-F. Woo, S.-S. Chng, Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. *Mol. Microbiol.* **98**, 1133–1146 (2015).

49. Mass spectrometry confirmed the presence of MlaA and OmpF in our purified sample. However, expression of OmpC and OmpF is known to be regulated by osmolarity, and the absence of OmpC may reflect differing culture conditions.

50. A. Konovalova, T. J. Silhavy, Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370** (2015), doi:10.1098/rstb.2015.0030.

51. V. R. F. Matias, A. Al-Amoudi, J. Dubochet, T. J. Beveridge, Cryo-Transmission Electron Microscopy of Frozen-Hydrated Sections of Escherichia coli and Pseudomonas aeruginosa. *J. Bacteriol.* **185**, 6112–6118 (2003).

52. S. Okuda, E. Freinkman, D. Kahne, Cytoplasmic ATP Hydrolysis Powers Transport of Lipopolysaccharide Across the Periplasm in E. coli. *Science*. **338**, 1214–1217 (2012).

53. M. D. L. Suits, P. Sperandeo, G. Dehò, A. Polissi, Z. Jia, Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis. *J. Mol. Biol.* **380**, 476–488 (2008).

54. B. Tefsen, J. Geurtsen, F. Beckers, J. Tommassen, H. de Cock, Lipopolysaccharide Transport to the Bacterial Outer Membrane in Spheroplasts. *J. Biol. Chem.* **280**, 4504–4509 (2005).

55. P. Sperandeo *et al.*, Characterization of lptA and lptB, Two Essential Genes Implicated in Lipopolysaccharide Transport to the Outer Membrane of Escherichia coli. *J. Bacteriol.* **189**, 244–253 (2007).

56.   P. Sperandeo *et al.*, Functional Analysis of the Protein Machinery Required for Transport of Lipopolysaccharide to the Outer Membrane of Escherichia coli. *J. Bacteriol.* **190**, 4460–4469 (2008).

57.   T. Wu *et al.*, Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proc. Natl. Acad. Sci.* **103**, 11754–11759 (2006).

58.   N. Ruiz, L. S. Gronenberg, D. Kahne, T. J. Silhavy, Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Proc. Natl. Acad. Sci.* **105**, 5537–5542 (2008).

59.   We thank C. Waddling from the UCSF Crystallography and SAXS Facility for support with crystallization; J. Holton, G. Meigs, and the staff of Advanced Light Source beamline 8.3.1 for beamline support; A. Gray, C. Gross, and M. Elvekrog (UCSF) for providing plasmids and bacterial strains; N. Stuurman and A. Williamson (UCSF) for assistance with fluorescence microscopy; G. Isom and I. Henderson (U. Birmingham), Y. Cheng, E. Palovcak, JP. Armache and other members of the Cheng lab (UCSF), and A. Brilot (UCSF) for helpful suggestions and discussions; M. Braunfeld for EM technical assistance; C. Kennedy for assistance with high performance computing; C. Hong and Z. Yu at Janelia Research Campus for EM data collection; the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley and the UC Davis Campus Mass Spectrometry Facilities for protein MS assistance; and M. Morrissey, X. Su, M. Tanenbaum, Y. Cheng, and I. Wilson for critical reading of the manuscript. This work was supported by NIH grants K99GM112982 (G.B.) and 1S10OD020054, and the Howard Hughes Medical Institute (R.D.V.). D.C.E. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (Grant DRG-2140-12). EM data were collected at the HHMI CryoEM Shared Facility at Janelia Research Campus. 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.

60.   A. Stivala, M. Wybrow, A. Wirth, J. C. Whisstock, P. J. Stuckey, Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics*. **27**, 3315–3316 (2011).

61.   T. Baba *et al.*, *Mol. Syst. Biol.*, in press, doi:10.1038/msb4100050.

62.   D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).

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

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

65.   T. C. Terwilliger *et al.*, Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).

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

67.   P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).

68.   A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

69.   X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods*. **10**, 584–590 (2013).

70.   D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

71.   H. Ru *et al.*, Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*.

163, 1138–1152 (2015).

72.   D. Elmlund, H. Elmlund, SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J. Struct. Biol.* **180**, 420–427 (2012).

73.   H. Elmlund, D. Elmlund, S. Bengio, PRIME: Probabilistic Initial 3D Model Generation for Single-Particle Cryo-Electron Microscopy. *Structure*. **21**, 1299–1306 (2013).

74.   S. H. W. Scheres, A Bayesian View on Cryo-EM Structure Determination. *J. Mol. Biol.* **415**, 406–418 (2012).

75.   S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

76.   S. H. W. Scheres, S. Chen, Prevention of overfitting in cryo-EM structure determination. *Nat. Methods*. **9**, 853–854 (2012).

77.   M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*. **9**, 173–175 (2012).

78.   UniProt Consortium, Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42**, D191-198 (2014).

79.   R. Leinonen *et al.*, *Nucleic Acids Res.*, in press, doi:10.1093/nar/gkq967.

80.   S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife*. **3**, e02030 (2014).

81.   Y. Song *et al.*, High-Resolution Comparative Modeling with RosettaCM. *Structure*. **21**, 1735–1742 (2013).

82.   M. L. Oldham, J. Chen, Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci.* **108**, 15152–15156 (2011).

83.   F. DiMaio *et al.*, Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature*. **473**, 540–543 (2011).

84.   D. C. Ekiert *et al.*, A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science*. **333**, 843–850 (2011).

85.   P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science*. **336**, 1030–1033 (2012).

Acknowledgements: (*59*)

Supplementary Materials

Materials and Methods

Fig S1 – S14

Table S1 – S3

References (60 – 85)

Supporting Online Material for

Architecture of a lipid transport system
for the bacterial outer membrane

Damian C. Ekiert[*,§], Gira Bhabha[*], Garrett Greenan, Sergey Ovchinnikov, Jeffery S. Cox, Ronald D. Vale

[*] These authors contributed equally to this work.

[§] To whom correspondence should be addressed. Email: damian.ekiert@ucsf.edu

Materials and Methods

Bacterial strains and growth conditions.
All *E. coli* strains used in this work are derivatives of the wild-type strain MG1655. Mutant strains were constructed by P1 transduction of Kanamycin resistance-marked alleles from an appropriate strain from the Keio collection (*61*).

TEM of *E. coli* outer membrane.
Overnight cultures of MG1655 and mutant strains were diluted 1:100 in LB and grown with shaking at 37˚C until an OD600 of ~0.7. 3µl of culture was spotted onto an electron microscopy grid (Quantifoil 2/2), and cells allowed to attach for ~30 seconds. The grid was blotted for 20 seconds and then plunge frozen in liquid ethane using an FEI Vitrobot Mark III. The grids were imaged using a FEI Polara microscope operating at 300keV, and equipped with a Gatan energy filter and K2 camera. UCSF tomography software was used to collect images of the wild-type and mutant cells.

Expression and purification of MlaD.
DNA corresponding to the "periplasmic domain" (residues 32-183) and "core MCE domain" (residues 32-140) of MlaD was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an N-terminal 6xHis tag followed by a TEV protease cleavage site just upstream of MlaD. The resulting plasmids (pBE1160 and pBE1161, respectively) were transformed into Rosetta 2 cells (Novagen). For expression, overnight cultures of Rosetta 2/pBE1160 or Rosetta 2/pBE1161 were grown at 37° C with shaking to an OD600 of ~0.9, then induced by addition of IPTG to a final concentration of 1 mM and continued incubation overnight shaking at 15° C. Cultures were harvested by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole). Cells were lysed by two passes through an Emulsiflex-C3 cell disruptor (Avestin), then centrifuged at 38,000g to pellet cell debris. The clarified lysates were incubated with NiNTA resin (Qiagen) at 4° C, which was subsequently washed with Ni Wash Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole) and bound proteins eluted with Ni Elution Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole). MlaD containing fractions eluted from the NiNTA column were pooled and concentrated before separation on a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0 and 150 mM NaCl.

Crystallization and structure determination of MlaD.
Gel filtration fractions containing purified MlaD proteins were concentrated to 20-60 mg/mL and sitting drop, vapor-diffusion crystallization trials were conducted using the JCSG Core I-IV screens (Qiagen). Crystals of the core MCE domain of MlaD (residues 32-140) grew from drops consisting of 100 nL protein plus 100 nL of a reservoir solution consisting of 0.1 M sodium acetate pH 4.5, and 40% 1,2 propanediol, and were cryoprotected by supplementing the reservoir solution with 25% ethylene glycol. Native diffraction data was collected at ALS beamline 8.3.1, and indexed to C2 and reduced using XDS (table S1) (*63*).

As MlaD has no homology to proteins of known structure, and has no Met residues in the core MCE domain, we set out to generate a panel of single mutants in which each Leu residue of the core domain was converted to Met (Leu42Met, Leu52Met, Leu73Met, Leu79Met, Leu84Met, Leu99Met, Leu106Met, Leu107Met, Leu112Met, Leu114Met, Leu123Met, and Leu128Met). Of these 12 target mutants, we successfully generated 11 by site directed mutagenesis (reaction for Leu73Met failed). Test expressions revealed that 5 of these 11 mutants expressed at near wild-type levels (Leu79Met, Leu84Met, Leu106Met, Leu107Met, Leu123Met). We then prepared selenomethionine-derivatived protein for each of these five mutants and screened for diffraction-quality crystals, which were obtained

19

for 4 of the 5 mutants. Multi-wavelength anomalous dispersion (MAD) datasets were collected from selenomethionine-derivatived crystals of the Leu84Met and Leu107Met mutants at ALS beamline 8.3.1. Anomalous signal from the Leu107Met crystals was weak despite high data quality and strong Se signal in X-ray fluorescence scans, and later this residue was found to be at the tip of a flexible/partially disordered loop. However, Leu84Met yielded a high quality data set, which was indexed to C2 and reduced using XDS (*63*), and phased by MAD using the SHELX C/D/E pipeline (*64*) (table S1 and S2). The polyalanine model from SHELXE was rebuild using the Autobuild module (*65*) of Phenix. The resulting model was adjusted in Coot (*66*) and refined using Phenix (*67*). The final model consists of 7 copies of the MCE domain forming a hepatmeric ring.

