Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol

Zhang, Lu; Zhao, Yao; Gao, Yan; Wu, Lijie; Gao, Ruogu; Zhang, Qi; Wang, Yinan; Wu, Chengyao; Wu, Fangyu; Gurcha, Sudagar S; Veerapen, Natacha; Batt, Sarah M; Zhao, Wei; Qin, Ling; Yang, Xiuna; Wang, Manfu; Zhu, Yan; Zhang, Bing; Bi, Lijun; Zhang, Xian'en

DOI:

[10.1126/science.aba9102](https://doi.org/10.1126/science.aba9102)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Zhang, L, Zhao, Y, Gao, Y, Wu, L, Gao, R, Zhang, Q, Wang, Y, Wu, C, Wu, F, Gurcha, SS, Veerapen, N, Batt, SM, Zhao, W, Qin, L, Yang, X, Wang, M, Zhu, Y, Zhang, B, Bi, L, Zhang, X, Yang, H, Guddat, LW, Xu, W, Wang, Q, Li, J, Besra, GS & Rao, Z 2020, 'Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol', *Science*, vol. 368, no. 6496, pp. 1211-1219.
<https://doi.org/10.1126/science.aba9102>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is the author's version of the work. It is posted here by permission of the AAAS for personal use, not for redistribution. The definitive version was published in *Science* 368, DOI: 10.1126/science.aba9102

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol

Lu Zhang^{1,2†}, Yao Zhao^{1,3,4†}, Yan Gao⁵, Lijie Wu¹, Ruogu Gao^{4,6}, Qi Zhang¹, Yinan Wang^{1,4}, Chengyao Wu¹, Fangyu Wu², Sudagar S. Gurucha⁷, Natacha Veerapen⁷,
5 Sarah M. Batt⁷, Wei Zhao², Ling Qin¹, Xiuna Yang¹, Manfu Wang¹, Yan Zhu¹, Bing Zhang¹, Lijun Bi⁶, Xian'en Zhang⁶, Haitao Yang¹, Luke W. Guddat⁸, Wenqing Xu¹,
Quan Wang^{1,6*}, Jun Li^{1*}, Gurdyal S. Besra^{7*}, Zihe Rao^{1,2,5,6*}

¹Shanghai Institute for Advanced Immunochemical Studies, iHuman Institute, School of Life Science and Technology, ShanghaiTech University, Shanghai, 201210, China.

10 ²State Key Laboratory of Medicinal Chemical Biology, Frontiers Science Center for Cell Response, College of Life Sciences, College of Pharmacy, Nankai University, Tianjin 300353, China.

³CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, 200031,
15 China

⁴University of Chinese Academy of Sciences, Beijing, 100101, China.

⁵Laboratory of Structural Biology, Tsinghua University, Beijing, 100084, China.

⁶National Laboratory of Biomacromolecules and Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, CAS,
20 Beijing 100101, China.

⁷Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

⁸School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.

25 † These authors contributed equally.

L.Q.'s current address is Division of Structural Biology, University of Oxford, Oxford, OX3 7BN, UK.

* Correspondence to wangq@ibp.ac.cn (Q.W.), lijun1@shanghaitech.edu.cn (J.L.), G.Besra@bham.ac.uk (G.S.B.) and raozh@tsinghua.edu.cn (Z.R.)

30

Abstract

The arabinosyltransferases EmbA, EmbB, and EmbC are involved in *Mycobacterium tuberculosis* cell wall synthesis and are recognized as the targets for the anti-tuberculosis drug ethambutol. We have determined cryo-electron microscopy and
35 x-ray crystal structures of mycobacterial EmbA-EmbB and EmbC-EmbC complexes, in the presence of their glycosyl donor and acceptor substrates and with ethambutol. These structures show how the donor and acceptor substrates bind in the active site and how ethambutol inhibits by binding to the same site as both substrates in EmbB and EmbC. The majority of drug-resistant mutations are located nearby to the
40 ethambutol-binding site. Collectively, our work provides a structural basis for understanding the biochemical function and inhibition of arabinosyltransferases and development of new anti-tuberculosis agents.

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*Mtb*), is one of the oldest
45 known diseases to infect humans, but remains a major cause of morbidity and mortality resulting in more than 1.5 million deaths each year (*1*). Ethambutol is one of the five first line anti-TB drugs that are currently in clinical use to treat TB (*1*). It is particularly effective in combination therapy against multidrug-resistant forms of this infectious disease (*1*). The membrane-embedded Emb proteins, EmbA, EmbB and

50 EmbC, which are involved in cell wall biosynthesis are regarded as the targets of ethambutol, since mutations to these proteins result in TB strains that are clinically resistant to this drug, noting that the majority of these resistance sites occur within EmbB (2-5).

Mtb has a complex cell wall compared to Gram-negative and Gram-positive bacteria. Its core structure, mycolyl-arabinogalactan-peptidoglycan (mAGP), is composed of three highly unusual elements covalently linked together: (i) long-chain mycolic acids (MA), (ii) a highly branched arabinogalactan (AG), and (iii) a cross-linked network of peptidoglycan (PG). The other key component of the cell wall, lipoarabinomannan (LAM), is a species of phosphatidyl-*myo*-inositol derived glycolipids containing mannan and arabinan domains and is an important virulence factor playing a key role in host-pathogen interactions, and in modulating the host immune response during infection (6-10). The composition of AG and LAM suggests at least seven different arabinosyltransferases (AraTs) are involved in the assembly of the arabinan domain (6). As belonging to the same family of AraTs, the Emb proteins have a high similarity in amino acid sequences (sharing ~40% identity). Both EmbA and EmbB have been shown to play key roles in the formation of the $\alpha(1\rightarrow3)$ linkage on the terminal hexaarabinofuranosyl motif of AG (11) (Fig. 1A). Furthermore, these proteins are suggested to function in a coordinating way by forming a heterodimer within cells (12), while the terminal $\beta(1\rightarrow2)$ linkage at the AG non-reducing end is catalyzed by AftB (13) (Fig. 1A). The product of the EmbA/EmbB and AftB catalyzed reactions plays a key role by serving as the mycolic acid attachment site for AG (14). EmbC functions by forming $\alpha(1\rightarrow5)$ glycosidic linkages leading to the linear elongation of the arabinan chain of LAM (Fig. 1B) (15). Decaprenyl-phosphate-arabinose (DPA) is the only known arabinose donor for these AraTs in mycobacterial species (16). In *Mtb* the *embABC* genes are clustered in the genome and have all been shown to be essential for *in vitro* growth of H37Rv by analysis of saturated Himar1 transposon libraries (17), and *embA* and *embC* have also

been shown to be essential in *Mtb* under normal growth conditions through the generation of deletion mutants (18, 19).

80 The Emb proteins belong to the glycosyltransferase C (GT-C) superfamily (15, 20). Despite their importance in cell wall synthesis, the three-dimensional structure for any component of an Emb protein is yet to be determined, with the exception being the C-terminal soluble domain of EmbC (21). Thus, it remains poorly understood how these enzymes function, how ethambutol exerts its mode of action, and how *emb* mutations lead to ethambutol resistance. To provide a foundation for understanding these phenomena, we have determined the three-dimensional structures of *Mtb* and *Mycobacterium smegmatis* (*Msm*) EmbA-EmbB, and *Msm* EmbC-EmbC (EmbC₂), with each in complex with ethambutol. Ethambutol is observed bound in the conserved active sites of both EmbB and EmbC. In addition, we have also solved the three-dimensional structures of *Msm* EmbA-EmbB and EmbC₂ in complex with the donor and acceptor substrates, DPA and di-arabinose, respectively. Unexpectedly, these structures also show that each Emb protomer has an acyl-carrier-protein, AcpM, bound to their cytoplasmic surface. These studies provide new molecular insights as to how arabinose is transferred by these enzymes and how ethambutol binds at location that overlaps with the region where the donor and acceptor substrates bind.

Results

Characterization and structure determination of EmbA-EmbB and EmbC₂ complexes

Mtb and *Msm* EmbA-EmbB complexes were expressed in *Msm* cultures and then purified to homogeneity (figs. S1A and S2E). Unexpectedly, the endogenous *Msm* acyl-carrier-protein AcpM (MSMEG_4326, 99 aa, ~11 kDa), involved in the biosynthesis of mycolic acids (6, 22), was identified to be co-purified with each of these samples by SDS-PAGE silver staining and mass spectrometry (figs. S1C, S1D,

S2B and S2J). SDS-PAGE showed the EmbA and EmbB have a 1:1 stoichiometry in
105 EmbA-EmbB complexes (**fig. S1B**), whose composition was further confirmed by
mass spectrometry (**figs. S1E-H**). *Msm* EmbC is shown as a single band of ~120 kDa
by SDS-PAGE (**fig. S2A**), and a single band between 242 kDa and 480 kDa by
BN-PAGE (**fig. S2F**). Taking into account the fact that detergents wrap around this
complex and the existence of AcpM, the molecular weight determined by BN-PAGE
110 is consistent with dimer formation of *Msm* EmbC observed from our solved structures
(see below).

All EmbA-EmbB and EmbC₂ samples were shown to be active in cell-free activity
assays. The established arabinosyltransferase activity assay of purified *Mtb* and *Msm*
EmbA-EmbB complexes (**Fig. 1C and fig. S1J**), together with 2D heteronuclear single
115 quantum correlation (HSQC) NMR analysis of EmbA and EmbB deletion mutants (23)
(**fig. S1K**), confirmed they function as arabinosyltransferases, which catalyze the
formation of an $\alpha(1\rightarrow3)$ -arabinofuranosyl linkage, in accordance with previous
studies (11, 12, 16). The *Msm* EmbC₂ sample was confirmed to catalyze the formation
of an $\alpha(1\rightarrow5)$ -arabinofuranosyl linkage (16) (**Fig. 1D and fig. S2G**). More
120 importantly, the cell-free arabinosyltransferase activity of both EmbA-EmbB and
EmbC₂ is inhibited by ethambutol, confirming that ethambutol targets these proteins
(**Figs. 1C, 1D and figs. S1J, S2G**).

Single particle cryo-EM was used to determine four structures of Emb proteins at
2.81~3.10 Å resolution: *Mtb* EmbA-EmbB-AcpM₂ bound with ethambutol, *Msm*
125 EmbA-EmbB-AcpM₂ bound with ethambutol, *Msm* EmbA-EmbB-AcpM₂ bound with
di-arabinose, and *Msm* EmbC₂-AcpM₂ bound with ethambutol (**Fig. 1E, figs. S3-S7
and Table S1**). For the Emb protomers, most regions of the polypeptide for EmbA,
EmbB and EmbC could be traced except for the cytoplasmic loop CL1 (residue range
248-268) in *Mtb* EmbB and a periplasmic segment (residue range 780-810) of *Msm*
130 EmbC. The density for AcpM bound to the *Msm* EmbB or to the *Mtb* EmbA surface

has a resolution in the 3~5 Å range, whereas the map for AcpM bound to the *Msm* EmbA or *Mtb* EmbB protomer was less clear, and was not built (Fig. 1E and figs. S3F, S4F, S5F).

For the crystallographic study of EmbC, homologs from *Mtb*, *Msm*, *M. marinum* and
135 *M. xenopi* were all expressed and purified (fig. S2A). However, only the crystal structure of *Msm* EmbC in complex with di-arabinose could be determined at 3.3 Å resolution (Fig. 1E, fig. S8 and Table S2). Based on the electron density, the structure of two molecules of AcpM (residues 3~86) could be built (Fig. 1E and fig. S8A).

Overall structures of the EmbA-EmbB-AcpM₂ and EmbC₂-AcpM₂ complexes

140 In our cryo-EM and crystal structures, EmbA-EmbB form a heterodimeric complex, whilst EmbC is a symmetric homodimer (Fig. 1E). In the EmbC dimer, two EmbC protomers are nearly identical which could well superimpose on each other (fig. S9A). The mode of dimerization is similar in the two complexes and is achieved by forming hydrophobic clusters between transmembrane (TM) domains close to the cytoplasmic
145 side and periplasmic side (figs. S10A and S10B). The individual EmbA, EmbB and EmbC proteins all have a similar fold (Figs. 1F, 1G and fig. S10C) and contain common features including a 15-helix TM domain, N-/C-terminal periplasmic domains (PDs) identified as PN and PC (Figs. 1F, 1G and figs. S9D, S9E, S9G) and periplasmic and cytoplasmic loops connecting the TM helices, some of which are of
150 structural and physiological important. Periplasmic loop 2 (PL2) contains two crossed helices (EH1 and EH2) and harbors the highly conserved catalytically relevant DDx motif (24) (Figs. 1F, 1G and fig. S13A); PL5 forms two tandem helices (EH3 and EH4) and contributes to the gating of DPA and dimerization of the complex (Figs. 1F, 1G and fig. S10B). The positively charged cytoplasmic loop 1 (CL1) forms extensive
155 interactions with AcpM (Fig. 1G and fig. S11).

In each structure the PN domain (also PL1 linking TM1 and TM2) adopts a jelly-roll-fold (fig. S9D), which is typical for polysaccharide binding units, such as plant lectins or where carbohydrates act as enzyme substrates (25). The PC domain can be divided into two subdomains, with subdomain-I displaying a mixed α/β structure and subdomain-II exhibiting a jelly-roll-fold (fig. S9E). This fold is similar to the previously reported C-terminal structure of *Mtb* EmbC (PDB code 3PTY), both containing a bound Ca^{2+} responsible for structural stability (21) (figs. S9C and S9F). The last 30 amino acids in the PC domain embrace the PN domain, which contributes to the stabilization of the entire PDs (Figs. 1F and 1G). The active site is located in a pocket at the junction between the TM domain and the periplasmic domains (Fig. 1F), composed by PL2-6, helix α_6 in PC domain and residue Trp965 in PC domain (represented by *Msm* EmbC) (fig. S9H). Substrates (di-arabinose analogue and DPA) or ethambutol are observed in this site (Fig. 1E), providing crucial information for understanding catalysis, drug inhibition and resistance.

By providing structural information that is distinct from other known glycosyltransferases, the structures of the Emb proteins, substantially broadens our understanding of the GT-C superfamily. A structural comparison with *Archaeoglobus fulgidus* AglB (26), *Campylobacter lari* PglB (27), *Cupriavidus metallidurans* ArnT (28), yeast STT3 in OST complex (29), yeast PMT1 in PMT1-PMT2 complex (30) and yeast ALG6 (31) show that the Emb proteins possess periplasmic architectures that are distinct in size and shape from other GT-C proteins (fig. S12). Common features to other GT-C members include: (1) TM regions sharing a common core of 11 TM helices with a similar fold (fig. S12); (2) and crossed helices resembling EH1 and EH2 in Emb proteins, bearing the conserved catalytically relevant D[N]D[E]x motif (fig. S12).

