

Structure of an endogenous mycobacterial MCE lipid transporter

<https://doi.org/10.1038/s41586-023-06366-0>

Received: 24 December 2022

Accepted: 22 June 2023

Published online: 26 July 2023

 Check for updates

James Chen¹, Alice Fruhauf¹, Catherine Fan¹, Jackeline Ponce², Beatrix Ueberheide^{2,3,4}, Gira Bhabha¹✉ & Damian C. Ekiert^{1,5}✉

To replicate inside macrophages and cause tuberculosis, *Mycobacterium tuberculosis* must scavenge a variety of nutrients from the host^{1,2}. The mammalian cell entry (MCE) proteins are important virulence factors in *M. tuberculosis*^{1,3}, where they are encoded by large gene clusters and have been implicated in the transport of fatty acids^{4–7} and cholesterol^{1,4,8} across the impermeable mycobacterial cell envelope. Very little is known about how cargos are transported across this barrier, and it remains unclear how the approximately ten proteins encoded by a mycobacterial *mce* gene cluster assemble to transport cargo across the cell envelope. Here we report the cryo-electron microscopy (cryo-EM) structure of the endogenous *Mce1* lipid-import machine of *Mycobacterium smegmatis*—a non-pathogenic relative of *M. tuberculosis*. The structure reveals how the proteins of the *Mce1* system assemble to form an elongated ABC transporter complex that is long enough to span the cell envelope. The *Mce1* complex is dominated by a curved, needle-like domain that appears to be unrelated to previously described protein structures, and creates a protected hydrophobic pathway for lipid transport across the periplasm. Our structural data revealed the presence of a subunit of the *Mce1* complex, which we identified using a combination of cryo-EM and AlphaFold2, and name *LucB*. Our data lead to a structural model for *Mce1*-mediated lipid import across the mycobacterial cell envelope.

M. tuberculosis, the causative agent of tuberculosis, is one of the leading causes of death due to infectious disease, resulting in over 1 million deaths annually⁹. *M. tuberculosis* establishes a niche within the phagosomal compartment of host macrophages, in which it can grow and replicate. To survive in the phagosome, *M. tuberculosis* must scavenge nutrients from the host cell^{1,2}, and uses an ensemble of active transporters to import iron^{10,11}, lipids^{1,2} and other metabolites¹². In particular, the MCE family of proteins has been implicated in the import of lipid substrates such as fatty acids^{4–7} and cholesterol^{1,4} across the cell envelope of *M. tuberculosis* and related species such as *M. smegmatis*^{8,13–16} (Fig. 1a). MCE proteins are critical for virulence in *M. tuberculosis*^{1,3} and other bacterial pathogens^{17–20}, underscoring their fundamental importance during infection. To mediate the uptake of lipids, MCE transporters must translocate substrates across the impenetrable cell envelope, which consists of (1) the inner membrane (IM); (2) the complex mycobacterial outer membrane (MOM); and (3) a periplasmic space between the IM and MOM, containing the cell wall²¹. In Gram-negative bacteria, many cargos are transported by large transenvelope protein-based machines that mediate the passage of substrates across membranes and the periplasmic space. These protein machines include the LPS-export system^{22–25}, antibiotic efflux pumps²⁶ and a variety of specialized protein secretion systems²⁷. By contrast, it is unclear how substrates are transported across the highly divergent mycobacterial cell envelope, and whether such periplasm-spanning complexes exist in mycobacteria.

In *M. tuberculosis*, MCE transport systems are encoded by four different gene clusters, *mce1*–*mce4*, which are among the largest operons in the genome (Extended Data Fig. 1a). Each cluster has a core module of eight conserved genes: (1) two *yrbE* genes encoding the transmembrane (TM) subunits of an ATP-binding-cassette (ABC) transporter and (2) six genes encoding MCE proteins. A variable number of accessory proteins are often found adjacent to the eight-gene core module²⁸. Additional proteins encoded elsewhere in the genome are also required for *M. tuberculosis* MCE transporter function, including those encoding an ATPase, *MceG*^{1,29,30} and the integral membrane protein *LucA*^{4–6}. This gene organization is conserved in other mycobacterial species, including *M. smegmatis*²⁸ (Fig. 1b and Extended Data Fig. 1b). The proteins from each gene cluster are thought to interact with each other to form large complexes¹³ that have been proposed to facilitate transport across the cell envelope. The *Mce1* system has been shown to be important for the uptake^{4–6,14,31} and use^{4–6,14} of fatty acids, leading to a model in which *Mce1* has a critical role in the acquisition of key nutrients from the host. *M. tuberculosis* *mce1* mutants also accumulate free mycolic acids in the cell envelope⁷, suggesting a transport defect and a possible role in cell envelope maintenance, reminiscent of roles of some MCE systems in Gram-negative bacteria^{32–34}. Unravelling the exact function and mechanism of *Mce1* has been challenging, as recombinant expression and purification of MCE complexes is hampered by the complexity of their genetic organization, and studies to date have been

¹Department of Cell Biology, New York University School of Medicine, New York, NY, USA. ²Proteomics Laboratory, Division of Advanced Research Technologies, New York University School of Medicine, New York, NY, USA. ³Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA. ⁴Department of Neurology, New York University School of Medicine, New York, NY, USA. ⁵Department of Microbiology, New York University School of Medicine, New York, NY, USA. ✉e-mail: gira.bhabha@gmail.com; damian.ekiert@ekiertlab.org

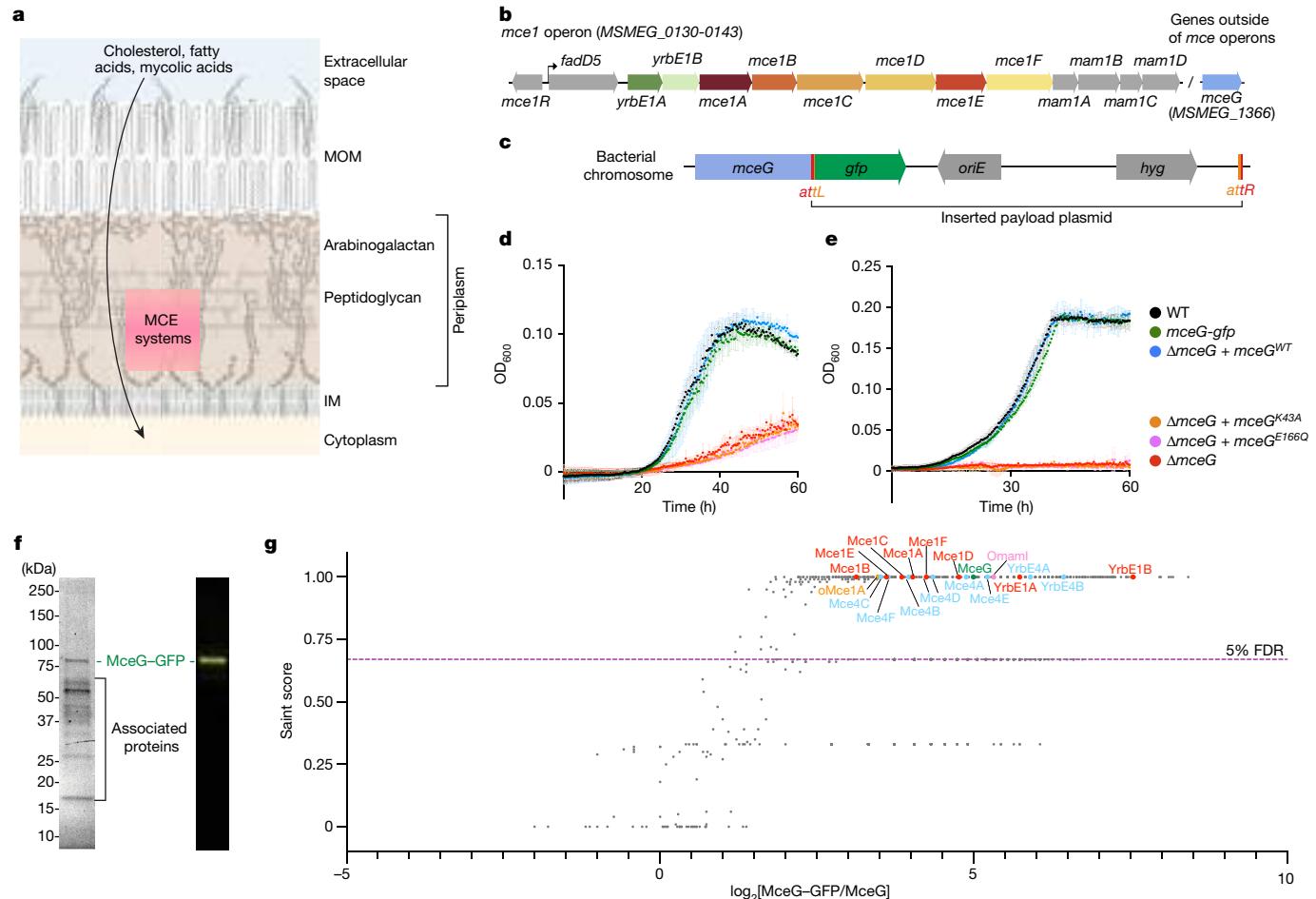


Fig. 1 | Construction and purification of MceG–GFP. **a**, The mycobacterial cell envelope, adapted from ref. 21. MCE systems are proposed to facilitate nutrient transport across the cell envelope. **b**, Schematic of the *M. smegmatis* *mce1* operon. **c**, The modified region of the *M. smegmatis* chromosome in the *mceG-gfp* strain. **d,e**, Growth curves in minimal medium containing palmitate (**d**) or cholesterol (**e**) as the sole carbon source for the following *M. smegmatis* strains: WT (*mc²155*), *mceG-gfp* and $\Delta mceG$, and $\Delta mceG$ complemented with a plasmid encoding WT MceG, MceG(K43A) (to inhibit ATP binding) or MceG(E166Q) (to inhibit ATP hydrolysis). Data are mean \pm s.d. of three biological replicates. OD₆₀₀, optical density at 600 nm. **f**, Analysis using SDS–PAGE after MceG–GFP affinity purification and size-exclusion chromatography (left) and the corresponding anti-GFP western blot analysis (right). Gel source data are provided in Supplementary Fig. 1a. For purifications, three biological replicates were performed with similar results. **g**, Proteins that were identified by MS that preferentially co-purify with MceG–GFP (GFP-tagged) compared with an untagged control. Each point corresponds to an individual protein; the fold enrichment (xaxis) versus the probability that a protein is an MceG interactor (SAINT score⁵¹; yaxis) is shown. A SAINT score of ≥ 0.67 yielded a false-discovery rate (FDR) of $\leq 5\%$ as indicated by the purple dotted line. Proteins related to MCE systems are indicated. Red, Mce1 proteins; blue, Mce4 proteins; orange, orphaned MCE proteins; pink, orphaned Mce-associated membrane proteins; and green, MceG. Plotted data are from three biological replicates.

limited to single subunits and smaller subcomplexes^{14,35}. Thus, how proteins are arranged in a complex to facilitate lipid transport across the cell envelope remains unclear, and elucidating the architecture of mycobacterial MCE systems is a key step towards understanding their transport mechanism.

Isolating endogenous *M. smegmatis* MCE complexes

To isolate intact complexes for structural studies in the absence of an established recombinant expression system, we purified endogenous MCE transporters from *M. smegmatis*, which have high sequence identity to their *M. tuberculosis* orthologues (around 68% identical²⁸) and similar functions^{8,13–16}. We inserted a GFP tag at the C terminus of MceG in the chromosome of *M. smegmatis* *mc²155* using homologous recombination (Fig. 1c, Methods and Supplementary Table 1). To assess the impact of the GFP tag on MceG function, we performed growth assays on the *mceG-gfp* strain in the presence of either palmitate or cholesterol as the sole carbon source. Palmitate and cholesterol are putative

substrates of Mce1 (refs. 4–6,14,31) and Mce4 (refs. 1,4,8), respectively, and both transport systems are dependent on the MceG ATPase^{1,15,29,30}. Deletion of *mceG* substantially reduces growth on palmitate (Fig. 1d) and cholesterol (Fig. 1e), consistent with previous results^{1,14,15,30}. By contrast, growth of our strain expressing chromosomally tagged MceG–GFP was indistinguishable from the wild type (WT), indicating that the MceG–GFP fusion is functional in both conditions (Fig. 1d,e). Mutating residues that are important for ATP binding and hydrolysis resulted in a complete loss of function, comparable to an *mceG* deletion^{14,30} (Fig. 1d,e). The GFP tag on MceG was used for affinity purification of endogenous MCE complexes from *M. smegmatis* cells (Fig. 1f and Extended Data Fig. 1c). As MceG is thought to be shared between multiple MCE systems in a given bacterial species^{15,29,30}, pulling down MceG–GFP may lead to the purification of a mixture of several MCE complexes expressed in *M. smegmatis* under our experimental conditions. To identify the protein subunits that form complexes with MceG and to assess the complexity of our sample, we used mass spectrometry (MS) analysis. These experiments revealed that MceG co-purifies with the eight core components

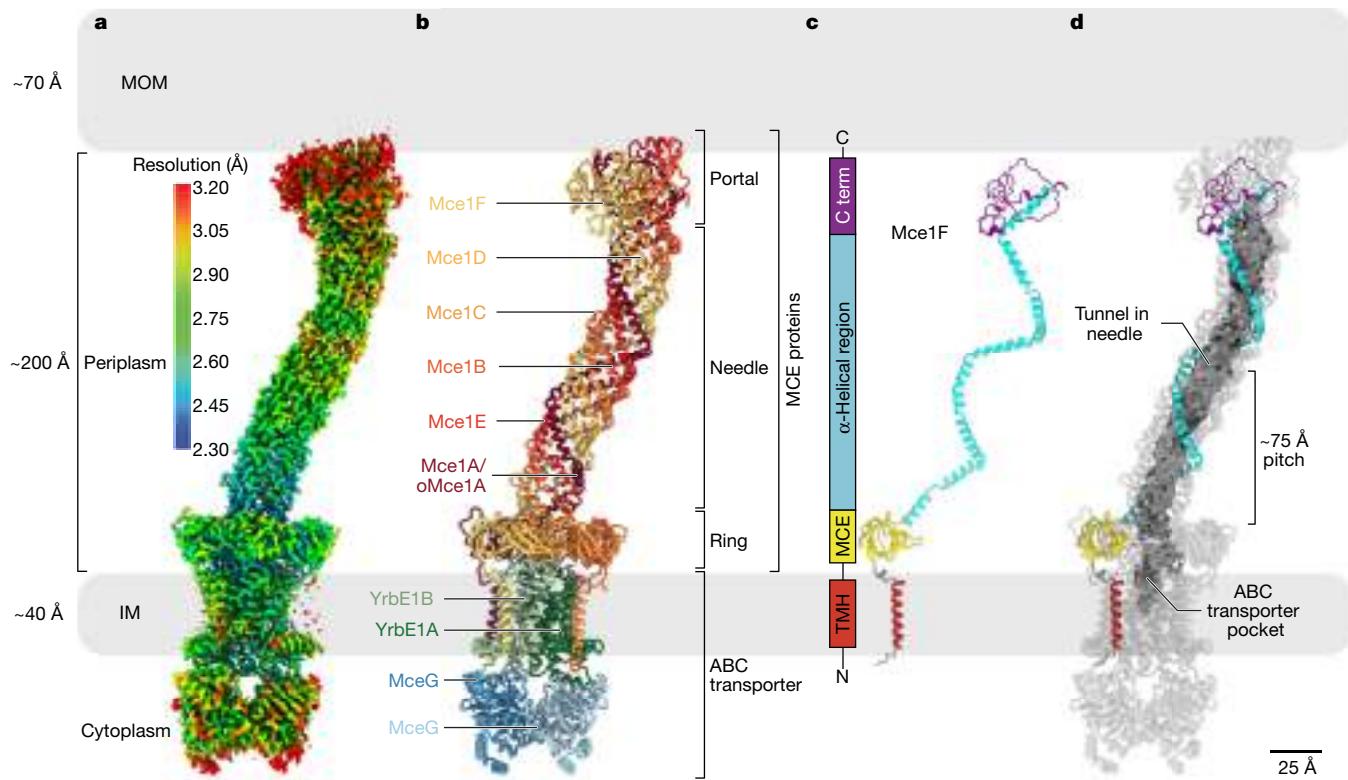


Fig. 2 | Cryo-EM structure of the endogenous MCE complex. **a**, Composite cryo-EM map (Map0) of the Mce1 complex coloured by local resolution as estimated using cryoSPARC⁵². Map0 was rendered with a contour level of 10.0 in ChimeraX⁵³. **b**, The structure of the Mce1 transport system, corresponding to the map shown in **a** and coloured by subunit as in Fig. 1b. The portal, needle and ring are formed from different regions of the six MCE proteins. **c**, Mce1F extracted from Fig. 2b to highlight the structure of an individual MCE protomer.

Colours are consistent in the 2D schematic and 3D structure, showing different regions of the MCE protein. Term, terminus; TMH, transmembrane helix. **d**, The structure of Mce1 shown as a cartoon with the tunnel within the needle assembly and the substrate-binding pocket in the ABC transporter rendered as a molecular surface and coloured grey (calculated in CASTp (v.3.0)⁵⁴). Mce1F is coloured as in **c**.

from the *mce1* and *mce4* operons, including the two YrbEs and all six MCE proteins from both operons (Fig. 1g and Supplementary Tables 2 and 3). Quantification of the relative protein abundance on the basis of peptide spectral matches shows that Mce1 subunits are most abundant (Supplementary Table 2). We did not observe any peptides corresponding to the *mce1*-encoded proteins Mce1R, Fadd5 or Mam1A–Mam1D, or the accessory factor LucA. MSMEG_6540, which is 84% identical to Mce1A, but encoded elsewhere in the genome, was also highly enriched in the MceG purifications and has recently been proposed to have a role in Mce1-mediated fatty acid uptake¹⁴. Although most other mycobacterial MCE proteins are encoded in six-gene modules, MSMEG_6540 is an ‘orphaned’ parologue of Mce1A found in a single-gene operon, which we therefore name oMce1A.

The structure of the Mce1 transporter

We determined the structure of the Mce1 transporter using single-particle cryo-EM (Extended Data Figs. 2a–c and 3a) to a resolution ranging from 2.30 Å to 3.20 Å (Map0; Fig. 2a, Extended Data Fig. 3b–d and Supplementary Table 4). Although our MS data indicate a mixture of Mce1 and Mce4 in the sample used for cryo-EM, side-chain density throughout our final map unambiguously corresponds to the Mce1 complex; we do not see any evidence of Mce4 subunits (Methods and Extended Data Fig. 4a,b). The Mce1 complex consists of ten protein subunits, including two copies of MceG and a single copy each of YrbE1A, YrbE1B, Mce1A/oMce1A, Mce1B, Mce1C, Mce1D, Mce1E and Mce1F (Fig. 2b). Several proteins encoded by the *mce1* operon were absent from the complex, including Fadd5 and Mam1A–Mam1D,

suggesting that these proteins may bind with lower affinity, transiently or may not interact directly. Density for the Mce1A subunit is ambiguous at residues that differ between Mce1A and oMce1A, suggesting that our reconstruction contains a mixture of these highly homologous proteins at the location of the Mce1A subunit (Methods). Our final model is nearly complete, apart from regions that are predicted to be unstructured near the C termini of Mce1C, Mce1D and Mce1F (Extended Data Fig. 5a).

Mce1 forms a highly elongated complex, around 310 Å in length, that can be divided into four main parts (Fig. 2b,c and Supplementary Video 1): (1) the portal, a globular domain formed by the C termini of the Mce1ABCDEF subunits, that lies proximal to the MOM; (2) the needle, which comprises a long central tunnel and is formed by the α-helical regions of the Mce1ABCDEF subunits; (3) the ring, formed by the MCE domains of the Mce1ABCDEF subunits; and (4) the ABC transporter in the IM, which consists of the YrbE1AB permease subunits and the MceG ATPase subunits. The Mce1 complex is anchored in the IM at one end, and the ring, needle and portal extend about 225 Å into the periplasmic space. As the periplasmic width of *M. smegmatis* is around 200 Å (ref. 36), the Mce1 complex is long enough to span the distance between the MOM and IM, with the potential to import fatty acids/mycolic acids through its central tunnel, shielded from the surrounding hydrophilic space (Fig. 2d). This is conceptually similar to molecular machines in Gram-negative bacteria that form tunnels and bridges to move small hydrophobic molecules across the periplasm^{24,26,34}. However, the elongated tunnel of Mce1 is structurally divergent from proteins characterized to date (Extended Data Fig. 5b) and is one of the first structures of such a periplasm-spanning transport system in mycobacteria (Extended Data Fig. 5c).