Crystals of the complete MlaD periplasmic domain (residues 32-183) grew from drops consisting of 100 nL protein plus 100 nL of a reservoir solution consisting of 0.2 M zinc acetate, 0.1 M MES pH 6.0, and 15% ethanol, and were cryoprotected by supplementing the reservoir solution with 20% ethylene glycol. Native diffraction data was collected at ALS beamline 8.3.1, and indexed to C2 and reduced using XDS (table S1) (*63*). This data set was phased by molecular replacement using Phaser (*68*), with the MlaD core MCE domain as a search model. The resulting model was adjusted in Coot (*66*) and refined using Phenix (*67*). This final model consists of 3 copies of the MlaD periplasmic domain. Application of the 2-fold symmetry operator results in the biologically relevant hexameric ring.

Purification of the MlaFEDB complex
DNA corresponding to the complete *mlaFEDCB* operon (from MlaF start codon to MlaB stop codon) was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an C-terminal 6xHis tag. The resulting plasmid (pBE1200) was transformed into Rosetta 2 cells (Novagen). For expression, overnight cultures of Rosetta 2/pBE1200 were grown at 37° C with shaking to an OD600 of ~0.9, then induced by addition of IPTG to a final concentration of 1 mM and continued incubation for ~4hr shaking at 37° C. Cultures were harvested by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole). Cells were lysed by two passes through an Emulsiflex-C3 cell disruptor (Avestin), then centrifuged at 35,000g to pellet cell debris. The clarified lysates were transferred to ultracentrifuge tubes and centrifuged at 200,000g to pellet membrane fraction. Membrane pellets were resuspended in lysis buffer, then n-Dodecyl-β-D-Maltopyranoside (DDM; Anatrace) was added to a final concentration of 25 mM. After incubation for ~1hr with rocking at 4° C, the DDM solublized membrane fraction was centrifuged at 200,000g. The supernatants were loaded onto a NiNTA column (Qiagen) at 4° C, which was subsequently washed with Ni Wash Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 0.5 mM DDM) and bound proteins eluted with Ni Elution Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol, 0.5 mM DDM). Eluate from the NiNTA column was concentrated before separation on a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, and 0.5 mM DDM. SDS-PAGE analysis of the resulting sample revealed the presence of four proteins, which were confirmed by trypsin digestion and tandem mass spectrometry (MS/MS) to be MlaF, MlaE, MlaD and MlaB. For cryo-EM, the sample in DDM was exchanged into amphipol A8-35 prior to size exclusion chromatography. Briefly, protein at ~1mg/mL was incubated for ~5 hours with 3-fold excess by mass of A8-35. Detergent from the buffer was then absorbed by incubation with Bio-Beads (Bio-Rad) overnight at 15 mg Bio-Beads per mL of solution. The sample was filtered to remove Bio-Beads and subjected to size exclusion chromatography. The protein quality was then assessed by negative stain EM prior to cryo-EM grid preparation.

Electron microscopy (EM) data collection
Negative stain EM data

To prepare grids for negative stain EM, sample after size exclusion chromatography was applied to a freshly glow discharged carbon coated 400 mesh copper grids and blotted off. Immediately after blotting, a 2% uranyl formate solution was applied for staining and blotted off. The stain was applied five times per sample. Samples were allowed to air dry before imaging. Data were collected on a Tecnai T12 microscope (FEI) equipped with a 4K x 4K CCD camera (UltraScan 4000, Gatan) at a nominal magnification of 52,000 corresponding to a pixel size of 2.21 Å^2/pixel on the specimen, and a defocus range of 1 – 1.7 uM underfocus.

Cryo EM data
For cryo-EM, 3µL of MlaFEDB complex at a concentration of ~1 mg/mL and containing amphipol A8-35 was applied to quantifoil holey carbon grid (1.2/1.3, 400 mesh). Grids were blotted for 6 s at 4 ˚C and plunge frozen in liquid ethane using a Vitrobot Mark III. Cryo-EM images were collected on a Titan Krios, FEI company (Janelia Research Campus, "Krios 2") operated at 300 kV and equipped with a K2 Summit direct electron detector camera (Gatan). Images were recorded using super-resolution counting mode following an established protocol (*69*). Specifically, images were recorded at a nominal magnification of 20,000 x, corresponding to a calibrated super-resolution pixel size of 1.02 Å on the specimen and 0.51 Å for super-resolution images. The dose rate on the camera was set to be › 8 counts (corresponding to › 9.9 electron) per physical pixel per second. The total exposure time was 6 s, leading to a total accumulated dose of 41 electrons per Å^2 on the specimen. Each image was fractionated into 30 subframes, each with an accumulation time of 0.2 s per frame. All dose-fractionated cryo-EM images were recorded using SerialEM (*70*). Images were recorded with a defocus in the range from 1.5 to 3.0 µm.

EM Data processing
Negative stain EM data
Images were binned, resulting in a pixel size of 4.42 Å^2/pixel. Particles were picked automatically using template-based picking. Simplified Application Managing Utilities for EM Labs (SAMUEL, http://liao.hms.harvard.edu/samuel) scripts were used for image preprocessing, particle picking and 2D classification, as described by Ru *et al* in Supplemental Experimental Procedures (*71*).
After a satisfactory stack of particles were obtained, an initial model was obtained using the software simple 2.0 (*72, 73*). 3D classification was performed on the stack in RELION (*74–76*) using the initial model generated from simple 2.0 filtered to 50 Å as an initial model. The data were divided into 4 classes, and each resulting class was refined using the auto-refine procedure in RELION.

Cryo EM data
Images were binned, resulting in a pixel size of 3.03 Å^2/pixel for cryo-EM data of the MlaFEDB complex. Particles were picked automatically using template-based picking. Simplified Application Managing Utilities for EM Labs (SAMUEL, http://liao.hms.harvard.edu/samuel) scripts were used for image preprocessing, particle picking and 2D classification, as described by Ru *et al* in Supplemental Experimental Procedures (*71*). After a satisfactory stack of particles were obtained, 2D classification was also carried out in RELION. The resulting stack of particles was subjected to iteratve 3D classification and the auto-refine procedure in RELION until the most homogeneous set of particles was obtained. Some compositional heterogeneity was present, which resulted in a large amount of top views of MlaD alone (not in complex with MlaE, F and B). These particles were separated out early in procedure. The remaining particles of the MlaFEDB complex also contain conformational heterogeneity, and possible compositional heterogeneity with respect to the bound nucleotide, as the nucleotide state was not controlled in this dataset. No exogenous ATP, ADP or ATP analog was added, so we expect the sample to be largely in the "apo" state. However, it is possible that endogenous

nucleotide may be bound in some particles. Regardless, the orientation between the permease subunit (MlaE), the ATP-binding domains (MlaF) and the potential regulatory subunits (MlaB) shows some flexibility. Therefore, while several 2D classes show detailed features, these are lost in the 3D reconstruction, even after 3D classification. Likely a much larger dataset with a more homogenous sample will be required to reach higher resolution and identify the pseudo-atomic details of the complex.

Identification of inter-subunit restraints in MlaFEDB complex using co-evolving residue pairs.
A multiple sequence alignment was generate using HHblits (*77*) with uniprot database (*78*) from 2015_06 for each of the MLA components. To prevent paralogous copies of each genes from corrupting the final alignments, stringent e-values were used. For MlaE, MlaF and MlaD evalues of 1E-20, 1E-40 and 1E-20 were used respectfully. The joint alignment was restricted to genes found within the same operon (maximum of two genes apart in the same contig). The genomic distance was determined by mapping the uniprot ids back to their respective contiguous sequences from the ENA database (*79*). The alignments were then filtered to remove sequences that don't cover at least 75 % of the query, positions that have more than 50% gaps and redundancy was reduced to 90%. GREMLIN was then run with default parameters to predict the top inter-co-evolving residue pairs (*80*). Residue pairs with predicted probability of 80% or higher are shown in SI figure X mapped on the model structures.

The previously predicted structure of MlaE dimer (*45*) was refined using the rosetta hybrid protocol (*81*) in context of MlaF dimer and co-evolution restraints. The MlaF dimer was modeled using outward-facing Maltose/maltodextrin import ATP-binding protein MalK conformation [3RLF; (*82*)].

Expression and purification of MlaC.
DNA corresponding to mature, signal peptide-cleaved MlaC (residues 22-211) was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an N-terminal 6xHis tag and TEV protease cleavage site. The resulting plasmid (pBE1203) was transformed into Rosetta 2 cells (Novagen). For expression, overnight cultures of Rosetta 2/pBE1203 were grown at 37° C with shaking to an OD600 of ~0.9, then induced by addition of IPTG to a final concentration of 1 mM and continued incubation overnight shaking at 15° C. Cultures were harvested by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole). Cells were lysed by two passes through an Emulsiflex-C3 cell disruptor (Avestin), then centrifuged at 38,000g to pellet cell debris. The clarified lysates were incubated with NiNTA resin (Qiagen) at 4° C, which was subsequently washed with Ni Wash Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole) and bound proteins eluted with Ni Elution Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole). MlaC containing fractions eluted from the NiNTA column were pooled and concentrated before separation on a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0 and 150 mM NaCl.

Crystallization and structure determination of MlaC.
Gel filtration fractions containing purified MlaC proteins were concentrated to 10-60 mg/mL and sitting drop, vapor-diffusion crystallization trials were conducted using the JCSG Core I-IV screens (Qiagen). Crystals of MlaC grew from drops consisting of 100 nL protein plus 100 nL of a reservoir solution consisting of 0.1 M citric acid pH 3.5 and 1.6 M ammonium sulfate, and were cryoprotected by supplementing the reservoir solution with 37.5% ethylene glycol. Native diffraction data was collected at ALS beamline 8.3.1, and indexed to $P2_12_12_1$ and reduced using XDS (table S1) (*63*). The structure was phased by molecular replacement using Phaser (*68*). Initial attempts to use the crystal

structures of MlaC homologs as search models (PDB IDs: 2QGU and 4FCZ) were unsuccessful. However, we were able to obtain a few clear solutions using the following approach. We created a homology model of MlaC with Phyre2 (*44*), based on 2QGU as a template. This homology model rebuilt using MR-Rosetta (*83*), and the resulting 1000 rebuilt models were tested as search models using Phaser (*68*). A small number of these models gave clear solutions, which were adjusted in Coot (*66*) and refined using Phenix (*67*). In the later stages of refinement, two phosphatidic acid molecules were built into clearly defined electron density in the core of each of the two MlaC monomers in the asymmetric unit.