The *Msm* AcpM binds to the cytoplasmic face of both EmbA-EmbB and EmbC₂ complexes, thus forming the EmbA-EmbB-AcpM₂ complexes and EmbC₂-AcpM₂

complexes, respectively (Fig. 1E). It has a four-helix topology arranged in a right-handed bundle (figs. S11A and S11H), which is similar to that of *Mtb* AcpM (PDB code: 1KLP) (32) (fig. S9B). The AcpMs bind to each Emb protomer through extensive electrostatic interactions (figs. S11B, S11C and S11I). This type of assembly has not been observed in any other glycosyltransferase. The binding mode of AcpM to EmbA-EmbB and to EmbC₂ is similar, whereby Helix α 2 of AcpM and the connecting loops at its N-/C-terminus are intimately engaged with the CLs of the Emb proteins (figs. S11A and S11H). This is consistent with the known role of α 2 in AcpM as a contact site with its target proteins (*e.g.* AcpS (33)). 4'-Phosphopantetheine (Ppant), is also observed covalently attached to the conserved Ser41_{AcpM} located on α 2 of AcpM and inserts into the gap between TM6-7 and CL1 of the Emb protein in the cryo-EM structures of both EmbA-EmbB and EmbC₂ complexes (figs. S11E, S11H and S11I). In contrast, in the crystal structure of EmbC, the side chain of Ser41 of AcpM interacts with the main chain of Arg247 of EmbC, on the CL1 with no Ppant observed (fig. S11D). Mutagenesis and functional studies showed that when the interactions between EmbC and AcpM are disrupted, the produced LAM species becomes smaller (except for R352A) (fig. S11G), although the formation of the EmbC₂-AcpM₂ complex with these mutants and cell-free arabinosyltransferase activity are largely preserved (figs. S2C, S2D and S11F). A likely explanation is that the mode of AcpM binding changes in the complex and this, in turn, affects the ability of AcpM to modulate LAM synthesis *in vivo*. Sequence alignment demonstrates that this region is not conserved across species (fig. S13A), indicating the variable binding abilities within the different Emb proteins. Since, AcpM plays an important role in the biosynthesis of fatty acid and cell wall (34), the association with proteins involved in AG assembly or LAM synthesis could imply some clues for the physiological function of this pattern. Further investigation into the exact function of AcpM in the Emb complexes is needed.

210 Several functionally important lipids are also observed in these cryo-EM structures. The substrate, DPA, is observed endogenously bound to EmbA in the conserved donor binding cavity in the *Msm* EmbA-EmbB complexes (Figs. 1E, 2B and fig. S7E); the reaction byproduct DP (the leaving group of DPA) is observed in a similar position in EmbB in the *Mtb* EmbA-EmbB complex (Fig. 1E and fig. S7F). The identity of DPA and DP was subsequently confirmed by mass spectrometry (fig. S11). In all the EmbA-EmbB complexes, native cardiolipins were observed binding in a similar manner (*i.e.* in the TM domain at the dimer interface) (Fig. 1E and figs. S7G, S10A). Thus, it plays an important role in helping stabilize the heterodimeric complex.

220 Substrate binding in the active sites of EmbA-EmbB and EmbC₂ complexes

The Emb proteins have two substrates, an arabinose donor DPA, and an acceptor arabinan. To understand how substrates bind, we determined the crystal structure of di-arabinose-bound *Msm* EmbC₂ and the cryo-EM structure of the di-arabinose-bound *Msm* EmbA-EmbB complex (Fig. 1E). These structures show that there are two substrate entrances leading to the active site, henceforth denoted as the donor entrance and the acceptor entrance (see below) (Fig. 2A). In the cryo-EM structure of *Msm* EmbA-EmbB, an endogenous donor substrate DPA is observed in the EmbA subunit. Its head part, the arabinose moiety and the phosphate group, binds in the active site, while the long decaprenyl tail extends out through the donor entrance and fits in a hydrophobic cavity created by TM7-9 (Fig. 2B). The arabinose moiety forms polar interactions with surrounding residues Asp261, Asn265, Glu289 on PL2, while the phosphate group is stabilized by polar bonds to Arg365, Gln409, Thr553, Trp555 and Gln558 (Fig. 2C). Asp261, the catalytic residue on the DDx motif conserved in GT-C family (fig. S13A and S13B), is 3.3 Å from the C1 atom of DPA (Fig. 2C). Note that a periplasmic Loop₇₆₆₋₈₀₆ (loop harbors residues 766-806) hangs down from the PC domain of *Msm* EmbA and blocks the acceptor entrance (fig. S14C). The Asn780 on

this loop inserts into the active site and interacts with the Asp261 and is also close to the arabinose group of DPA (Fig. 2F). Whilst in EmbB and EmbC, the corresponding loop is either shorter and folded on the PC domain of EmbB, or flexible in the solvent for EmbC (fig. S14C).

Consistent with the above observations, in the crystal structure of di-arabinose-bound *Msm* EmbC₂, an endogenous phosphate ion appears to be trapped in the active site by di-arabinose (part of Ara₂OC8) and maltose (part of detergent DDM) (Fig. 2D and figs. S8B, S8C). It is bound to a positively charged region that includes Arg383, His574 and His575 and is near Thr570, Trp572 and the catalytic Asp279 (Fig. 2E). We propose that this phosphate represents the phosphate group of DPA, as it is superimposable with the phosphate group of DPA we observe in *Msm* EmbA (Fig. 2F and fig. S14B). To verify this proposal, we measured the K_d values of DPA for the wild-type and mutant *Msm* EmbC proteins. The K_d for wild-type EmbC is 3.0 μ M, while for the EmbC mutants, the affinity is greatly reduced to 122 μ M and 137 μ M for H574A and H575A mutants, or undetectable for R383A and T570S mutants (Table S3). Furthermore, enzymatic activity is completely lost for these phosphate binding site mutants (Fig. 2G). Consistent with this, the *in vivo* LAM synthesis was almost completely inhibited when the *embC* knock-out strain was complemented with the corresponding mutant alleles (for the W572A mutant, the LAM species become smaller) (Fig. 2H). This phosphate binding site for DPA is thus crucial for EmbC function.

The diarabinoside group, which is identified as part of the substrate analogue Ara₂OC8 that was added during crystallization, binds between Asp279 and the phosphate (Fig. 2D and figs. S8B, S8C). For clarity, we denote the positions of the two-arabinofuranose rings as D site (arabinose from donor) and A₀ site (arabinose from the terminal residue of aceptor), with the D-site being deeper in the pocket while the A₀-site is close to the acceptor entrance (Fig. 2D). The arabinofuranose in

the D-site is sandwiched by Tyr282 and the phosphate group. Its hydroxyl groups
265 interact with Asn298, Glu307 and Tyr314 through hydrogen bonds (Fig. 2I). The
A₀-site arabinofuranose is clamped by Trp572 and Trp965 (Fig. 2I). Additionally,
Trp302 and Val1004 form van der Waals interactions with the D-site and A₀-site
arabinofuranoses, respectively (Fig. 2I). The side-chain of the catalytic Asp279 points
towards the $\alpha(1\rightarrow5)$ glycosidic bond between the two arabinofuranose groups (Figs.
270 2D and 2I). The K_d value of Ara₂OC8 with *Msm* EmbC is 36.7 μ M (fig. S2H). When
these residues are mutated to Ala or Asn, the enzymatic activity of *Msm* EmbC is
reduced significantly or cannot be detected (Fig. 2J), confirming that they are
essential for the function of EmbC. Consistent with this, these mutations almost
completely abolish production of LAM species in the *Msm* EmbC knock-out strain
275 complemented with the same alleles (for the Y314A mutant, the LAM species
become smaller) (Fig. 2K). This is in accord with a previous report that showed the
D279A mutant could not produce LAM (24). Considering that the two arabinoside
groups are located on different sites of the catalytic site Asp279, the diarabinoside
appears to represent the reaction product after Asp279 catalyzes the formation of the
280 $\alpha(1\rightarrow5)$ linkage between the two arabinofuranose rings. The arabinofuranose in the
D-site is most likely mimicking the newly added residue in the product, which is
superimposable with the arabinose group of DPA in *Msm* EmbA (Fig. 2F) thus
originated from donor DPA, while arabinofuranose in the A₀-site resembles an
arabinose from the terminal residue in the acceptor.

285 Based on the above structural and functional data, the arabinose transfer mechanism
that completes one cycle of elongation can be proposed for EmbC (Fig. 2L): (i) Firstly,
the terminal arabinofuranose of a LAM precursor from a previous round of reaction
stays in the catalytic pocket and binds at the A₀-site. The hydrophilic head group of
DPA inserts into the pocket from the donor entrance, with its phosphate group and the
290 donor arabinofuranose binds at the phosphate binding site and the D-site, respectively.
(ii) Next, the catalytic Asp activated the transfer of donor arabinofuranose to the

arabinofuranose residue in the A₀-site by forming an α(1→5) linkage. As a result, the donor arabinofuranose becomes the new terminal residue of the elongated product, which binds to the D-site. The di-arabinose binding mode in the crystal structure of EmbC represents the terminal di-arabinose of the elongated product. (iii) The byproduct of the reaction, DP, leaves the catalytic pocket. (iv) In the final step, the elongated product is re-loaded, or re-positioned by an unknown mechanism, to the active site that the terminal arabinofuranose occupy the A₀-site again, ready for the next elongation reaction.

In the cryo-EM map of *Msm* EmbA-EmbB complex, the density of the di-arabinose (part of the incubated tetra-arabinose) was found within the active site of EmbB subunit (fig. S14D), a location similar to the di-arabinose observed in the crystal structure of *Msm* EmbC₂ (fig. S14C). The di-arabinose bound in *Msm* EmbB is due to interactions with Trp578, His580, Trp972 and Trp1012 with the A₀-site group and Tyr288, Asn304, Glu313 and Arg495 with the D-site group, the catalytic residue Asp285 forms polar interaction with the oxygen atom on the glycosidic bond (fig. S14D). These residues are conserved amongst the EmbB and EmbC family of proteins (fig. S13A). Similar to the di-arabinose in the crystal structure of *Msm* EmbC, the D-site arabinose group in *Msm* EmbB is proposed to be provided by the donor DPA (i.e. the arabinose has been cleaved from the donor DPA) upon superimposing with DPA of *Msm* EmbA (fig. S14A). It could be concluded that the A₀-site arabinose represents the sugar moiety from the acceptor substrate so that the di-saccharide mimics the product in the reaction center. Given the donor substrate DPA is observed in its pre-catalytic state in EmbA, it is plausible that the binding of substrates is a sequentially coupled process with DPA binding followed by acceptor binding. Furthermore, the likely one-step reaction catalyzed by the EmbA-EmbB complex suggests that the newly formed glycosidic bond occurs followed by release of all of the products. The next cycle needs to be initiated by the binding of a new molecule of DPA. Future investigations will be needed to elucidate how the entire catalytic cycle

320 works.

Structural basis for ethambutol inhibition on EmbB and EmbC

Our functional data confirmed that ethambutol inhibits the enzymatic transferase activity of both the EmbA-EmbB and EmbC₂ complexes *in vitro* (Figs. 1C, 1D and figs. S1J, S2G). These results are in agreement with the observation that branching of the terminal hexa-arabinan motif in AG, as well as synthesis of LAM, can be inhibited by ethambutol (23, 35). The K_d for ethambutol binding was measured to be 0.42 μ M for *Mtb* EmbA-EmbB (Fig. 3B and Table S3) and 0.31 μ M for *Msm* EmbA-EmbB (Fig. 4C and Table S3), thus strong binding is observed in both cases. For *Msm* EmbC, the K_d value was measured to be 11.1 μ M (fig. S2I and Table S3), thus a relatively weaker binding affinity for this class of Emb protein.

In order to elucidate how ethambutol binds, cryo-EM structures of the *Mtb* and *Msm* EmbA-EmbB complexes, as well as *Msm* EmbC₂, all in complex with ethambutol, were determined at 2.97 Å, 2.90 Å and 2.81 Å resolution, respectively (Fig. 1E). Analysis of these three maps showed that density consistent with that of ethambutol is located within the active site of the EmbB and EmbC subunits (Figs. 3A, 3C and fig. S15D). In contrast, no density for ethambutol was observed in the EmbA subunit. However, there is density for the endogenous donor substrate, DPA, whose hydrophilic moiety bound to *Msm* EmbA is in a similar location as ethambutol when bound to EmbB (Figs. 3F and 3G). We thus suggest that ethambutol preferentially binds to EmbB and EmbC rather than to EmbA, a conclusion in line with clinical drug resistance studies (3-5, 36-40). Considering the similarity in ethambutol binding in all the determined complexes (fig. S15F), we focus our analysis of its binding mode based on the *Mtb* EmbA-EmbB complex. In this case, the two charged imino groups of ethambutol, NH1 and NH2, play key roles in binding. They form three electrostatic interactions with EmbB (the distances of NH1-Asp299, NH2-Asp299 and NH2-DP

are 3.2 Å, 3.3 Å and 3.4 Å, respectively) by positioning themselves between Asp299 and the phosphate group of the DP (leaving group from donor DPA) (Figs. 3C and 3D). In addition, π -cation interactions are formed by NH1 with Trp988 and NH2 with Tyr302 (Figs. 3C and 3D). The two hydroxyl groups of ethambutol form hydrogen bonds with His594 and Glu327, respectively, while the two hydroxybutanyl groups form van der Waals interactions with Ile303, Met306 on PL2, with Trp592 on PL6 and with Trp1028 on the PC domain (Figs. 3C and 3D).

When superimposing the EmbB (or EmbC) subunits in complex with ethambutol and disaccharide, we found that they overlapped within a high degree of similarity (Fig. 3E and fig. S15E). Since the di-saccharide represents the arabinose group from both donor and acceptor, we infer that ethambutol inhibits the arabinose transfer reaction by competing with the binding of both substrates in the active site, in accordance with the hypothesis that ethambutol interferes with transfer of arabinose as evidenced by the rapid accumulation of DPA in ethambutol treated *Msm* cells (41).

Superimposing the structure of *Msm* EmbA onto *Msm* EmbB reveals the conformation of the Loop₇₆₆₋₈₀₆ of *Msm* EmbA not only locks DPA in the active site, but also hinders ethambutol from binding in the active site (Figs. 3F and 3G). Considering structural similarity between *Msm* and *Mtb* and sequence homology (fig. S13A) amongst mycobacterial EmbA proteins, this long loop is identified as a conserved and unique feature of EmbA that not only responsible for substrate trapping, but also prevents it from being targeted by ethambutol.