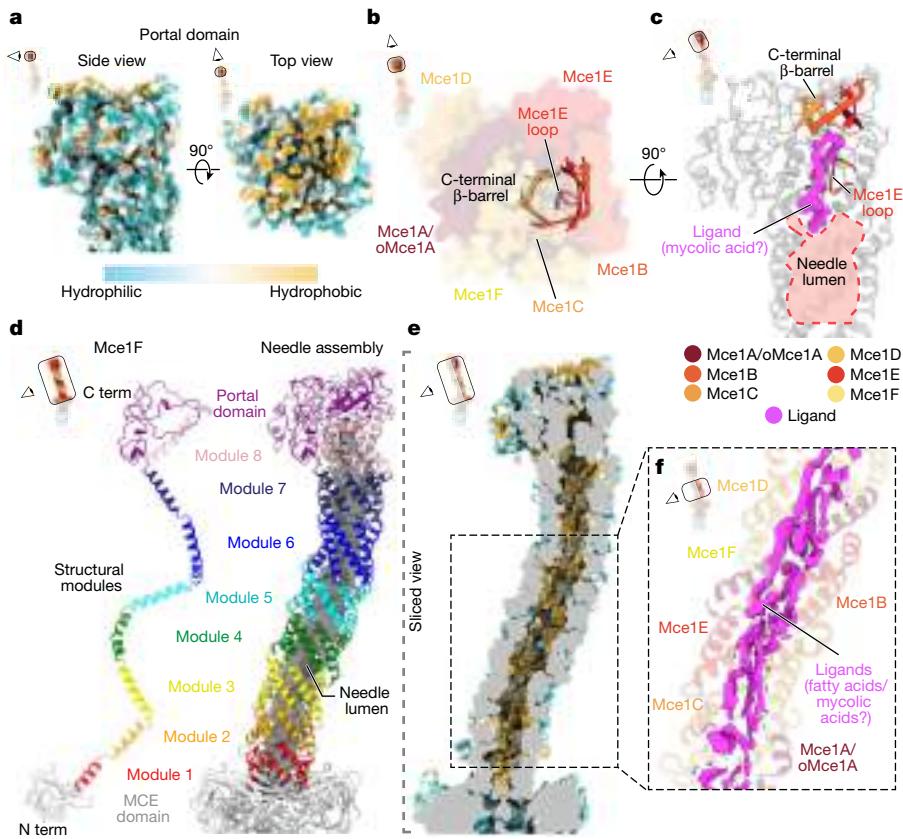


Fig. 3 | The architecture of the portal and needle domains of the Mce1 system. **a**, The portal domain. Side and top views, as indicated by the inset, are both shown as molecular surfaces and coloured by hydrophobicity. **b**, Top view of the portal domain, as in **a**. The C-terminal β-barrel and Mce1E loop are shown as a cartoon representation with bright colours, and other portions of the proteins are shown as transparent molecular surfaces. **c**, Side view of portal domain, as indicated by the inset at the top left. The Mce1E loop and C-terminal β-barrel are shown in colour, and the surrounding protein regions are shown in grey. Ligand density (magenta) from Map0 was rendered using ChimeraX⁵³ volume zone, with a 3.0 Å distance cut-off around UNL1 and a contour level of 7.6. The needle lumen is outlined and coloured (light red). **d**, One MCE protomer,

Mce1F, extracted from the hexamer (left). The MCE needle is coloured by module and MCE domains are shown in light grey. Right, the Mce1 needle domain, coloured by module. The lumen of the needle is rendered as a grey molecular surface (calculated using CASTp (v.3.0)⁵⁴). **e**, Slice through the needle domain showing the lining, viewed as indicated by the inset. The protein is rendered as a molecular surface and coloured by hydrophobicity. The hydrophobicity scale is the same as in **a**. **f**, Ligand density (magenta) in the region of the needle indicated by the box in **e**, viewed as indicated by the inset at the top left. Ligand densities from Map0 were rendered using ChimeraX volume zone with a 3.0 Å distance cut-off around UNL1–31 and a contour level of 7.0.

Portal forms an entrance to the transporter

Substrates for import from the MOM may enter the Mce1 complex through the portal domain (Fig. 3a), which is composed of a small six-stranded β-barrel (Fig. 3b) surrounded by non-canonically structured regions (Extended Data Fig. 6a,b). Apart from the β-barrel motif, the portal domain has no apparent homology to any known protein domains. The C terminus of each MCE protein contributes a single β-strand to the formation of the β-barrel, and also provides a portion of the surrounding non-canonical regions. Despite being formed from six homologous MCE proteins (Mce1A–Mce1F), the C-terminal regions of each MCE subunit are structurally distinct and vary widely in length (Extended Data Fig. 6a,b). The exterior surface of the portal domain is largely hydrophilic and this domain is therefore unlikely to be directly embedded in the MOM (Fig. 3a). However, the MOM-facing surface is hydrophobic (Fig. 3a), and may function as an interface for interacting with other proteins in the MOM that remain to be identified. The lumen of the β-barrel is aligned with the tunnel and has a hydrophobic interior, potentially acting as an entry point for substrates (Fig. 3c). Although this β-barrel is formed from just six strands, the high tilt of its β-strands results in a barrel diameter that is similar to that of the eight-stranded fatty-acid-binding phospholipase PagP found in the *Escherichia coli* outer membrane³⁷ (Extended Data Fig. 6c). In our structure, passage

through the β-barrel is blocked by a few loosely packed hydrophobic side chains that protrude into the lumen. If and how opening may occur is unclear, but relatively subtle side-chain rearrangements may be sufficient to open a pore large enough for a fatty acid/mycolic acid to thread through.

The needle forms a tunnel for transport

The portal feeds directly into a tunnel created by the needle, a unique α-helical structure that is strikingly curved. Our EM data for Mce1 suggest that the curved needle is fairly rigid, and we do not observe straight or alternatively curved states. The needle curvature probably arises from the asymmetric, heterohexameric assembly of the MCE proteins, but its functional role is not immediately clear. Each MCE subunit contains eight copies of a helical repeat motif, separated by well-defined kinks (Fig. 3d and Extended Data Fig. 6a). The helical segments from Mce1ABCDEF twist around each other to form a left-handed superhelix with a pitch of around 75 Å and almost exactly two complete turns (Fig. 2d). The first helical repeats from each MCE subunit associate to form a six-helix bundle. Similarly, repeats 2, 3, 4, 5, 6, 7 and 8 associate to form separate six-helix bundles, for a total of eight structurally similar modules (Extended Data Fig. 6d). These eight modules stack on top of each other to make a long, needle-like tube, and are connected by

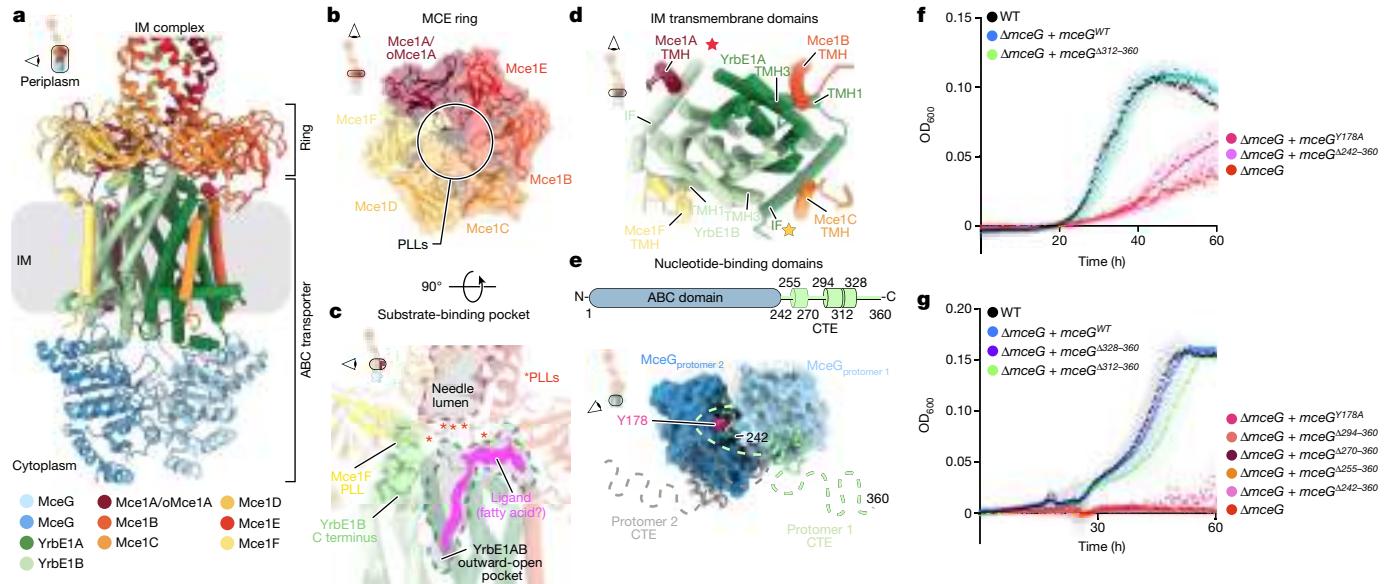


Fig. 4 | Architecture of the ring and ABC transporter complex at the IM.

a, IM complex of Mce1, viewed as indicated by the inset. The structure is coloured according to the key. **b**, The Mce1ABCDEF heterohexameric MCE ring viewed from the periplasm, as indicated by the inset. Proteins are coloured as in **a**. The MCE domains of Mce1ABCDEF are shown as cartoons with a transparent molecular surface. **c**, The interface between the MCE ring and YrbE1AB heterodimer. Proteins are coloured as in **a** and PLLs are indicated by red asterisks. The C terminus of YrbE1B is shown as spheres. The needle lumen and ABC transporter pocket are outlined. Ligand density (magenta) is present in the YrbE1AB cavity. Ligand density from Map0 rendered using ChimeraX³³ volume zone with a 2.5 Å distance cut-off around UNL19 and a contour level of 5.0. **d**, The TM region of Mce1, composed of YrbE1AB and Mce1ABCDEF TM helices (TMHs), viewed as indicated by the inset. Helices are shown as cylinders

short linkers (Fig. 3d). The six-helix bundles appear to be unrelated to previously described folds, such as six-helix coiled-coils³⁸.

The inside of the needle contains a long tunnel, around 7,000 Å³ in volume, with an inner diameter ranging from 7–11 Å. The tunnel is lined with hydrophobic residues, potentially providing a sheltered passage-way for lipids to cross the periplasm (Fig. 3e and Extended Data Fig. 6e). Numerous strong densities are present in the needle, which may correspond to bound substrates (Fig. 3f). The resolution of these densities is too low to unambiguously identify the ligand, but the size and shape are consistent with fatty acid chains that range from 10 to 49 carbons in length (Extended Data Fig. 4c). In many places, 3–5 fatty-acid-like densities appear to run parallel to each other along the long axis of the needle, suggesting that multiple substrates may be transported in bulk through the tunnel. One of the largest and most prominent densities is located in the needle just below the portal domain, where a loop from Mce1E protrudes into the lumen and partially occludes the otherwise broad and featureless tunnel (Fig. 3b,c). The constriction in the tunnel formed by this loop may create a substrate-binding site that is reminiscent of the high-affinity site in the long-chain fatty acid transporter FadL³⁹. In our structure, strong density for a possible approximately 49-carbon-long, mycolic-acid-like substrate fills the area surrounding this loop (Fig. 3c), consistent with a possible role of Mce1 in mycolic acid recycling and MOM maintenance⁷. This binding site, just beyond the β-barrel entrance, may be involved in substrate selection, occurring before transport through the tunnel.

MCE ring links needle to ABC transporter

The hydrophobic tunnel through the needle leads to a pore through the ring, which is formed by six MCE domains (Fig. 4a). Each MCE domain

and are coloured as in **a**. The stars represent the expected position of the unresolved Mce1E lipid anchor and Mce1D TMH based on the *E. coli* homologue MlaFEDB (PDB: 6XBD)⁴¹. **e**, Diagram of the MceG domain (top). Bottom, the MceG homodimer, viewed as indicated by the inset. The ABC domains are shown as surfaces, and the C-terminal extension (CTE) is shown as green ($\text{MceG}_{\text{protomer}1}$) or grey ($\text{MceG}_{\text{protomer}2}$) cartoons. Unresolved regions are indicated by dotted lines. Tyr178 of $\text{MceG}_{\text{protomer}2}$ (pink spheres) interacts with the C-terminal extension of $\text{MceG}_{\text{protomer}1}$ (green cartoon). **f,g**, Growth curves in minimal medium containing palmitate (**f**) or cholesterol (**g**) as the sole carbon source for the following *M. smegmatis* strains: WT (mc¹155), ΔmceG and ΔmceG complemented with a plasmid encoding the indicated MceG mutant. Data are mean ± s.d. of three biological replicates.

in the ring is structurally similar (Extended Data Fig. 7a) but the domains are only around 17% identical to one another at the sequence level (Extended Data Fig. 7b), leading to a heterohexameric ring with the following arrangement: Mce1A/oMce1A–Mce1E–Mce1B–Mce1C–Mce1D–Mce1F (Fig. 4b). This contrasts with the rings observed in other MCE protein assemblies, including LetB, PqiB and MlaD, which are homohexameric and approximately six-fold symmetric³⁴. The pore of the Mce1 ring is formed by a pore-lining loop (PLL) from each MCE domain (Fig. 4b and Extended Data Fig. 7c). The arrangement of the PLLs may form a gate between the periplasmic needle assembly and the substrate-binding pocket of the ABC transporter below (Fig. 4c). In our structure, the pore through the ring is closed, and a conformational change is probably required to allow the passage of substrates into the ABC transporter. Opening and closing of the tunnel through MCE rings has been observed previously in LetB and PqiB³⁴, and may also occur in the Mce1 ring.

The ABC transporter completes the pathway

The pore through the MCE ring leads to the ABC transporter in the IM, which consists of a heterodimer of the permease proteins YrbE1A and YrbE1B, and a homodimer of the ATPase MceG (Fig. 4a). YrbE1A and YrbE1B are around 29% identical to one another, and each consist of an N-terminal interfacial helix and five TM helices, which are homologous to the TM domains of the recently described type VIII ABC transporter MlaFEDB^{40–42} (Extended Data Fig. 7d,e). The TM helices of Mce1A, B, C and F are well resolved and clearly interact around the periphery of the ABC transporter TM domains and anchor the MCE ring in place (Fig. 4d). The TM helix of Mce1D and the lipid anchor of lipoprotein Mce1E are not resolved in our structure (Extended Data Fig. 5a) but may also have similar roles. The MCE ring is slightly tilted with respect to the YrbE

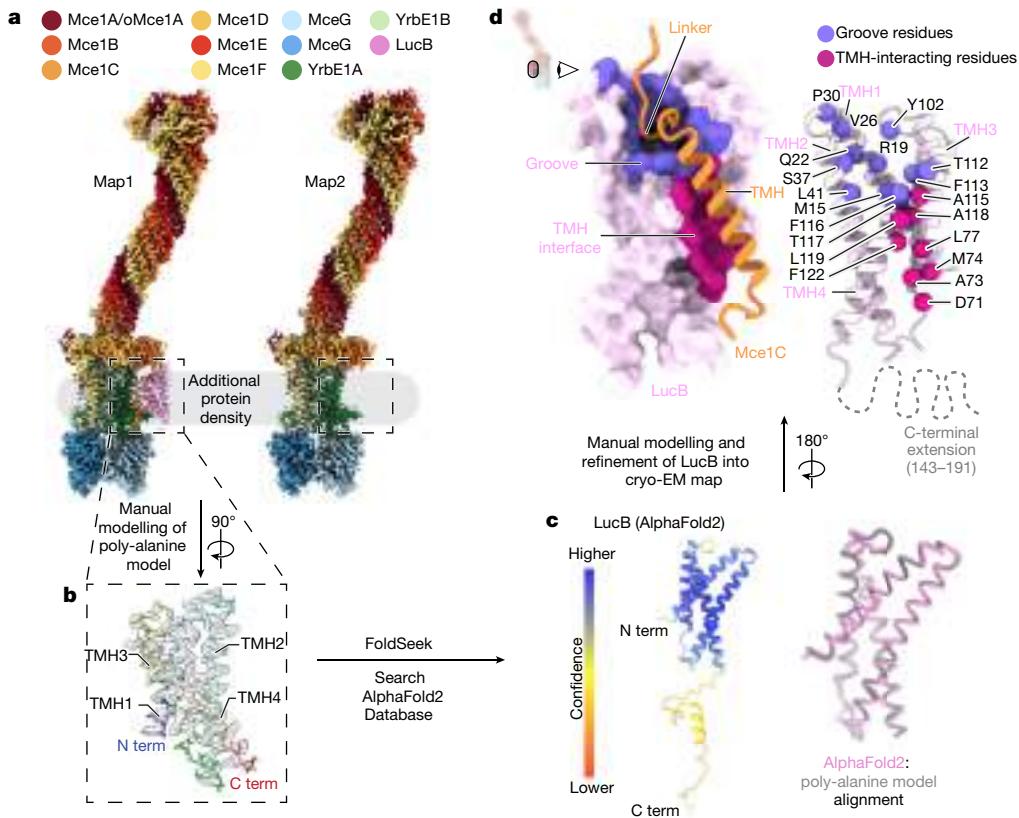


Fig. 5 | LucB is an accessory factor that binds to the Mce1 complex. **a**, Composite density maps for class 1 (Map1) and class 2 (Map2) coloured by protein subunit. Map1 was rendered with a contour level of 10.0 and Map2 was rendered with a contour level 10.0 in ChimeraX⁵³. The colour key is shown at the top. Pink density is observed only in class 1. **b**, The poly-alanine model was manually built into extra density observed in class 1 and used as a search model for Foldseek⁴⁹. Protein is coloured in rainbow colours (N terminus (blue); C terminus (red)), and corresponding cryo-EM density is shown as a transparent grey surface. Protein density from Map1 was rendered using ChimeraX volume zone with a 2.5 Å distance cut-off around the poly-alanine model and a contour level of 8.0. Term, terminus; TMH, transmembrane helix. **c**, AlphaFold2

prediction of LucB identified from a Foldseek⁴⁹ search (left). The model is coloured by prediction confidence; the N-terminal domain is predicted with high confidence⁵⁵. Right, structural alignment of the Cα backbone and LucB AlphaFold2 prediction for the N-terminal domain. **d**, The mode of interaction between LucB and the Mce1 complex, viewed as indicated by the inset. LucB (pink) was rendered as a molecular surface and Mce1C (orange) is shown in cartoon representation (left). LucB-binding interfaces are coloured as follows: groove (purple) and TMH interface (magenta). Right, LucB and Mce1C are shown as cartoons. Residues that make up the LucB groove (purple) and TMH interface (magenta) are shown as spheres.

subunits (about 4°) (Extended Data Fig. 7f), reminiscent of conformations previously described in the homologous MlaFEDB MCE transporter from *E. coli*⁴¹. The C terminus of YrbE1B wedges into the space between the MCE ring and the YrbEs, making contacts with the Mce1F PLL (Fig. 4c and Extended Data Fig. 7g). This extension may stabilize the tilted state, possibly playing a role in coupling conformational changes in the ABC transporter to MCE ring opening/closing. In contrast to the homodimer found in most bacterial ABC transporters, the YrbE1AB heterodimer could facilitate the recognition of asymmetric substrates.

In our structure, YrbE1AB adopts an outward-open state, with a narrow substrate-binding pocket of approximately 150 Å³ that is formed between the YrbE subunits (Figs. 2d and 4c). Density for an elongated ligand, resembling a fatty acid, is observed extending upwards from the substrate-binding pocket (Fig. 4c). An MceG ATPase is bound to each YrbE subunit, forming a homodimer (Fig. 4e). No notable density was observed in the MceG ATP-binding site and the dimer is open, enabling nucleotide exchange. Our structure suggests that the resting state of the Mce1 complex is outward-open, similar to the MlaFEDB phospholipid transporter^{41,43,44} and the LptBFG LPS transporter^{45,46}. Each MceG contains an approximately 120-amino-acid C-terminal extension that is much longer than canonical ABC transporters. This extension consists of several α-helices connected by flexible linkers that interact with the neighbouring MceG subunit (Fig. 4e). Growth assays

in medium with palmitate (Fig. 4f) or cholesterol (Fig. 4g) as the sole carbon source demonstrated that mutations in the C-terminal extension of MceG, and its interface with the neighbouring subunit through residue Tyr178, are important for function (Fig. 4f,g), consistent with previous findings¹⁵. We used western blotting to assess the expression of MceG mutants and their effect on the cellular levels of Mce1F and Mce4F (representative subunits of Mce1 and Mce4, respectively); both complexes are thought to be stabilized by their interaction with MceG³⁰. Expression levels of MceG mutants are similar to each other, and to WT MceG. MceG mutants that were non-functional in cells resulted in a reduction in Mce1F levels, and almost a complete loss of Mce4F expression (Extended Data Fig. 8a,b), consistent with recent findings in *M. tuberculosis*³⁰. The MceG C-terminal extension may facilitate the formation of the MceG homodimer and its association with the rest of the transport complex, as recently proposed for the ATPase subunit of a homologous MCE transporter⁴⁷.