Expression and purification of MlaA-OmpF complex.
DNA corresponding to mature, signal peptide-cleaved MlaA (residues 1-251) was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an C-terminal 6xHis tag. The resulting plasmid (pBE1242) was transformed into Rosetta 2 cells (Novagen). For expression, overnight cultures of Rosetta 2/pBE1242 were grown at 37° C with shaking to an OD600 of ~0.9, then induced by addition of IPTG to a final concentration of 1 mM and continued incubation for ~4hr shaking at 15° C. Cultures were harvested by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol). Cells were lysed by two passes through an Emulsiflex-C3 cell disruptor (Avestin), then centrifuged at 38,000g to pellet cell debris. The clarified lysates transferred to ultracentrifuge tubes and centrifuged at 200,000g to pellet membrane fraction. Membrane pellets were resuspended in lysis buffer by sonication, then n-Dodecyl-β-D-Maltopyranoside (DDM; Anatrace) was added to a final concentration of 25 mM. After incubation for 1hr with rocking at 4° C, the DDM solublized membrane fraction was centrifuged at 200,000g. The supernatants were loaded onto a NiNTA column (Qiagen) at 4° C, which was subsequently washed with Ni Wash Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol, 0.5 mM DDM) and bound proteins eluted with Ni Elution Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol, 0.5 mM DDM). MlaA containing fractions eluted from the NiNTA column were pooled and concentrated before separation on a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, and 0.5 mM DDM. SDS-PAGE analysis of the resulting sample revealed the presence of two proteins in roughly 1:1 stoichometry, which were confirmed by trypsin digestion and tandem mass spectrometry (MS/MS) to be MlaA and OmpF.

Protein-protein interactions by biolayer interferometry.
DNA corresponding to mature, signal peptide-cleaved MlaC (residues 22-211) was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an N-terminal 6xHis tag, biotinylation site (amino acid sequence: GGGLNDIFEAQKIEWHE), and TEV protease cleavage site. The resulting plasmid (pBE1230) was transformed into Rosetta 2 cells (Novagen), and tagged MlaC was expressed and purified essentially as described above. Purified MlaC at ~2 mg/mL was biotinylated by the addition of 25μg BirA enzyme/mg total protein, purified essentially as previous described (*84*). The reactions were carried out in the following buffer: 100mM Tris pH 8.0, 10mM ATP, 10mM MgOAc, 50μM biotin. The biotinylation reactions were incubated at 4 °C overnight. Biotinylated MlaC was purified by size exclusion chromatography, and concentrated to ~5-20 mg/mL.

Biotinylated MlaC was diluted to 50 μg/mL in 1x kinetics buffer (1x PBS, pH 7.4, 0.01% BSA, and 0.002% Tween 20) and loaded onto streptavidin coated biosensors (ForteBio) and incubated with varying concentrations of MlaFEDB complex, MlaD, or MlaA-OmpF. All binding data were collected on an Octet RED384 instrument at 25 °C. In each case, an empty sensor (no MlaC loaded) was included as a negative control to test for non-specific interactions of MlaFEDB complex, MlaD, and

23

MlaA-OmpF.

Expression and purification of PqiB.

DNA corresponding to mature, signal peptide-cleaved PqiB (residues 39-546) was amplified from the MG1655 genome and cloned by Gibson assembly (*62*) into a custom pET vector resulting in an N-terminal 6xHis tag and TEV protease cleavage site. The resulting plasmid (pBE1163) was transformed into Rosetta 2 cells (Novagen). For expression, overnight cultures of Rosetta 2/pBE1163 were grown at 37° C with shaking to an OD600 of ~0.9, then induced by addition of IPTG to a final concentration of 1 mM and continued incubation overnight shaking at 15° C. Cultures were harvested by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole). Cells were lysed by two passes through an Emulsiflex-C3 cell disruptor (Avestin), then centrifuged at 38,000g to pellet cell debris. The clarified lysates were incubated with NiNTA resin (Qiagen) at 4° C, which was subsequently washed with Ni Wash Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole) and bound proteins eluted with Ni Elution Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole). PqiB containing fractions eluted from the NiNTA column were pooled and concentrated before separation on a Superdex 200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0 and 150 mM NaCl.

24

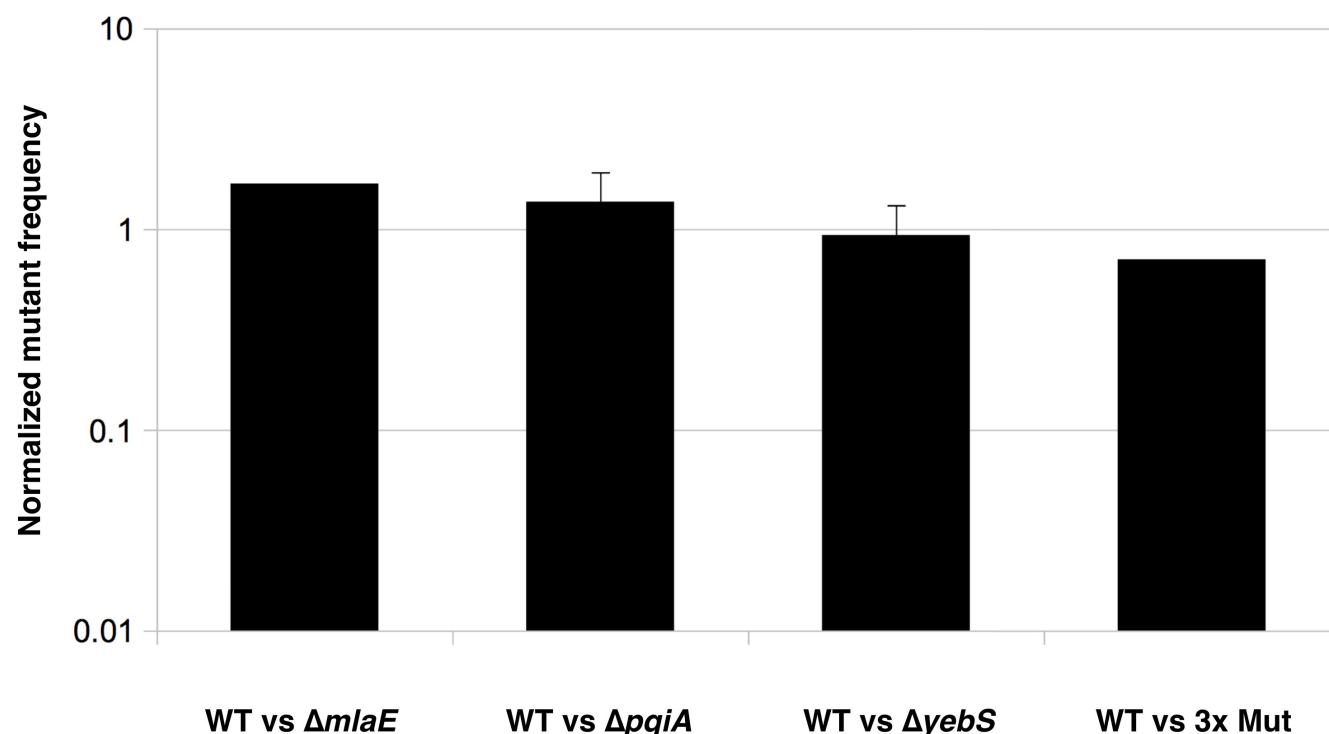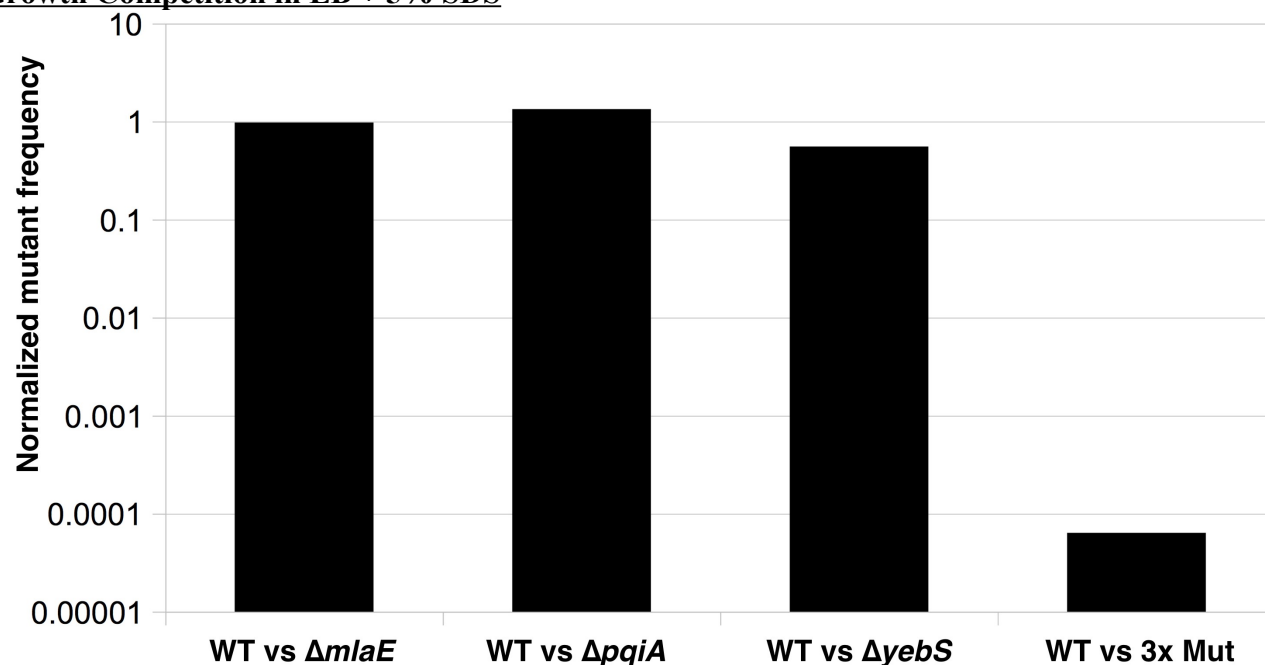**Fig. S1. Growth competition between WT and MCE mutant *E. coli* strains in rich media.** After mixing equal volumes of WT and MCE mutant overnight cultures, mixed cultures were diluted 1:1000 in fresh LB and incubated ~16hr at 37° C with shaking. The fraction of mutant cells (mutant cells divided by total cells) was estimated before and after growth competition by plating dilutions on LB (total cells) and LB+Kan (mutant cells).   The ratios presented here represent the change in the fraction of mutant cells between the initial mixed culture and the post-competition culture [(fraction mutant T_{16hr}) / (fraction mutant T_0)]. Values above 1 indicate a growth advantage in the mutant, while values below 1 indicate a growth defect in the mutant. For all the mutants, the frequency of mutant cells changed by less than 2-fold between the input and post-competition cultures, suggesting MCE genes are dispensable for in vitro growth in rich media.