Structural interpretation for ethambutol resistance

Numerous mutations in the *embCAB* locus have been reported in ethambutol resistant *Mtb* strains by showing increased minimal inhibitory concentration (MIC) to that of wild-type strain, which are likely due to stress under ethambutol exposure. The majority are changes to the *embB* gene. Mutations to three sites, Met306, Gly406, and

Gln497 in *Mtb* EmbB (3-5, 36) (Fig. 4A) are regarded as resistance hotspots, with Met306, which is conserved in all EmbB proteins (fig. S13A), being the most frequently observed of all mutants (2, 3, 42). In the ethambutol-bound *Mtb* EmbB structure, the side chain of Met306 is directly involved in ethambutol binding through van der Waals contacts with the hydroxybutanyl group on NH₂ of ethambutol (Fig. 4B). Any changes in this site could affect drug binding. Clinical resistant isolates have favored mutations, such as M306V and M306I to *Mtb* EmbB (3, 4, 35). Mutation of the equivalent methionine in *Msm* EmbB (Met292) to Val results in a ~13-fold decrease in binding affinity to ethambutol (Fig. 4C), whereas arabinosyltransferase activity of the mutants (equivalent *Msm* M292V and M292I) remains unaffected and resistant to ethambutol (Fig. 4D). Met306 is also involved in nonpolar interaction with the surrounding residues Tyr302 and Glu327 (Fig. 4B), the latter residue also interacts with ethambutol (Fig. 3C). Mutations on Met306 would thus likely also change the interaction network to affect ethambutol binding. Two other mutant hotspot residues, Gly406 and Gln497, are both more than 10 Å from ethambutol, thus have no direct interaction with the drug. Mutation of Gln497 will no longer permit an interaction with Glu328, as a result may lead to a disruption of the Glu327-ethambutol interaction (Fig. 4F). It is unclear how mutation to Gly406 leads to drug resistance, given its location at the junction of PL3 and TM6 (Fig. 4A), it is likely a bulkier side chain mutation could result in steric hindrance that transmits to conformational change at the ethambutol binding site. Mutations on Ile289 of *Msm* EmbB (*i.e.* I289M and I289F), which is conserved in ethambutol-sensitive mycobacterial organisms (*i.e.* *Msm* and *M. bovis* (43)) (fig. S13A), are also known to be responsible for ethambutol resistance in *Msm* (44). Analysis of the structure show that the I289M and I289F mutations would sterically hinder ethambutol binding whilst not affecting enzymatic activity (Fig. 4D and 4E), and therefore result in resistance.

To further investigate the resistance to EmbB, a collection of 1,814 resistant sites from 61 studies documented in MycoResistance database (Database URL:

400 <http://www.hmulinglab.org/MycoResistance/>) (42) were manually picked with the top 16 mutation sites mapped onto the *Mtb* EmbB (fig. S15A). Among these resistance mutations (including hotspots mentioned above), approximately 73.7% mutations occur to residues on PL2, which has been shown to be pivotal to both drug binding and catalysis, thus further emphasizing the significance of PL2 as a hotspot for the
405 binding of new drug candidates. PL3 and PL5 harbor 8.7% and 8.1% mutations, respectively, with the remainder found on TM4, CL2 and TM5 (fig. S15B). Amongst the mutation sites on PL2, four D299E mutated strains, were found to be related to ethambutol resistance. In some GTs, it has been reported that the corresponding Glu can serve as the catalytic residue (20, 45-47), thus this change can occur and at the
410 same time the longer side-chain can continue to act as a catalytic nucleophile.

The clinically relevant ethambutol resistance mutation sites in *Mtb* EmbC (3, 4, 36-40), are also highly conserved (fig. S15C) in the *Msm* EmbC₂ structure. The results showed that major clustering occurs within PL2. These sites are close to the ethambutol-binding site (fig. S15C). Thus, the resistance data maps well to the
415 structure of *Msm* EmbC and consists with resistance analysis on *Mtb* EmbB.

Conclusion

We have demonstrated that a heterodimeric EmbA-EmbB and a homodimeric EmbC form functional arabinosyltransferase complexes both associated with two copies of AcpM and that ethambutol targets to the active site of EmbB and EmbC. The DPA
420 and di-arabinose bound structure of the EmbA-EmbB and EmbC₂ complexes allow us to understand the structural features required for catalysis. Contrastingly, the EmbA-EmbB complex catalyzes a branching reaction, whereas EmbC₂ catalyzes a multi-circle elongation reaction (Figs. 1A and 1B). Based on these structures and supporting data, we are able to propose that ethambutol functions by competing with
425 the substrates for binding to the EmbB and EmbC subunits. Its binding mode almost

precisely overlaps with the di-saccharide product analogue, and as a result *Mtb* may have limited opportunity to develop resistance without compromising its own ability to successfully construct its elaborate cell wall. This property is attractive since it is well known that co-administration of multiple drugs that have different targets is an effective way to slow development of resistance. Understanding how ethambutol interacts with other possible targets will allow the complete definition of the mode-of-action of ethambutol. Since EmbB/EmbC and its orthologs are well conserved across mycobacteria, the development of drugs that are broadly effective against these and other human pathogenic infectious diseases, including leprosy (caused by *Mycobacterium leprae*), is feasible.

References and Notes:

1. WHO, Global Tuberculosis Report 2019. (2019).
2. H. Safi, B. Sayers, M. H. Hazbon, D. Alland, Transfer of embB codon 306 mutations into clinical *Mycobacterium tuberculosis* strains alters susceptibility to ethambutol, isoniazid, and rifampin. *Antimicrob Agents Chemother* **52**, 2027-2034 (2008).
3. Q. Sun *et al.*, Mutations within embCAB Are Associated with Variable Level of Ethambutol Resistance in *Mycobacterium tuberculosis* Isolates from China. *Antimicrob Agents Chemother* **62**, (2018).
4. L. L. Zhao *et al.*, Analysis of embCAB mutations associated with ethambutol resistance in multidrug-resistant *mycobacterium tuberculosis* isolates from China. *Antimicrob Agents Chemother* **59**, 2045-2050 (2015).
5. F. Brossier *et al.*, Molecular Analysis of the embCAB Locus and embR Gene Involved in Ethambutol Resistance in Clinical Isolates of *Mycobacterium tuberculosis* in France. *Antimicrob Agents Chemother* **59**, 4800-4808 (2015).
6. M. Jankute, J. A. Cox, J. Harrison, G. S. Besra, Assembly of the Mycobacterial Cell Wall. *Annu Rev Microbiol* **69**, 405-423 (2015).
7. I. Vergne, M. Gilleron, J. Nigou, Manipulation of the endocytic pathway and phagocyte functions by *Mycobacterium tuberculosis* lipoarabinomannan. *Front Cell Infect Microbiol* **4**, 187 (2014).
8. A. K. Mishra, N. N. Driessen, B. J. Appelmelk, G. S. Besra, Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host-pathogen interaction. *Fems Microbiol Rev* **35**, 1126-1157 (2011).
9. V. Briken, S. A. Porcelli, G. S. Besra, L. Kremer, Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Mol Microbiol* **53**, 391-403 (2004).

10. M. Gilleron, J. Nigou, D. Nicolle, V. Quesniaux, G. Puzo, The acylation state of mycobacterial lipomannans modulates innate immunity response through toll-like receptor 2. *Chem Biol* **13**, 39-47 (2006).
- 465 11. V. E. Escuyer *et al.*, The Role of the embA and embB Gene Products in the Biosynthesis of the Terminal Hexaarabinofuranosyl Motif of Mycobacterium smegmatis Arabinogalactan. *J Biol Chem* **276**, 48854-48862 (2001).
12. S. Khasnobis *et al.*, Characterization of a specific arabinosyltransferase activity involved in mycobacterial arabinan biosynthesis. *Chem Biol* **13**, 787-795 (2006).
- 470 13. M. Seidel *et al.*, Identification of a novel arabinofuranosyltransferase AftB involved in a terminal step of cell wall arabinan biosynthesis in Corynebacteriaceae, such as Corynebacterium glutamicum and Mycobacterium tuberculosis. *J Biol Chem* **282**, 14729-14740 (2007).
14. M. McNeil, M. Daffe, P. J. Brennan, Location of the mycolyl ester substituents in the cell walls of mycobacteria. *J Biol Chem* **266**, 13217-13223 (1991).
- 475 15. S. Berg, D. Kaur, M. Jackson, P. J. Brennan, The glycosyltransferases of Mycobacterium tuberculosis - roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. *Glycobiology* **17**, 35R-56R (2007).
16. R. E. Lee, P. J. Brennan, G. S. Besra, Mycobacterial arabinan biosynthesis the use of synthetic arabinoside acceptors in the development of an arabinosyl transfer assay. *Glycobiology* **7**, 1121-1128 (1997).
- 480 17. M. A. DeJesus *et al.*, Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *mBio* **8**, (2017).
18. A. G. Amin *et al.*, EmbA is an essential arabinosyltransferase in Mycobacterium tuberculosis. *Microbiology* **154**, 240-248 (2008).
- 485 19. R. Goude, A. G. Amin, D. Chatterjee, T. Parish, The critical role of embC in Mycobacterium tuberculosis. *J Bacteriol* **190**, 4335-4341 (2008).
20. L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555 (2008).
- 490 21. L. J. Alderwick *et al.*, The C-Terminal Domain of the Arabinosyltransferase Mycobacterium tuberculosis EmbC Is a Lectin-Like Carbohydrate Binding Module. *PLoS Pathog* **7**, e1001299 (2011).
22. M. L. Schaeffer, G. Agnihotri, H. Kallender, P. J. Brennan, J. T. Lonsdale, Expression, purification, and characterization of the Mycobacterium tuberculosis acyl carrier protein, AcpM. *Biochim Biophys Acta, Mol Cell Biol Lipids* **1532**, 67-78 (2001).
- 495 23. R. E. Lee, W. Li, D. Chatterjee, R. E. Lee, Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed. *Glycobiology* **15**, 139-151 (2005).
- 500 24. S. Berg *et al.*, Roles of conserved proline and glycosyltransferase motifs of EmbC in biosynthesis of lipoarabinomannan. *J Biol Chem* **280**, 5651-5663 (2005).
25. A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781 (2004).
26. S. Matsumoto *et al.*, Crystal structures of an archaeal oligosaccharyltransferase provide

- insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci U S A* **110**, 17868-17873 (2013).
27. C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355 (2011).
28. V. I. Petrou *et al.*, Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608-612 (2016).
29. R. Wild *et al.*, Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* **359**, 545-550 (2018).
30. L. Bai, A. Kovach, Q. You, A. Kenny, H. Li, Structure of the eukaryotic protein O-mannosyltransferase Pmt1-Pmt2 complex. *Nat Struct Mol Biol* **26**, 704-711 (2019).
31. J. S. Bloch *et al.*, Structure and mechanism of the ER-based glucosyltransferase ALG6. *Nature* **579**, 443-447 (2020).
32. H. C. Wong, G. Liu, Y.-M. Zhang, C. O. Rock, J. Zheng, The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*. *J Biol Chem* **277**, 15874-15880 (2002).
33. K. D. Parris *et al.*, Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883-895 (2000).
34. L. Kremer *et al.*, Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of *Mycobacterium tuberculosis* fatty acid synthase II. *J Biol Chem* **276**, 27967-27974 (2001).
35. B. Cuevas-Cordoba *et al.*, Mutation at embB codon 306, a potential marker for the identification of multidrug resistance associated with ethambutol in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **59**, 5455-5462 (2015).
36. C. Plinke *et al.*, embCAB sequence variation among ethambutol-resistant *Mycobacterium tuberculosis* isolates without embB306 mutation. *J Antimicrob Chemother* **65**, 1359-1367 (2010).
37. Y. Xu, H. Jia, H. Huang, Z. Sun, Z. Zhang, Mutations Found in embCAB, embR, and ubiA Genes of Ethambutol-Sensitive and -Resistant *Mycobacterium tuberculosis* Clinical Isolates from China. *Biomed Res Int* **2015**, 951706 (2015).
38. S. V. Ramaswamy *et al.*, Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **44**, 326-336 (2000).
39. W. L. Huang, T. L. Chi, M. H. Wu, R. Jou, Performance assessment of the GenoType MTBDRsl test and DNA sequencing for detection of second-line and ethambutol drug resistance among patients infected with multidrug-resistant *Mycobacterium tuberculosis*. *J Clin Microbiol* **49**, 2502-2508 (2011).
40. A. Sandgren *et al.*, Tuberculosis drug resistance mutation database. *PLoS Med* **6**, e2 (2009).
41. B. A. Wolucka, M. R. McNeil, E. de Hoffmann, T. Chojnacki, P. J. Brennan, Recognition of the lipid intermediate for arabinogalactan/arabinomannan biosynthesis and its relation to the mode of action of ethambutol on mycobacteria. *J Biol Chem* **269**, 23328-23335 (1994).
42. E. Dai *et al.*, MycoResistance: a curated resource of drug resistance molecules in *Mycobacteria*. *Database (Oxford)* **2019**, (2019).
43. Ethambutol. *Tuberculosis* **88**, 102-105 (2008).

44. M. A. Lety, S. Nair, P. Berche, V. Escuyer, A Single Point Mutation in the embB Gene Is Responsible for Resistance to Ethambutol in Mycobacterium smegmatis. *Antimicrob Agents Chemother* **41**, 2629-2633 (1997).
45. H. Ihara *et al.*, Crystal structure of mammalian alpha1,6-fucosyltransferase, FUT8. *Glycobiology* **17**, 455-466 (2007).
46. T. Kubota *et al.*, Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol* **359**, 708-727 (2006).
47. T. A. Fritz, J. Raman, L. A. Tabak, Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J Biol Chem* **281**, 8613-8619 (2006).
48. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
49. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331 (2017).
50. K. Zhang, Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
51. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290 (2017).
52. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
53. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
54. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
55. W. Kabsch, XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
56. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).
57. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485 (2010).
58. C. Vonrhein *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302 (2011).
59. B. Zhang *et al.*, Crystal Structures of Membrane Transporter MmpL3, an Anti-TB Drug Target. *Cell* **176**, 636-648.e613 (2019).
60. I. Wittig, H. P. Braun, H. Schagger, Blue native PAGE. *Nat Protoc* **1**, 418-428 (2006).
61. W. DeLano, The PyMOL molecular graphics system. *Schrödinger, LLC, New York*, (2010).
62. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).