LucB is a subunit of Mce1

Notably, we observed density for an additional unknown subunit associated with the ABC transporter within a subpopulation of our particles (Extended Data Fig. 2a). Focused three-dimensional (3D) classification led to the emergence of two classes (Fig. 5a)—class 1 (Map1,

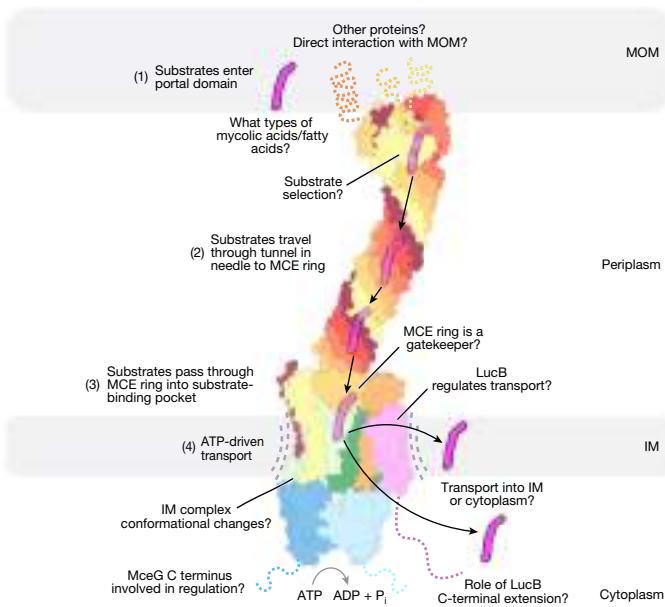


Fig. 6 | Model for Mce1-mediated transport. The model for Mce1-mediated transport of fatty acids or mycolic acids in mycobacteria, highlighting our current understanding and open questions: (1) substrates (magenta) enter the portal domain; (2) substrates travel down the hydrophobic tunnel through the needle domain; (3) substrates pass through the ring into the ABC transporter; (4) ATP hydrolysis in MceG drives conformational changes in the ABC transporter, enabling substrates to be transported. The nature of the conformational changes that drive transport and how LucB regulates Mce1 remain unclear. P_i, inorganic phosphate.

around 2.76 Å (Extended Data Fig. 3e–h and Supplementary Table 4) and class 2 (Map2, approximately 2.90 Å; Extended Data Fig. 3i–l and Supplementary Table 4). The additional subunit, found only in class 1, lies almost entirely within the TM region, and consists of 4 TM helices (Fig. 5b). Examination of our MceG–GFP MS data did not suggest an obvious candidate protein consistent with our EM density (Supplementary Tables 2 and 3). To identify this unknown subunit, we built a poly-alanine model into the density and used these coordinates to perform a structure-based search of the Protein Data Bank (PDB) and AlphaFold Protein Structure Database⁴⁸ using Foldseek⁴⁹ (Fig. 5b). Although no proteins with a similar structure were identified in the PDB, a search of the AlphaFold database revealed predicted structures that matched our poly-alanine model well, including MSMEG_3032 and its *M. tuberculosis* homologue Rv2536 (ref.⁵⁰) (approximately 61% identical) (Fig. 5c). Fitting the AlphaFold2 MSMEG_3032 model into our EM density required minimal adjustment apart from a few side-chain rotamer changes, supporting the assignment of MSMEG_3032/Rv2536 as a component of the Mce1 system (Fig. 5d and Extended Data Fig. 4d). On the basis of a possible role as a lipid uptake coordinator, analogous to the proposed role of LucA⁴, we rename MSMEG_3032/Rv2536 to LucB. To validate the interaction identified from our structure, we assessed whether MCE transporter components co-purify with LucB. We constructed an *M. smegmatis* strain with chromosomally tagged LucB–GFP, and purified the protein by anti-GFP affinity and size-exclusion chromatography (Extended Data Fig. 9a,b). Negative-stain EM analysis of the resulting sample reveals particles with the characteristic shape and features of the Mce1 system (Extended Data Fig. 9c,d). MS analysis of purified LucB–GFP (Extended Data Fig. 9e and Supplementary Tables 2 and 5) showed significant enrichment of Mce1 subunits, whereas Mce4 subunits were not significantly enriched. Consistent with these results, growth of a *lucB*-deletion mutant was indistinguishable from the WT in medium with cholesterol as the sole carbon source (Extended Data Fig. 9f), and deletion of *lucB* leads to decreased levels of Mce1F but not

Mce4F in cells (Extended Data Fig. 9g,h), suggesting that LucB may be functionally linked to Mce1 but not to Mce4.

In our structure, LucB interacts almost exclusively with Mce1C, primarily through interactions with the Mce1C TM helix and the linker connecting the TM helix to the MCE domain (Fig. 5d). The Mce1C linker sits in a conserved cleft formed between the TM2 and TM4 helices of LucB (Extended Data Fig. 10a), and the Mce1C TM helix packs against TM3 and TM4 of LucB. Binding to Mce1C positions the LucB C-terminal extension towards the cytoplasm where it could potentially interact with MceG or recruit other proteins (Fig. 5d). The C-terminal extension is not resolved in our map and is predicted to be disordered (Fig. 5c), but may become ordered after interacting with a binding partner. The conformation of the Mce1 complex is very similar in both classes, apart from a better definition of density for the Mce1C TM helix and interacting loop in the presence of LucB (overall root mean squared deviation (r.m.s.d.) = 0.50 for class 1 versus class 2), suggesting that there is no global conformational change in the Mce1 system after LucB binding.

LucB, for which there is a single parologue in *M. smegmatis* and *M. tuberculosis*, is a protein of unknown function. Concurrent with the release of the pre-print corresponding to this paper, a recent report independently linked LucB function to MCE systems in *M. tuberculosis*³⁰. Orthologues of this protein can be found in bacteria of the Actinomycetales order, particularly in the families Gordoniaceae, Mycobacteriaceae, Nocardiaceae, Pseudonocardiaceae and Tsukamurellaceae (Extended Data Fig. 10b). Notably, LucB orthologues appear to be found predominantly in diderm bacteria containing *M. tuberculosis*-like *mce* operons²⁸, with a conserved eight-gene cluster encoding two distinct YrbE and six distinct MCE proteins. By contrast, orthologues of LucB are not apparent in genomes containing simpler MCE gene clusters encoding single YrbE and MCE proteins subunits, such as those found in *E. coli*. This observation, coupled with our data, suggests that LucB may have evolved to function specifically with the hetero-oligomeric MCE transporters that arose in the actinobacterial lineage, and may be involved in the regulation of activity in these transporters.

Discussion

The mycobacterial cell envelope is highly complex and divergent from its Gram-negative counterparts. Mechanisms for how substrates are transported across the mycobacterial cell envelope have remained unclear. Mce1 has been proposed to be a fatty acid and/or mycolic acid importer^{4–7,14,31}, and our high-resolution structure of an endogenous Mce1 transport complex enables us to propose a model for how this important virulence factor may work to import substrates (Fig. 6 and Supplementary Video 2). First, fatty acids or mycolic acids from the MOM may enter through the β-barrel of the portal domain, either directly or mediated by additional unknown factors in the MOM. How the Mce1 complex recognizes specific substrates is unclear, but one possibility is that substrate selection occurs at the apparent substrate-binding site noted just below the β-barrel of the portal domain. After entering the complex, the substrates travel across the periplasm through the hydrophobic tunnel created by the curved Mce1ABCDEF needle, in which several substrates may be accommodated simultaneously. At the base of this needle, the ring of MCE domains must undergo a conformational change, opening the central pore to allow substrate entry into the IM ABC transporter. As there is no clear energy source available in the periplasm, movement of substrates from the MOM to the IM may be largely energy-independent and bidirectional. However, ATP hydrolysis by MceG probably drives unidirectional translocation of substrates into the cytoplasm or IM, by triggering conformational changes in the YrbE1AB subunits. In another type VIII ABC transporter^{40,42}, MlaFEDB, ATP binding leads to a collapse of the lipid-binding pocket within the TM domains, and is thought to drive expulsion of the substrate from the transporter⁴³. Similar conformational changes may occur in Mce1, and future studies focused on directly probing the conformational

change cycle, substrate specificity and directionality will shed light on these mechanisms. LucB, which we show binds to Mce1C, may have a role as a regulator, or a scaffold protein to recruit other parts of the system that are not yet known. Although LucB is not structurally related to LucA⁴, both are small TM proteins that may regulate MCE systems. Our data provide a structural framework for how mycobacteria may use MCE systems to scavenge resources from the host cell, by providing a tunnel for the transport of substrates across the cell envelope without compromising the protective nature of this barrier.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06366-0>.

1. Pandey, A. K. & Sassetti, C. M. Mycobacterial persistence requires the utilization of host cholesterol. *Proc. Natl Acad. Sci. USA* **105**, 4376–4380 (2008).
2. Lee, W., VanderVen, B. C., Fahey, R. J. & Russell, D. G. Intracellular *Mycobacterium tuberculosis* exploits host-derived fatty acids to limit metabolic stress. *J. Biol. Chem.* **288**, 6788–6800 (2013).
3. Gioffré, A. et al. Mutation in mce operons attenuates *Mycobacterium tuberculosis* virulence. *Microbes Infect.* **7**, 325–334 (2005).
4. Nazarova, E. V. et al. Rv3723/LucA coordinates fatty acid and cholesterol uptake in *Mycobacterium tuberculosis*. *eLife* **6**, e26969 (2017).
5. Nazarova, E. V. et al. The genetic requirements of fatty acid import by *Mycobacterium tuberculosis* within macrophages. *eLife* **8**, e43621 (2019).
6. Laval, T. et al. De novo synthesized polyunsaturated fatty acids operate as both host immunomodulators and nutrients for. *eLife* **10**, e71946 (2021).
7. Cantrell, S. A. et al. Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of *Mycobacterium tuberculosis*. *J. Microbiol.* **51**, 619–626 (2013).
8. García-Fernández, J., Papavinasasundaram, K., Galán, B., Sassetti, C. M. & García, J. L. Molecular and functional analysis of the mce4 operon in *Mycobacterium smegmatis*. *Environ. Microbiol.* **19**, 3689–3699 (2017).
9. Cohen, A., Mathiasen, V. D., Schön, T. & Wejse, C. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. *Eur. Respir. J.* **54**, 1900655 (2019).
10. Rodriguez, G. M. & Smith, I. Identification of an ABC transporter required for iron acquisition and virulence in *Mycobacterium tuberculosis*. *J. Bacteriol.* **188**, 424–430 (2006).
11. Arnold, F. M. et al. The ABC exporter IrtAB imports and reduces mycobacterial siderophores. *Nature* **580**, 413–417 (2020).
12. Rempel, S. et al. A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature* **580**, 409–412 (2020).
13. Rank, L., Herring, L. E. & Braunstein, M. Evidence for the mycobacterial Mce4 transporter being a multiprotein complex. *J. Bacteriol.* **203**, e00685–20 (2021).
14. Chen, Y. & Chng, S.-S. A conserved membrane protein negatively regulates Mce1 complexes in mycobacteria. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.06.08.495402> (2022).
15. García-Fernández, J., Papavinasasundaram, K., Galán, B., Sassetti, C. M. & García, J. L. Unravelling the pleiotropic role of the MceG ATPase in *Mycobacterium smegmatis*. *Environ. Microbiol.* **19**, 2564–2576 (2017).
16. Klepp, L. I. et al. Impact of the deletion of the six mce operons in *Mycobacterium smegmatis*. *Microbes Infect.* **14**, 590–599 (2012).
17. Nakamura, S. et al. Molecular basis of increased serum resistance among pulmonary isolates of non-typeable *Haemophilus influenzae*. *PLoS Pathog.* **7**, e1001247 (2011).
18. Zhang, L. 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).
19. Senaratne, R. H. et al. *Mycobacterium tuberculosis* strains disrupted in mce3 and mce4 operons are attenuated in mice. *J. Med. Microbiol.* **57**, 164–170 (2008).
20. Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T. & Riley, L. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261**, 1454–1457 (1993).
21. Dulberger, C. L., Rubin, E. J. & Boutte, C. C. The mycobacterial cell envelope—a moving target. *Nat. Rev. Microbiol.* **18**, 47–59 (2020).
22. Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. *Nature* **567**, 486–490 (2019).
23. Owens, T. W. et al. Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550–553 (2019).
24. Sherman, D. J. et al. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science* **359**, 798–801 (2018).
25. Okuda, S., Freinkman, E. & Kahne, D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in *E. coli*. *Science* **338**, 1214–1217 (2012).
26. Du, D. et al. Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).
27. Costa, T. R. D. et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).
28. Casali, N. & Riley, L. W. A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genom.* **8**, 60 (2007).
29. Joshi, S. M. et al. Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc. Natl. Acad. Sci. USA* **103**, 11760–11765 (2006).
30. Fieweger, R. A. et al. MceG stabilizes the Mce1 and Mce4 transporters in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **299**, 102910 (2023).
31. Forrellad, M. A. et al. Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis* **94**, 170–177 (2014).
32. Malinvernini, J. C. & Silhavy, T. J. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc. Natl. Acad. Sci. USA* **106**, 8009–8014 (2009).
33. Thong, S. et al. Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. *eLife* **5**, e19042 (2016).
34. Ekert, D. C. et al. Architectures of lipid transport systems for the bacterial outer membrane. *Cell* **169**, 273–285 (2017).
35. Asthana, P. et al. Structural insights into the substrate-binding proteins Mce1A and Mce4A from. *IUCrJ* **8**, 757–774 (2021).
36. Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. M. & Engelhardt, H. Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc. Natl. Acad. Sci. USA* **105**, 3963–3967 (2008).
37. Ahn, V. E. et al. A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J.* **23**, 2931–2941 (2004).
38. Rhys, G. G. et al. Navigating the structural landscape of de novo α-helical bundles. *J. Am. Chem. Soc.* **141**, 8787–8797 (2019).
39. van den Berg, B., Black, P. N., Clemons, W. M. Jr & Rapoport, T. A. Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304**, 1506–1509 (2004).
40. Thomas, C. et al. Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett.* **594**, 3767–3775 (2020).
41. Coudray, N. et al. Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *eLife* **9**, e62518 (2020).
42. Ekert, D. C., Coudray, N. & Bhabha, G. Structure and mechanism of the bacterial lipid ABC transporter, MlaFEDB. *Curr. Opin. Struct. Biol.* **76**, 102429 (2022).
43. Chi, X. et al. Structural mechanism of phospholipids translocation by MlaFEDB complex. *Cell Res.* **30**, 1127–1135 (2020).
44. Tang, X. et al. Structural insights into outer membrane asymmetry maintenance in Gram-negative bacteria by MlaFEDB. *Nat. Struct. Mol. Biol.* **28**, 81–91 (2021).
45. Luo, Q. et al. Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nat. Struct. Mol. Biol.* **24**, 469–474 (2017).
46. Dong, H., Zhang, Z., Tang, X., Paterson, N. G. & Dong, C. Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG. *Nat. Commun.* **8**, 222 (2017).
47. Kolich, L. R. et al. Structure of MlaFB uncovers novel mechanisms of ABC transporter regulation. *eLife* **9**, e60030 (2020).
48. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
49. van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01773-0> (2023).
50. Garcia, J. et al. *Mycobacterium tuberculosis* Rv2536 protein implicated in specific binding to human cell lines. *Protein Sci.* **14**, 2236–2245 (2005).
51. Choi, H. et al. Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Curr. Protoc. Bioinform.* **8**, 8.15.1–8.15.23 (2012).
52. Punjabi, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
53. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
54. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
55. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Methods

Bacterial strain construction

M. smegmatis strains were generated by oligonucleotide-mediated recombineering followed by Bxb integrase targeting (ORBIT)⁵⁶. Plasmids for ORBIT were obtained from Addgene. pKM444 was a gift from K. Murphy (Addgene plasmid, 108319)⁵⁶. pKM461 was a gift from K. Murphy (Addgene plasmid, 108320)⁵⁶. pKM464 (deletions) was a gift from K. Murphy (Addgene plasmid, 108322)⁵⁶. Then, 380 µl of electrocompetent *M. smegmatis* cells (mc^2 155 strain⁵⁷) was mixed with 200 ng of a recombinase expression plasmid (pKM444 or pKM461) containing the Che9c phage RecT annealase and Bxb1 integrase in ice-cooled electroporation cuvettes (Bio-Rad MicroPulser Electroporation Cuvettes, 0.2 cm gap, 1652092) and were electroporated using a Bio-Rad MicroPulser Electroporator (settings: Ec2, 2.5 kV). Transformed cells were allowed to recover for around 16 h in Middlebrook 7H9 broth (Difco, 271310), containing 0.05% (v/v) Tween-80 (Sigma-Aldrich, P1754), before plating on 7H10 (Difco, DF0627-17-4) plates, containing 0.05% (v/v) Tween-80 and 20 µg ml⁻¹ kanamycin (Genesee Scientific, 18–211). Colonies were picked, cultured in 7H9 medium and protein expression was induced with 500 ng ml⁻¹ anhydrotetracycline (ATc, Sigma-Aldrich, 31741) at an OD₆₀₀ of 0.5. For chromosomal tagging, the induced cells were made electrocompetent and 380 µl of cell were subsequently co-transformed with 200 ng of pBEL2108 (a derivative of payload plasmid pKM468 (Addgene, 108434)⁵⁶ containing a 3C protease cleavage site upstream of the eGFP tag) and 1 µg of targeting oligonucleotide using electroporation. The MceG-GFP strain (bBEL591) was generated with a 3C-eGFP-4xGly-TEV-Flag-6xHis tag on the C terminus of MceG (*MSMEG_1366*) using the following oligo (IDT Ultramer DNA Oligo): 5'-GTTGCCCGCGCCGGCCCTTGAGACACGTCAGGCCGGCCGT GACGGCCGGCGCTATCGCGCAAACCTCAGGTTGTACCGTACACCA CTGAGACCCGGTGTTGACCAGACAAACCCGCTGTTGGCACCT CGATGACGCCCGTCGGCGAGTCGCTAGTTCTGACGGCGCGGTG GCGGCC-3'. The LucB-GFP (bBEL595) strain was generated with a 3C-eGFP-4xGly-TEV-Flag-6xHis tag on the C terminus of LucB (*MSMEG_3032*) using the following oligo (IDT Ultramer DNA Oligo): 5'-CACGATCTGTGACCGTACTCGCTACGCTGCCCCCATGAGCAAGT GGTTACTGGCGGAGTGGTGTTCGCAGGTTGTCTGGTACACCACCG CGGTCTCAGTGGTGTACGGTACAAACCCGCTGGAGAATCCGGACCA GCCCGCTCAGAGCTGATCCGGCTCAGCTTACAAACGAGAGTTGTT TGTT-3'. Transformants were recovered in 7H9 overnight and plated on either LB + agar (Difco, DF0446-07-5) or 7H10 plates containing 50 µg ml⁻¹ hygromycin (GoldBio, H-270) and incubated at 37 °C for 3–5 days. Colonies were verified for insertion of the payload plasmid by PCR and subsequently confirmed by whole-genome resequencing (SeqCenter).

For knockout strains, 380 µl of electrocompetent induced cells were co-transformed with 200 ng of pKM464 and 1 µg of targeting oligo using electroporation. The $\Delta mceG$ strain (bBEL594) containing a deletion of *mceG* (*MSMEG_1366*) was generated using the following oligo (IDT Ultramer DNA Oligo): 5'-CCGTGACGGCCGGCTGATCGCGCAAACCTCAGCCTGTTGGCACCTCGATGACGCCGGTTGTACCGTACACCACT GAGACCCGGTGGTTGACCAGACAAACCCAACCCCGTCACGTCGATT TGGACGCCATCAAAGATCCTTCCGCTACGCCCTACCCACAC-3'.

The $\Delta mce1$ strain (bBEL603) containing a deletion of the *mce1* operon (*MSMEG_0132-0143*) was generated using the following oligo (IDT Ultramer DNA Oligo): 5'-CTACAGCTCGGGAGAGGTCC AACTCGCTCAGGGCAGGCATCTGAGTGAATCCGCACTGCGGGTTGACAGACAAA CCTGGACGTGATCCACGCCCTCCT-3'. The $\Delta mce4$ strain (bBEL604) containing a deletion of the *mce4* operon (*MSMEG_5902-5893*) was generated using the following oligo (IDT Ultramer DNA Oligo): 5'-CACCCCCGCCCCAAATGAACGGAGCTGAGGTTGATGAAACAGCTT CGGGTTCCAGCCCCGGGTTGCTGGTCAACCACCGCGGTCTCAGTG

GTGTACGGTACAACCTGGCTATCGACGCCATCAAGCCGATCTAGTG TCGCGCGTCATTAATTAGTTACAGTGT-3'. The $\Delta lucB$ strain (bBEL608) containing a deletion of *lucB* (*MSMEG_3032*) was generated using the following oligo (IDT Ultramer DNA Oligo): 5'-CACGATCTGTGACGCT ACTCGCTACGCTGTGCCCCCATGAGCAAGTGGTTACTGCGCGGAGTGG TGTTCGCAGGTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGG TACAAACCCCGCTGGAGAATCCGGACCAGCGCGTCAGAGCTGATCC GGGCTCAGCTTACAAACGAGAGTTGTTGTGGT-3'. Transformants were recovered in 7H9 overnight and plated on 7H10 plates containing 50 µg ml⁻¹ hygromycin and incubated at 37 °C for 3–5 days. Colonies were verified for insertion of the payload plasmid by PCR and subsequently confirmed by whole-genome resequencing (SeqCenter).

Complementation plasmid construction

For complementation of the ORBIT-constructed *mceG* knockout (bBEL594), a derivative of pMV261zeo (a gift from J. Cox) was cloned containing WT *mceG* (pBEL2759). The coding sequence of *mceG* was amplified genomic DNA extracted from *M. smegmatis* cells using AccuPrime Pfx DNA Polymerase (Invitrogen, 12344032) and cloned into pMV261zeo using Gibson assembly. TOP10 cells (Invitrogen, C404010) were transformed with the assembled vector using heat shock and plated on LB + agar plates containing 25 µg ml⁻¹ zeocin (Gibco, R25001). Colonies were screened for the correct DNA sequences using Sanger sequencing (Azenita). Complementation plasmids containing MceG mutants and LucB were generated in a similar manner (pBEL2708, MceG(Δ 328–360); pBEL2709, MceG(Δ 312–360); pBEL2710, MceG(Δ 294–360); pBEL2711, MceG(Δ 270–360); pBEL2712, MceG(Δ 255–360); pBEL2713, MceG(Δ 242–360); pBEL2714, MceG(K43A); pBEL2717, MceG(E166Q); pBEL2719, MceG(Y178A); LucB, pBEL2836).

Complementation plasmids were electroporated into either electrocompetent WT *M. smegmatis* cells, $\Delta mceG$ *M. smegmatis* cells or $\Delta lucB$ *M. smegmatis* cells. Cells were plated on 7H10 plates containing the appropriate antibiotics (such as 25 µg ml⁻¹ zeocin, 50 µg ml⁻¹ hygromycin). Colonies were selected, cultured in Middlebrook 7H9 (Difco, 271310) containing 0.05% (v/v) Tween-80 (Sigma-Aldrich, P1754) and appropriate antibiotics, frozen as 20% glycerol stocks for future use.