Fig. S2. Growth competition between WT and MCE mutant *E. coli* strains in the presence of SDS or CaCl$_2$. After mixing equal volumes of WT and MCE mutant overnight cultures, mixed cultures were diluted 1:1000 in fresh LB+5% SDS or LB+100 mM CaCl$_2$, incubated ~16 hr at 37C with shaking, and analyzed as described in fig. S1. Addition of 5% SDS or 100 mM CaCl$_2$ had little effect on the growth of any of the single mutants, but dramatically reduces the frequency of the triple mutant.

Growth Competition in LB + 5% SDS

Growth Competition in LB + 100 mM CaCl$_2$
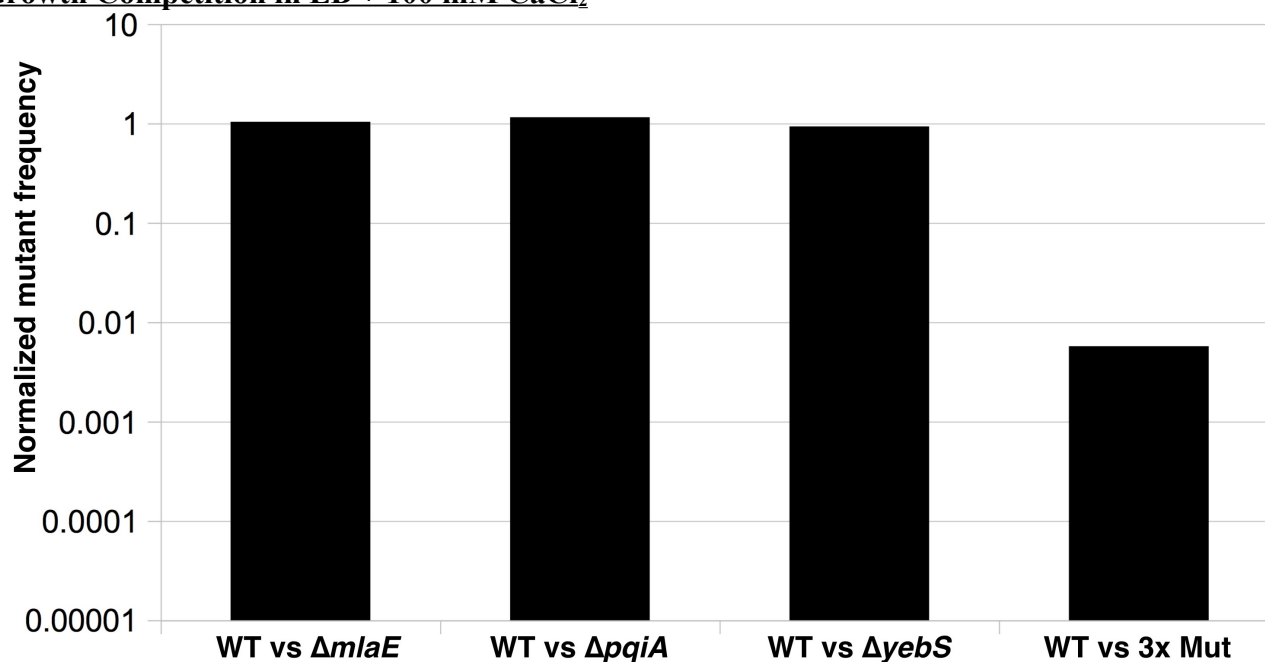
Fig. S3. Vital staining of wt and MCE mutant cells. Live cells growing exponentially in LB were co-stained with DAPI and PI for 30 minutes then imaged. The culture of the triple mutant had more PI-positive cells than the wild-type, though the vast majority of cells in both cultures were PI-negative, indicating that the inner membrane remains intact. DAPI staining of the triple mutant was brighter than the wild-type, with clear nucleoid localization, while the staining of wild-type cells under these conditions was strongest at the cell periphery (cell wall/membrane staining). Consistent with the role of the OM as a barrier to the entry of small molecules into the cell, DAPI entry into WT cells appears to be inefficient, while in the triple mutant the OM is compromised, allowing efficient nucleoid staining.
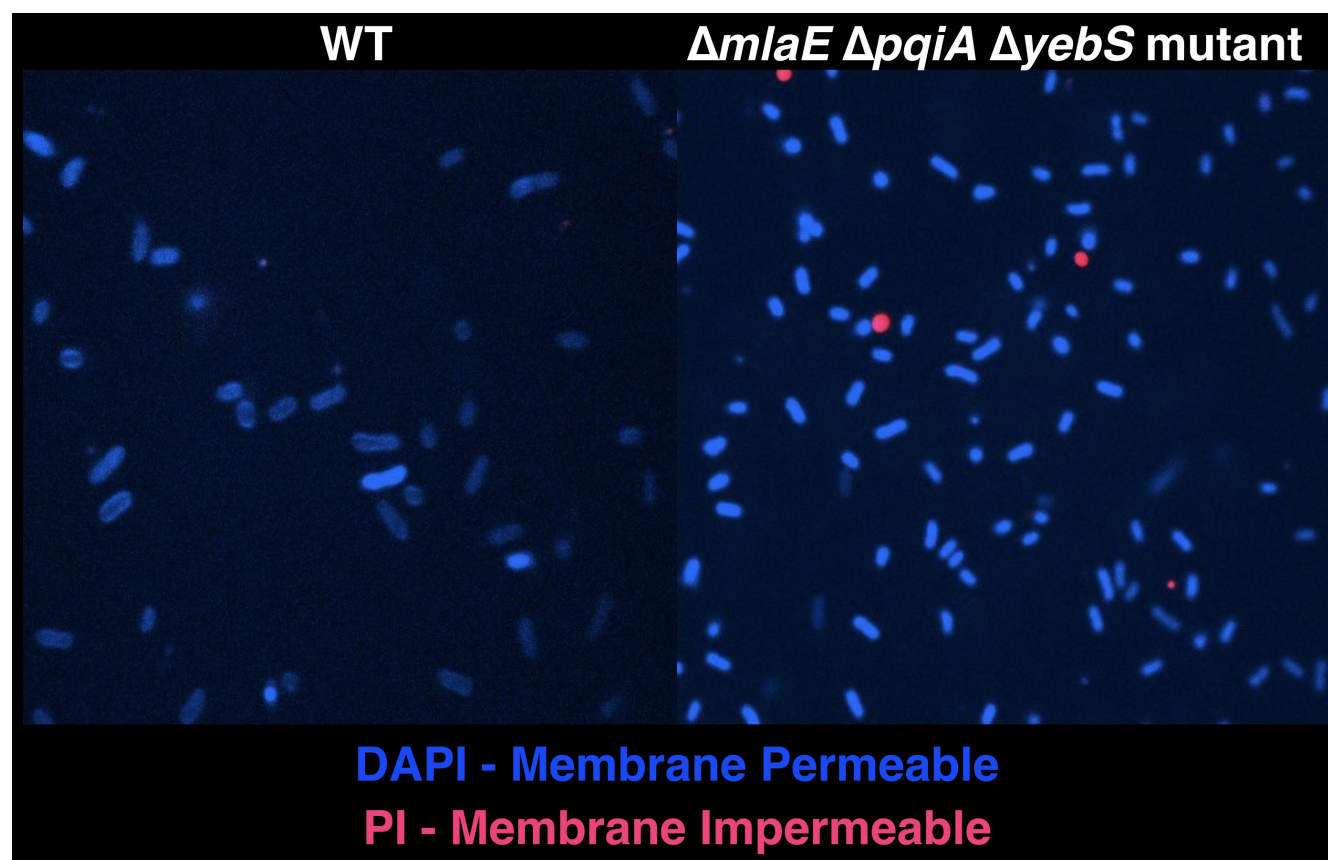
Fig. S4. Initial electron density map for MlaD[core] [32-140] after phasing and density modification using SHELX C/D/E pipeline. $2F_O$-F_C electron density map displayed a blue mesh, contoured at 2σ. The Cα trace of the final model is shown in red for comparison.
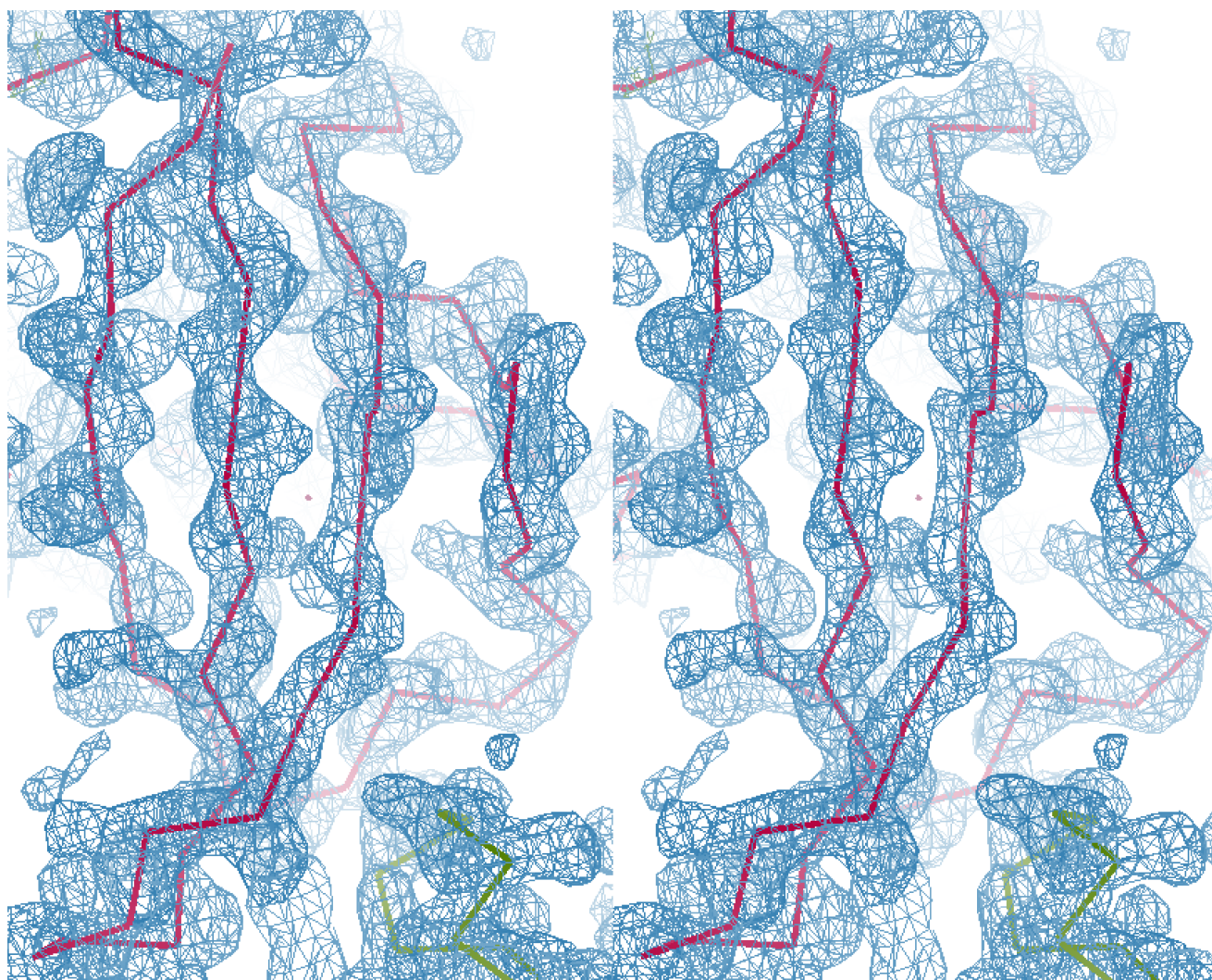
Fig. S5. Closest structural neighbors of MCE domain identified by DALI. Topology diagrams highlighting the differences between the MCE domain and the closest structural neighbors in the PDB. Adapted from diagrams generated by Pro-Origami (*60*)).