Acknowledgments:

We thank Prof. Suwen Zhao, Shuhui Wang and Meng Wu (ShanghaiTech University), Zhanqiang Du, Hongzhen Jin (Nankai University), and Dr. Xi Cheng (Shanghai Institute of Materia Medica, Chinese Academy of Sciences) for assisting
590 with computational and functional experiments; Wenzhang Chen, Jiakang Chen, and Wei Zhu (Analytical Chemistry Platform of Shanghai Institute for Advanced Immunochemical Studies), Chao Peng from (Mass Spectrometry System of NCPSS) for their technical support on mass spectrometry analysis; James Bennett (University of Oxford) for the cryo grids preparation; National Centre for Protein Science
595 Shanghai (Protein Expression and Purification system) and The Molecular and Cell Biology Core Facility of the School of Life Science and Technology (ShanghaiTech University) for use of their instrumentation and technical assistance; Prof. Xinshan Ye, Dr. Decai Xiong, Feng Xia, Jin Bai and Yong Wu (Peking University) and Prof. Yanmei Li (Tsinghua University) for their help on carbohydrate chemistry; the
600 Bio-Electron Microscopy Facility of ShanghaiTech University; Dr. Qianqian Sun for her help of cryo-EM technical support; and staff from beamlines BL17U1, BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility (China), BL41XU at SPring-8 (Japan), beamline I04-1 at Diamond light source (UK), beamline PX at Swiss Light Source (Switzerland) for assistance during data collection. **Funding:** This
605 work was supported by grants from National Key R&D Program of China (grant No. 2017YFC0840300) and Project of International Cooperation and Exchanges NSFC (grant No. 81520108019) to Z.R.; Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29020000) to L.B. and Medical Research Council MR/S000542/1 to G.S.B. **Author contributions:** Z.R. conceived the project. Z.R.,
610 G.S.B., J.L., Q.W., L.Z., and Yao Z. designed the experiments; Yao Z., L.Z., Q.Z., Y.W., C.W., F.W. and L.Q. cloned and purified the Emb proteins and their mutants; Y.G., L.Z., Yao Z. and Y.W. prepared cryo-EM samples; L.Z., Yao Z., Y.G., R.G. L.Q. and Y.W. collected and processed the cryo-EM data; Y.G. and R.G.

reconstructed all the cryo-EM maps; L.Z., Yao Z., R.G. and Y.G. built and refined the
 615 structure models; Yao Z. and L.Z. grew the crystals and collected the diffraction data;
 J.L., L.W. and Yao Z. solved the crystal structure of EmbC; S.S.G., N.V. and S.M.B.
 performed the enzymatic activity assay, synthesized chemical compound for activity
 assay; L.B. and X.Z. provided the *emb* knock out *M. smegmatis* strain; L.Z., Yao Z.,
 M.W., Yan Z., W.Z., Y.W., F.W., B.Z. and X.Y. performed mass spectrum
 620 experiment and other biochemical experiments. All the authors analyzed and
 discussed the results. L.Z., J.L., Q.W., Yao Z, R.G., H.Y., L.W.G., W.X., G.S.B and
 Z.R. prepared the manuscript with the help of all the authors. **Competing interests:**
 The authors declare no competing interests. **Data and materials availability:** All
 data are available in the manuscript or the supplementary materials. The accession no.
 625 for the 3D cryo-EM density maps reported in this paper is 30218 for *Mtb*
 EmbA-EmbB-AcpM₂ in complex with ethambutol, 30216 and 30219 for *Msm*
 EmbA-EmbB-AcpM₂ in complex with ethambutol and di-saccharide, and 30217 for
Msm EmbC₂-AcpM₂ in complex with ethambutol. The PDB accession number for the
 coordinates of the *Msm* EmbA-EmbB-AcpM₂ in complex with ethambutol and
 630 di-saccharide are 7BVC and 7BVG, *Mtb* EmbA-EmbB-AcpM₂ in complex with
 ethambutol is 7BVF, *Msm* EmbC₂-AcpM₂ in complex with ethambutol and
 di-arabinose are 7BVE and 7BVH.

Supplementary Materials:

Materials and Methods

635 Figures S1-S15

Tables S1-S3

References (48-62)

Figures and Legends

640

Fig. 1. Activity and overall structure of EmbA-EmbB-AcpM₂ and EmbC₂-AcpM₂ complexes.

645 (A) Schematic representation of the enzyme reaction catalyzed by the EmbA-EmbB-AcpM₂ complex, which transfers an arabinose residue from DPA in an $\alpha(1\rightarrow3)$ -linkage to an arabinan acceptor *e.g.* NV1. The extended product then serves as a precursor for subsequent extension by a $\beta(1\rightarrow2)$ -arabinosyltransferase catalyzed by AftB, resulting in the synthesis of the terminal branching hexaarabinofuranosyl motif found in AG. DPA, decaprenyl-phosphate-arabinose; DP, decaprenyl-phosphate; AG, arabinogalactan.

(B) Schematic representation of the $\alpha(1\rightarrow5)$ arabinosyltransferase reaction catalyzed by EmbC₂-AcpM₂ complex, leading to elongation of the arabinan chain in the LAM precursor. The structure of LAM precursor may contain several mannose groups and arabinose groups, simplified here for clarity. LAM, lipoarabinomannan.

(C) (up) The designed reaction scheme illustrating $\alpha(1\rightarrow3)$ arabinosyltransferase (EmbA-EmbB) and $\beta(1\rightarrow2)$ arabinosyltransferase (AftB) activity assays. (down) Cell-free $\alpha(1\rightarrow3)$ -arabinosyltransferase activity of the purified wild-type EmbA-EmbB complexes and catalytic site mutations in the presence and absence of ethambutol (see also [fig. S1J](#)). NV1 was used as the acceptor and DP[¹⁴C]A as the donor as described in previous studies (11, 12, 16). *Msm* membrane contains ethambutol-resistant arabinosyltransferase AftB. EMB, ethambutol. Data presented are the mean values +SD calculated from three independent experiments.

(D) (up) The designed reaction scheme illustrating an $\alpha(1\rightarrow5)$ arabinosyltransferase activity assay to characterize EmbC₂ activity. (down) Cell-free $\alpha(1\rightarrow5)$ -arabinosyltransferase activity of the purified EmbC₂ complex in the presence and absence of ethambutol (see also [fig. S2G](#)). NV6 was used as the acceptor and DP[¹⁴C]A as the donor as described in previous studies (13, 16). *Msm* membrane contains ethambutol-resistant arabinosyltransferase AftB. EMB, ethambutol. Data

670 presented are the mean values +SD calculated from three independent experiments.

(E) Overall view of cryo-EM structures of *Mtb* and *Msm* EmbA-EmbB-AcpM₂ complexes in complex with ethambutol or di-arabinose, and cryo-EM and crystal structures of *Msm* EmbC₂-AcpM₂ complexes in complex with ethambutol or di-arabinose. The AcpM protomer binds to each Emb protein in all the complexes, 675 un-modeled AcpM protomer bound to *Mtb* EmbB or *Msm* EmbA are colored in grey. The drug EMB, the substrate DPA and Ara₂, the lipids are shown as spheres. Ara₂, di-arabinose.

(F) The overall fold of Emb proteins represented by *Mtb* EmbB. The PN, PC and TM domains are colored differently. The functional important PL2 and PL5 are also 680 highlighted. The location of active site is marked by a dashed circle. PN/PC, N-/C-terminal periplasmic domain; TM transmembrane domain; PL, periplasmic loop connecting two transmembrane helices.

(G) Topological diagram of Emb proteins colored as in (F), DDx motif is shown as three red spheres. CL, cytoplasmic loop connecting two transmembrane helices; EH, 685 extra-cellular helix.

690 **Fig. 2. Substrate binding in the active sites of EmbA-EmbB and EmbC₂ complexes.**

(A) An overview of the substrate entrances to the active site in Emb proteins represented by *Msm* EmbC. *Msm* EmbC is shown as electrostatic surface representation. The arrows indicate two entrances to the active site, one for donor and one for acceptor. The dashed line indicates the clipping position for (D).

(B) An overview of *Msm* EmbA complexed with donor substrate DPA. DPA is shown as spheres. The head groups (arabinose and phosphate) bind in the active site through the donor entrance, while the tail groups bind to the TM region.

700 (C) Zoom-in view of DPA binding in the active site of *Msm* EmbA. Polar interactions are indicated by black dashed lines. The distance between the catalytic Asp261 and C1 atom of DPA is marked by a salmon dashed line.

(D) A clipped view of the active site with acceptor entrance and donor entrance in *Msm* EmbC. Ara₂ (part of Ara₂OC8), maltose (part of detergent DDM) and P_i as well as the catalytic residue, Asp279, are shown as sticks. Ara₂ could mimic the terminal two arabinose groups of the reaction elongated product by EmbC. The red box and green box indicate the binding position (D-site) for arabinose from donor and binding position (A₀-site) for arabinose from the terminal residue of acceptor, respectively.

(E) The phosphate binding site of *Msm* EmbC. P_i and surrounding residues are shown as sticks. Polar interactions are indicated by dashed lines.

(F) Superposition between *Msm* EmbC (yellow) and *Msm* EmbA (magenta) in the active site. The D-site arabinose group of Ara₂ in *Msm* EmbC binds in a same position as that of DPA in *Msm* EmbA. The phosphate in *Msm* EmbC binds a similar position as the phosphate group of DPA in *Msm* EmbA. In *Msm* EmbA, the Asn780 on Loop₇₆₆₋₈₀₆ interacts with the catalytic Asp261 and is close to the arabinose group of DPA.

(G) Effect of mutated residues in the phosphate binding site on $\alpha(1\rightarrow5)$ -arabinosyltransferase activity of *Msm* EmbC₂ (13, 16). Data presented are the

mean values +SD calculated from three independent experiments.

720 **(H)** Effect of mutated residues in the phosphate binding site of *Msm* EmbC on LAM synthesis. LAM was extracted from recombinant *Msm* strains and was analyzed by SDS-PAGE (on a denaturing, 16% acrylamide gel) and periodic acid-Schiff staining. *embCA*, the *Msm embC* knock-out mutant; *embC* / *embC_R383A et al.*, the *Msm embC* knock-out mutant complemented with plasmid carrying *embC* wild-type
725 wild-type / R383A *et al.* mutant alleles.

(I) Close-up view of the Ara₂ binding site of *Msm* EmbC. Ara₂ and interacting residues are shown as sticks. Polar interactions are indicated by dashed lines.

(J) Effect of mutated residues in the Ara₂ binding site on $\alpha(1\rightarrow5)$ -arabinosyltransferase activity of *Msm* EmbC₂ (13, 16). Data presented are the
730 mean values +SD calculated from three independent experiments.

(K) Effect of mutated residues in the Ara₂ binding site of *Msm* EmbC on LAM synthesis. LAM was extracted from recombinant *Msm* strains and was analyzed by SDS-PAGE (on a denaturing, 16% acrylamide gel) and periodic acid-Schiff staining. *embCA*, the *Msm embC* knock-out mutant; *embC* / *embC_D279N et al.*, the *Msm embC* knock-out mutant complemented with plasmid carrying *embC* wild-type
735 wild-type / D279N *et al.* mutant alleles.

(L) Proposed mechanism of arabinose transfer and chain elongation for EmbC. Helix EH1 and EH3 are indicated as cylinders. The catalytic Asp residue (Asp279 in *Msm* EmbC) is represented as blue sticks. Arabinose and phosphate groups are shown as a
740 pentagon and circle, respectively.

Fig. 3. Structural basis for ethambutol inhibition of EmbB and EmbC.

(A) Cartoon representations of ethambutol-bound *Msm* EmbB (left), *Mtb* EmbB (middle) and *Msm* EmbC (right). Ethambutol (EMB) is shown as cyan spheres.

(B) Binding affinity of ethambutol with the *Mtb* EmbA-EmbB complex measured by the MST assay. The K_d value is provided and the data are representative mean values + SD calculated from three independent experiments.

(C) Structural details of ethambutol binding to *Mtb* EmbB, PL2, PL6 and the PC domain are colored in purple, blue and red, respectively. Interacting residues are shown as sticks. Polar interactions are indicated by dashed lines. The cryo-EM map density for ethambutol (threshold 0.4) is shown as grey mesh.

(D) Schematic diagram of the interaction between ethambutol and *Mtb* EmbB.

760 **(E)** Superposition of the active site region of the ethambutol (cyan) bound *Msm* EmbB to the di-saccharide (yellow) bound *Msm* EmbB.

(F) Superposition of *Msm* EmbA (salmon) to *Msm* EmbB (blue) on their periplasmic regions. Loop₇₆₆₋₈₀₆ in *Msm* EmbA which is longer, blocks the acceptor entrance to the active site. Whilst, the corresponding Loop₇₇₆₋₈₀₆ in *Msm* EmbB is folded inside
765 the PC domain.

(G) Zoom-in view of the active site upon superposition in (F). The ethambutol (cyan) in *Msm* EmbB clashes with the arabinose group in DPA and Asn780 in *Msm* EmbA.

Fig. 4. Structural interpretation for ethambutol resistance.

775

(A) Cartoon representation of ethambutol binding pocket in *Mtb* EmbB. Ethambutol (EMB) is shown as sticks. Residue sites bearing the drug resistant hotspot mutants are labeled with dark red spheres.

780

(B) Structural details of the Met306 interaction environment in *Mtb* EmbB. The dashed lines indicate the distance of surrounding residues or ethambutol to Met306.

(C) Binding affinity of ethambutol to *Msm* EmbA-EmbB or its ethambutol resistance associated mutants equivalent to that of *Mtb* EmbB, measured by the MST assay. The K_d values are provided and the data are representative mean values + SD calculated from three independent experiments.

785

(D) Cell-free $\alpha(1\rightarrow3)$ -arabinosyltransferase activity of the ethambutol resistant mutations on EmbB (*Msm* M292V, M292I and I289F) in the absence and presence of ethambutol (11, 12, 16). The wild type *Msm* and *Mtb* EmbA-EmbB complexes are used as controls. Data presented are mean values +SD calculated from three independent experiments.

790

(E) Mutations I289M (grey color) and I289F (dark grey color) affect ethambutol binding by forming steric clashes that ultimately lead to ethambutol resistance. The cryo-EM map for ethambutol (threshold 0.3) is shown as cyan mesh.

(F) The drug resistant hotspot Gln497, surround the drug-binding pocket.