Cholesterol growth assay

The cholesterol growth assay was adapted from previous studies^{13,15}. In brief, *M. smegmatis* strains were streaked on 7H10 plates supplemented with 0.05% (v/v) Tween-80 and the appropriate antibiotics from frozen glycerol stocks. Colonies were used to seed M9 medium (1 l distilled H₂O, 12.8 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 g NH₄Cl, 25 µl 1 M CaCl₂, 500 µl 1 M MgSO₄) supplemented with 0.5% glycerol and 0.05% (v/v) tyloxapol (Ty, Sigma-Aldrich, T0307) with appropriate antibiotics. M9 cultures were grown to an OD₆₀₀ of around 0.7–1.0 at 37 °C and collected. Strains were washed twice by pelleting cells by centrifugation at 4,000 rcf for 5 min at 22 °C and resuspended in M9 medium with 0.05% tyloxapol. After the wash steps, strains were resuspended in M9 medium with 0.05% tyloxapol to an OD₆₀₀ of 0.1 and were used to seed 200 µl cultures (starting OD₆₀₀ of 0.005) for growth in 96-well plates. For each strain, the following media were used: (1) M9 + 0.05% Ty + 0.5% (v/v) glycerol (carbon source positive control); (2) M9 + 0.05% Ty + 0.009 g ml⁻¹ methyl-β-cyclodextrin (MBC, Sigma-Aldrich, C4555) (no carbon source control); and (3) M9 + 0.05% Ty + 0.009 g ml⁻¹ MBC + 0.69 mM cholesterol (Sigma-Aldrich, C8667). Cultures were grown at 37 °C with shaking and the OD₆₀₀ was monitored for each strain using a plate reader (BioTek). At least three biological replicates were conducted and plotted using Prism (GraphPad).

Palmitate growth assay

The palmitate growth assay was adapted from a previous study¹⁴. In brief, *M. smegmatis* strains were streaked on 7H10 plates supplemented with 0.05% (v/v) Tween-80 and the appropriate antibiotics from frozen glycerol stocks. Colonies were used to inoculate 7H9 medium

supplemented with 0.05% (v/v) tyloxapol and appropriate antibiotics. Cultures were grown to stationary phase ($>2 \text{ OD}_{600}$) at 37 °C and used to subculture Sauton's medium (0.5 g l⁻¹ K₂HPO₄, 0.25 g l⁻¹ MgSO₄, 2 g l⁻¹ citric acid, 0.4 g l⁻¹ L-asparagine, 0.05 g l⁻¹ ferric ammonium citrate, 5 g l⁻¹ fatty acid-free BSA (Goldbio, A-421-100), 0.05% tyloxapol, pH adjusted to 7.0 using 1 M NaOH) containing 300 μM palmitic acid (Sigma-Aldrich, P9767-10G) as the sole carbon source. To prepare this medium, palmitic acid was dissolved in 0.1 M NaOH, heated to 80 °C with shaking and slowly added into the prewarmed Sauton's medium (37 °C) with constant stirring until a final concentration of 300 μM. After addition of palmitic acid, the medium was incubated at 37 °C with constant stirring for 1 h to allow the fatty acid to conjugate with the BSA. Subcultures were also prepared with Sauton's medium containing either 0.5% glycerol (carbon source positive control) or 0.5 g l⁻¹ fatty acid-free BSA (no carbon source control). The starting OD₆₀₀ was 0.0001 for all the growth conditions. Cultures were grown at 37 °C with shaking and the OD₆₀₀ was monitored for each strain using a plate reader (BioTek). At least three biological replicates were conducted and plotted using Prism (GraphPad).

Bacterial growth and protein purification

M. smegmatis was grown in Middlebrook 7H9 supplemented with 0.05% (v/v) Tween-80 and additional antibiotics as needed (such as 50 μg ml⁻¹ hygromycin). For protein expression and purification of chromosomally GFP-tagged MceG (bBEL591) or GFP-tagged LucB (bBEL595), overnight cultures of each strain were diluted 1:1,000 and grown with shaking at 37 °C and 200 rpm until 0.8–1.2 OD₆₀₀. Cells were collected by centrifugation at 4,000 rcf, 4 °C. Pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM MgSO₄, 5 mM 6-aminocaproic acid (Sigma-Aldrich, A2504), 5 mM benzamidine (Sigma-Aldrich, B6506) and 1 mM phenylmethylsulfonyl fluoride (PMSF, Sigma-Aldrich, 10837091001)) and stored at –80 °C. Cells were thawed at room temperature and lysed by four passes through a chilled Emulsiflex-C3 cell disruptor (Avestin) at an output pressure of 20 kpsi. Unlysed cells and debris were removed by centrifugation at 39,000 rcf for 30 min at 4 °C. The membranes from the resulting supernatant were pelleted by ultracentrifugation in a Fiberlite F37L-8 x 100 Fixed-Angle Rotor (Thermo Fisher Scientific, 096-087056) at 37,000 rpm (182,460 rcf) for 90 min at 4 °C and resuspended in membrane resuspension (MR) buffer (50 mM Tris-HCl pH 7.5, 15% (v/v) glycerol, 5 mM MgSO₄, 150 mM NaCl, 5 mM 6-aminocaproic acid, 5 mM benzamidine and 1 mM PMSF). Resuspended membranes were stored at –80 °C. For affinity purification, membranes were thawed and solubilized overnight with the addition of 20 mM n-dodecyl-β-D-maltoside (DDM, Inalco, D310S) at 4 °C, and insoluble material was removed by centrifugation at 37,000 rpm (182,460 rcf) for 60 min. GFP affinity resin was prepared using a method adopted from a previous study⁵⁸. Plasmids used to make the GFP affinity resin were obtained from Addgene. pTP396 was a gift from R. Voorhees (Addgene plasmid, 149336)⁵⁸. pTP264 was a gift from D. Görlich (Addgene plasmid, 149334)⁵⁸. pAV0286 was a gift from D. Görlich (Addgene plasmid, 149333)⁵⁸. In brief, purified His14-Avi-SUMO^{Eul} anti-GFP nanobody (expressed from pTP396) was biotinylated using BirA (expressed from pTP264) and further purified using the Superdex 200 16/60 gel-filtration column (Cytiva, 28-9909-44) equilibrated in GF1 buffer containing 50 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM dithiothreitol (DTT, Amresco, M109). The biotinylated anti-GFP nanobody was added to Pierce High Capacity Streptavidin Agarose Resin (Thermo Fisher Scientific, 20359) equilibrated in GF1 buffer and allowed to incubate with the resin overnight at 4 °C. 0.6 ml bed volume of resin was washed three times with GF1 buffer and blocked by incubation with 100 μM biotin (Sigma-Aldrich, B4501) in 50 mM HEPES/KOH pH 7.5 for 5 min on ice with occasional mixing. Beads were washed three times with GF1 Buffer and subsequently washed three times with MR buffer containing 20 mM DDM before use. Solubilized membranes were incubated with the equilibrated GFP affinity resin

at 4 °C for 6 h and then washed three times with 125 column volumes of membrane wash buffer (50 mM Tris-HCl pH 7.5, 15% (v/v) glycerol, 5 mM MgSO₄, 150 mM NaCl, 5 mM 6-aminocaproic acid, 5 mM benzamidine, 1 mM DDM and 1 mM PMSF). Immobilized proteins were eluted by incubation with 1 ml of 250 nM SENP^{Eul} protease (expressed and purified from pAV286) overnight at 4 °C. Eluted proteins were pooled and concentrated using the Amicon Ultra-0.5 Centrifugal Filter Unit concentrator (MWCO 100 kDa, UFC510096) before separation on the Superdex 200 16/60 column (GE Healthcare) equilibrated in GF2 Buffer (50 mM Tris-HCl pH 7.5, 5 mM MgSO₄, 150 mM NaCl, 1 mM DDM and 1 mM DTT). Fractions containing GFP-tagged MceG or GFP-tagged LucB were buffer-exchanged in storage buffer (50 mM Tris-HCl pH 7.5, 20% (v/v) glycerol 5 mM MgSO₄, 150 mM NaCl, 1 mM DDM and 1 mM DTT) and stored separately at –80 °C.

Western blot analysis to detect GFP

Purified protein fractions were separated on the Mini-PROTEAN TGX Stain-Free protein gel (Bio-Rad Laboratories). Separated protein bands were visualized using the stain free gel application mode on the ChemiDoc MP Imaging System (Bio-Rad Laboratories). Protein gel was transferred to a nitrocellulose membrane (Bio-Rad, 1704271) using the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories). Membranes were blocked in PBST containing 5% milk for 30 min at 22 °C. The membranes were then incubated with primary antibodies against GFP (custom polyclonal rabbit anti-GFP antibody (provided by the Foley laboratory) at a dilution of 1:5,000) and His (monoclonal mouse anti-penta-His antibody (Qiagen, 34660) at a dilution of 1:10,000) in PBST + 5% BSA overnight at 4 °C. The membranes were washed three times with PBST and were incubated with goat anti-rabbit IgG polyclonal antibody (IRDye 800CW, LI-COR Biosciences, 925-32211, at dilution of 1:10,000) and goat anti-mouse IgG polyclonal antibody (IRDye 680RD, LI-COR Biosciences, 926-68070, at a dilution of 1:10,000) as the secondary antibodies in PBST + 5% BSA for 1 h at 22 °C. The membranes were washed three times with PBST and imaged using the LI-COR system (LI-COR Biosciences) and analysed using ImageJ⁵⁹. Gel source data are provided in Supplementary Fig. 1a,c.

Negative-stain EM

To prepare grids for negative-stain EM analysis, a fresh sample of either MceG-GFP or LucB-GFP was applied to a freshly glow-discharged (30 s) carbon-coated 400 mesh copper grid (Ted Pella, 01754-F) and blotted off. Immediately after blotting, a 2% uranyl formate solution was applied for staining and blotted off on filter paper. Application and blotting of stain was repeated five times. The samples were allowed to air dry before imaging. Data were collected on the Talos L120C TEM (FEI) equipped with the 4k x 4k OneView camera (Gatan) at a nominal magnification of $\times 73,000$ corresponding to a pixel size of 2.00 Å px⁻¹ on the sample, and a defocus range of –1 to –2 μm defocus. For negative stain-EM, no statistical methods were used to predetermine sample size. Negative-stain dataset size was determined to be sufficient by the ability to see features in the 2D classes of picked particles. For LucB-GFP data, 60 micrographs were acquired and data processing was performed in cryoSPARC (v.3.3.1)⁵². Micrographs were imported and CTF was estimated using Patch CTF (amplitude contrast: 0.4, minimum resolution: 60 Å). In total, 82 particles were picked manually in the Manual Picker module, extracted with a 256 pixel box size and sorted using 2D classification to generate 2D templates for Template Picker. Using these templates, 7,351 particles were picked with template picking, extracted with a 256 pixel box size and sorted using 2D classification to further refine 2D templates for Template Picker. The 2D classes with the best details were used as 2D templates for a final round of picking. In the final round, 6,937 particles were picked, extracted with a 256 pixel box size and sorted by 2D classification. The most populated 2D class is shown in Extended Data Fig. 9d, with 927 particles.

Sample preparation for MS

Protein samples from WT *M. smegmatis* cells (strain mc²155, bBEL246), MceG–GFP strain (bBEL591), LucB–GFP (bBEL595) strain were purified using the protein purification method described above. Three biological replicates were performed for each strain and analysed by MS. Affinity-purified proteins were reduced with DTT at 57 °C for 1 h (2 µl of 0.2 M) and subsequently alkylated with iodoacetamide at room temperature in the dark for 45 min (2 µl of 0.5 M). To remove detergents and other buffer components, the samples were loaded onto a NuPAGE 4–12% Bis-Tris Gel 1.0 mm (Life Technologies). The gel was run for approximately 25 min at 200 V. The gel was stained using GelCode Blue Stain Reagent (Thermo Fisher Scientific). The entire protein band was excised, extracted and analysed in a single MS analysis per gel lane. The excised gel pieces were destained in 1:1 (v/v) solution of methanol and 100 mM ammonium bicarbonate solution using at least three exchanges of destaining solution. The destained gel pieces were partially dehydrated with an acetonitrile rinse and further dried in a SpeedVac concentrator until dry. Then, 200 ng of sequencing-grade modified trypsin (Promega) was added to each sample. After the trypsin was absorbed, 250 µl of 100 mM ammonium bicarbonate was added to cover the gel pieces. Digestion proceeded overnight on a shaker at room temperature. The solution was removed and placed into a separate Eppendorf tube. The gel pieces were covered with a solution of 5% formic acid and acetonitrile (1:2; v:v) and incubated with agitation for 15 min at 37 °C. The extraction buffer was removed and placed into the Eppendorf tube with the previously removed solution. This was repeated three times and the solution dried in the SpeedVac concentrator. The samples were reconstituted in 0.5% acetic acid and loaded onto equilibrated Micro spin columns (Harvard apparatus) using a micro centrifuge. The bound peptides were washed three times with 0.1% TFA followed with one wash with 0.5% TFA. Peptides were eluted by the addition of 40% acetonitrile in 0.5% acetic acid followed by 80% acetonitrile in 0.5% acetic acid. The organic solvent was removed using the SpeedVac concentrator and the sample was reconstituted in 0.5% acetic acid and kept at –80 °C until analysis.

MS data collection

Liquid chromatography separation was performed online on the EASY-nLC 1200 (Thermo Fisher Scientific) system using the Acclaim PepMap 100 (75 µm × 2 cm) precolumn and PepMap RSLC C18 (2 µm, 100 Å × 50 cm) analytical column. Peptides were gradient-eluted directly to an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) using a 95 min acetonitrile gradient from 5 to 35% B over 60 min followed by a ramp to 45% B over 10 min and 100% B over another 10 min with a hold at 100% B for 10 min (where A is 2% acetonitrile in 0.5% acetic acid; and B is 80% acetonitrile in 0.5% acetic acid). The flow rate was set to 200 nL min^{–1}. High-resolution full MS spectra were acquired every 3 s with a resolution of 120,000, an AGC target of 4 × 10⁵, with a maximum ion injection time of 50 ms and a scan range of 400 to 1,500 *m/z*. After each full MS scan, data-dependent HCD MS/MS scans were acquired in the Orbitrap using a resolution of 30,000, an AGC target of 2 × 10⁵, a maximum ion time of 200 ms, one microscan, 2 *m/z*/isolation window, normalized collision energy (NCE) of 27 and dynamic exclusion of 30 s. Only ions with a charge state of 2–5 were allowed to trigger an MS2 scan.

Analysis of MS data

The MS/MS spectra were searched against the NCBI *M. smegmatis* database with common laboratory contaminants and the sequence of the tagged bait proteins were added using SEQUEST within Proteome Discoverer (v.1.4; Thermo Fisher Scientific). The search parameters were as follows: mass accuracy better than 10 ppm for MS1 and 0.02 Da for MS2, two missed cleavages, fixed modification carbamidomethyl on cysteine, variable modification of oxidation on methionine and deamidation on asparagine and glutamine. The data were filtered using

a 1% FDR cut-off for peptides and proteins against a decoy database and only proteins with at least two unique peptides are reported in Supplementary Table 2.

To obtain a probabilistic score (SAINT score) that a protein is an interactor of either MceG or LucB, the data were analysed using the SAINT Express algorithm⁵¹. A one-sided volcano plot was generated showing fold change (Tag/WT) versus SAINT score. Proteins with a SAINT score ≥0.67 yielded an FDR of ≤5% and were considered to be potential interactors. Analysed data are annotated in Supplementary Table 3 (for MceG) and in Supplementary Table 5 (for LucB) and plotted in Fig. 1g (for MceG) and Extended Data Fig. 9e (for LucB) using Prism (GraphPad).

Cryo-EM sample preparation

The MceG–GFP complex was freshly purified as described above. Gel-filtration fractions corresponding to higher-molecular-mass complexes containing MceG were screened by negative-stain EM. Fractions of interest were then concentrated to around 1.7 mg ml^{–1} in cryo-EM buffer (50 mM Tris-HCl pH 7.5, 5 mM MgSO₄, 150 mM NaCl, 1 mM DDM and 1 mM DTT). Continuous carbon grids (Quantifoil R 2/2 on Cu 300 mesh grids + 2 nm Carbon, Quantifoil Micro Tools, C2-C16nCu30-01) were glow-discharged for 5 s in an easiGlow Glow Discharge Cleaning System (Ted Pella). In total, 3.5 µl sample was added to the glow-discharged grid. Using a Vitrobot Mark IV (Thermo Fisher Scientific), grids were blotted for 3–3.5 s at 22 °C at 100% chamber humidity and then plunge-frozen into liquid ethane. Grids were clipped for screening.

Cryo-EM screening and data collection

Clipped cryo-EM grids were screened at the NYU cryo-EM laboratory on the Talos Arctica (Thermo Fisher Scientific) system equipped with a K3 camera (Gatan). Images of the grids were collected at a nominal magnification of ×36,000 (corresponding to a pixel size of 1.0965 Å) with total dose of 50 e[–] Å^{–2}, over a defocus range of –2.0 to –3.0 µm. The grids were selected for data collection on the basis of ice quality and particle distribution. Selected cryo-EM grids were imaged at the Pacific Northwest Center for Cryo-EM on Krios 2, a Titan Krios G3 electron microscope (Thermo Fisher Scientific) equipped with a K3 BioContinuum direct electron detector (Gatan). Super-resolution videos were collected at 300 kV using SerialEM⁶⁰ at a nominal magnification of ×105,000, corresponding to a super-resolution pixel size of 0.41275 Å (or a nominal pixel size of 0.8255 Å after binning by 2). Videos were collected over a defocus range of –0.8 to –2.4 µm and each video consisted of 60 frames with a total dose of 60 e[–] Å^{–2}. A total of 43,925 videos were collected, consisting of 21,915 videos at 0° tilt and 22,010 videos at –30° tilt. Further data collection parameters are shown in Supplementary Table 4. Cryo-EM sample size was chosen as a maximum possible with practical limitation for data collection and processing. No statistical methods were used to predetermine sample size for cryo-EM. The cryo-EM dataset size was determined to be sufficient by the ability to reach resolutions beyond 4 Å in 3D construction.

Cryo-EM data processing

The dataset was split up into batches of 1,000 videos (45 batches total) and processed in cryoSPARC (v.3.3.1)⁵², as described in Extended Data Figs. 3 and 4. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using the cryoSPARC Patch Motion module. The contrast transfer function was estimated for each summed image using cryoSPARC Patch CTF.

From the first batch of 1,000 images, 27 particles were manually picked in cryoSPARC that were then extracted (box size = 480 px) and used to train within the Topaz Train module⁶¹ in cryoSPARC (expected number of particles = 50, use pre-trained initialization, ResNet16). After training, particles were picked using the trained Topaz model and extracted (10,618 particles, box size = 480 px). CryoSPARC 2D

Article

classification (n = number of classes = 50) was performed and particles from 2D classes with high resolution detail were selected (1,051 particles) for Topaz Train (expected number of particles = 300, use pretrained initialization, ResNet16). The trained Topaz model was used to pick and extract 105,604 (box size = 480) particles that were curated by 2D classification (n = 50). Particles from the well-defined classes were selected (14,402 particles after removing duplicates) and further curated using 2D classification (n = 50).

Particles from classes representing the top, side and tilted views were selected (2,887 particles) and processed using cryoSPARC ab initio reconstruction to generate an initial 3D model (Ref1 complex (1,268 particles), Ref2 (919 particles), Ref3 (700 particles)). To generate decoys for downstream particle curation, 50,927 'junk' particles were selected from the 2D classification and processed using cryoSPARC ab initio reconstruction to generate three decoy models (decoy 1 (17,094 particles), decoy 2 (16,915 particles) and decoy 3 (16,918 particles)). For a more isotropic reconstruction in 3D, the 1,268 particles from Ref1 were sorted in 2D (n = 10) and different views of the particles were selected individually: side (588 particles), tilted (505 particles), top (43 particles). These select particles were used to generate Topaz models to specifically pick side, tilted and top views of the particle using the Topaz Train module (expected number of particles = 300, use pretrained initialization, ResNet16).

Using these Topaz picking models, separate Topaz Extract jobs were performed for each view, particles were extracted (box size = 480, binned by 4) and combined. The combined particles were curated by cryoSPARC 2D classification (n = 50), processed for duplicate removal (alignments2D) and curated in 3D using cryoSPARC heterogenous refinement (n = 4, templates = (1) decoy 1, (2) decoy 2, (3) decoy 3, (4) model). Particles sorted into template 4 (model) were selected for further processing. This curation scheme was performed on each batch of micrographs resulting in 2,869,223 curated particles, in which 1,820,584 particles came from the non-tilted images and the remaining 1,048,639 particles came from the -30° tilted images.

Particles were re-extracted (box size = 360 px, unbinned) and were further curated by running six rounds of heterogeneous refinement (n = 4, templates = (1) decoy 1, (2) decoy 2, (3) decoy 3, (4) Model), in which particles that were sorted into template 4 (Model) were used as the input for the next round. After multiple rounds of heterogeneous refinement (round 1, 992,273 particles; round 2, 637,446 particles; round 3, 510,255 particles; round 4, 468,001 particles; round 5, 437,324 particles; round 6, 414,343 particles) and removing remaining duplicates (alignment3D), the 341,566 curated particles were refined using cryoSPARC non-uniform refinement⁶², generating a consensus map at 2.83 Å resolution. Particles images were randomly assigned into odd/even groups for resolution assessment (gold-standard Fourier shell correlation (FSC)). Blinding is not needed for cryo-EM analysis due to the automated handling of the data.