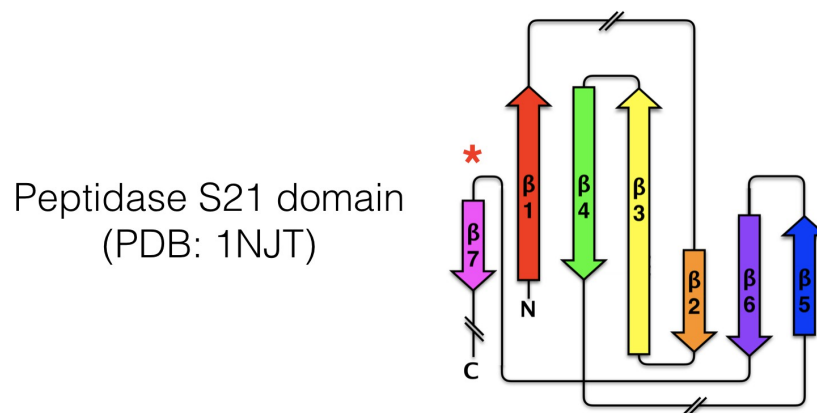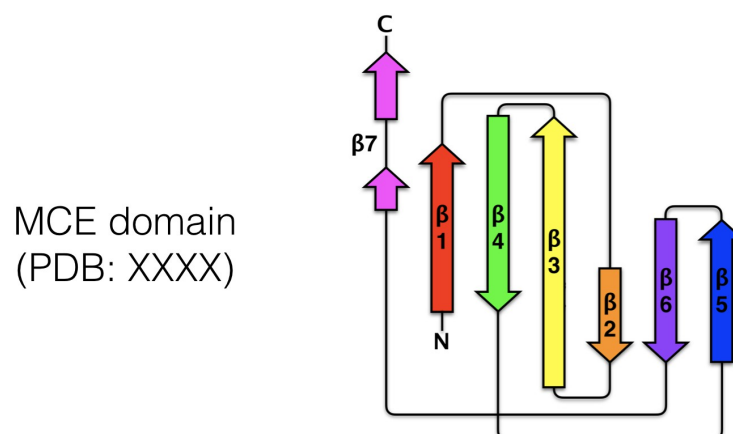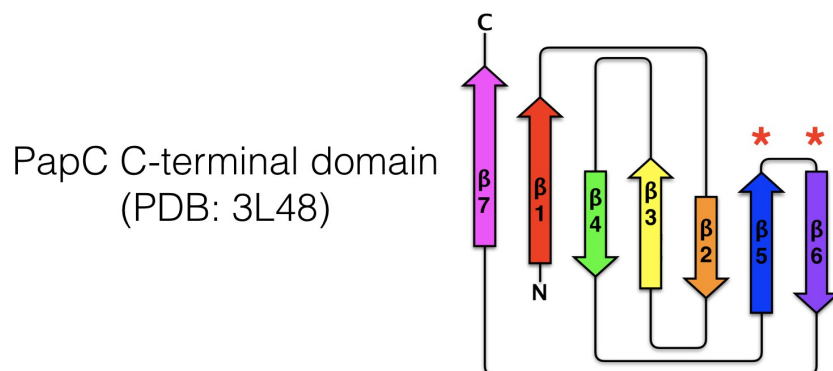
Fig. S6. **Comparison of hexameric and 14-mer MlaD structures.** **(A)**, the full-length MlaD ectodomain, which includes an additional ~45 amino acids at the C-terminus that associate to form a helical bundle, forms a single hexameric ring. **(B)**, in contrast, the core MCE domain forms a tetradecamer, consisting of two rings of seven MCE domains stacked on top of one another with their pores roughly aligned (subunits of upper ring colored, bottom ring gray). **(C)**, 90° rotated view of the full-length, hexameric MlaD ectodomain. **(D)**, 90° rotated view of the tetradecameric MlaD core MCE domain, colored as in **(B)**. **(E)**, 2D class averages from single particle, negative stain electron microscopy are also consistent with a 14 subunit, stacked ring organization for the MCE core domain from MlaD, suggesting that the 14-mer observed in the crystal is simply due to crystal packing. While it is conceivable that MlaD can assume more than one oligomeric state, we believe the hexamer is the predominant and perhaps the only functional form *in vivo*, for three reasons. First, the longer, more complete periplasmic domain crystal structure is hexameric, as is the full-length protein solubilized from membranes, and it is also hexameric in the context of the larger MlaFEDB complex. Second, PqiB, another MCE protein from *E. coli*, is only hexameric. Third, the "head-to-head" arrangement of the two stacked rings in the 14-mer places the C-termini from each protomer near the interface, which would result in major steric overlap between the helical bundles observed in the hexameric and MlaFEDB structures.
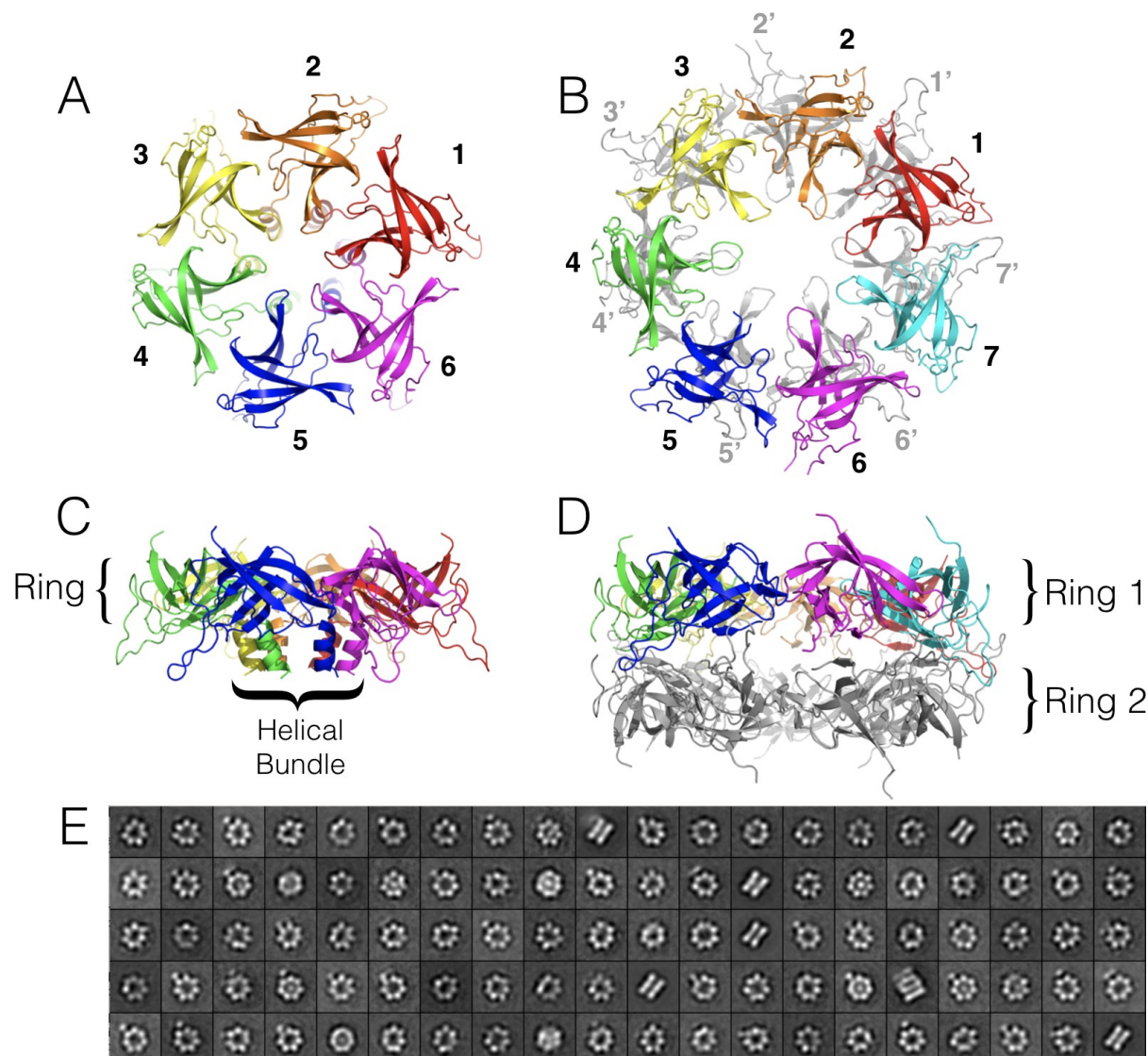
Fig. S7. Organization of MCE gene cluster in *M. tuberculosis*. The M. tuberculosis genome has 4 MCE operons, each encoding 6 MCE genes.