795

800

805

Supplementary Materials

Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol

Lu Zhang^{1,2†}, Yao Zhao^{1,3,4†}, Yan Gao⁵, Lijie Wu¹, Ruogu Gao^{4,6}, Qi Zhang¹, Yinan
Wang^{1,4}, Chengyao Wu¹, Fangyu Wu², Sudagar S. Gurucha⁷, Natacha Veerapen⁷,
815 Sarah M. Batt⁷, Wei Zhao², Ling Qin¹, Xiuna Yang¹, Manfu Wang¹, Yan Zhu¹, Bing
Zhang¹, Lijun Bi⁶, Xian'en Zhang⁶, Haitao Yang¹, Luke W. Guddat⁸, Wenqing Xu¹,
Quan Wang^{1,6*}, Jun Li^{1*}, Gurdial S. Besra^{7*}, Zihe Rao^{1,2,5,6*}

* Correspondence to wangq@ibp.ac.cn (Q.W.), lijun1@shanghaitech.edu.cn (J.L.),
820 G.Besra@bham.ac.uk (G.S.B.) and raozh@tsinghua.edu.cn (Z.R.)

This PDF file includes:

Materials and Methods

Figs. S1 to S15

825 Table S1 to S3

References (48-62)

Materials and Methods

Protein expression and membrane preparation

830 The cluster of *embA-embB* genes (*Rv3794-3795*) from *Mtb* strain *H37Rv* genome and (*MSMEG_6388-6389*) from *Msm* strain *mc²155* genome were cloned into the engineered pMV261 vector fused with a flag tag attached to the N-terminus of EmbA and a 10 × His tag to the C-terminus of EmbB, under the control of an acetamide promoter. The *embC* gene (*MSMEG_6387*) from *Msm* strain *mc²155* genome was
835 cloned into the same vector fused with a 10 × His tag attached to the C-terminus of EmbC. Recombinant plasmid was introduced into *Msm mc²155* competent cells by electroporation. For large scale production, cells were cultured in 1 L Luria-Broth (LB) medium supplemented with 50 µg/mL kanamycin, 20 µg/mL carbenicillin, and 0.1% (v/v) Tween80 (to avoid cell aggregation) at 37 °C with shaking at 220 rpm
840 until the OD₆₀₀ reached 1.0. The EmbA-EmbB-ethambutol complexes were prepared by adding ethambutol at twice minimum inhibitory concentration (MIC) whilst the target protein was overexpressed in *Msm* cells. Four days after induction with 0.2% (w/v) acetamide at 16 °C, the cells were collected in Buffer A containing 20 mM HEPES, 150 mM NaCl and 5% (v/v) glycerol, pH 7.4 for EmbA-EmbB complex, or
845 buffer B containing 20 mM Tris-HCl, 150 mM NaCl, and 5% (v/v) glycerol, pH 8.0 for EmbC. Cells were lysed by French Press at 1,200 bar and 4 °C. Cell debris was cleared by centrifugation at 10,000 g for 10 min at 4 °C. The membrane pellet was collected by ultracentrifugation (150,000 g, 1 h) at 4 °C then resuspended in Buffer A for EmbA-EmbB complex, or buffer C containing 20 mM Tris-HCl, 150 mM NaCl, 5
850 mM MgCl₂, 5% (v/v) glycerol, pH 8.0 for EmbC and stored at -80 °C until use. All mutants were expressed using the same protocol as the wild-type protein.

Protein purification

Thawed membrane fractions were solubilized with 1% (w/v) *n*-dodecyl- β -D-maltopyranoside (DDM; Anatrace) by gently agitating for 1.5 h at 4 °C. Detergent-insoluble material was removed by ultracentrifugation (18,000 rpm, 30 min). For EmbA-EmbB complex (*Mtb* EmbA: 1094 aa, ~116 kDa; *Mtb* EmbB: 1098 aa, ~118 kDa; *Msm* EmbA: 1080 aa, ~115 kDa; *Msm* EmbB: 1082 aa, ~117 kDa), supernatant was purified by nickel affinity resin (Qiagen) and then anti-FLAG (Sigma) affinity followed by size-exclusion chromatography using a Superose 6 Increase column (GE Healthcare) pre-equilibrated with Buffer D containing 20 mM HEPES, 150 mM NaCl, pH 7.4, and 0.04% (w/v) glyco-diosgenin (GDN, Anatrace). The peak fraction corresponding to the EmbA-EmbB complex was concentrated to 5 mg/mL for cryo grid preparation. For *Msm* EmbC (1074 aa, ~115 kDa), supernatant was purified by Co-NTA agarose beads and then applied to a size exclusion chromatography column (Superose-6 increase, GE Healthcare) pre-equilibrated with Buffer E containing 10 mM Tris-HCl, 100 mM NaCl, 1 mM MgCl₂ 1% (v/v) glycerol, pH 8.0 supplemented with 0.02% (w/v) DDM for crystallization or 0.04% (w/v) GDN for Cryo-EM study.

Protein for activity assays was purified using the same protocol except the gel filtration buffer was exchanged to an assay buffer (Buffer F) containing 50 mM MOPS, 10 mM MgCl₂, pH 7.9, 5 mM β -mercaptoethanol, 5% (v/v) glycerol and DDM at twice critical micelle concentration (CMC).

Grid preparation and data collection

For the ethambutol-bound EmbC₂ complex and di-arabinose-bound EmbA-EmbB complex, drug/ligand were added to concentrated target proteins just prior to transferring to the cryo-grid. Aliquots of the freshly purified sample were applied to glow-discharged holey carbon grids (Quantifoil Cu R0.6/1.0, Solarus Gatan Plasma System H₂/O₂ for 25 s). Grids were blotted for either 2.5 s for EmbA-EmbB or 3 s for

EmbC₂, flash-frozen in liquid ethane and cooled in liquid nitrogen using an FEI Mark
880 IV Vitrobot (humidity 100%, temperature 281 K, blotting paper TED PELLA 595
filter paper). Images were taken using an FEI Titan Krios electron microscope
operating at 300 kV with a Gatan K3 Summit direct electron detector at a nominal
magnification of 29,000. Images were recorded in super-resolution mode and binned
to a pixel size of 0.82 Å. Automated single-particle data acquisition was performed
885 with SerialEM data collection software (1). Defocus values varied from 0.8 to 2.5 µm
for EmbA-EmbB or 1.5 to 2.5 µm for EmbC₂. Each stack was exposed for 2 s with a
total dose of 50 e⁻/Å², with 40 frames per stack. The details of electron microscopy
data collection parameters for each batch of EmbA-EmbB or EmbC₂ complexes are
listed (Table S1).

890 EM image processing

All dose-fractioned images were motion-corrected and dose-weighted by
MotionCorr2 software (2) and their contrast transfer functions were estimated by Gctf
(resolution range: 4~25 Å; search defocus: 0.1~4 µm) (3). For the *Msm*
EmbA-EmbB-AcpM₂ in complex with ethambutol dataset, 1,855,947 particles were
895 picked automatically from 5,100 images (Particle diameter: 220 Å; Minimum
separation distance: 110 Å) and extracted with a box size of 384 pixels using
cryoSPARC (4). The subsequent 2D, 3D classification and refinement steps were all
performed in cryoSPARC. The instructions of data processing and refinement
protocols are available at <https://cryosparc.com/docs/reference/jobs>. 256,328 particles
900 were selected after two rounds of 2D classification (Number of 2D classes: 100;
Window inner radius: 0.85; Maximum resolution: 6 Å; Iterations: 20). 100,000
particles were used to do *ab-initio* reconstruction in two classes (Classes: 4;
Maximum resolution: 12 Å; Initial resolution: 35 Å; Class similarity: 0.1), which were
used as 3D volume templates for heterogeneous refinement with all selected particles
905 (Refinement box size: 128 pixels). 227,206 particles were converged into one class,

yielding a 5.04 Å initial map. Next, this particle set was used to perform homogeneous refinement, yielding 2.99 Å. After non-uniform (NU) refinement, the final resolution reached 2.90 Å (fig. S4 and Table S1). The datasets for *Msm* EmbA-EmbB-AcpM₂ in complex with di-arabinose, *Mtb* EmbA-EmbB-AcpM₂ in complex with ethambutol and *Msm* EmbC₂-AcpM₂ in complex with ethambutol were processed in the same way (figs. S3, S5, S6 and Table S1).

Model building and refinement

For *Mtb* EmbA-EmbB-AcpM₂ in complex with ethambutol, all residues in each subunit were modelled as alanine in the initial building and assigned subsequently with the guidance of secondary structure prediction of Phyre2 (5). Manual adjustment of the complete model was first performed in COOT 0.8.8 (6), followed by iterative rounds of real-space refinement in PHENIX 1.12 (7) and manual adjustment in COOT. Structures for the *Msm* EmbA-EmbB-AcpM₂ in complex with ethambutol or di-saccharides, and for *Msm* EmbC₂-AcpM₂ in complex with ethambutol were built and refined in the same way (Table S1). Refinement strategies used included, "minimization_global", "local_grid_search" and "atomic displacement parameters (ADP)". Restraints included rotamer restraints, Ramachandran restraints and NCS (non-crystallographic symmetry) constraints.

Crystallization

Crystallization trials were performed by the hanging-drop vapor diffusion method at 16 °C. The *Msm* EmbC protein solution in buffer E supplemented with 0.02% (w/v) DDM, diluted to 1-3 mg/mL, was mixed in a 1:1 (v/v) ratio with the reservoir solution. Crystals were grown from the condition containing 50 mM HEPES (pH 6.8~7.5), 100 mM NaCl, 5~10% (v/v) polyethylene glycol 4000 (PEG4000) and 20~30% (v/v) polyethylene glycol 200 (PEG 200). After optimization, crystals were harvested, flash-cooled and stored in liquid nitrogen for data collection. To obtain the

Terbium (Tb)-derivative crystals for phasing, proteins were incubated with Tb-Xo4 from a Polyvalan Crystallophore No1 kit (Molecular Dimensions) prior to co-crystallization in the same reservoir condition. To obtain crystals with bound
935 di-arabinose, proteins were incubated with 1 mM Ara₂OC8 prior to co-crystallization in the same reservoir condition.

X-ray data collection and structure determination

X-ray data were collected on beamlines BL17U1, BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility (SSRF), beamline I04-1 at Diamond Light Source
940 (DLS), beamline PX at Swiss Light Source (SLS) and beamline BL41XU at SPring-8. Multiple data sets were processed, merged and scaled using XDS (8) and the CCP4 suite (9) to obtain the final data set. The Tb-derivative anomalous data were collected at a wavelength of 1.6491 Å. Six Tb heavy-atom sites were found using the program SHELXD (10). The initial phases were determined by the single anomalous
945 dispersion (SAD) method using phenix.autosol, with the figure of merit (FOM) and BEYES CC score to be 0.3 and 41, respectively. After density modification, densities of helices and strands could be clearly observed in the map and the initial protein model was successfully traced and manually built in COOT. The model was further completed and refined against the native data, which was processed by anisotropic
950 correction in phenix.autosol. After several cycles of iterative manual building and refinement, the $R_{\text{work}}/R_{\text{free}}$ of the final model is obtained to be 23.2%/26.5%, by refinement with autoBUSTER using default parameters (11). Data collection and structure refinement statistics are summarized in **Table S2**.

Arabinosyltransferase activity assays

955 Assays were essentially performed as described previously (12-14) using NV1 for EmbA-EmbB or NV6 for EmbC₂ (1 mM in water), DP[¹⁴C]A (100,000 cpm, stored in 1% IgePal), 1 mM ATP, 1 mM NADP, purified EmbA-EmbB / EmbC₂ complexes (4

μM) in buffer F or *Msm* membrane and P60 fractions (1 mg each) and in some cases ethambutol, with the appropriate amounts of buffer F. All samples were made to a final volume of 80 μL. These were incubated at 37 °C for 1 h, quenched by the addition of 533 μL of chloroform/methanol (1:1, v/v) and mixed overnight at 4 °C. The supernatant was recovered following centrifugation and dried. The residue was resuspended in 2 mL of ethanol/water (1:1, v/v) and loaded onto a 1 mL SAX SepPak and washed with 2 mL of ethanol and the eluate collected and dried. The sample was resuspended in a mixture of water-saturated *n*-butanol (2 mL) and water (2 mL) and the organic phase recovered. The aqueous phase was re-extracted using water-saturated *n*-butanol (2 mL) and the organic phases pooled and re-washed with water (2 mL). The organic layer was dried and resuspended in *n*-butanol. The incorporation of [¹⁴C]arabinose from DP[¹⁴C]A was determined by scintillation counting and by subjecting samples to TLC using silica gel plates developed in chloroform/methanol/water/ammonium hydroxide (65:25:3.6:0.5, v/v/v/v) and visualized by autoradiography using Kodak BioMAX MR films. Each assay was repeated three times.

Preparation of AG from the mAGP complex and two-dimensional ¹H/¹³C-nuclear magnetic resonance (2D-NMR) spectroscopy

Bacterial cells were resuspended in phosphate-buffered saline containing 2% (v/v) Triton X-100 (pH 7.2), disrupted by sonication and centrifuged. The pelleted material was extracted three times with 2% (w/v) SDS in phosphate-buffered saline at 95 °C for 1 h, washed with water, 80% (v/v) acetone in water, and acetone, and subsequently lyophilised to yield a highly purified mAGP preparation. This was then subjected to mild base hydrolysis for 4 days to remove mycolic acids using 5% (w/v) KOH in methanol at 37°C. The insoluble residue was recovered by centrifugation at 27,000 g. The sample was washed repeatedly with methanol, followed by diethyl-ether, and the resulting AGP treated with 2 M NaOH for 16 h at 80°C. The supernatant, which

985 contained base-solubilised AG, was recovered by centrifugation at 27,000 g for 30 min. The crude AG preparation was neutralised with acetic acid and dialysed to remove salt (MWCO 3500). The supernatant was diluted in cold ethanol (80%, v/v) and left at -20 °C overnight to precipitate the base-solubilised AG, which was then recovered by centrifugation and lyophilised. 2D-NMR spectra of AG samples were
990 recorded using a Bruker DMX-500 instrument as described previously (15). Samples were repeatedly exchanged in deuterium oxide (99.9 atom % D) with intermediate lyophilisation and analysed at 313 K. The ¹H and ¹³C NMR chemical shifts were referenced relative to the solvent signal D₂O at δ 4.79.