Heterogeneity was observed around the IM region of the complex so particles were processed for a round of heterogeneous refinement (n = 4, templates = (1–4) consensus map). Class A (48,786 particles) and class B (113,261 particles) both contained additional density corresponding to extra protein density in the IM region and were combined, whereas the additional density was not observed in class C (59,724 particles) and class D (119,795 particles). Class C and Class D were very similar when compared by visual inspection, and these two classes were therefore combined. Non-uniform refinement was performed on the combined sets of particles, resulting in two major classes (both containing density for MceG, YrbE1AB and Mce1ABCDEF): class 1, which contains the extra protein density (162,047 particles, 2.94 Å), and class 2, which lacks this density (179,519 particles, 3.04 Å).

Local refinements were performed for each class by recentring the particles on the region of interest using the cryoSPARC Volume Alignment Tool, re-extracting the particles with the new centre (box size = 360 px, unbinned), refining the particles on the re-centred 3D

template using non-uniform refinement, performing particle subtraction in cryoSPARC using a mask around the region of interest, followed by refinement using cryoSPARC local refinement of the subtracted particles. This procedure was performed on each class to generate locally refined maps for the following regions: (1) MceG₂, (2) YrbE1AB + Mce1ABCDEF (TM helix + TM domains + MCE ring) ± extra factor, (3) Mce1ABCDEF (MCE ring + first half of C-terminal MCE needle), (4) Mce1ABCDEF (second half of C-terminal MCE needle). For class 1, the following maps were generated for corresponding regions: (1) Map1a (161,434 particles, 3.05 Å), (2) Map1b (162,004 particles, 2.89 Å), (3) Map1c (158,508 particles, 2.97 Å), (4) Map1d (156,741 particles, 3.16 Å). For class 2, the following maps were generated for each region: (1) Map2a (178,844 particles, 3.13 Å), (2) Map2b (179,480 particles, 2.99 Å), (3) Map2c (175,490 particles, 3.06 Å), (4) Map2d (173,315 particles, 3.19 Å). To generate a composite map, particles from each class were re-extracted with a box size of 640 px (unbinned) and refined using non-uniform refinement to generate maps that included the entire complex (Map1e for class 1 and Map2e for class 2). These maps were used as a template to stitch the locally refined maps together to generate a composite density map. In regions aside from the extra density (later assigned as LucB/MSMEG_3032), these maps were lower resolution compared with the map from the consensus set of particles before classification, but did not show any notable differences compared with the consensus map. Thus, local refinements were performed on the consensus set of particles in a similar manner as used to generate maps for model building, but with masking out the MSMEG_3032/LucB density.

Local refinements were performed using the same approach that was applied to class 1 and class 2 on the set particles from the consensus refinement. This procedure was used on the following regions: (1) MceG₂, (2) YrbE1AB + Mce1ABCDEF (TM helix + TM domains + MCE ring) masking out density for the extra factor, (3) Mce1ABCDEF (MCE ring + first half of C-terminal MCE needle), (4) Mce1ABCDEF (second half of C-terminal MCE needle). For the consensus map, the following locally refined maps were generated for each region: (1) Map0a (340,238 particles, 2.91 Å), (2) Map0b (341,490 particles, 2.73 Å), (3) Map0c (332,050 particles, 2.75 Å), (4) Map0d (330,104 particles, 3.00 Å). To generate a composite map, the consensus set of particles was also re-extracted with a box size of 640 px (unbinned) and refined using non-uniform refinement to generate a map that included the entire complex (Map0e). This map was used as a template to stitch the locally refined maps together to generate a composite density map. These maps were of much higher quality compared to local refined maps of class 1 and class 2, and were therefore used for initial model building.

For each map, the overall resolution reported in cryoSPARC was estimated using the gold-standard FSC criterion (FSC = 0.143). Directional FSCs were computed using 3DFSC⁶³. Local-resolution maps were computed using the cryoSPARC Local Resolution Estimation module. Locally refined maps were combined into composite maps for the consensus map, class 1 and class 2 using the PHENIX (v.1.20.1) Combine Focused Maps module⁶⁴. Composite maps were generated for sharpened maps and half maps (for calculating FSC and estimating local resolution of the composite maps). For the consensus composite map, maps 0a, 0b, 0c and 0d were combined using Map0e as a template to generate Map0. For the class 1 composite map, maps 1a, 1b, 1c and 1d were combined using Map1e as a template to generate Map1. For the class 2 composite map, maps 3a, 3b, 3c and 3d were combined using Map2e as a template to generate Map2. Global FSCs were calculated by importing composite half maps into the Validation FSC cryoSPARC module and the local resolution was estimated using the Local Resolution cryoSPARC module. The nominal global resolution was estimated to be 2.71 Å for Map0, 2.76 Å for Map1 and 2.90 Å for Map2. Directional FSCs for the composite maps were computed using 3DFSC in cryoSPARC.

Model building and refinement

The MS data indicated a mixture of Mce1 and Mce4 proteins in the cryo-EM sample. To assess which proteins were present in the cryo-EM reconstruction, their stoichiometry and position in the complex, we generated AlphaFold2 (ref.⁵⁵) predictions for each MCE-related protein and assessed their fit into the consensus reconstruction, which contains the ATP-binding cassette (ABC) transporter and the MCE ring. Using ColabFold⁶⁵, AlphaFold⁵⁵ predictions were generated for MceG (AFpdb1), Mce1 proteins (AFpdb1–9), Mce4 proteins (AFpdb10–17), and orphaned MCE protein (AFpdb18). Predictions are summarized in Supplementary Table 6. We performed rigid-body fits of the predicted structures into the cryo-EM map using UCSF Chimera (v.1.16)⁶⁶, and determined that the complex consisted of two protomers of MceG, two protomers of YrbEs and six MCE proteins. The two protomers of MceG (AFpdb1) fit unambiguously into the density that corresponded to the ATPase component of the ABC transporter. For YrbE and MCE proteins, we further refined the rigid-body-fitted models using real-space refinement in PHENIX (v.1.20.1)⁶⁴. We then examined regions of each protein in which the sequences are divergent between candidate proteins and used side-chain density to assign the correct subunit. The YrbE subunits (AFpdb2–3,10–11) were fit as rigid bodies into the TM region of the cryo-EM map using UCSF Chimera and refined in real space using PHENIX. The refined models were manually inspected in COOT (v.0.8.9.2)⁶⁷ to assess the overall fit for the Cα backbone and side chains of each protein into the map. On the basis of manual inspection, we assigned the cryo-EM density to YrbE1A and YrbE1B. The MCE domains of each Mce1 (AFpdb4–9) or Mce4 (AFpdb12–17) protein were fitted into each position of the MCE ring (positions 1–6) using UCSF Chimera. Once fit into the density, the MCE domains were real-space-refined in PHENIX and manually inspected in COOT. On the basis of this analysis, Mce1 proteins fit best into the map and were assigned the following positions in the MCE ring (going clockwise): (1) Mce1A, (2) Mce1E, (3) Mce1B, (4) Mce1C, (5) Mce1D, (6) Mce1F. Thus, using this approach, we were able to unambiguously assign Mce1 protein subunits into the cryo-EM map (Extended Data Fig. 4a,b). Notably, oMce1A (AFpdb18), which was identified in the MS data and is 84% identical to Mce1A, fits well into the cryo-EM map at the same position as Mce1A, suggesting a possible mixture of Mce1A and oMce1A in the reconstruction. Focused 3D classification around regions that differ between the two proteins did not produce classes of which the density was resolved enough to unambiguously assign Mce1A versus oMce1A. Mce1A was used for modelling the Mce1 complex as it belongs in the same operon as the other Mce1 proteins.

As a starting point for model building of the entire complex, AlphaFold2 (ref.⁵⁵) and AlphaFold-Multimer⁶⁸ were used to predict 3D structures of Mce1 proteins and subcomplexes as summarized in Supplementary Table 6. Predictions were performed on ColabFold⁶⁵ and COSMIC⁶⁹. The C-terminal region of AFpdb20 was trimmed starting at the following residues: Mce1A (residue 167), Mce1B (residue 151), Mce1C (residue 149), Mce1D (residue 160), Mce1E (residue 169) and Mce1F (residue 149). For initial model building, AFpdb19, AFpdb20 (trimmed) and AFpdb21 were combined in PyMOL Molecular Graphics System (v.2.5.1 Schrödinger). In brief, chains were renamed for each prediction: Mce1A (chain A), Mce1B (chain B), Mce1C (chain C), Mce1D (chain D), Mce1E (chain E), Mce1F (chain F), MceG (chain G and H), YrbE1A (chain I), YrbE1B (chain J). The predicted models were aligned in PyMOL using the align command: (1) AFpdb19 and AFpdb20 were aligned based on chain I and J; and (2) AFpdb3 was aligned to AFpdb2 based on the first α-helical module of the MCE proteins (chain A 150–167, chain B 134–151, chain C 134–149, chain D 145–160, chain E 151–169, chain F 135–149). Overlapping residues were trimmed, and aligned models were stitched to produce a composite PDB file of the Mce1 complex on the basis of AlphaFold2 predictions.

From the three cryo-EM maps (Map0, Map1, Map2), Map0 has the highest resolution and the most featureful density. Thus, modelling of the Mce1 complex was performed on the locally refined maps corresponding to Map0 (Map0a–d), except for model building of LucB, which was performed using Map1b. Note that Map0 includes Mce1 complex particles with and without LucB. However, as there is no conformational change in the Mce1 complex at the resolutions we are at, the higher number of particles results in better-quality density for the Mce1 complex minus LucB. The starting models were fitted into their corresponding locally refined maps using the Fit in Map function in UCSF Chimera. For each map, the PDB file was trimmed to remove regions of the protein that were not defined in the map. Rigid-body fitting into the cryo-EM maps was performed using PHENIX. Fitted models were visually inspected and manually adjusted in COOT. Real-space refinement with Ramachandran and secondary structure restraints was performed in PHENIX using 5 cycles and 100 iterations to optimize the fit and reduce clashes. These models were iteratively inspected, manually rebuilt in COOT and refined in PHENIX until completion. Models built into the locally refined maps were aligned and combined in PyMOL. These models served as templates to generate a composite density map (Map0) for the consensus set of particles using the PHENIX Combine Focused Maps module.

In Map0, poly-carbon chain unknown ligands (UNLs) were manually built into extra densities corresponding to substrates, and real-space-refined in COOT. Elongated ligands (LIG, chemical string: CCCCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC) were generated using PHENIX eLBOW⁷⁰. Planar ligands were derived from BNZ (benzene) and DKM (5-[{(3S,4S)-3-(dimethylamino)-4-hydroxypyrrolidin-1-yl}-6-fluoro-4-methyl-8-oxo-3,4-dihydro-8H-1-thia-4,9b-diazacyclopent[a]cd]phenalene-9-carboxylic acid). The composite model (containing ligands) was real-space-refined into Map0 using PHENIX with global minimization, Ramachandran, secondary structure and ligand restraints. We use UNLs because the resolution of our density clearly indicates the presence of additional molecules, but is not high enough to unambiguously define these molecules.

Our final consensus model for Map0 is nearly complete, apart from regions in Mce1A (residues 1–17), Mce1C (residues 310–524), Mce1D (residues 1–41 and 268–547), Mce1E (residues 1–32), Mce1F (residues 400–518), MceG protomers (residues 1,256–280 and 326–360), YrbE1A (residues 1–13), and YrbE1B (residues 1–26), which are predicted to be flexible or unstructured (Extended Data Fig. 5a). Notably, no TM helix was observed for Mce1E (MSMEG_0138; Rv0173/LprKin *M. tuberculosis*). Mce1E has been proposed to be a lipoprotein due the presence of a possible signal peptide and lipobox at its N terminus⁷¹. Notably, the first resolvable residue for Mce1E is Cys33, the cysteine that would be lipidated; however, density around this region was not sufficient to resolve this modification. In our MS data, we do not detect N-terminal peptides for Mce1E, suggesting that this region may indeed be cleaved.

Models for Map1 and Map2 were built using the model for Map0 as the starting model. The Map0 model was fitted and trimmed into the locally refined maps of each class in UCSF Chimera and PyMOL. Real-space refinement with Ramachandran and secondary structure restraints was performed in PHENIX. Models were iteratively inspected, manually rebuilt in COOT and refined in PHENIX until completion. For class 1, extra protein density was observed near the TM helix of Mce1C in the IM region of Map1b that corresponded to an additional subunit bound to the complex, LucB. To determine the identity of this unknown protein, we used a combination of model building and AlphaFold2. The Cα backbone of the polypeptide was traced manually in COOT. This Cα model was used to search structural databases (AlphaFold/Swiss-Prot v2, AlphaFold/Proteome v2, PDB100 211201, GMGCL 2204) using TM-align mode in Foldseek⁴⁹. One of the highest-ranking hits from this search (TM-score of 0.9509) was a putative, conserved, integral membrane protein from *M. tuberculosis* (Rv2536, AF-P95017-F1-model_v2.pdb) found from the AlphaFold Protein Structure Database. The structure

Article

of the *M. smegmatis* orthologue of this protein (MSMEG_3032/LucB, AFpdb22) was predicted in ColabFold, docked into the cryo-EM density using Chimera, stitched into the model of Map1 using PyMOL and refined in PHENIX. Completed locally refined models were then aligned and combined in PyMOL and used to generate a composite density map for class 1 (Map1) and class 2 (Map2) in PHENIX. Ligands were added to stitched models for Map1 and Map2 and models were real-space-refined using PHENIX.

Statistics for the final models (Supplementary Table 4) were extracted from the results of the real_space_refine algorithm in PHENIX⁶⁴ as well as MolProbity⁷² and EMRinger⁷³. Structural alignments and associated r.m.s.d. values were calculated using UCSF Chimera (v.1.16)⁶⁶ and PyMOL (Schrödinger). FSCs that were calculated in cryoSPARC were plotted in GraphPad Prism (v.9.3.1). Mce1 tunnel volume was calculated using CASTp (v.3.0)⁵⁴ with a probe radius of 2.2 Å and the inner diameter was calculated using MOLE v.2.5 pore mode⁷⁴. The cavity of the ABC transporter substrate-binding pocket was calculated by CASTp (v.3.0) using a probe radius of 2.2 Å. Figures and Supplementary Videos were generated with PyMOL (Schrödinger), UCSF Chimera and ChimeraX⁵³.

Western blot analysis of *M. smegmatis* lysates

Bacterial strains were cultured in 7H9 medium supplemented with 0.05% (v/v) Tween-80 and the appropriate antibiotics and grown to stationary phase (~2 OD₆₀₀) at 37 °C with shaking. 2.5 ml of culture at an OD₆₀₀ = 2 (or the equivalent) was centrifuged at 4,000 rcf for 10 min, resuspended in 200 µl 1× SDS loading dye and transferred to 1.5 ml Bioruptor Plus TPX microtubes (Diagenode, C30010010-300). Cells were lysed by sonication at 4 °C (high setting, 20 min total, 15 s on, 30 s off) using a Bioruptor (Diagenode). The lysates were heated to 100 °C for 10 min before separation on a Mini-PROTEAN TGX Stain-Free protein gel (Bio-Rad Laboratories). Separated protein bands were visualized using the stain free gel application mode on the ChemiDoc MP Imaging System (Bio-Rad Laboratories). The protein gel was transferred to a nitrocellulose membrane (Bio-Rad, 1704271) using the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories). Membranes were blocked in PBST containing 5% milk for 30 min at 22 °C. The membranes were then incubated with 1 µg ml⁻¹ primary polyclonal antibodies against MceG, Mce1F and Mce4F (Capra Science) in PBST + 5% BSA overnight at 4 °C. The membranes were washed three times with PBST and were incubated with goat anti-rabbit IgG polyclonal antibody (IRDye 800CW, LI-COR Biosciences, 925-32211, at dilution of 1:10,000) and donkey anti-chicken IgG polyclonal antibody (IRDye 680RD, LI-COR Biosciences, 926-68075, at a dilution of 1:10,000) as the secondary antibodies in PBST + 5% BSA for 1 h at 22 °C. The membranes were washed three times with PBST and imaged using the stain-free gel application mode on the ChemiDoc MP Imaging System and using the LI-COR (LI-COR Biosciences) system. ImageJ⁵⁹ was used to quantify the total protein loaded onto the stain-free image of the blot and band intensity from the western blots. Gel source data are provided in Supplementary Fig. 1b,d.

Quantification and statistical analysis

The local resolution of the cryo-EM maps was estimated using cryoSPARC Local Resolution⁵². Directional 3DFSCs were calculated using 3DFSC⁶³. The quantification and statistical analyses for model refinement and validation on deposited models were performed using PHENIX⁶⁴, MolProbity⁷² and EMRinger⁷³. Structural alignments and associated r.m.s.d. values were calculated using UCSF Chimera⁶⁶ and PyMOL (Schrödinger). Tunnel and cavity volumes were calculated using CASTp (v.3.0)⁵⁴ and the tunnel diameter was estimated using MOLE (v.2.5)⁷⁴. Multiple sequence alignments were generated using MUSCLE⁷⁵ and JalView⁷⁶. Phenotypic assays were replicated at least three times ($n = 3$). The mean and s.d. of three replicates were plotted using Prism (GraphPad). Protein purifications and grid preparations

were replicated at least three times ($n = 3$). Western blots of bacterial lysates were replicated at least three times ($n = 3$) and were quantified in ImageJ⁵⁹. The mean and s.d. of replicates were plotted using Prism (GraphPad). MS data were analysed using Proteome Discoverer (v.1.4; Thermo Fisher Scientific) and the SAINT Express algorithm⁵¹ and plotted using Prism (GraphPad).

Materials availability

All unique/stable reagents generated in this study are available without restriction from the corresponding authors.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM maps have been deposited at the Electron Microscopy Data Bank under the following accession codes: Map0 (EMD-29025), Map0a (EMD-29228), Map0b (EMD-29229), Map0c (EMD-29230), Map0d (EMD-29231), Map0e (EMD-29232), Map1 (EMD-29023), Map1a (EMD-29233), Map1b (EMD-29234), Map1c (EMD-29235), Map1d (EMD-29236), Map1e (EMD-29237), Map2 (EMD-29024), Map2a (EMD-29238), Map2b (EMD-29239), Map2c (EMD-29240), Map2d (EMD-29241) and Map2e (EMD-29242). The coordinates of the atomic models have been deposited at the PDB under the following accession codes: 8FEF (Map0), 8FED (Map1) and 8FEE (Map2). Coordinates for the atomic models used to make figures were obtained from the PDB under the following accession codes: 6XBD (MlaFEDB), 6MIT (LptBFGC), 2R19 (LptA), SIV9 (LptDE), 5NIK (MacAB-TolC), 6VOD (LetB) and 1THQ (PagP). Cryo-EM data were deposited in Electron Microscopy Public Image Archive (I1343). The MS files are available at MassIVE under dataset identifier MSV000090807 and ProteomeXchange under identifier PXD038456. A list of the bacterial strains and plasmids that have been deposited at Addgene is provided in Supplementary Table 1 with their identifiers.

56. Murphy, K. C. et al. ORBIT: a new paradigm for genetic engineering of mycobacterial chromosomes. *MBio* **9**, e01467-18 (2018).
57. Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**, 1911–1919 (1990).
58. Pleiner, T. et al. Structural basis for membrane insertion by the human ER membrane protein complex. *Science* **369**, 433–436 (2020).
59. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
60. Mastroserio, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
61. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryo-EM and cryoET. *Nat. Commun.* **11**, 5208 (2020).
62. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
63. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
64. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
65. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
66. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
67. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
68. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv <https://doi.org/10.1101/2021.10.04.463034> (2022).
69. Cianfracco, M. A., Wong-Barnum, M., Youn, C., Wagner, R. & Leschziner, A. COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination. In *Proc. PEARC17: Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact 1–5* (Association for Computing Machinery, 2017).
70. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
71. Sutcliffe, I. C. & Harrington, D. J. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol. Rev.* **28**, 645–659 (2004).

72. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
73. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
74. Pravda, L. et al. MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.* **46**, W368–W373 (2018).
75. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
76. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
77. Suits, M. D. L., Sperandeo, P., Dehò, G., Polissi, A. & Jia, Z. Novel structure of the conserved Gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J. Mol. Biol.* **380**, 476–488 (2008).
78. Botos, I. et al. Structural and functional characterization of the LPS transporter LptDE from Gram-negative pathogens. *Structure* **24**, 965–976 (2016).
79. Fitzpatrick, A. W. P. et al. Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump. *Nat. Microbiol.* **2**, 17070 (2017).
80. Isom, G. L. et al. LetB structure reveals a tunnel for lipid transport across the bacterial envelope. *Cell* **181**, 653–664 (2020).
81. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
82. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

Acknowledgements We thank the members of the Bhabha and Ekiert laboratories for discussions; N. Coudray, J. Ilmain, G. Isom, M. Macrae, F. Rubino and J. Sudar for feedback on our manuscript; H. Darwin for supplying the *M. smegmatis* strain (mc²155); J. Cox, C. Vieni, K. Murphy, R. Voorhees and D. Görlich for sharing plasmids; the members of the Foley laboratory (Memorial Sloan Kettering Cancer Center) for providing purified rabbit GFP antibody; the members of the Chng laboratory (National University of Singapore) for discussion on the palmitate growth assay; K. Dancel-Manning for the illustration of the mycobacterial cell envelope; F. Lang and K. Dancel-Manning for overseeing the use of the Talos L120C microscope and the facility in which the Talos L120C is housed; A. Paquette, W. Rice and

B. Wang for assistance with cryo-EM grid screening and microscope operation; the members of the Central Lab Services team at NYU School of Medicine for preparation of medium and buffers; and S. Mulligan and L. Vega for assistance with cryo-EM data collection. EM data processing has used computing resources at the HPC Facility at NYU, and we thank the members of the HPC team for high-performance computing support. This work was supported by the following funding sources: Schmidt Science Fellows (to J.C.), Jane Coffin Childs Memorial Fund for Medical Research (to J.C.), pilot funding from the NYU Langone Health Antimicrobial-resistant Pathogen Program (to G.B. and D.C.E.) and NIH/NIAID 1R01AI174646 (to G.B. and D.C.E.). G.B. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts (PEW-00033055). The NYU Microscopy Center is partially supported by NYU Cancer Center Support Grant NIH/NCI P30CA016087. The MS experiments were supported in part by NYU Grossman School of Medicine and with a shared instrumentation grant from the NIH (1S10OD010582-01A1) for the purchase of an Orbitrap Fusion Lumos. A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

Author contributions J.C., D.C.E. and G.B. conceived the project. D.C.E. and G.B. supervised and administered the project. J.C. performed cloning, protein purifications and biochemistry. J.C. prepared cryo-EM samples, collected and processed cryo-EM data. J.C., D.C.E. and G.B. built models and performed structural analysis. J.C., A.F. and C.F. performed phenotypic assays. J.P. and B.U. performed MS experiments and analyses. J.C., B.U., D.C.E. and G.B. acquired funding for the project. J.C., D.C.E. and G.B. wrote the original draft of the manuscript. J.C., A.F., C.F., B.U., D.C.E. and G.B. revised and edited manuscript.