Operon 1 mce1A mce1B mce1C mce1D mce1E mce1F

Operon 2 mce2A mce2B mce2C mce2D mce2E mce2F

Operon 3 mce3A mce3B mce3C mce3D mce3E mce3F

Operon 4 mce4A mce4B mce4C mce4D mce4E mce4F

Fig. S8. **Cryo EM of the MlaFEDB complex**. **(A)** Representative raw micrograph of MlaFEDB using a K2 camera. **(B)** Gold standard FSC curve of final reconstruction. Horizontal line indicates 0.143 FSC.
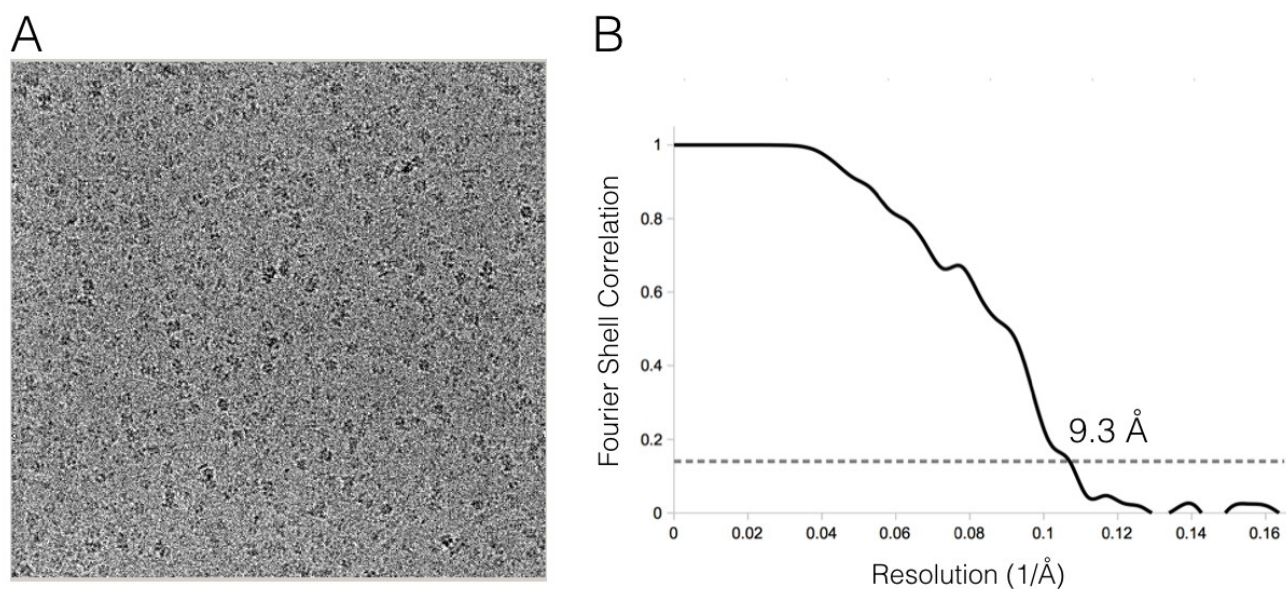
Fig. S9. Location of strongly co-evolving residue pairs in the MlaFEDB complex. Only co-evolving pairs with a probability > 0.8 are shown. **(A),** Co-evolving residue pairs suggestive of potential contacts between MlaD (left column) and MlaE (right column). **(B)**, Co-evolving residue pairs suggestive of potential contacts between MlaE (left column) and MlaF column). N-terminal helix of MlaE not shown for clarity and the complex was pulled apart to reveal the contacts.
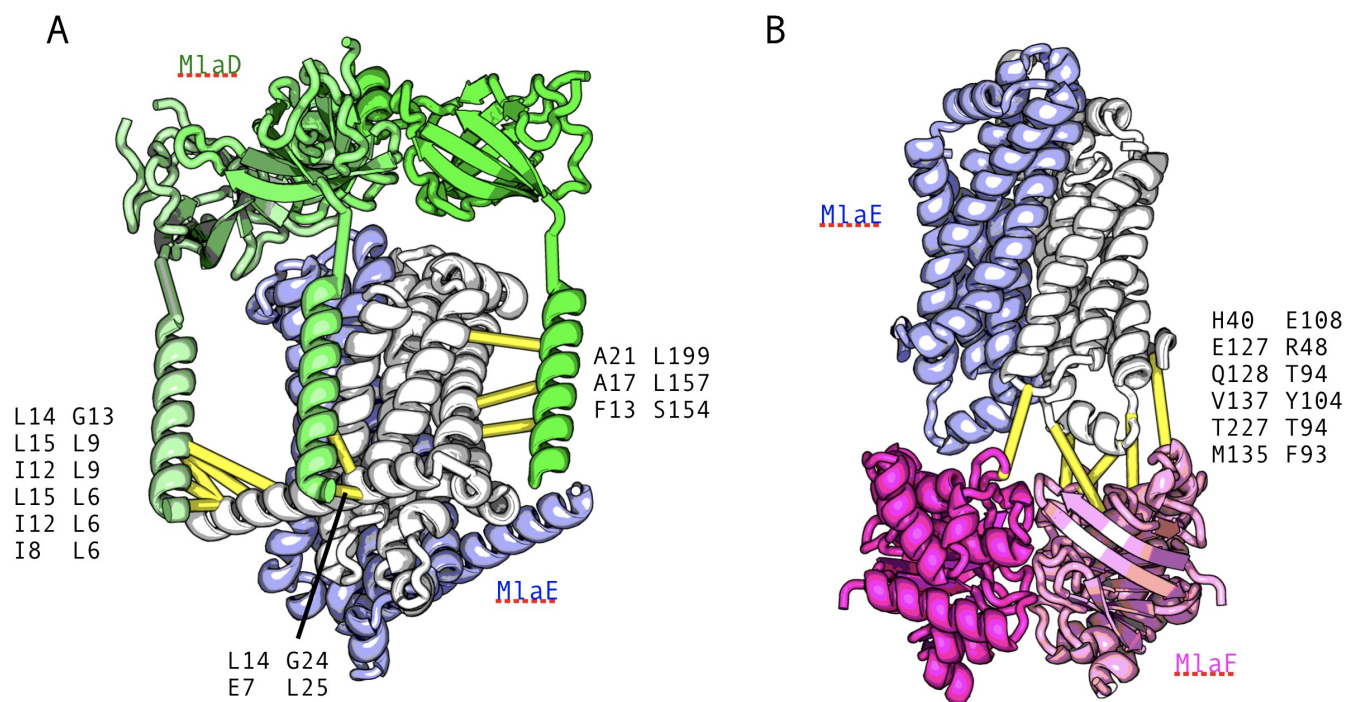
Fig. S10. Comparison of MlaC with cystatin and other NTF2 superfamily members. (A), In contrast to cystatins, where a single α-helix (red) packs tightly against one face of the β-sheet, the NTF2 superfamily extends the sheet with 2 additional β-strands while adding two α-helices, creating a hydrophobic pocket between the sheet and helices. MlaC adds several addition helices to this core cystatin/NTF2 fold. **(B)**, The hydrophobic pocket varies in volume across the diverse members of the NTF2 superfamily, from a modest depression in NTF2 itself that mediates interaction with Ran-GDP (dark blue bars), to a prominent active site cleft in many enzymes where it can accommodate small hydrophobic substrates, such as steroids in the ketosteroid isomerases (green bars). Compared with other members of the NTF2 superfamily, MlaC and the related 2QGU and 4FCZ protein structures have some of the largest cavities observed to date (red bars).
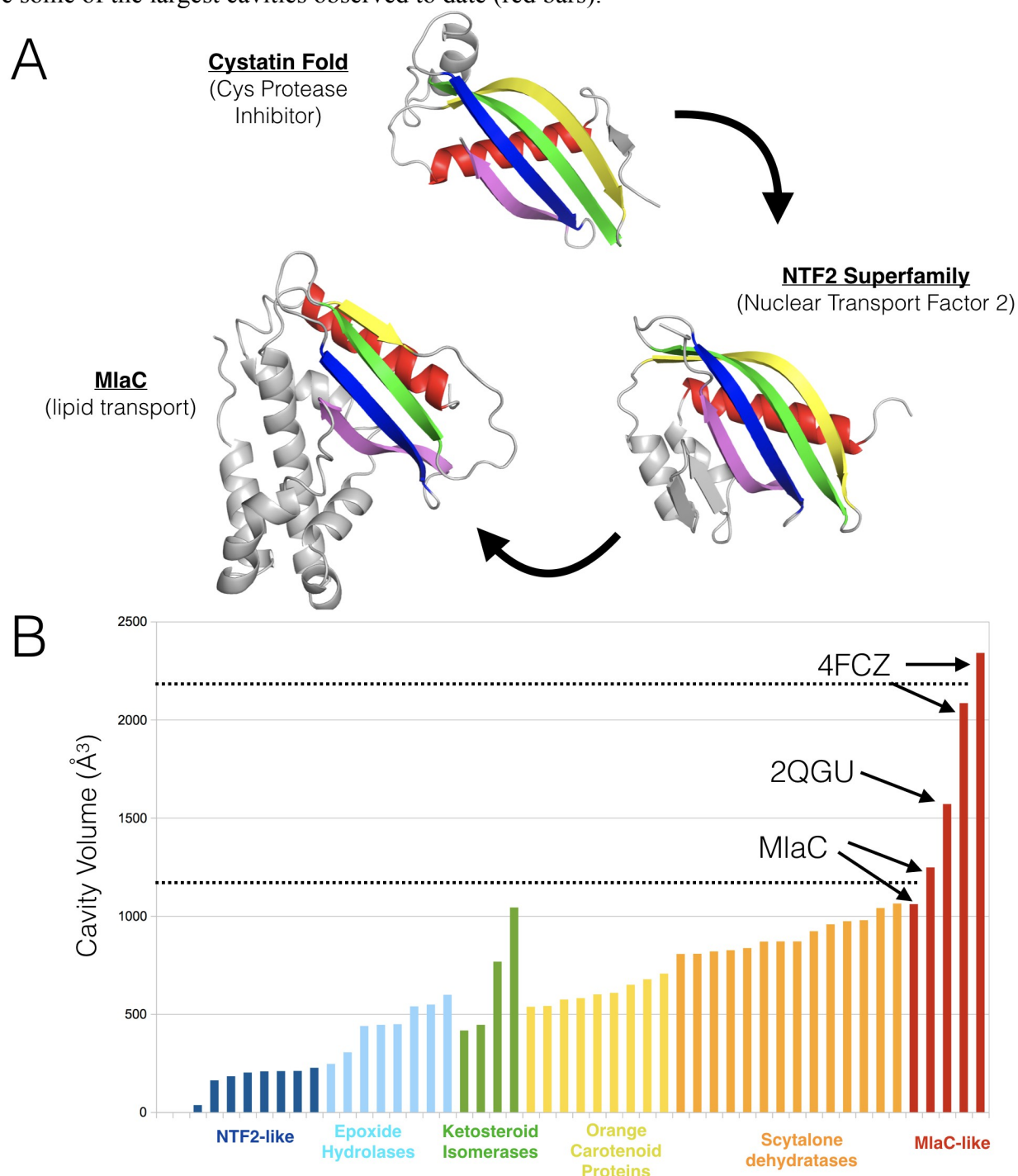