Mass spectrometry

995 For DPA and DP identification, 50 µL of DDM purified EmbA-EmbB complex from *Mtb* and *Msm*, with or without ethambutol was incubated with 350 µL chloroform/methanol (1:1, v/v) then left overnight on ice. The suspension was converted to a bilayer by adding 250 µL chloroform/water (7:3, v/v) the next day. The lower organic phase was pooled after centrifugation and then dried in a speed vacuum
1000 concentrator. The dried lipids were re-dissolved in 20 µL chloroform/methanol. 1 µL of the sample was injected into QTOF (SCIEX 4600) MS coupled with UPLC (Shimadzu, 30A). After loading the sample onto the chromatography column (Waters Bioresolve Polyphenyl, 450 Å, 2.7 µm, 2.1 × 150 mm) the product was eluted by gradient as followed: Buffer G (0.1% (v/v) formic acid and 1% (v/v) acetonitrile) for
1005 1 min, then 5% to 95% Buffer H (0.1% (v/v) formic acid in acetonitrile) in 3 min, then 95% Buffer H for 3.5 min. The flow rate was 50 µL/min. The mass spectrometer was operated in negative mode. The source voltage, the curtain gas, and the source temperature were set to 4500 V, 30 psi and 350 °C, respectively. A SIM scan (m/z: 909.6, window width: 2 Da) followed by a MS2 scan was used to detect the targeted
1010 lipid. The collision energy was set to 35 eV.

Extraction and analysis of LAM

The experiment was carried out according to a method reported previously (16). In brief, *Msm* was grown in 7H9 liquid media supplemented with 50 µg/mL kanamycin and 20 µg/mL carbenicillin. Cells were harvested at mid-log phase and washed twice
1015 in phosphate buffered saline. Pellet was resuspended in 4 mL 50% ethanol (v/v in water) and disrupted using probe sonication. The mixture was then refluxed at 85 °C for 6 h followed by centrifugation and recovery of the supernatant. The ethanol extraction process was repeated five times, and the combined supernatants dried. Crude lipoglycans were then subjected to 90% phenol treatment (w/v in water) at 65
1020 °C for 1 h. After cooling, the sample was centrifuged and the upper aqueous layer recovered and dialyzed against water (MWCO 3,500 Da), dried and subjected to 10% Tricine SDS-PAGE and then visualized using a Pierce Glycoprotein Staining Kit (Thermo Scientific).

Microscale thermophoresis assay

1025 The Microscale thermophoresis (MST) assay was accomplished according to a previously reported method (17). The binding affinity of the detergent purified wild-type EmbA-EmbB complex or mutants from both *Mtb* and *Msm* to ethambutol was measured using a Monolith NT.115 (Nanotemper Technologies). The His-tagged protein was labeled with RED fluorescent dye NT-647 according to the
1030 manufacturer's procedure. For each assay, the labeled protein at 200 nM was incubated with the same volume of unlabeled ligands at 16 different concentrations in the same buffer as the protein at room temperature for 10 min. The samples were then loaded into capillaries (NanoTemper Technologies) and measured at 25 °C by using 40% LED and medium MST power. Binding affinities of ethambutol, DPA,
1035 Ara₂OC₈, with the wild-type EmbC and mutants were measured under the same parameter. Each assay was repeated three to five times. In DPA-protein MST

experiments, where additional detergent was required to solubilize the DPA substrate, the same concentration of detergent was added to both the protein and DPA substrate stocks (0.02% (w/v) n-dodecyl- β -D-maltoside). Initial DPA-protein samples were measured thrice with the sample thoroughly mixed prior to incubation by pipetting, over a time course of approximately 50 min, to ensure that there was no change to the binding curve over time. K_d values were calculated using the MO. Affinity Analysis v.2.2.4 software. All of the final plots were made using GraphPad Prism 8.0.

Electrophoresis

In order to determine the native oligomerization state of the full-length EmbC in solution, we used the Blue Native PAGE technique. The purified protein sample was mixed with 10 $\times$ loading buffer (0.1% (w/v) Ponceau S, 50% (w/v) glycerol) and loaded onto a 4-16% Blue Native PAGE mini gel (1.5 $\times$ 8.3 $\times$ 7.3 mm) at 4 $^{\circ}$ C. The gel was run at 100 V for 10 min. Cathode buffer B (50 mM Tricine, 7.5 mM imidazole, 0.02% (w/v) Coomassie brilliant blue G-250) was then changed to cathode buffer B/10 (50 mM Tricine, 7.5 mM imidazole, 0.002% (w/v), Coomassie Brilliant Blue G-250). The run continued at 150 V for 1.5 h. All buffers and procedures are based on a standard Blue Native PAGE protocol (18).

Creation of figures

Figures of molecular structures were generated using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC.) (19) and UCSF ChimeraX (20).

A

B

C

D

E

1 VPHDGNERSH RIRLARAAVS GIAGLLLCGI VPLLPNQTT ATIFWPQGST
 51 ADGNITQITA PLVSGAPRAL DISIPCSAIA TLPANGGLVL STLPAGGVD
 101 GKAGLFVRAN QDTVVVAFR SVAAVAARST IAAGGCCALH IWADTGAGAGA
 151 DFMGIPGGAG TLPPEKRPQV GGIFTDLKVG KQPGLSAIVD IDTRFITTPG
 201 ALKKAVMLLG VLAVLVMVG LAALDRLSRG RTRLRWLTRY RPRVRVGFAS
 251 RLADAAVIAT LLLMHVIGAT SSDDGYLLTV ARVAPKAGP RNYTHYFGTT
 301 EAPFDWYTSV LAQLAAVSTA GVMRLPATL AGIACWLIVS RFVLRRLGPG
 351 PGGLASNRVA VFTAGAVFLS AMLPFNNGLR PEPLIALGVL VTWLVRSI
 401 ALGRLAPAAV AIIIVATLTAT LAPQGLIALA PLLTGARAI QRIRRRATD
 451 GLLAPLAVLA AASLITVVV FRDQTLATVA ESARIKYKVG PTIAMYQDFI
 501 RYYFLTVESN VEGSMRFRFA VLVLLFCLFG VLFVLLRRGR VAGLASPAW
 551 RLIGTTAVGL LLLTFTPTKW AVQGFAGFL AGVLGAVTAF TFARIGLHSH
 601 RNLTLYVTAL LFLVLAWSG INGWFYVQNY GVPWYDIQPV IASHPVMSF
 651 LTLILTGLL AAWYHFRDQ KQTEPQZWR RNRILASTPL LVAIVIMVAG
 701 EVGSMAKAV FRVPLYTTAK ANLTALSTGL SSCAMADDVL AEPDNAGML
 751 QPVPQAGFP DGPLGGISPV GFKPEGVED LKSDPVVSKP GLVNSDASP
 801 KPNAAITDSA GTAGGKGPVG TNGSHAALPF GLDPARTPVV GSYGENNLAA
 851 TATSAWYQLP PRSPDRPLVV YSAAGAIWSY KEDGDFIYG SKLQWGVTG
 901 PDGRIQPLQG VFPIDIGPQP AWRNRLFPLA WAPPEADVAV IVAYDPNLS
 951 EQWFAFTPPR VPVLESRLR IGSAATPVMD IATAANFPCC RPFSEHLGIA
 1001 ELPQYRILPD HKQTAASSNL WQSSTGGPF LFTQALLRTS TIATYLRGDV
 1051 YRDNGSVEVY RLVPADQAP DAVVEEGVIT VPGWGRPGPI RALP

G

1 MTEPSRIARL IAVVAGIAGV LLCGLVPLL VEETTATVLV PQGVGADGNV
 51 TELTAPLVAG APPALDVTIP CRAVAELPAD GGVVFSTNPA GQTEAGRNGM
 101 FIRANADVY VAFRDTVAAV APREAVDSGA CSEIHWADV SAVGADFAGI
 151 PDASGTLVD KQPDSCVF DEVDYDQV RQIDIDTRF ITSPTLLKTA
 201 VMVLGLACVI GSIVALALLD RGWRRRPRT RGRAGLWTWI TDTGVIGLL
 251 IWHIVGAPTS DDGYNMTIAR VASEAGYTN YRYFYGASEA PFDWQSVLS
 301 HLASISTAGV WMRLPATAAA IATWLIIISRC VLPRIGRRVA ANRVAMLTAG
 351 ATFLAALWPF NNGLRPEPLI AFAVITVWML VENSIGTRRL WPAAVAIVIA
 401 MFSVTLPQGG LIALAPLLVG ARAIRGVVTA RRAGTGILAS LAPLAASVAV
 451 VFVIIFRDQT LATVAESVRI KYVVGPTIPW YQEFRLYYFL TVEDSVDGSL
 501 TRRFAVLVLL LCLFGLIMVL LRRGRVPQAV SGPLWRICGS TAIGLLLLIL
 551 TPTKWAIQFG AFAGLAGALG GVTAFAFARV GLHSRRNLAL YVTALLFILA
 601 WATSGLNGWF YVGNVGPWF DKQPVIAHYP VTTIFVLAI VGGLLAGWLH
 651 FRDYGAGHTE VADTGNRAL ASTPLLIVAT IMVVELGSM VKATVGRYPV
 701 YTVGSANIAA RSAGDSCAM ADAVLVEADP NEGMLQPVPG QRFGEYGPLG
 751 GEDPVGFTPN GVSDTLEPAE PVAANPGTPN SDGPVDKPMI GIGYAAGTGG
 801 GYGPEGVNGS RFLPFGLDF SRTPVNGSYG ENKLAARKTS AWYQLPPRTP
 851 DRPLVTVAAG GAIWYEEEDG SFNYGQSLKL QWGVHRPDGT YQALSEVQPI
 901 DIFQKAWRN LRFLAWAPP EANVARIVAD DPNLSEQWF AFTPPRPVVL
 951 QTAQQLGSG TPVLMIDIATA ANFPCCRPFA ERLGVAELPE YRIIPNFKQM
 1001 VVSSNQWQSA ADGGPFLFIQ ALLRTEATIP YLRDWMYRDN GSIEFYIRIV
 1051 PQEQAPTAAI EGSTRVFGV SRGGPIRALP

F

1 MTQCASRRKS TPNRAILGAF ASARGTRWVA TIAGLIGFVL SVATPLLPVV
 51 QTTAMLDMWQ RQQLGSVTAP LLSLTPVDFI ATVPCDVVRA MPPAGGVVLG
 101 TAPKQKGRM LGALFVUVSR RRVVDTRNV VTLSPVREOV TSPQCORIEV
 151 TSTHAGTFAN FVGLKDPSCA PLRSGFPDPN LRPQIVGVFT DLTGPAPPGL
 201 AVSATIDTRF STRPTTLKLL AIIGAIVATV VALIALWR LRDLGNGSIAQ
 251 LLLRPFPRPAS SPGGMRRLR RQVFTFLTD AVVIFGFLW HVIGANSSDD
 301 GYILGMARVA DHAGYMSNYF RWFSGPEDPF GWYNNLLALM THVSASLWM
 351 RLPDLAAGLV CWLLLSREVL PRLGPAVEAS KPAYWAAAMV LLTAMMPFNN
 401 GLRPEGIAL GSVTVYVIE RSMYRSRLTP AALAVVTAAF TLGVQPTGLI
 451 AVAALVAGGR PMLRILVRRH RLVGTLPLVS PMLAAGTVIL TVVFADQTLS
 501 TVLEATRVRA KIGPSQAWYT ENLRYYYLIL PTVDGSLSR FGFLITALCL
 551 FTAVFIMLR KRLPSVARGP AWRLMGVIFG TMFFLMFTPT KWHHFGFLA
 601 AVGAAMAALT TVLVSPSVLR WSRNRMAFLA ALFFLLALCW ATTNWWMYVS
 651 SYGVFPNSAM PKIDGITVST IFFALFAIAA GYAALWHPAF RGAGEGRLIR
 701 ALTTAPVPIV AGFMAAVFA SMVAGIVRQY PTYSNGWSNV RAFVGGCGLA
 751 DDVLEPDTN AGFMKPLDGD SGSWGPLGLP GGVNVGPTF NGVPEHTVAE
 801 AIVMKPNQPG TDYDWDAPTK LTPSGINGST VPLPYGLDPA RVPLAGTYTT
 851 GAOQSTLVS AWYLLPKPDD GHPLVVVTA GKTAGNSVLH GYTPGQTVVL
 901 EYAMPGGGAL VPAGRVPPDD LYGEQPKAWR NLRFAKRR ADAVAVRVVA
 951 EDLSLTPEDW IAVTPPRVPD LRSLQEYVGS TQPVLDWAV GLAFPCQPM
 1001 LHANGIAEIP KFRITPDYSA KKLDTDTWED GTNGLLGIT DLLLRARVMA
 1051 TYLSRDWARD WGSRLRKFDTE VDAPPAQLEI GSTATSGLWS PGKIRIGP

H

1 MSQNMDEAVS ENMDEAVSAG KQVRIARWVA TIAGLIGFVL SVSIPLLPV
 51 QTTATLNMWQ QGRLDNVAP LISQAPLELT ATVPCSVVR LPPEGLLVFG
 101 TAPAEGRDAA LNAMLNVTE TRVDVIVRNV VVASWRRDR RQDQGRIEI
 151 TSNLDGTYAD FVGLTQISGE DAGLKQRTGY PDPNLRPAIV GVFTDLTGPA
 201 PQGLSVSAEI DTRFTTHPTA LKLAAMLLAI VSTVIALALL WRLDRLDGRR
 251 MHRILPTRWR TVTAVDGVVV GMAIWWYVIG ANSSDDGYIL QMARTEAHAG
 301 YMANFYRMFG SPEDPFGWYV NVLALMTKYS DASIWIRLPD LICALICWLL
 351 LSREVLPRLG PAVAGSRAM WAAGLVLLGA WMPFNGLRP EGQIATGALI
 401 TYVLIERAVT SGRLTPAALA ITTAFTLGI QPTGLIAVAA LLAGGRPILR
 451 IVMRRRLVG TWPLIAPLLA AGTVILAVFV ADQTIATVLE ATRIRTAIGP
 501 SQEWWTENLR YYYLILPTTD GAISRVAVFV FTAMCLFPLS FMMLRRKHIA
 551 GVARGPAWRL MGIIFATMFF LMFTPTKWIH HFGFAAVGG AMAALATVLV
 601 SPTVLRSAVN RMAFLSLVLF VLAFCFASTN GWMYVSNFGA PFNNVSPKVG
 651 GVQISAIFFA LSAIAALWAF WLHLTRRTES RVVDRLTAAP IPVAAGFMV
 701 VMASMAIGV VRQYPTYNSG WANIRAFAGG CGLADDVLVE PDSNAGFLTP
 751 LPGAYGLGP LGGEDPGQFS PDGVDPRLIA FAIRLUNPQP GTDYDMNRPI
 801 KLDPEGINGS TVPLPYGLDP KRVMPVAGTYS TEAQQESR RANVCLPMD
 851 ETERAAHPLV VITAAGTITG ESVANGLTTG QTVDELAYTR GPDGTLVPAG
 901 RVTPTYDVGPT PSMRNLRYPR SEIPDDAVAV RVVAEDLSLS QGDWIAVTPP
 951 RVPQLQSVQE YVGSQDPVLM DWAVGLAFPC QQMLHANGV TEVPKFRISP
 1001 DYYAKLQSTD TWQDGINGGL LGITDLLLLA SVMSTYLSQD WQGDWGSRLK
 1051 FDTVVEATPA ELDFGSQTHS GLYSPGPLRI RP

Fig. S1. Characterization of EmbA-EmbB-AcpM₂ complex.