Competing interests The authors declare no competing interests.

Additional information

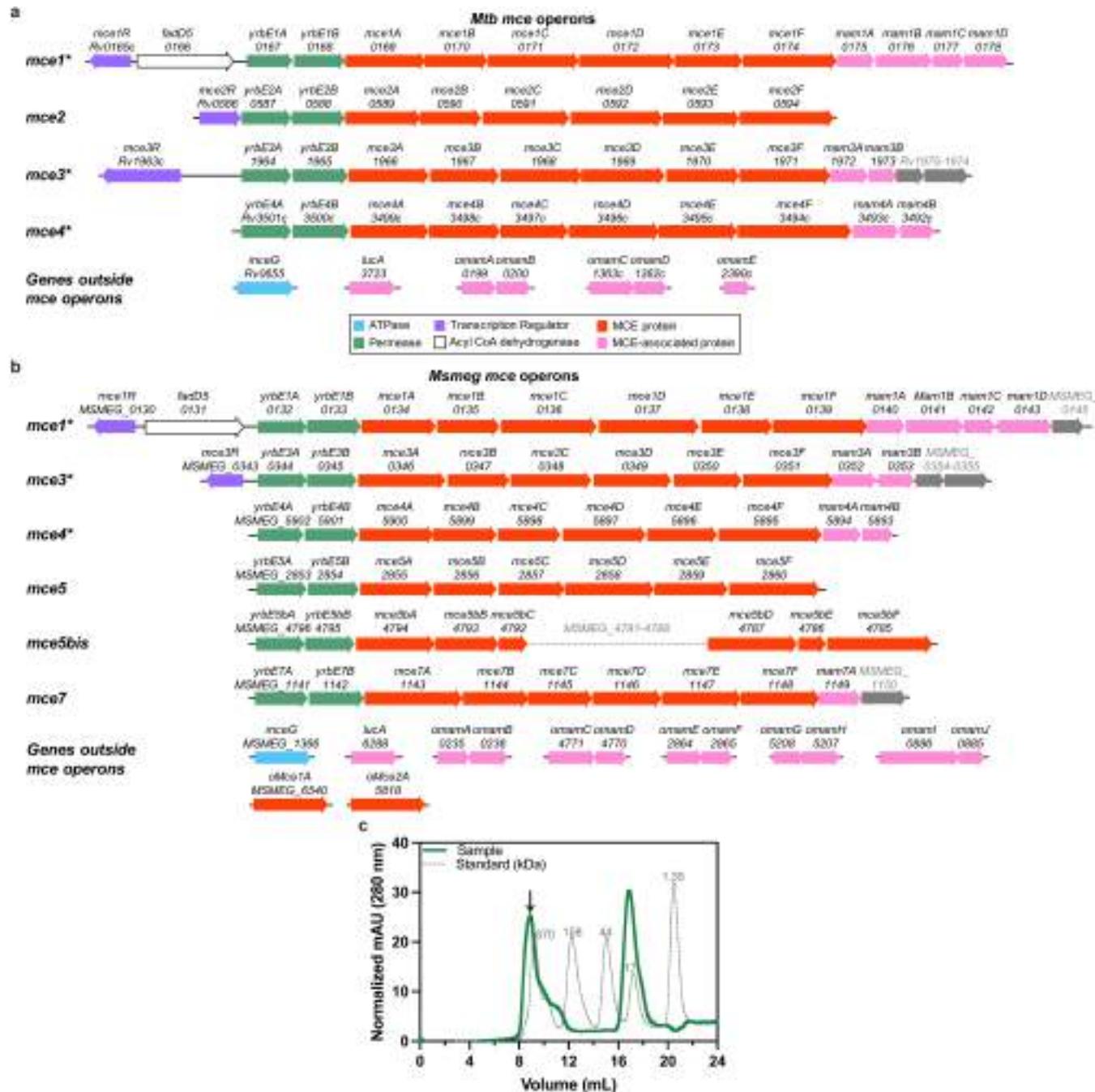
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-06366-0>.

Correspondence and requests for materials should be addressed to Gira Bhabha or Damian C. Ekiert.

Peer review information *Nature* thanks Ben Luisi, Jochen Zimmer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

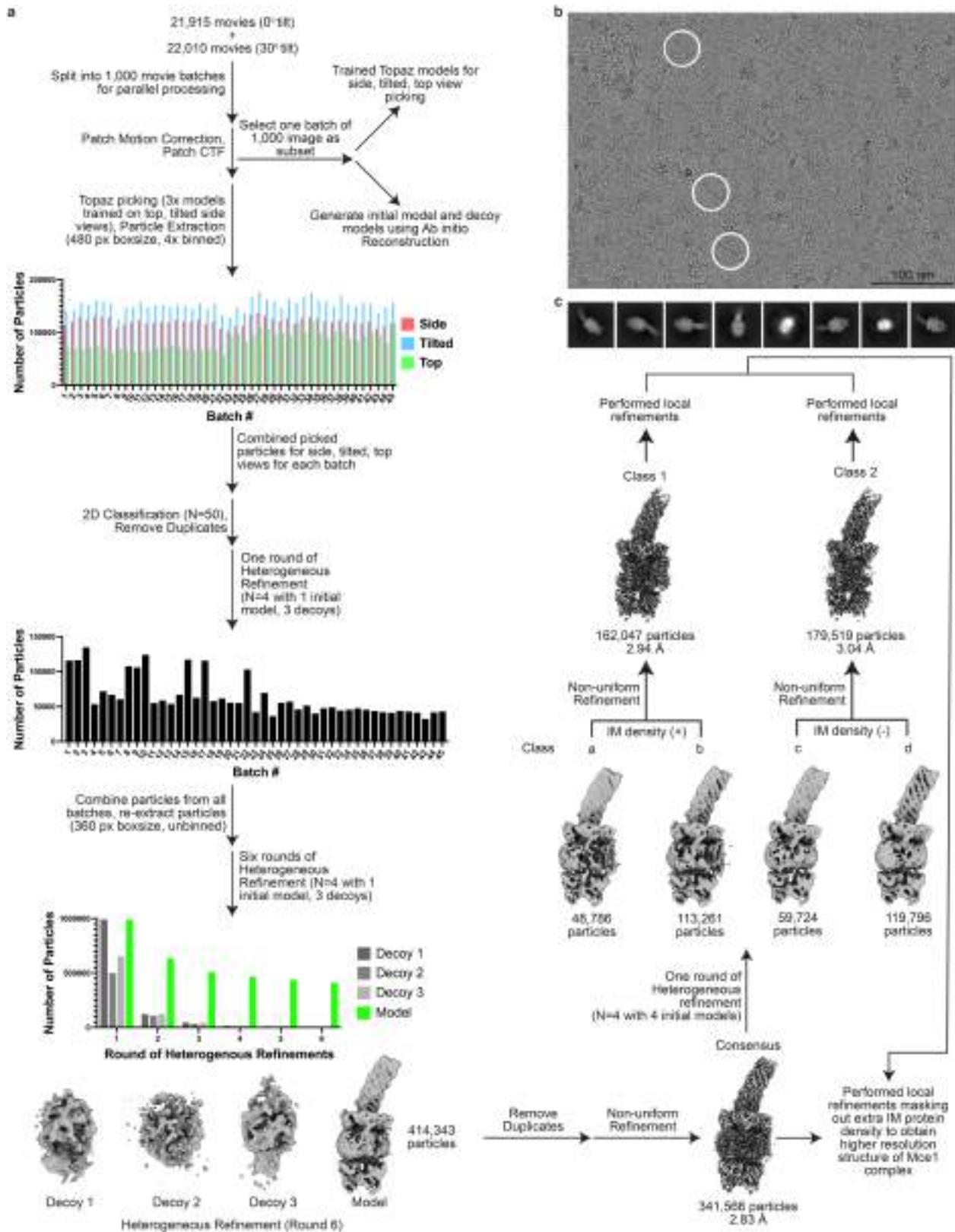
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



Extended Data Fig. 1 | MCE systems in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. **a**, Schematic of the four Mammalian Cell Entry (MCE) operons in the *Mycobacterium tuberculosis* (*Mtb*) genome. Arrows represent individual genes and their direction. Genes coloured as indicated the key below. Operons annotated with an asterisk are conserved between *Mtb* and

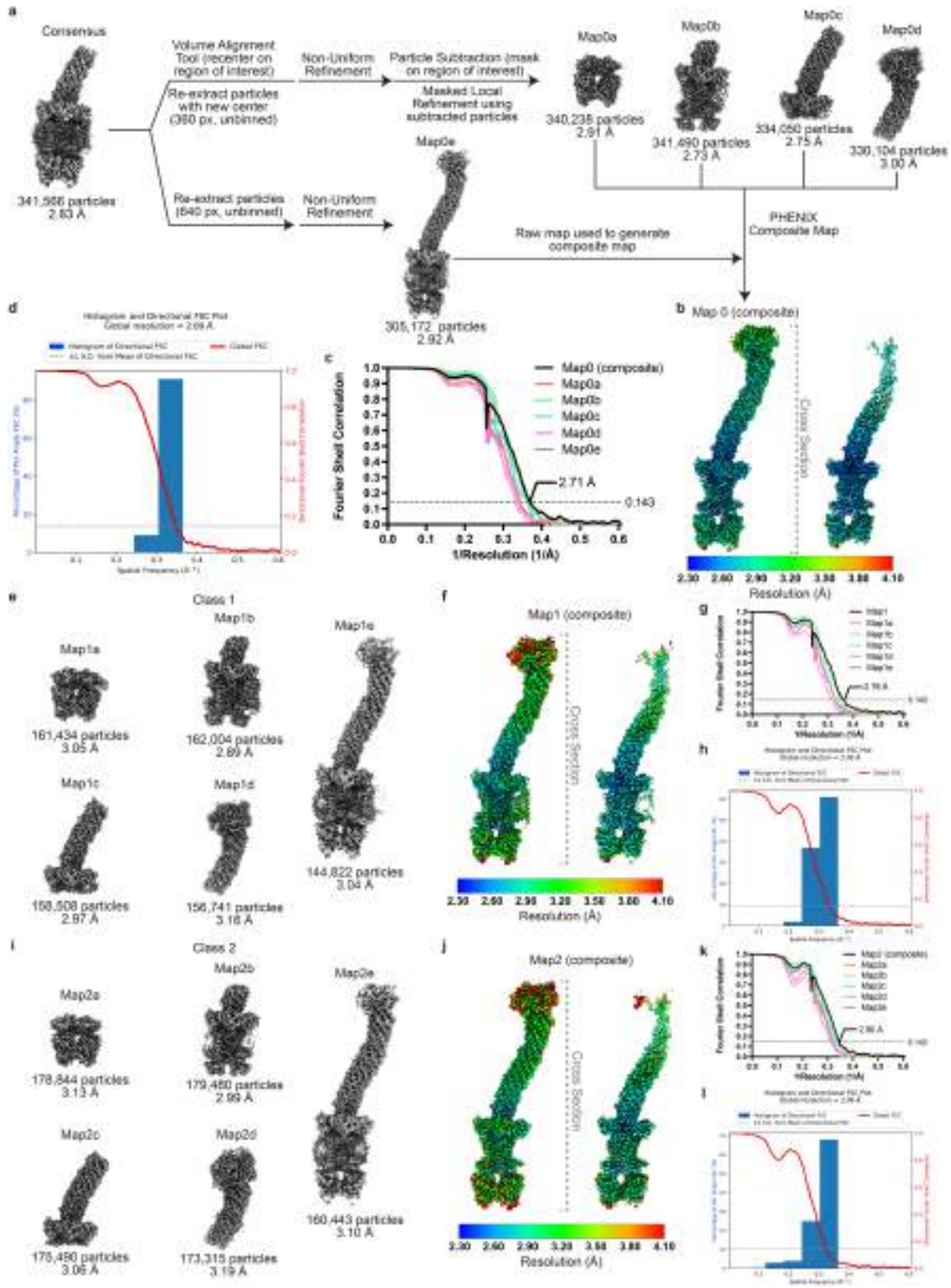
Mycobacterium smegmatis (*Msmeg*). **b**, Schematic of the six MCE operons in the *Msmeg* genome. Genes coloured as in Extended Data Fig. 1a. **c**, Size exclusion chromatogram of MceG-GFP purification (green) superimposed with protein standards (grey dotted line). Black arrow indicates protein sample shown in Fig. 1f and analysed by mass spectrometry in Fig. 1g.



Extended Data Fig. 2 | Cryo-EM data processing workflow (Part 1). **a**, Cryo-EM data processing pipeline. **b**, Representative cryo-EM micrograph from 43,925 micrographs. Particles of interest are circled in white. Scalebar (100 nm) is indicated on the bottom right of the micrograph. Three biologically independent grid preparations of purified MceG-GFP ($n = 3$) yielded similar

micrographs. **c**, Representative 2D classes of complex. Eight 2D class averages showing different views of the particles were generated in cryoSPARC⁵² using the final set of ‘consensus’ particles extracted with box size of 360 pixels and no binning.

Article



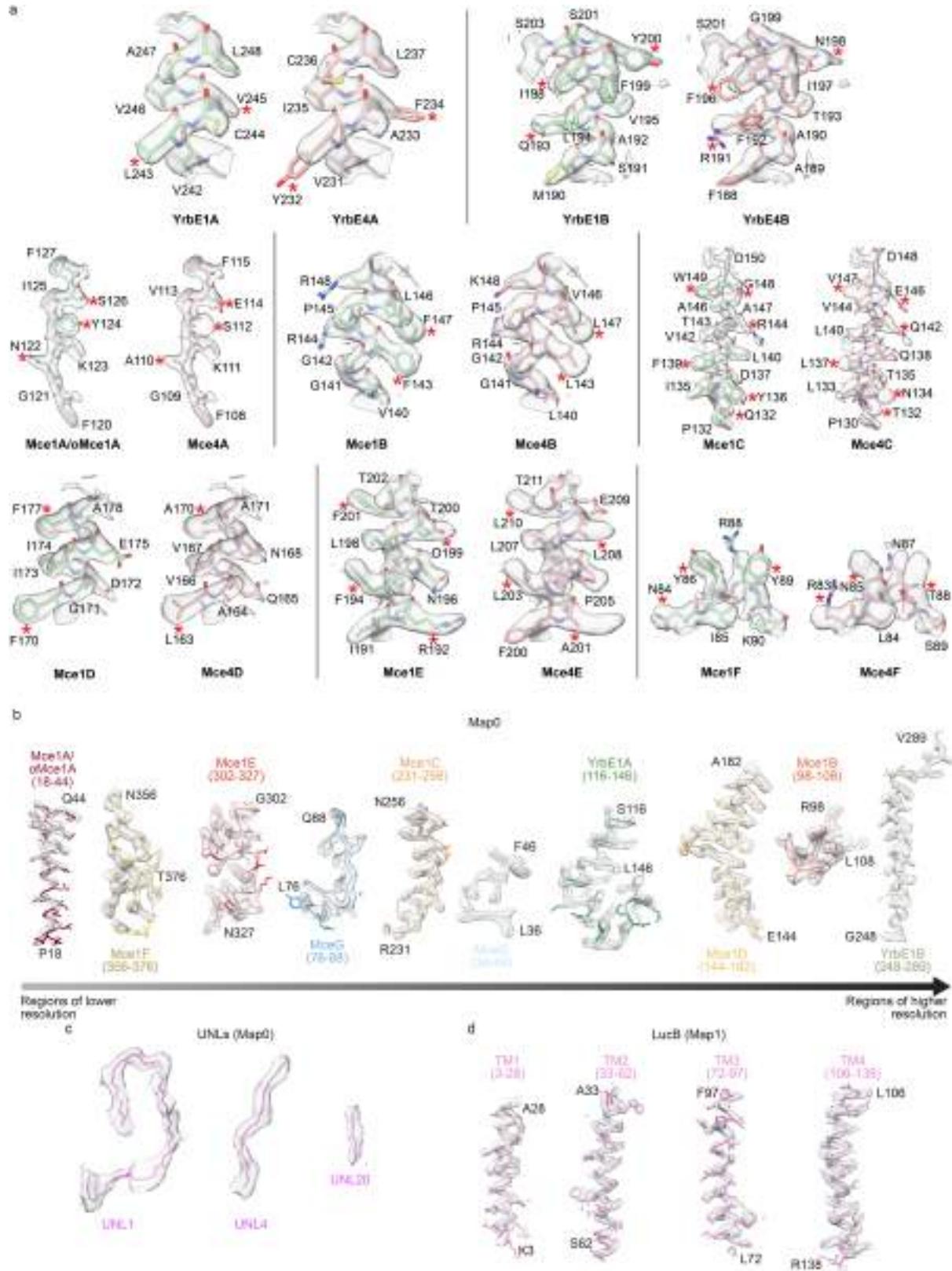
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Cryo-EM data processing workflow (Part 2).

a, Processing pipeline for local refinements performed in cryoSPARC⁵² and composite map generation. Locally refined maps and raw map for consensus set of particles are shown. Locally refined maps for the consensus set of particles were contoured with the following levels: Map0a (0.281), Map0b (0.257), Map0c (0.259), Map0d (0.199), Map0e (0.17) in ChimeraX⁵³. Map labels are indicated above the map and particle count and average resolution as reported by cryoSPARC are shown below. Composite density maps were generated in PHENIX⁶⁴. **b**, Composite Map0 coloured by local resolution that was estimated in cryoSPARC. (left) Whole structure view. Map0 contoured to 12.7. (right) Cross-sectional view. (bottom) key for local resolution colouring, ranging from 2.30 Å (blue) to 4.10 Å (red). **c**, Gold-standard FSC curve calculated in cryoSPARC for composite, locally refined, and raw maps for consensus set of particles. The dotted line represents the 0.143 FSC cutoff. **d**, Directional 3DFSC⁶³ calculated for Map0 (composite). **e**, Locally refined and raw maps for Class 1. Locally refined maps for Class 1 were contoured with the following levels: Map1a (0.172), Map1b (0.201), Map1c (0.185), Map1d (0.167), Map1e (0.15). Map labels are indicated above the map and particle count and

average resolution as reported by cryoSPARC are shown below. **f**, Composite density map for Class 1 (Map1) coloured by local resolution that was estimated in cryoSPARC. (left) Whole structure view. Map1 contoured to 10.1. (right) Cross-sectional view. (bottom) key for local resolution colouring, ranging from 2.30 Å (blue) to 4.10 Å (red). **g**, Gold-standard FSC curve calculated in cryoSPARC for composite, locally refined, and raw maps for Class 1. The dotted line represents the 0.143 FSC cutoff. **h**, Directional 3DFSC calculated for Map1 (composite). **i**, Locally refined and raw maps for Class 2. Locally refined maps for Class 2 were contoured with the following levels: Map2a (0.177), Map2b (0.148), Map2c (0.163), Map2d (0.126), Map2e (0.15). Map label is indicated above the map and particle count and resolution are shown below. **j**, Composite density map for Class 2 (Map2) coloured by local resolution that was estimated in cryoSPARC. (left) Whole structure view. Map2 contoured to 10.2. (right) Cross-sectional view. (bottom) key for local resolution colouring, ranging from 2.30 Å (blue) to 4.10 Å (red). **k**, Gold-standard FSC curve calculated in cryoSPARC for composite, locally refined, and raw maps for Class 2. The dotted line represents the 0.143 FSC cutoff. **l**, Directional 3DFSC calculated for Map2 (composite).