Fig. S11. All three MlaC-like structures determined to date have bound lipid. 2Fo-Fc electron density map contoured at 1 sigma for MlaC **(A)**, 2QGU **(B)**, and 4FCZ **(C)**, showing only the density corresponding to the bound lipids. In **(C)**, the density was segmented and colored red and blue to indicate the regions corresponding to the two putative diacylglycerol-like groups. The bound lipids are similarly oriented in each structure [white spheres for MlaC **(D)** and 2QGU **(E)**; and green and cyan spheres for the two diacylglycerol-like groups of 4FCZ **(F)**].
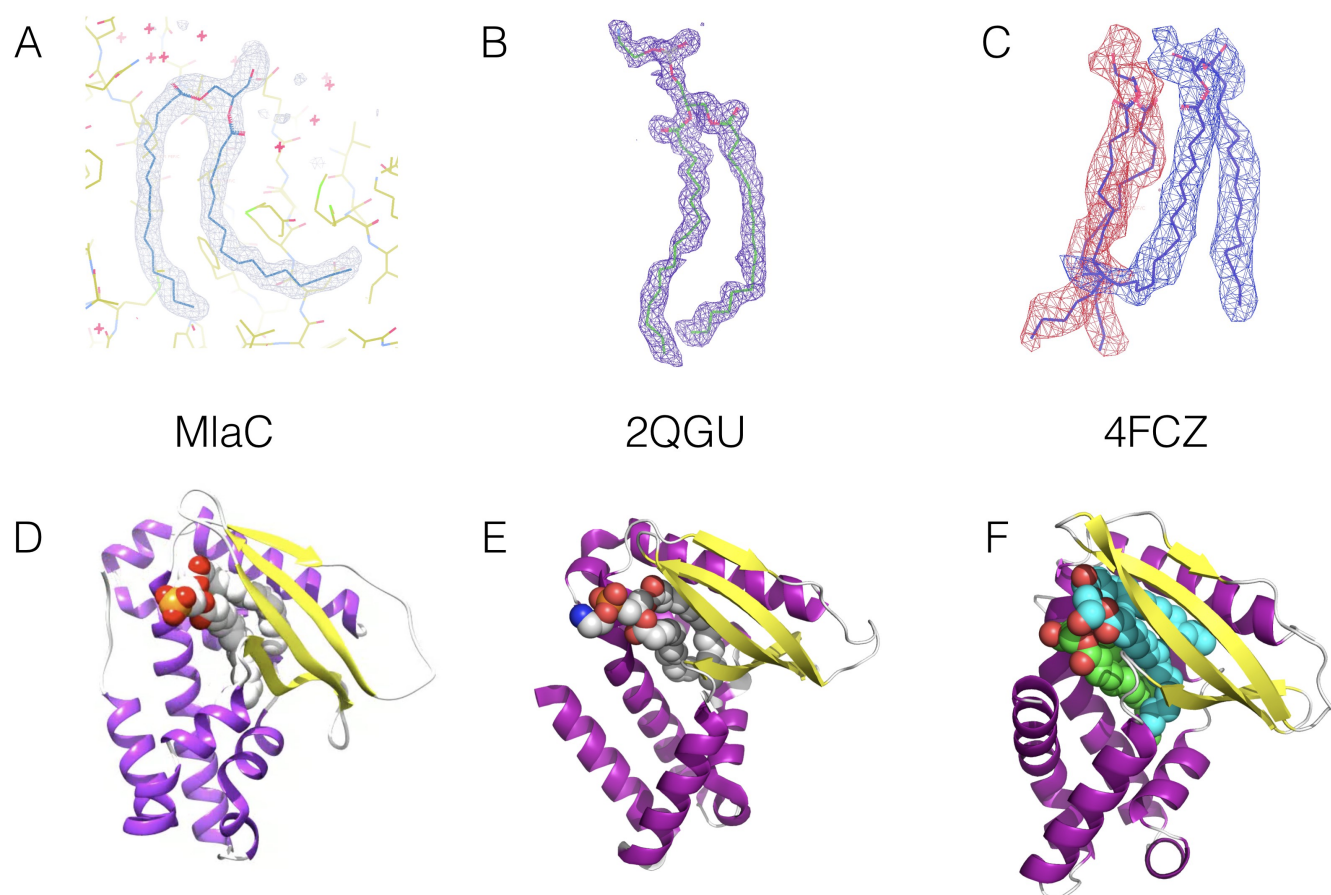
Fig. S12. Binding of MlaD to MlaC. MlaD interacts directly with biotinylated MlaC protein immobilized on a streptavidin-coated biosensor over a range of MlaD concentrations, as assayed by biolayer interferometry.
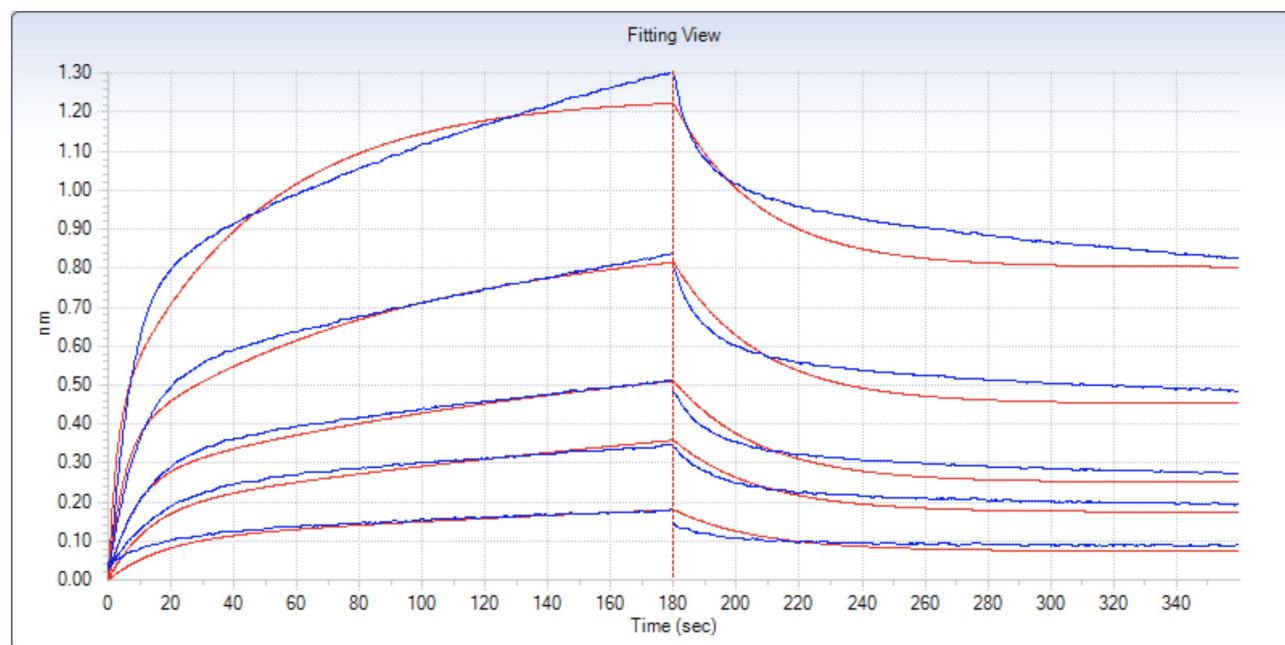
Fig. S13. **The C-terminal domain of PqiB forms a needle-like projection**. **(A)**, Schematic of constructs that were used. Top, full length construct of PqiB periplasmic domain; Bottom, construct in which the C-terminal domain, predicted to be helical, has been truncated. **(B)**, Representative single particles from negative stain EM data for the constructs shown in **(A)**. Yellow arrows point to the density corresponding to a needle-like projection, clearly identifiable in Fig. 5B. This density is missing from the C-terminal truncated construct, suggesting that the C-terminal domain forms this needle-like projection.
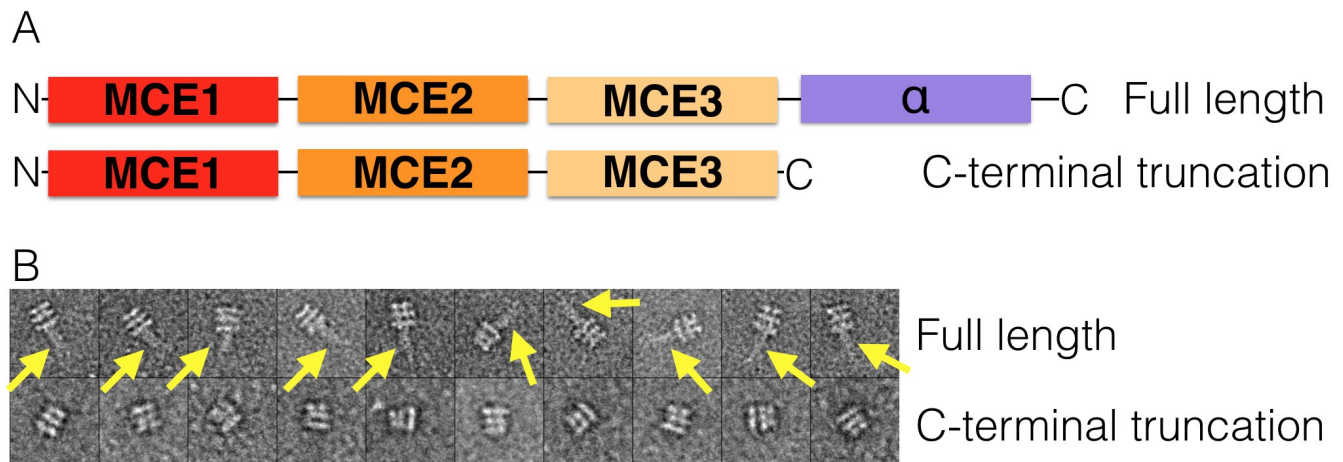
Fig. S14. Possible architectures of MCE domains proteins from M. tuberculosis and chloroplasts.
(A), The chloroplasts of plants and other photosynthetic eukaryotes contain a single MCE gene encoding a protein with a single MCE domain followed by a long C-terminal helical region (called TGD2 in *A. thaliana*). Based on the architecture of PqiB, we predicted that TGD2 will form a single homo-hexameric ring with a long needle-like projection. **(B)**, Like the chloroplast proteins MCEs from *M. tuberculosis* also encode a single MCE domain followed by a long C-terminal helical region, and likely adopt a similar architecture. However, since mycobacterial MCE operons usually encode 6 MCE proteins, we predict that these proteins will form hetero-hexamers with 6 unique chains.
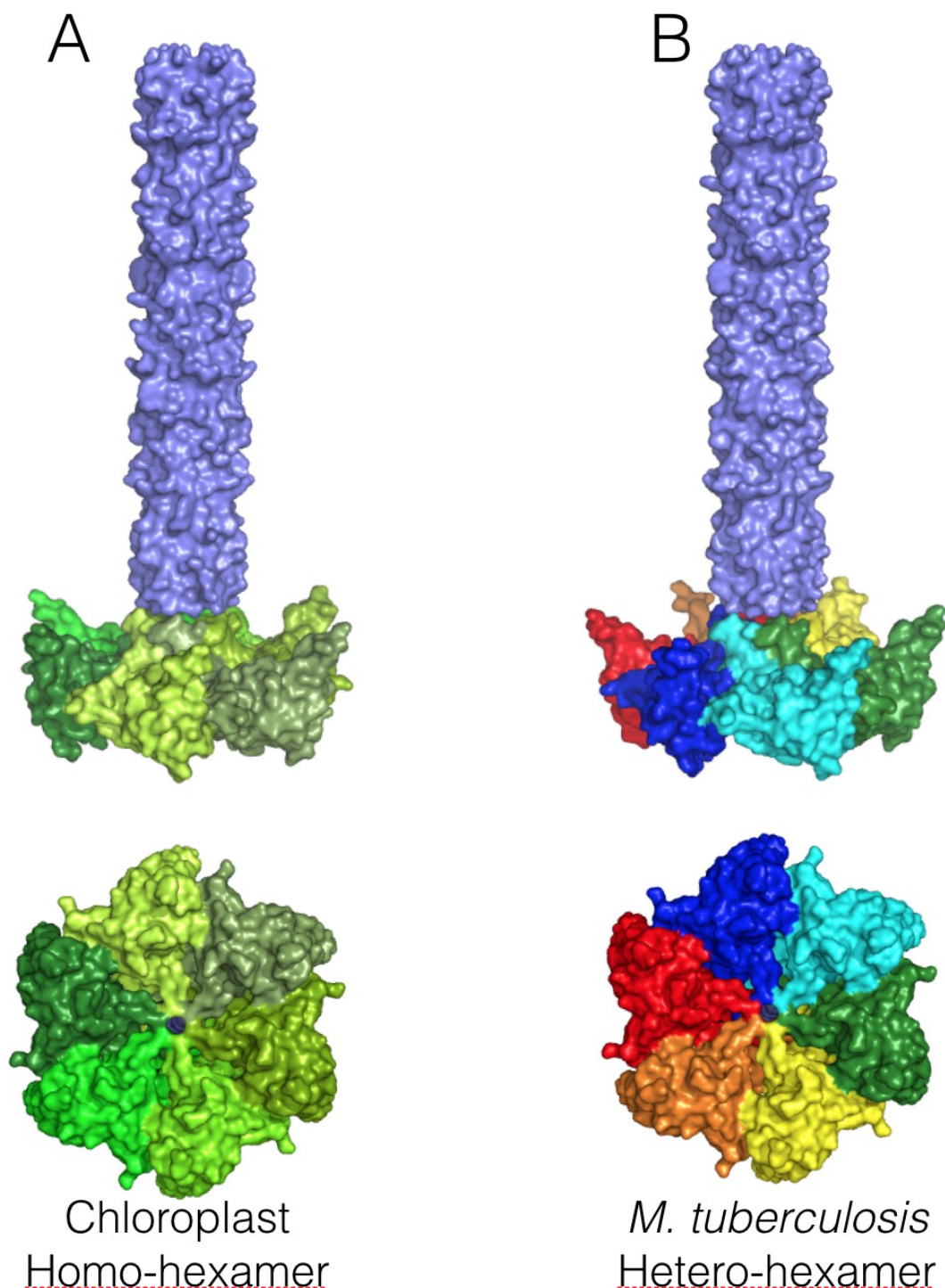