(A) Size-exclusion chromatography on a Superose 6 gel filtration column (GE healthcare) for the EmbA-EmbB complex from *Mtb* (left) and *Msm* (right) purified with detergent GDN.

(B) SDS-PAGE of the main peak fraction from size-exclusion chromatography corresponding to (A) as imaged by Coomassie Brilliant Blue. The upper and lower bands correspond to EmbB and EmbA, respectively.

1070 **(C)** Sample from (B) run on a Tricine gel and imaged by silver staining. The lower band at around 10 kDa (marked as a red star) was identified by mass spectrometry analysis as *Msm* AcpM as shown in (D).

(D) Mass spectrometry analysis identified the star labeled band in (C) as *Msm* AcpM (MSMEG_4326, ~10.7 kDa). The detected peptides of *Msm* AcpM are highlighted in

1075 different colors.

(E-H) Tryptic digestion mass spectrometry of purified *Mtb* EmbA(**E**)-EmbB(**F**) complex and *Msm* EmbA(**G**)-EmbB(**H**) complex. The detected peptides are highlighted in different colors. Darker shades indicate more overlaps of the peptides detected.

1080 **(I)** Mass spectrometry analysis of solvent extracted DPA (up) from purified *Msm* EmbA-EmbB and DP (down) from purified *Mtb* EmbA-EmbB treated with ethambutol.

(J) (up) The designed reaction scheme illustrating $\alpha(1\rightarrow3)$ arabinosyltransferase (EmbA-EmbB) and $\beta(1\rightarrow2)$ arabinosyltransferase (AftB) activity assays (12, 13, 21).

1085 (down) Radiometric-TLC analysis of arabinosyltransferase activity using NV1. Radiolabeled reaction products, containing DP[^{14}C]A and the product NV10 catalyzed by EmbA-EmbB were resolved by TLC as shown, and NV11 catalyzed firstly by EmbA-EmbB, and then by AftB (resistant to ethambutol and sourced from of *Msm* membranes).

1090 **(K)** 2D-HSQC NMR spectra of purified cell wall AG preparations from (left) wild type *Msm* (*Msm*-WT), (middle) *Msm* $\Delta embA$ and (right) *Msm* $\Delta embB$. Both *embA* and *embB* knockouts lack the 3-arm branching at the terminus of AG (15).

Fig. S2. Characterization of EmbC₂-AcpM₂ complex.

(A) Coomassie SDS-PAGE of the purified EmbC₂ samples from *Msm*, *Mtb*,
1100 *Mycobacterium marinum* (*Mm*) and *Mycobacterium xenopi* (*Mx*) as imaged by
Coomassie Brilliant Blue. The band located between 100 kDa and 130 kDa for each
lane corresponds to the EmbC protein.

(B) SDS-PAGE of the same samples in (A) shows an additional band (labelled with a
star) between 10 kDa and 15 kDa, imaged by silver staining. This was identified as
1105 *Msm* AcpM by mass spectrometry analysis shown in (J).

(C) SDS-PAGE of *Msm* EmbC mutants disrupting salt bridges with *Msm* AcpM as
imaged by Coomassie Brilliant Blue.

(D) SDS-PAGE of the same samples in (C) and imaged by silver staining shows
association between *Msm* EmbC and *Msm* AcpM is preserved for these mutants. The
1110 band corresponding to *Msm* AcpM is labelled with a red star.

(E) Size-exclusion chromatography of *Msm* EmbC₂-AcpM₂ purified with DDM
detergent by a Superose 6 gel filtration column (GE healthcare).

(F) Blue Native PAGE (BN-PAGE, analysis shows a band between marker 242 kDa
and 480 kDa, suggesting that the *Msm* EmbC-AcpM complex is wrapped in detergent
1115 and exists as an oligomer rather than a monomer in solution.

(G) (up) The designed reaction scheme illustrating an $\alpha(1\rightarrow5)$ arabinosyltransferase
activity assay to characterize EmbC₂ activity. (down) The thin-layer chromatography
(TLC) results show that *Msm* EmbC₂ catalyzes arabinose transfer from DP[¹⁴C]A to
Ara₂OC8 (NV6), forming an $\alpha(1\rightarrow5)$ -arabinofuranosyl linkage to produce
1120 [¹⁴C]AraAra₂OC8 (NV4), which co-migrated with a NV4 synthetic standard (R_F =
0.49). The synthesis of NV4 was inhibited by ethambutol. Whilst, the
ethambutol-resistant AftB from a source of *Msm* membranes, which catalyzes a
 $\beta(1\rightarrow2)$ -arabinofuranosyl linkage, produced a slower-migrating [¹⁴C]AraAra₂OC8
(NV9) product (R_F = 0.45) (12, 14) by TLC. R_F , retention factor. The silica gel plates

1125 on the left and right were exposed by autoradiography for 7 days and 3 days, respectively.

(H) MST curve of the binding affinity between *Msm* EmbC₂ and Ara₂OC8. Data presented are mean values +SD calculated from three independent experiments.

(I) MST curve of binding affinity between *Msm* EmbC₂ and ethambutol. Data
1130 presented are mean values +SD calculated from three independent experiments.

(J) Mass spectrometry analysis identified the star labeled band from *Msm* EmbC sample in (B) as *Msm* AcpM (MSMEG_4326, ~10.7 kDa), which contains at least the *apo* form (no covalent modification on Ser41).

A **B** **C** **D** **E** **F** **G**

Fig. S3. Cryo-EM data processing and validation of *Mtb* EmbA-EmbB-AcpM₂ in complex with ethambutol.

1140

(A) Flow chart for the processing of cryo-EM data. The density observed around the TM region of the final model is the signal of detergent that wraps around the membrane protein.

1145

(B) Representative electron micrograph.

(C) Selected reference-free 2D class averages.

(D) Gold-standard fourier correlation curves of 3D reconstructions.

(E) Posterior precision directional distributions of all particles used in the final 3D reconstruction reported by cryoSPARC.

1150

(F) Model to map fourier correlation curves reported by PHENIX.

(G) The density map colored according to the local resolution estimation using cryoSPARC.

A

B

C

D

E

F

G

Fig. S4. Cryo-EM data processing and validation of *Msm* EmbA-EmbB-AcpM₂ in complex with ethambutol.

- (A) Flow chart for the processing of cryo-EM data. The density observed around the
1160 TM region of the final model is the signal of detergent that wraps around the
membrane protein.
- (B) Representative electron micrograph.
- (C) Selected reference-free 2D class averages.
- (D) Gold-standard fourier correlation curves of 3D reconstructions.
- 1165 (E) Posterior precision directional distributions of all particles used in the final 3D
reconstruction reported by cryoSPARC.
- (F) Model to map fourier correlation curves reported by PHENIX.
- (G) The density map colored according to the local resolution estimation using
cryoSPARC.
- 1170

A **B** **C** **D** **E** **F** **G**

Fig. S5. Cryo-EM data processing and validation of *Msm* EmbA-EmbB-AcpM₂ in complex with di-arabinose.

1175

(A) Flow chart for the processing of cryo-EM data. The density observed around the TM region of the final model is the signal of detergent that wraps around the membrane protein.

(B) Representative electron micrograph.

1180 (C) Selected reference-free 2D class averages.

(D) Gold-standard fourier correlation curves of 3D reconstructions.

(E) Posterior precision directional distributions of all particles used in the final 3D reconstruction reported by cryoSPARC.

(F) Model to map fourier correlation curves reported by PHENIX.

1185 (G) The density map colored according to the local resolution estimation using cryoSPARC.

A **B** **C** **D** **E** **F** **G**

1190 **Fig. S6. Cryo-EM data processing and validation of *Msm* EmbC₂-AcpM₂ in complex with ethambutol.**

(A) Flow chart for the processing of cryo-EM data.

(B) Representative electron micrograph.

1195 (C) Selected reference-free 2D class averages.

(D) Gold-standard fourier correlation curves of 3D reconstructions.

(E) Posterior precision directional distributions of all particles used in the final 3D reconstruction reported by cryoSPARC.

(F) Model to map fourier correlation curves reported by PHENIX.

1200 (G) The density map colored according to the local resolution estimation using cryoSPARC.

Fig. S7. Example regions of cryo-EM maps of EmbA-EmbB complex and EmbC₂ complex.

(A-B) The cryo-EM map (threshold 0.4) of TM1-5 and EH4 helix in dimer interface from the EmbA protomer of *Msm* EmbA-EmbB-AcpM₂ in complex with ethambutol.

(C-D) The cryo-EM map (threshold 0.4) of TM9-13 and EH4 helix in dimer interface from the EmbB protomer of *Mtb* EmbA-EmbB-AcpM₂ in complex with ethambutol.

(E-F) Local resolution of cryo-EM map densities of DPA *Msm* in EmbA, DP in *Mtb* EmbB. The C1~C25 atoms of DPA and C1~C9 atoms of DP are missing in the cryo-EM maps possibly due to flexibility.

(G) The cryo-EM map (threshold 0.2) of cardiolipins in dimer interface *Msm* EmbA-EmbB-AcpM₂ in complex with ethambutol.

(H-I) The cryo-EM map (threshold 0.3) of TM1-5, EH2 helix and ethambutol binding site in *Msm* EmbC₂-AcpM₂ in complex with ethambutol.

1220

Fig. S8. Electron densities for the proteins and ligands from crystal structure of *Msm* EmbC₂-AcpM₂.

1225

(A) The $2F_o - F_c$ electron density (light blue mesh, contoured at 1σ) for the structure of *Msm* EmbC (dark blue ribbon) complexed with *Msm* AcpM (magenta ribbon). The density of AcpM is highlighted in the zoom-in insert.

1230

(B) The $2F_o - F_c$ annealing omit density map (light blue mesh, contoured at 1σ) and $F_o - F_c$ annealing omit density map (light green mesh, contoured at 3σ) for Ara₂, P_i and maltose in the active site. Ara₂, P_i, maltose and Asp279 of *Msm* EmbC are shown as sticks.

(C) The molecular structures of Ara₂OC8 and detergent DDM. The Ara₂ and maltose in the active site in (B) correspond to the boxed regions of Ara₂OC8 and DDM.

1235

Fig. S9. Structural domains of Emb proteins represented by the crystal structure of *Msm* EmbC.

(A) Structural superposition of the two EmbC-AcpM halves of the *Msm* EmbC₂-AcpM₂ complex.

(B) Structural alignment between *Msm* AcpM in our structure (cyan) and *Mtb* AcpM (PDB code 1KLP) (yellow). N/C, N/C terminus.

1245 (C) The coordination of Ca²⁺ in the PC domain of *Msm* EmbC. Ca²⁺ and interacting residues are shown as a sphere and sticks, respectively. The 2F_o-F_c annealing omit density map (contoured at 1σ) and F_o-F_c annealing omit density map (contoured at 3σ) are shown as blue and green meshes, respectively.

(D) The structure of the PN domain.

1250 (E) The structure of the PC domain. The red dashed line represents the missing flexible region of residue 780-810.

(F) Structure superposition of the PC domain of *Msm* EmbC (magenta) and the C-terminal domain of *Mtb* EmbC (PDB code 3PTY) (grey).

1255 (G) Arrangement of the 15 TM helices in *Msm* EmbC shown as a slice through the TM domain. The distances between TM2 and TM3, and between TM9 and TM10 are indicated.

(H) Detailed view of the active site of *Msm* EmbC. The structural elements (PL2-6, α6 and Trp965) composing the active site are shown in different colors. Ara₂, P_i and maltose are shown as ball and sticks. Trp965 and Asp279 are shown as sticks.

1260

Fig. S10. Structural comparison for Emb proteins.

1265 **(A)** Dimerization interfaces of *Mtb* EmbA-EmbB. The dimer interfaces (dashed boxes) between EmbA and EmbB are located near the periplasmic and cytoplasmic sides of the TM region. The lower left and lower right zoomed-in inlets show details of the dimer interactions. Residues participating in dimer formation are shown as sticks. The upper right zoomed-in inlet shows cardiolipins (CDL, yellow) in the dimer interface
1270 viewed from EmbA side. cryo-EM map densities (threshold 0.2) of cardiolipins are shown as yellow meshes.

(B) Dimerization interfaces of *Msm* EmbC₂. The dimer interfaces (dashed boxes) of EmbC₂ are located near the periplasmic and cytoplasmic sides of the TM region. The cavity between the two TM regions is shown as a light orange surface. Zoomed-in
1275 inlets show details of the dimer interactions. The two EH4 helices are anti-parallel to each other. Residues participating in dimer formation are shown as sticks. The hydrogen bond between the two serine residues is displayed as a dashed line.

(C) Superposition of EmbA, EmbB and EmbC structures, represented by the cryo-EM structures of *Mtb* EmbA, *Msm* EmbB and *Msm* EmbC.

1280

Fig. S11. The interface between Emb proteins and AcpM.

1285

(A) *Msm* AcpM (purple) binds to the cytoplasmic face of *Msm* EmbC (grey). The CL1, CL2, CL3 and CL5 of *Msm* EmbC at the interface are highlighted in green. The

modification site Ser41_{AcpM} and the interacting residue Arg247_{EmbC} are shown in sticks.

1290 **(B)** The salt bridges between *Msm* EmbC and *Msm* AcpM. Interacting residues are shown as stick models.

(C) (up) Electrostatic surface representation of *Msm* EmbC from the cytoplasmic view. Interacting segments of *Msm* AcpM are shown as purple tubes with the side-chains as thin stick models. The dashed circle indicates a positively charged area
1295 interacting with *Msm* AcpM. (down) Electrostatic surface representation of *Msm* AcpM. Interacting segments with *Msm* EmbC are shown as green tubes and side-chains as thin stick models. The dashed circle indicates a negatively charged area interacting with *Msm* EmbC.

(D) Zoom-in view of the interaction between Ser41_{AcpM} and Arg247_{EmbC} in the crystal
1300 structure of the *Msm* EmbC₂-AcpM₂ complex. The side chain of Ser41_{AcpM} and main chain of Arg247_{EmbC} are shown as sticks. The dashed line indicates the hydrogen bond.