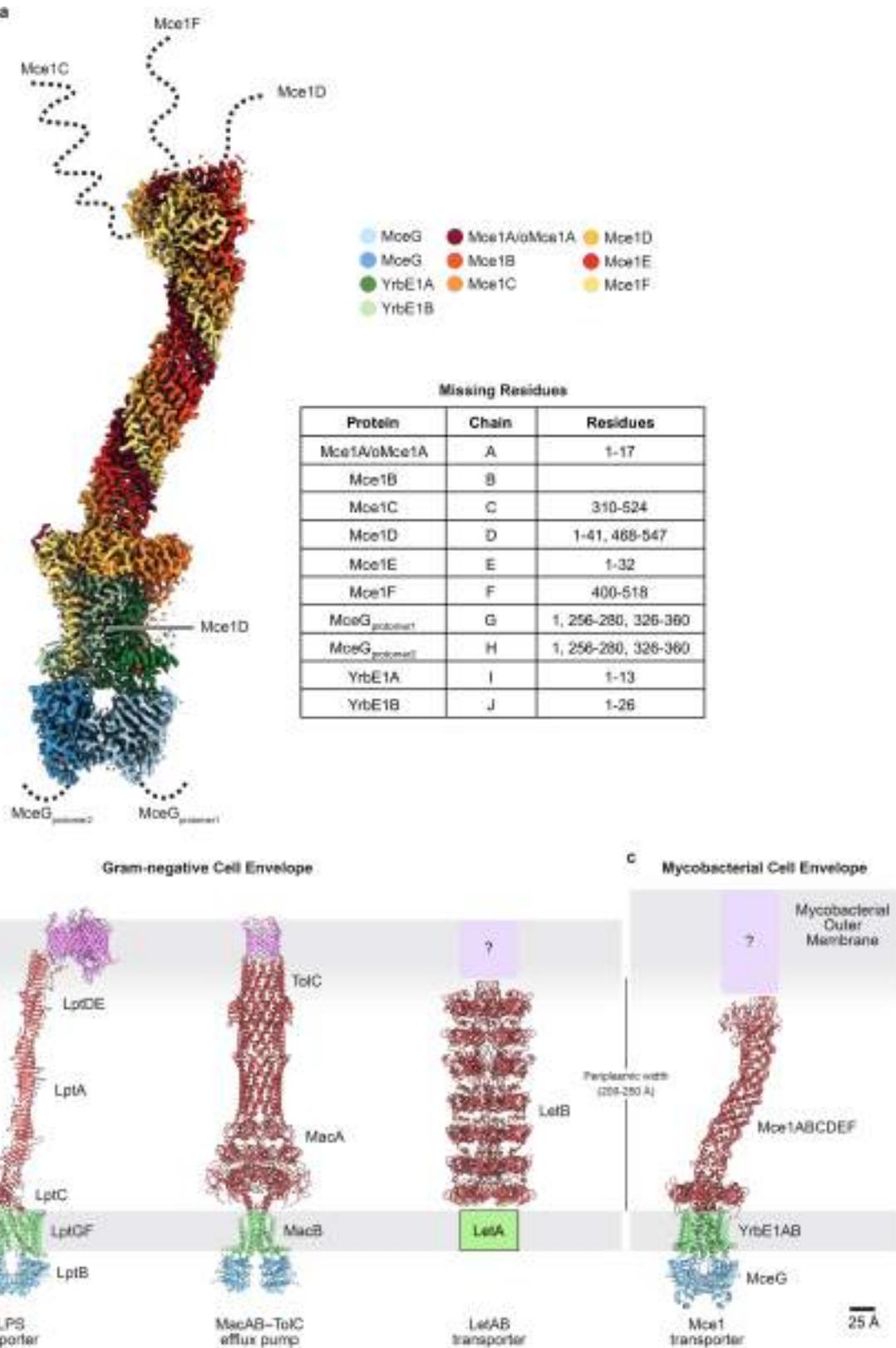
Article



Extended Data Fig. 4 | See next page for caption.

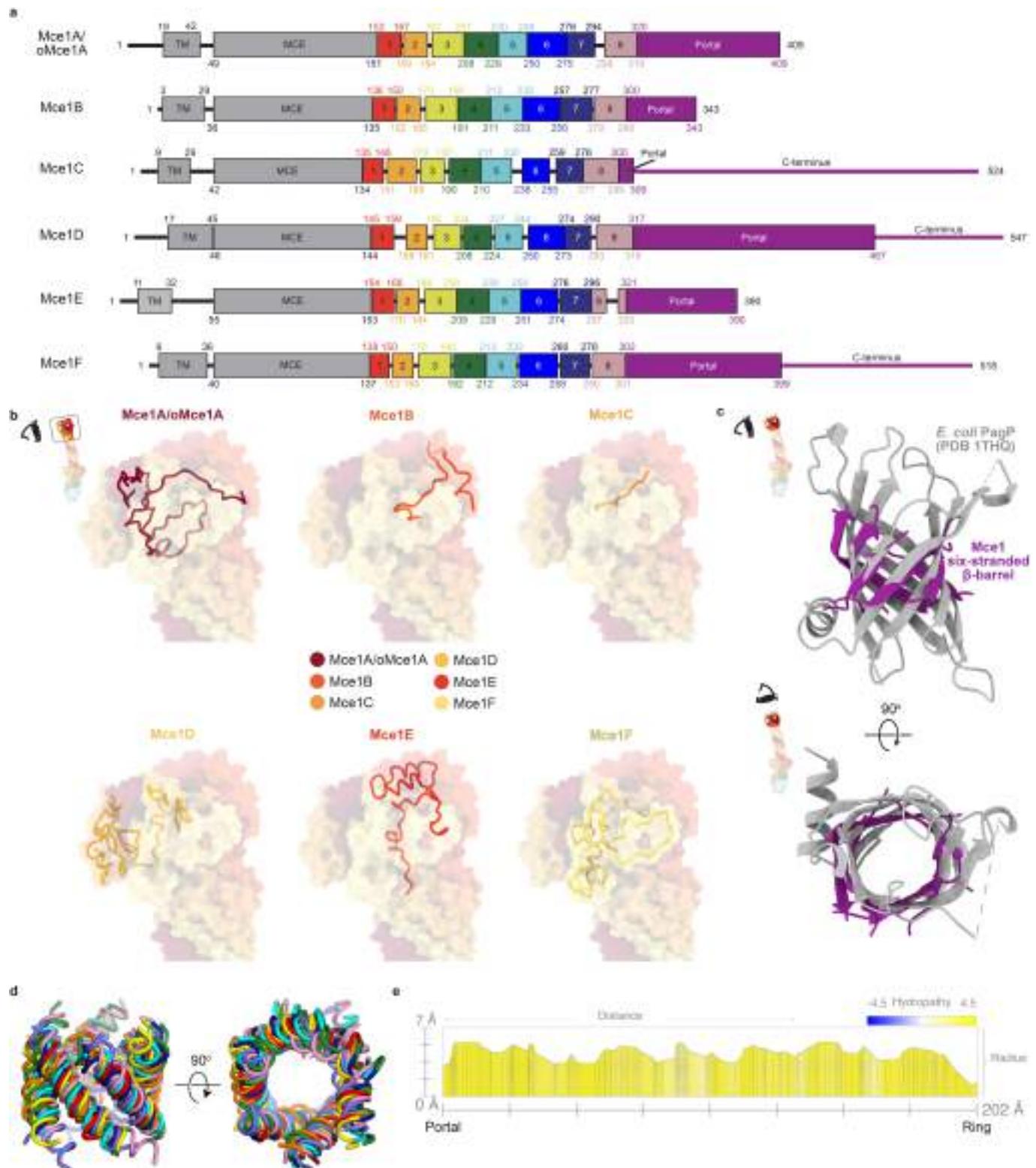
Extended Data Fig. 4 | Model to map fits and representative density from areas at different resolutions. **a**, Examples of model to map fits comparing Mce1 (green) and Mce4 (red) proteins fit into our density. Proteins are shown as sticks and superimposed on to the cryo-EM density from Map0, shown as a transparent grey surface. Protein densities rendered using ChimeraX⁵³ ‘volume zone’ with 2.0–3.0 Å distance cutoff around the indicated protein residues with the following contour levels: YrbE1A (12.0); YrbE1B (10.0) Mce1A/oMce1A (10.0), Mce1B (10.0), Mce1C (10.0), Mce1D (10.0), Mce1E (10.0), Mce1F (8.0). Red asterisks highlight examples of key residues that differ between Mce1 and Mce4 proteins, which led to the identification of Mce1 as the predominant complex represented in our cryo-EM density. **b**, Examples from each subunit

for Map0. Examples are arranged left to right from lower to higher resolution. Protein densities rendered using ChimeraX⁵³ ‘volume zone’ with 2.0 Å distance cutoff around the indicated protein residues with the following contour levels: Mce1A/oMce1A (6.0), Mce1F (14.0), Mce1E (10.0), MceG protomer 2 (10.0), Mce1C (8.0), MceG protomer 1. YrbE1A (12.0), Mce1D (8.0), Mce1B (8.0), YrbE1B (10.0). **c**, Examples of ligand density from Map0. Ligand densities rendered using ChimeraX ‘volume zone’ with 2.5 Å distance cutoff around UNLs and with the following contour levels: UNL1 (8.0), UNL4 (6.0), UNL20 (8.0). **d**, Examples of fits for LucB transmembrane helices (labeled TM1–TM4) from Map1. Protein densities rendered using ChimeraX ‘volume zone’ with 2.5 Å distance cutoff around each helix of LucB and contour level 7.0.



Extended Data Fig. 5 | Final Mce1 model and comparison with periplasm-spanning transporters in double-membraned bacteria. **a,** Missing regions in consensus composite density map (Map0). Map0 contoured to 10.0 and Map0 colour-coded based on the key on the right. Regions of the proteins that were not resolvable in the cryo-EM map are indicated by dotted

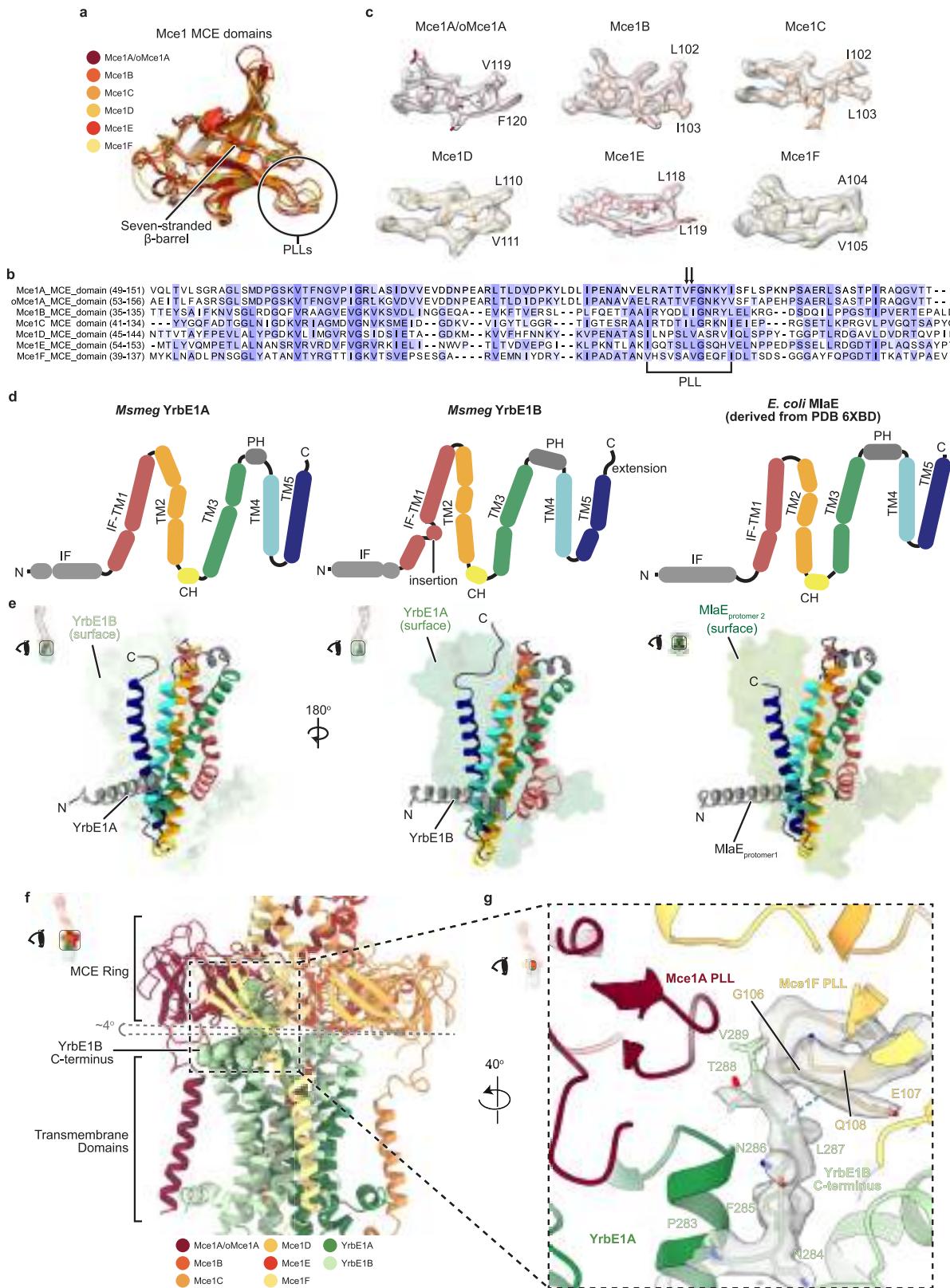
lines, drawn approximately to scale, and indicated in the chart. **b,** Protein complexes that span the cell envelope in Gram-negative bacteria. (left) LPS exporter (modelled based on PDBs 6MIT²³, 2R19⁷⁷, SIV9⁷⁸); (middle) MacAB-TolC efflux pump (PDB 5NIK⁷⁹); (right) LetAB transporter (PDB 6V0D⁸⁰). **c,** Structure of mycobacterial Mce1 transporter.



Extended Data Fig. 6 | Structural features of Mce1ABCDEF portal and needle modules in the Mce1 complex. **a**, Schematic of Mce1ABCDEF proteins showing boundaries for the transmembrane helix (TM), MCE domain (MCE), eight needle modules, and portal domain/C-terminus. **b**, C-termini of Mce1ABCDEF that form the portal domain, viewed as indicated by inset. Proteins are shown as transparent molecular surfaces and coloured as in the key. Superimposed in each panel are the residues of the indicated protein, shown as cartoon. **c**, Alignment of *E. coli* PagP (PDB 1THQ)³⁷ and the Mce1 six-stranded β-barrel of the portal. Proteins were aligned in ChimeraX⁵³.

PagP is coloured grey and the Mce1 six-stranded β-barrel is coloured purple. **d**, Structural alignment of the eight structural modules in PyMOL using ‘cealign’ command (Schrödinger, LLC.). Proteins are superimposed, and coloured as in Extended Data Fig. 6a. **e**, Hydrophobicity and radius of tunnel in Mce1 needle calculated using MOLE 2.5⁷⁴. X-axis shows distance going down the tunnel (from left to right); Y-axis shows the distance away from the central axis of the tunnel. Plot shows hydrophobicity, where blue is hydrophilic and yellow is hydrophobic.

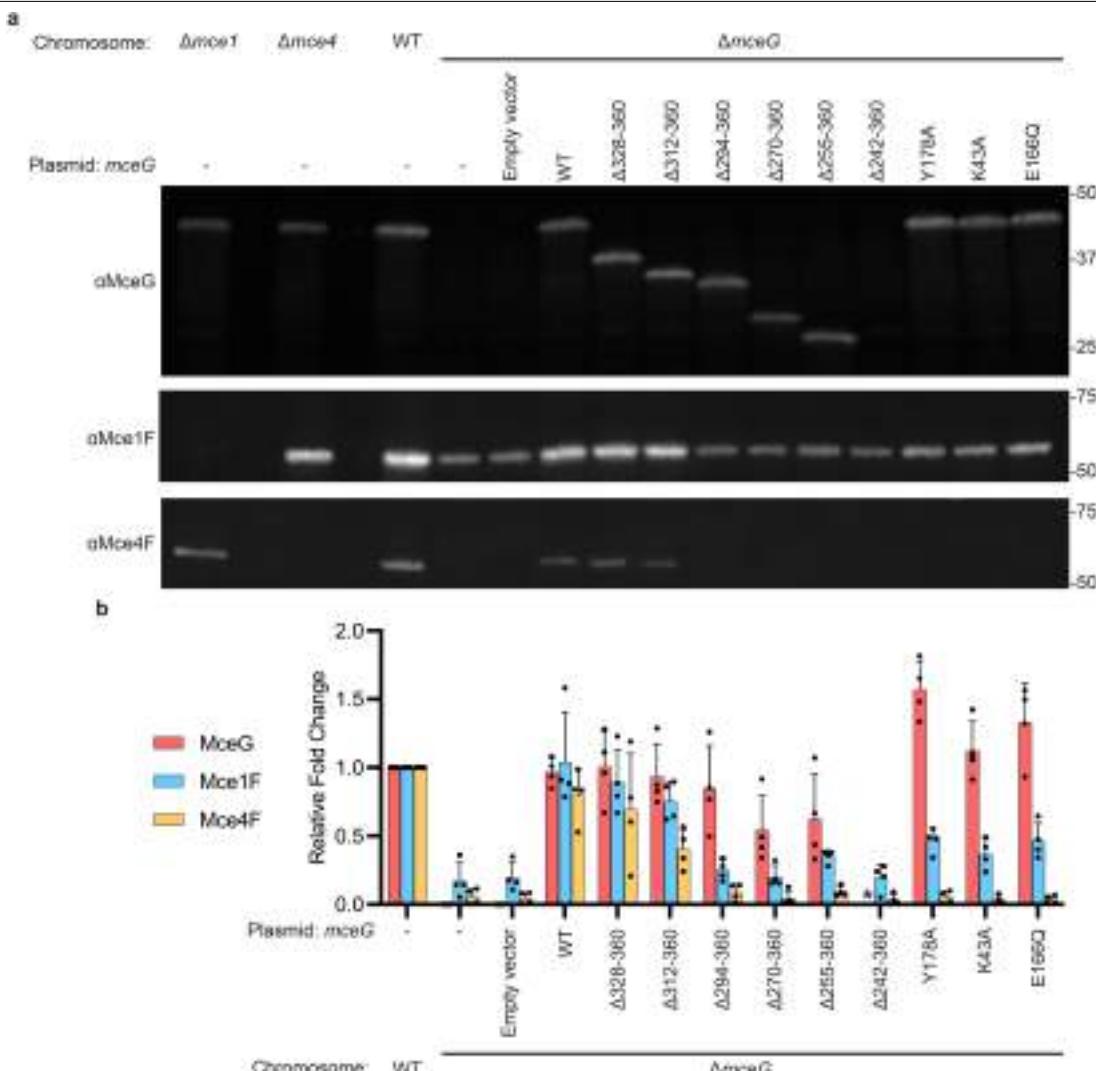
Article



Extended Data Fig. 7 | See next page for caption.

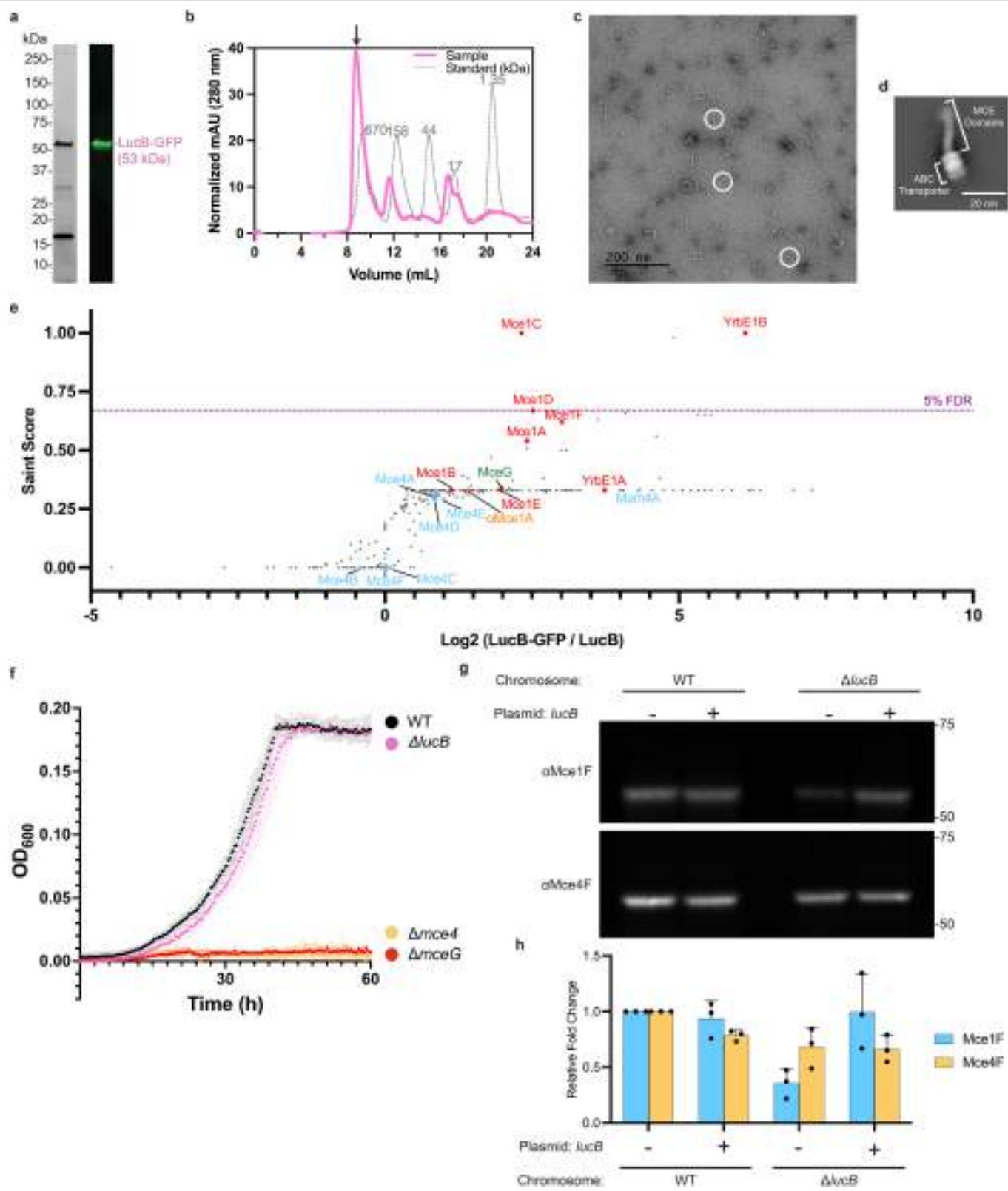
Extended Data Fig. 7 | Structural features of MCE domains and ABC transporter transmembrane domains, YrbE1A and YrbE1B. **a**, Structural alignment of Mce1 MCE domains colour-coded based on key. Domains were aligned to the MCE domain of Mce1A. Pore-lining loops (PLLs) are circled. **b**, Protein sequence alignment of the MCE domains from *Msmeg* Mce1 proteins using MUSCLE⁷⁵ and visualized using JalView⁷⁶. Sequence alignment is coloured by BLOSUM62 score (not conserved, white; conserved, blue). PLL region is highlighted with black bracket and pore facing residues are indicated with black arrows. **c**, Gallery of Mce1 PPLs. Protein backbones are shown as cartoon tubes with residues shown as sticks and coloured as Extended Data Fig. 7a. Cryo-EM density for the PLLs is shown as a grey transparent surface. Protein densities rendered using ChimeraX⁵³ ‘volume zone’ with 2.0 Å distance cutoff around each PLL at contour level 10.0. Pore facing residues are annotated. **d**, Topology diagram of YrbE1A, YrbE1B and *E. coli* homologue MlaE (derived from PDB 6XBD⁴¹). CH, coupling helix; PH, periplasmic helix; IF, interfacial helix; TM,

transmembrane helix. **e**, Structures of YrbE1A, YrbE1B and MlaE. YrbE1A and YrbE1B form a heterodimer, while MlaE forms a homodimer. One protomer is shown as a cartoon and the other as a molecular surface in each representation. **f**, View of Mce1IM complex as indicated by inset on upper left. Model coloured as in the key. The C-terminus of YrbE1B (residues 280–289) is shown as spheres. Grey-dotted lines indicate the MCE ring tilt relative to YrbE1B. **g**, Zoom-in view of region boxed in Extended Data Fig. 7f, oriented as indicated by inset, highlighting interaction between YrbE1B C-terminus and Mce1F PLL. Proteins are shown as cartoon ribbons and coloured as Extended Data Fig. 7f. Cryo-EM density is shown as a transparent grey surface for the YrbE1B C-terminus (chain J, residues 280–289) and Mce1F PLL (chain F, residues 99–110). Protein densities rendered using ChimeraX⁵³ ‘volume zone’ with 2.0 Å distance cutoff around YrbE1B C-terminus and Mce1F PLL and 8.7 contour level. Hydrogen bonds are shown as cyan dotted lines.