Chloroplast
Homo-hexamer

M. tuberculosis
Hetero-hexamer

Table S1. Crystallographic data collection and refinement statistics.

	MlaD^{core} [32-140]	MlaD^{peri} [32-183]	MlaC
Data collection			
Space group	C2	C2	P2$_1$2$_1$2$_1$
Cell dimensions			
a, b, c (Å)	139.7, 103.8, 77.2	122.4, 63.8, 91.2	33.6, 115.9, 133.2
α, β, γ (°)	90.0, 111.0, 90.0	90.0, 130.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	1.116	1.116	1.116
Resolution (Å)	50-2.15 (2.21-2.15)[1]	50-2.85 (2.92-2.85)[1]	50-1.50 (1.54-1.50)[1]
Observations	421,073	46,094	599,263
Unique reflections	55,590	12,551	83,295
Redundancy	7.6 (7.6)	3.7 (3.7)	7.2 (7.1)
Completeness (%)	99.3 (98.6)	97.8 (98.9)	98.5 (96.1)
CC*[2]	1.00 (0.80)	1.00 (0.79)	1.00 (0.85)
CC1/2[2]	0.99 (0.46)	1.00 (0.47)	1.00 (0.57)
I / σI	15.4 (0.9)	14.9 (1.0)	22.2 (1.5)
R_{meas}	0.09 (2.46)	0.05 (1.76)	0.05 (1.40)
R_{pim}	0.04 (0.88)	0.03 (0.91)	0.02 (0.51)
Refinement			
Resolution (Å)	50-2.15	50-2.85	50-1.50
Reflections (work)	48,118	11,005	77156
Reflections (free)	1810	589	1,894
R_{work} (%) / R_{free} (%)	20.7 / 24.4	23.3 / 28.8	16.9 / 18.5
No. atoms			
Protein	5,552	2,513	3,147
Ligand/ion	0	6	80
Water	126	0	264
B-factors			
Protein	70.3	124.7	29.9
Ligand/ion	na	170.8	52.4
Water	53.6	na	39.3
R.m.s. deviations			
Bond lengths (Å)	0.003	0.003	0.009
Bond angles (°)	0.94	0.79	1.23
Ramachandran			
Favored	97.8	97.5	99.0
Outliers	0.7	0.3	0.0
PDB Code [3]	aaaa [3]	bbbb [3]	cccc [3]

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

[2] See Karplus and Diederichs (*85*).

[3] Coordinates and structure factors will be deposited in the Protein Databank and released upon publication.

Table S2. Crystallographic data collection and MAD phasing statistics for MlaD^{core} [32-140] crystal structure.

	MlaD^{core} [32-140]	
Data collection		
Space group	C2	
Cell dimensions		
a, b, c (Å)	139.3, 103.6, 77.6	
α, β, γ (°)	90.0, 111.1, 90.0	
	Peak	Remote
Wavelength (Å)	0.980	0.957
Resolution (Å)	50-2.80 (2.87-2.80) [1]	50-2.80 (2.87-2.80) [1]
Observations	962,235	962,435
Unique reflections	49,671	49,681
Redundancy	19.4 (19.5)	19.4 (19.5)
Completeness (%)	99.6 (99.9)	99.6 (99.9)
CC1/2 [2]	1.00 (0.67)	1.00 (0.72)
$I / \sigma I$	11.9 (1.6)	11.8 (1.7)
R_{meas}	0.32 (2.39)	0.30 (2.25)
FOM	0.55	

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

[2] See Karplus and Diederichs (*85*).

Table S3. Crystallographic data collection and refinement statistics for re-refined 4FCZ.

	4FCZ (original)	4FCZ (re-refined)
Data collection		
Space group	P2$_1$	
Cell dimensions		
a, b, c (Å)	79.9, 40.9, 82.4	
α, β, γ (°)	90.0, 106.5, 90.0	
Wavelength (Å)	0.9790	
Resolution (Å)	39.1-2.60 (2.69-2.60)[1]	
Observations	154,940	
Unique reflections	28,693	
Redundancy	5.4 (na)	
Completeness (%)	94.7 (56.0)	
CC*[2]	na	
CC1/2[2]	na	
I / σI	na (2.2)	
R_{merge}	0.11 (na)	
R_{pim}	na	
Refinement		
Resolution (Å)	39.1-2.60	39.1-2.60
Reflections (work)	13,229	14,060
Reflections (free)	1,574	741
R_{work} (%) / R_{free} (%)	23.3 / 29.8	22.3 / 26.7
No. atoms		
Protein	2,862	2,947
Ligand/ion	0	160
Water	41	0
B-factors		
Protein	39.4	49.7
Ligand/ion	-	71.3
Water	31.9	-
R.m.s. deviations		
Bond lengths (Å)	0.009	0.003
Bond angles (°)	1.19	0.60
Ramachandran		
Favored	93.7	97.1
Outliers	1.4	0.8
PDB Code [3]	4FCZ	bbbb [3]

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

[2] See Karplus and Diederichs (*85*).

[3] Coordinates and structure factors will be deposited in the Protein Databank and released upon publication.