Extended Data Fig. 8 | Protein expression of MceG, Mce1F, Mce4F for *mceG* bacterial strains used in growth assays. **a**, Representative western blots showing protein levels of MceG, Mce1F, Mce4F in the lysates of *mceG* bacterial strains used in growth assays. Labels above blots denote the *Msmeg* genotype and *mceG* plasmid used for complementation and molecular weight markers (kDa) are indicated on the right. Lysates of $\Delta mce1$ and $\Delta mce4$ were loaded as controls for anti-Mce1F and anti-Mce4F antibody specificity. Lysates were derived from the same experiment and blots were processed in parallel. For gel source data, see Supplementary Fig. 1b. Western blots were performed with

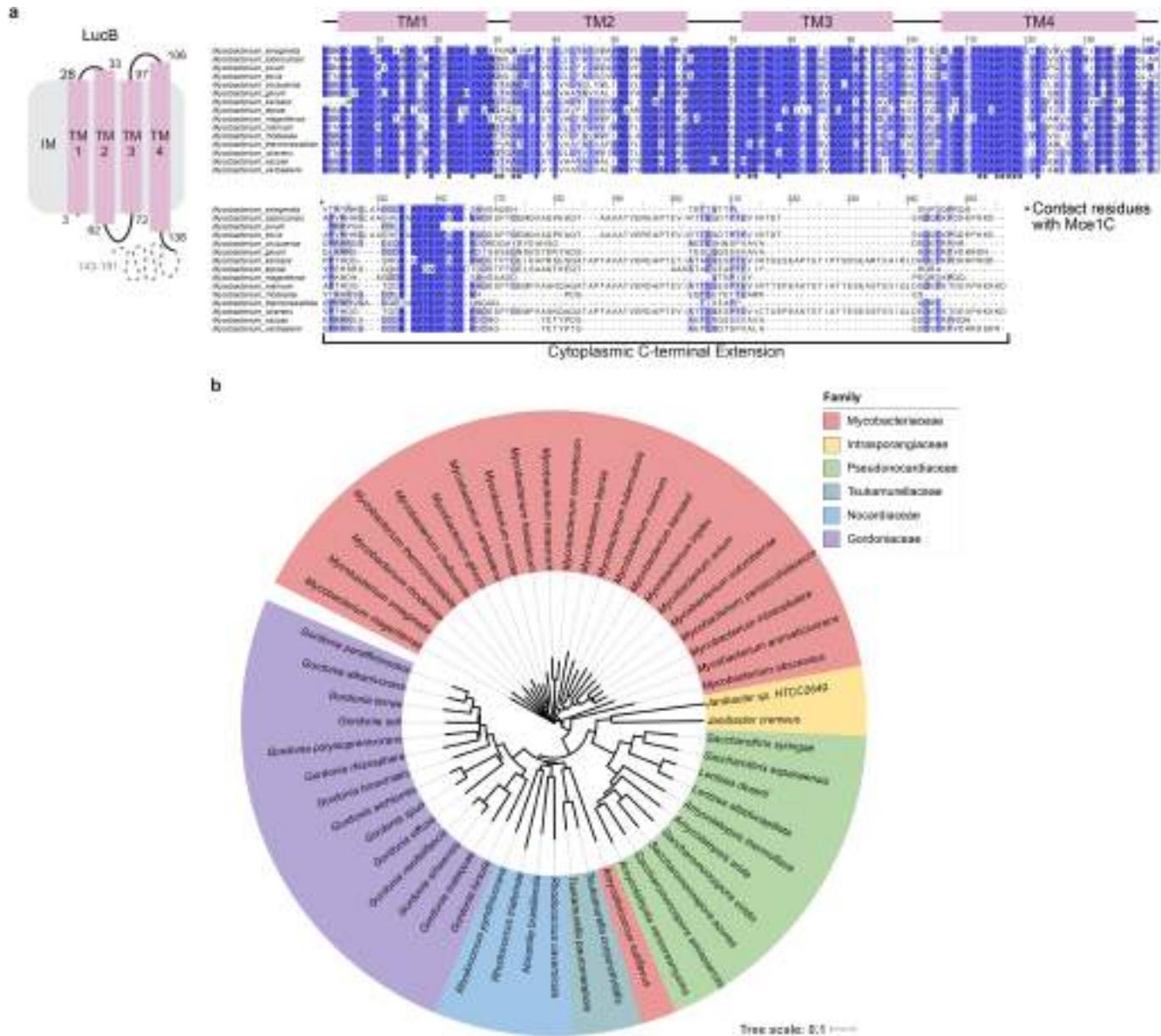
four independent biological replicates ($n = 4$), each with similar result. **b**, Plot showing fold change of protein levels of MceG (red), Mce1F (blue), Mce4F (yellow) in *mceG* mutants compared to Wild-type *Msmeg* (WT) based on the quantification of the western blots. Custom anti-MceG antibody was raised against the peptide 241-NGRRIGPIGMSEEKD-255, therefore cannot be used to detect MceG($\Delta 242\text{-}360$) (denoted with black asterisk on plot). For gel source data, see Supplementary Figure 1b. Plotted data are the mean of four biological replicates ($n = 4$) and error bars representing standard deviations are shown.



Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | LucB purification, proteomics, growth assay, and western blot results. **a**, SDS-PAGE of LucB-GFP affinity purification (left) and Western blot using an anti-GFP antibody against purified LucB-GFP (right). For gel source data, see Supplementary Fig. 1c. Purifications were performed with three independent biological replicates ($n = 3$), with similar results. **b**, Size exclusion chromatogram of LucB-GFP purifications (pink line) superimposed with protein standards (grey dotted line). Black arrow indicates protein sample further analysed by negative stain electron microscopy (Extended Data Fig. 9c,d) and mass spectrometry (Extended Data Fig. 9e). **c**, Negative stain electron microscopy micrograph of purified LucB-GFP from 60 micrographs. Particles of interest, which resemble the Mce1 complex, are circled in white. Scalebar (200 nm) is indicated on the bottom left of the micrograph. Three biologically independent grid preparations of purified LucB-GFP ($n = 3$) yielded similar micrographs. **d**, 2D class average of ‘MCE-like’ particles from negative stain electron microscopy of endogenously purified LucB-GFP. Scale bar = 20 nm. **e**, Plot of proteins identified by mass spectrometry that co-purify with LucB-GFP. Each point corresponds to an individual protein plotted by fold change difference after purification of LucB-GFP from *Msmeg* strain harbouring tagged LucB versus control wild-type *Msmeg mc²155* (x-axis) and the probability that a protein is a LucB interactor (SAINT score; y-axis). SAINT score = 1 identifies proteins with the highest probability of being a LucB interactor⁵¹. SAINT score ≥ 0.67 yielded an FDR (false discovery rate) of $\leq 5\%$ as indicated by the purple dotted line. Proteins related to the mycobacterial MCE systems are highlighted and annotated as shown in the key. Plotted data are from three biological replicates. **f**, Growth curves in minimal media containing cholesterol as the sole carbon source for $\Delta lucB$ (pink), WT *Msmeg mc²155* strain (black), $\Delta mceG$ (red), and $\Delta mce4$ (yellow). Plotted data are the mean of three biological replicates ($n = 3$) and error bars representing standard deviations are shown. **g**, Representative western blots showing protein levels of Mce1F and Mce4F in *Msmeg* strains as indicated in the figure. Labels above blots denote the *Msmeg* genotype and *lucB* plasmid used for complementation and molecular weight markers (kDa) are indicated on the right. Lysates were derived from the same experiment and blots were processed in parallel. For gel source data, see Supplementary Fig. 1d. Western blots were performed with three independent biological replicates ($n = 3$), each with similar result. **h**, Plot showing fold change of protein levels of Mce1F (blue) and Mce4F (yellow) in *lucB* bacterial strains compared to Wild-type *Msmeg* (WT) based on the quantification of the western blots. For gel source data, see Supplementary Figure 1d. Plotted data are the mean of three biological replicates ($n = 3$) and error bars representing standard deviations are shown.



Extended Data Fig. 10 | LucB orthologs are found in other bacteria with MCE systems. a, (left) 2D topology diagram of LucB. (right) Protein sequence alignment of LucB orthologs from fifteen different mycobacterial species generated using MUSCLE⁷⁵ and visualized using JalView⁷⁶. Sequence alignment is coloured by percent identity (not identical, white; identical, blue). Transmembrane helices of LucB (labeled TM1-TM4) are depicted above and residues that interact with Mce1C are highlighted below with asterisks.

b, Circular phylogenetic tree showing spread of orthologs of LucB in other

bacteria. LucB orthologs were compiled from eggNOG v5.0⁸¹ and AlphaFold Protein Structure Database⁴⁸. Protein sequences were aligned to generate a phylogenetic tree using MUSCLE⁷⁵ and the tree was visualized in iTOL⁸². Leaves indicate individual species and are coloured by bacterial family: Mycobacteriaceae (red), Intrasporangiaceae (yellow), Pseudonocardiaceae (green), Tsukamurellaceae (cyan), Nocardiaceae (blue), Gordoniaceae (purple). Tree scale indicated on bottom right.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection SerialEM v4

Data analysis AlphaFold2 Multimer v1, AlphaFold2 v2.3.0, CASTP v3.0, ChimeraX v1.4, COOT v0.8.9.2, cryoSPARC v3.3.1, EMRinger v1, FoldSeek v2-8bd520, GraphPad Prism v9.3.1, ImageJ v1.53a, Jalview v2.11.2.6, MOLE v2.5, MUSCLE v5, MolProbit v4.5, PHENIX v1.20.1, Proteome Discoverer v1.4, PyMOL Molecular Graphics System v2.5.1, SAINTexpress v3.6, Topaz v0.2.3, 3DFSC v3.0, UCSF Chimera v1.16

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All unique/stable reagents generated in this study are available without restriction from the corresponding authors, Gira Bhabha (gira.bhabha@gmail.com) and Damian C. Ekiert (damian.ekiert@ekiertlab.org). The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes: Map0

(EMD-29025), Map0a (EMD-29228), Map0b (EMD-29229), Map0c (EMD-29230), Map0d (EMD-29231), Map0e (EMD-29232), Map1 (EMD-29023), Map1a (EMD-29233), Map1b (EMD-29234), Map1c (EMD-29235), Map1d (EMD-29236), Map1e (EMD-29237), Map2 (EMD-29024), Map2a (EMD-29238), Map2b (EMD-29239), Map2c (EMD-29240), Map2d (EMD-29241), and Map2e (EMD-29242). The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes: PDB 8FEF (model for Map0), PDB 8FED (model for Map1), PDB 8FEE (model for Map2). Coordinates for atomic models used to make figures were obtained from the Protein Data Bank under accession codes: PDB 6XBD (MlaFEDB), PDB 6MIT (LptBFGC), PDB 2R19 (LptA), PDB 5IV9 (LptDE), PDB 5NIK (MacAB-TolC), PDB 6VOD (LetB), PDB 1THQ (PagP). Cryo-EM data was deposited in Electron Microscopy Public Image Archive: EMPIAR-11343. The mass spectrometry files are available at MassIVE (<https://massive.ucsd.edu>) with dataset identifier MSV000090807 and ProteomeXchange (proteomexchange.org) with identifier PXD038456. Bacterial strains and plasmids that have been deposited in Addgene are listed in Supplementary Table 1 with their identifiers.

Human research participants

Policy information about [studies involving human research participants](#) and [Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Cryo-EM sample size was chosen as a maximum possible with practical limitation for data collection and processing. No statistical methods were used to predetermine sample size for cryo-EM. The cryo-EM dataset size was determined to be sufficient by the ability to reach resolutions beyond 4 Å in 3D construction. For negative stain-EM, no statistical methods were used to predetermine sample size. Negative-stain dataset size was determined to be sufficient by the ability to see features in the 2D classes of picked particles.

Data exclusions

Particles from EM datasets were excluded during 2D and 3D classifications by removing classes with poor alignment and/or no high-resolution features (pre-established criterion), which is standard in the EM field. For mass spectrometry analysis, data was filtered using a 1% FDR cut off for peptides and proteins against a decoy database and only proteins with at least 2 unique peptides were reported. Data was not excluded for growth assays and western blots.

Replication

At least three biological replicates were performed for growth assays (Figs. 1d, 1e, 4f, 4g), protein purifications and grid preparations (Ext Data Figs. 2b, 7c) western blots (Fig. 1f, Ext Data Figs. 8a, 9a, 9g) and mass spectrometry experiments (Fig 1g, Ext Data Fig. 9e). See figure legends and methods for details. All attempts of replication were successful. High-resolution cryo-EM determination was performed once for given condition.

Randomization

Particles images were randomly assigned into odd/even groups for resolution assessment (gold-standard FSC).

Blinding

Blinding is not needed for cryo-EM analysis due to the automated handling of the data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
	<input checked="" type="checkbox"/> ChIP-seq
	<input checked="" type="checkbox"/> Flow cytometry
	<input checked="" type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

The primary antibodies used in this study are as follows: monoclonal Mouse anti-His antibody (commercially available, Qiagen cat. #34660), polyclonal Rabbit anti-GFP antibody (provided by Foley Lab, Memorial Sloan Kettering Cancer Center), polyclonal Chicken anti-MceG antibody (custom-raised, Capra Science Lot No. 1439.1406), polyclonal Rabbit anti-Mce1F antibody (custom-raised, Capra Science Lot No. 1477.1315), polyclonal Rabbit anti-Mce4F antibody (custom-raised, Capra Science Lot No. 1478.1315). The secondary antibodies used in this study are as follows: Goat anti-Rabbit IgG polyclonal antibody IRDye 800CW (commercially available, LI-COR Biosciences cat. #925-32211), Donkey anti-Chicken IgG polyclonal antibody IRDye 680RD (commercially available, LI-COR Biosciences cat. #926-68075), Goat anti-Mouse IgG polyclonal antibody IRDye 680RD (commercially available, LI-COR Biosciences cat. #926-68070).

Validation

Mouse anti-His antibody (Qiagen cat#34660, <https://www.qiagen.com/us/products/discovery-and-translational-research/protein-purification/tagged-protein-expression-purification-detection/anti-his-antibodies-bsa-free/?cmpid=QVenSPP1404quartzySC>) - Many different proteins have been detected, via their His tags, using QIAexpress Anti-His Antibodies (6xHis antibodies and 4xHis antibodies) (see table "Performance of Anti-His Antibodies"). For the sensitivity of the three Anti-His Antibodies in detecting this panel of proteins see the figure "Sensitivity of QIAexpress Anti-His Antibodies". Anti-His Antibodies enable identification of positive expression clones by colony blotting (see figure "Identification of positive expression clones by colony blotting") and detection of His-tagged thioredoxin (see figure "His-tagged thioredoxin detected by Penta-His Antibody in yeast cells"). The antibodies are also ideally suited for use in immunohistochemistry procedures (see figure "Immunohistochemistry with Penta-His Antibodies").

Rabbit anti-GFP antibody (provided by Foley Lab, Memorial Sloan Kettering Cancer Center) was validated by Western blot. Lysates of *Mycobacterium smegmatis* m2(155) and *Msmeg* MceG-GFP strain (bBEL591) were run on a SDS-PAGE gel, transferred to a nitrocellulose membrane, and blotted with Rabbit anti-GFP antibody. Signal for GFP-tagged MceG in bBEL591 was successfully detected with minimal background signal in m2(155) control.

The Capra Science custom-raised Chicken anti-MceG antibody, Rabbit anti-Mce1F antibody, and Rabbit anti-Mce4F antibody were validated by Western blot. Lysates of *Mycobacterium smegmatis* m2(155), *Msmeg* ΔmceG (bBEL594), *Msmeg* Δmce1 (bBEL603), and *Msmeg* Δmce4 (bBEL604) were run on a SDS-PAGE gel, transferred to a nitrocellulose membrane, and blotted with each antibody. For the Chicken anti-MceG antibody, signal for MceG was successfully detected in m2(155), *Msmeg* Δmce1, and *Msmeg* Δmce4 with minimal background signal in the *Msmeg* ΔmceG control. For the Rabbit anti-Mce1F antibody, signal for Mce1F was successfully detected in m2(155), *Msmeg* ΔmceG, and *Msmeg* Δmce4 with minimal background signal in the *Msmeg* Δmce1 control. For the Rabbit anti-Mce4F antibody, signal for Mce4F was successfully detected in m2(155), *Msmeg* ΔmceG, and *Msmeg* Δmce1 with minimal background signal in the *Msmeg* Δmce4 control.

Goat anti-Rabbit IgG polyclonal antibody IRDye 800CW (LI-COR Biosciences cat. #925-32211, <https://www.licor.com/bio/reagents/irdye-800cw-goat-anti-rabbit-igg-secondary-antibody>) - Isolation of specific antibodies was accomplished by affinity chromatography using pooled rabbit IgG covalently linked to agarose. Based on ELISA and flow cytometry, this antibody reacts with the heavy and light chains of rabbit IgG, and with the light chains of rabbit IgM and IgA. This antibody was tested by dot blot and and/or solid-phase adsorbed for minimal cross-reactivity with human, mouse, rat, sheep, and chicken serum proteins, but may cross-react with immunoglobulins from other species. The conjugate has been specifically tested and qualified for Western blot and In-Cell Western™ Assay applications.

Donkey anti-Chicken IgG polyclonal antibody IRDye 680RD (LI-COR Biosciences cat. #926-68075, <https://www.licor.com/bio/reagents/irdye-680rd-donkey-anti-chicken-secondary-antibody>) - The antibody was isolated from antisera by immunoaffinity chromatography using antigens coupled to agarose beads. Based on immunoelectrophoresis and/or ELISA, this antibody reacts with whole molecule chicken IgY, and with the light chains of other chicken immunoglobulins. No reactivity was detected against non-immunoglobulin serum proteins. This antibody was tested by ELISA and/or solid-phase adsorbed to ensure minimal cross-reactivity with bovine, goat, guinea pig, Syrian hamster, horse, human, mouse, rabbit, rat, and sheep serum proteins, but may cross-react with immunoglobulins from other species. The conjugate has been specifically tested and qualified for Western blot and In-Cell Western™ Assay applications.

Goat anti-Mouse IgG polyclonal antibody IRDye 680RD (LI-COR Biosciences cat. #926-68070, <https://www.licor.com/bio/reagents/irdye-680rd-goat-anti-mouse-igg-secondary-antibody>) - Isolation of specific antibodies was accomplished by affinity chromatography using pooled mouse IgG covalently linked to agarose. Based on ELISA and flow cytometry, this antibody reacts with the heavy and light chains of mouse IgG1, IgG2a, IgG2b, and IgG3, and with the light chains of mouse IgM and IgA. This antibody was tested by dot blot and and/or solid-phase adsorbed for minimal cross-reactivity with human, rabbit, goat, rat, and horse serum proteins, but may cross-react with immunoglobulins from other species. The conjugate has been specifically tested and qualified for Western blot and In-Cell Western™ Assay applications.