(E) Superposition between crystal structure of *Msm* EmbC₂-AcpM₂ (grey) in complex with di-arabinose and cryo-EM structure of *Msm* EmbC₂-AcpM₂ in complex (colored)
1305 with ethambutol at the EmbC-AcpM interface. In the cryo-EM structure, Ser41_{AcpM} is modified by the Ppant group, which inserts into TM region close to TM6-7. CL1 of EmbC shifts away from TM domain and encircles this group. Density for the Ppant group from the cryo-EM map (threshold 0.2) is shown in blue mesh.

(F) Effect of mutated residues in the interface between EmbC and AcpM on
1310 arabinosyltransferase activity of *Msm* EmbC₂. Data presented are means +SD calculated from three independent experiments.

(G) Effect of mutated residues of *Msm* EmbC in the interface interface between EmbC and AcpM on LAM synthesis. *embCΔ*, the *Msm embC* knock-out mutant; *embC* / *embC_R243A et al.*, the *Msm embC* knock-out mutant complemented with
1315 plasmid carrying *embC* wild-type / R243A *et al.* mutant alleles.

(H) The association between *Msm* AcpM and *Msm* EmbB. *Msm* AcpM interacts with CLs of *Msm* EmbB. A covalent lipid modification of a 4' phosphopantetheine (Ppant) moiety (sphere) on *Msm* AcpM also contributes to the association.

(I) Zoom-in view from (H) shows the *Msm* EmbB and *Msm* AcpM interactions, three pairs of salt bridges between are shown as dashed lines.

.

1325

Fig. S12. Structural comparison of Emb proteins (represented by EmbC) with other glycosyltransferases in the GT-C family.

1330 (up) Cartoon representations of Emb proteins represented by *Msm* EmbC, PglB (PDB code 3RCE), AglB (PDB code 3WAJ), ArnT (PDB code 5F15), STT3 (PDB code 6EZN), ALG6 (PDB code 6P25) and PMT1 (PDB code 6SNH). For each structure, the transmembrane region is colored in rainbow style with the N- and C-termini colored blue and red, respectively; the periplasmic region is colored in magenta.

1335 Crossed helices, resembling EH1 and EH2 in EmbC and bearing the conserved catalytically relevant D[N]D[E]x motif, is marked as a red circle. (down) Arrangement of the TM helices for the structures is shown as a slice through the TM domain. 11 TM helices inside the dashed circles have a common arrangement. TM9 (grey) of STT3 is missing in the model.

A

```

1      10      20      30      40      50
M.tub_EmbB ...MTQCASRRKSTPNRAILGAFASARGTRWATIAGLCFVLSVATPLLPVVTTAMLD
M.smg_EmbB ...MSGNMDEAVSGNMDEAVSAGKDVRIARWATIAGLCFVLSVSIPLLPVVTTATLND
M.bov_EmbB ...MTQCASRRKSTPSRAILGAFASARGTRWATIAGLCFVLSVATPLLPVVTTAMLD
M.mar_EmbB .....MSVTNETEQDTAATTASAREVRVTWATIAGLCFVLSVATPLLPVVTTAMLN
M.lep_EmbB .....MSVIYRAHRVAIANRTASARNVWATIAGLCFVSSVVTPLLPVVTTATLN
M.tub_EmbA .....VPHDGNERSHRIARLAAVVSGIACLLLCGIVPLLPVNQTTATIF
M.smg_EmbA .....MTEPSRIARLAAVVSGIACLLLCGIVPLLPVVTTTATVL
M.bov_EmbA .....MPHDGNERSHRIARLAAVVSGIACLLLCGIVPLLPVNQTTATIF
M.mar_EmbA .....VSPDGNERRIARLAAVVSGIACLLLCGIVPLLPVVTTTATIL
M.tub_EmbA .....VPHDGHEPPRIIRLTAVGAITGLLLCAVVPLPVVTTTATIR
M.lep_EmbA MATEAAPPRIAVRLPSTSVRDAGANYRIARVAVAGLICAVLAIATPLPVNQTTAQLN
M.smg_EmbA .....MTGPHAAGGSNHRIARLVAIAGLICTLMAIATTPLPVVTTTAELN
M.bov_EmbA MATEAAPPRIAVRLPSTSVRDAGANYRIARVAVAGLICAVLAIATPLPVNQTTAQLN
M.mar_EmbA .MATEAPGALEQLPSTSVSDSGAKYRIARLVAVVTGLCTLLALATPLPVVTTTAKLN
M.lep_EmbC .....VSGAGANYRIARLVAVIAGLCALLAMATPLPVNQNTAQLN

```

```

60      70      80      90      100
M.tub_EmbB WPQ...RGLLGSVTAPLISLTPVDFTATVPCDVVRAMPPAGG...VVLGTAPKQGDA
M.smg_EmbB WPQ...QRLDNVTAPLISQAPLELTATVPCSVVRDLPPEGG...LVFGTAPAEGRDA
M.bov_EmbB WPQ...RGLLGSVTAPLISLTPVDFTATVPCDVVRAMPPAGG...VVLGTAPKQGDA
M.mar_EmbB WPQ...NGLLNSVTAPLISLTPVNLTASVPCSVVRDMPAKGG...VVLGTAPKQGDA
M.lep_EmbB WPQ...NGLLNSVTAPLISLTPVDITATVPCAVVAALPPSGG...VVLGTAPKQGDA
M.tub_EmbA WPQGSTADGNITQITAPLVSGAPRALDISIPCSAIATLPANGG...LVLSTTPAGGVDT
M.smg_EmbA WPQGVGADGNVTELTAPLVAGAPRALDVTIPCRAVAELPADGG...VVFSTTPAGGIEA
M.bov_EmbA WPQGSTADGNITQITAPLVSGAPRALDISIPCSAIATLPANGG...LVLSTTPAGGVDT
M.mar_EmbA WPQGASADGDISQITAPLVSGAPRALDISIPCSAIATLPATGG...LVLSTTPAGGVDT
M.lep_EmbA WPQSATRDGWVTQITAPLVSGTPRALDISIPCSAMATLPDSVG...LVVSTTPSGGVDT
M.tub_EmbB WPQ...NGTFASVEAPLIGYVATDLNITVPCAAAGLAGSQNTGKTVLLSTTPKQAPKA
M.smg_EmbC WPQ...NGVWQSVDAPLIGYVATDLTITVPCAAAGLVGPENRNRSVLLSTTPKQAPKA
M.bov_EmbC WPQ...NGTFASVEAPLIGYVATDLNITVPCAAAGLAGSQNTGKTVLLSTTPKQAPKA
M.mar_EmbC WPQ...NGTFASVEAPLIGYVATDLDITVPCAAAGLAGFQNKGKTVLLSTTPKQAPNA
M.lep_EmbC WPQ...NSTFESVEAPLIGYVATGLNITVPCAAAGLTGFQSAGQTVLLSTTPKQAPKA

```

```

110      120      130      140      150      160
M.tub_EmbB NLQALFVVVSAQRVDVTDRNVILSVPREQVT...SPQCQRIEVTSTHAGTFANFVGLK
M.smg_EmbB ALNALFVVVSTETRVDVIVNRNVVASVNRDVAG...PDCQRIEITSNLDGTYADFVGLT
M.bov_EmbB NLQALFVVVSAQRVDVTDRNVILSVPREQVT...SPQCQRIEVTSTHAGTFANFVGLK
M.mar_EmbB NLQALFVVVNSKRVNVTDRNVILSVPREQVD...SPQCQRIEITSSTHAGTFATFVGLK
M.lep_EmbB NLNALFIDVNSQRVDVTDRNVILSVPRNQVAGDAGAPCSIEVTSTHAGTFATFVGLT
M.tub_EmbA GKAGLFVVRANQDTVVVAERDSVAAVAARSTIAG...GCSALHIWADTGGAGADFVGIP
M.smg_EmbA GRNGMFIRANADVVVAERDTVAAVAPREAVDSG...ACSETHVWADVSAVAGADFVGIP
M.bov_EmbA GKAGLFVVRANQDTVVVAERDSVAAVAARSTIAG...GCSALHIWADTGGAGADFVGIP
M.mar_EmbA GKNGLFVVRADKDSVVVAERDTVAAVAPRAAIEG...RCSALHIWADATGADFVGIP
M.lep_EmbA GKSGLFVVRANKNAVVVAERDSVAAVAPRAPAAG...NCSVLHIWANTRGAGANFVGIP
M.tub_EmbC VDRGLLQRANDLVLVVRNVPLVTAPLSQVLG...PTCQRLTFTAHADVAAEFVGLV
M.smg_EmbC IDRGLLIERINNDLTVIVRNTPVVSAPLEQVLS...PDCRYLTFTAHADKVIGEFVGLT
M.bov_EmbC VDRGLLQRANDLVLVVRNVPLVTAPLSQVLG...PTCQRLTFTAHADRVAAEFVGLV
M.mar_EmbC VDRGLLQRANDLVLVVRNVPVVSAPLSQVLS...PACQRLTFTAHAESVTAEFVGLK
M.lep_EmbC VDRGLLQRANDLVLVVRNVPVVSAPMSQVLS...PACQRLTFAAYFDKITAEFVGLT

```

```

170      180      190      200      210      220
M.tub_EmbB DPSG...APLRSGFPDPNLRQIVGVFTDLTGPAPPGLAVSATIDTRFSTRPTTLKLL
M.smg_EmbB QISGEDAGKLQRTGYPDPNLRQIVGVFTDLTGPAPPGLAVSATIDTRFSTRPTTLKLL
M.bov_EmbB DPSG...APLRSGFPDPNLRQIVGVFTDLTGPAPPGLAVSATIDTRFSTRPTTLKLL
M.mar_EmbB DPSG...APLRSGYPDPNLRQIVGVFTDLTGPAPDGLRLSATIDTRFSTRPTTLKLL
M.lep_EmbA DSAG...NPLRGGFPDPNLRQIVGVFTDLTGGAPSGLRLSATIDTRFSSSPTTLKRF
M.tub_EmbA GGAG...TLPPEKKQOVGGCHFTDLKVGAQPGLSARDIDTRFIITPGAIKKA
M.smg_EmbA DASG...TLPVDKKQOVSGCHFTDLKVPAQPGLSARDIDTRFIITSPTLKKTA
M.bov_EmbA GGAG...TLPPEKKQOVGGCHFTDLKVGAQPGLSARDIDTRFIITPGAIKKA
M.mar_EmbA GGAG...TLPPEKKQOVGGCHFTDLKVQAQPGLSARDIDTRFIITAPTVKKTI
M.lep_EmbA GGAG...ILTAEKKQOVGGCHFTDLKVPAQPGLSARDIDTRFIITAPTAKKII
M.tub_EmbC QGPNAEHPGAPLRGERSGYDFRQOVICGVFTDLAGPAPPGLSFSASVDTRYSSPTPKKMA
M.smg_EmbC QGPDDDDPGEAVRGERSGYDFRQOVICGVFTDLSGPAPEGLLSSATIDTRYSSSPTLLKLL
M.bov_EmbC QGPNAEHPGAPLRGERSGYDFRQOVICGVFTDLAGPAPPGLSFSASVDTRYSSPTPKKMA
M.mar_EmbC QGPNAEHPGEPLRGERSGYDFRQOVICGVFTDLTGTPPPGLSFSATIDTRYSSSPTPKMA
M.lep_EmbC YGPNAEHPGVPLGERSGYDFRQOVICGVFTDLSGPTPTGLNFSATIDTRYSSSPTLLKTI

```

```

230      240      250      260      270
M.tub_EmbB ATIGAIVATVALIALWRLDQLDGRGS...IAQLLLRPFRPASSPGGMRLIPASWRFTTL
M.smg_EmbB AMLLAIVSTVIALIALWRLDLDGR...IAQLLLRPFRPASSPGGMRLIPASWRFTTL
M.bov_EmbB ATIGAIVATVALIALWRLDQLDGRGS...IAQLLLRPFRPASSPGGMRLIPASWRFTTL
M.mar_EmbB ATIGAILATVALIALWRLDLDGR...IAQLLLRPFRPASSPGGMRLIPASWRFTTL
M.lep_EmbB AMMLAITITVALIALWRLDQLDGR...IAQLLLRPFRPASSPGGMRLIPASWRFTTL
M.tub_EmbA VMLLGVLAVLVAMVGLALDRLSRG...RTLRDWLTRYRP...RVRVGFASRL
M.smg_EmbA VMLLGLACVIGSIVLALDRLGWRR...RTLRDWLTRYRP...RVRVGFASRL
M.bov_EmbA VMLLGVLAVLVAMVGLALDRLSRG...RTLRDWLTRYRP...RVRVGFASRL
M.mar_EmbA AMILGAIAVLGAIIVLALDRLSRGGDALRDWRSPAWLSRYRPRLPLSRWVRVGFAWL
M.lep_EmbA AVGVGAAAVLIAILALSALDRRNRNGHRLINVRSMAWLAQWRVILATP...PRAGGSASI
M.tub_EmbC AMILGVALTGAALVALHLDTADGMR...HRRFLPARWWSIG
M.smg_EmbC AMIVGVAMTVIALGALHLDTADGRR...HRRFLPARWWSMTP
M.bov_EmbC AMILGVALTGAALVALHLDTADGMR...HRRFLPARWWSIG
M.mar_EmbC AMIIGLLSTAVALVALHLDTADGTR...HRRILPARWWSIGA
M.lep_EmbC AMILGVVLTIVALVALHLDTADGTQ...HRRLLPARWWSIGC

```

+

DDx

	280	290	300	310	320	330
M.tub_EmbB	TDAVVF	FGFLWHV	IGANSSDDGYILGMARVAD	HAGYMSNYRWFGSP	EDPFGWYVNNLLA	
M.smg_EmbB	VDGVVV	GGMAIYWV	IGANSSDDGYILQMARTAE	HAGYMANYSRWFGSP	EDPFGWYVNNVLA	
M.bov_EmbB	TDAVVF	FGFLWHV	IGANSSDDGYILGMARVAD	HAGYMSNYRWFGSP	EDPFGWYVNNLLA	
M.mar_EmbB	VDAVVF	FGFLWHV	IGANSSDDGYILGMARVAD	HAGYMSNYRWFGSP	EDPFGWYVNNLLA	
M.lep_EmbB	VDVAVF	FGFLWHV	IGANSSDDGYQMAMARTADHS	GYMANYSRWFGSP	EDPFGWYVNNLLA	
M.tub_EmbA	ADAAVI	ATLLWHV	IGATSSDDGYLLTVARVAPKA	GYVANYRYFGTTEAP	FDWYTSVLA	
M.smg_EmbA	TDTGVV	IGLLIWHV	IGAPSSDDGYNMTIARVASEA	GYTNYRYFGASEAP	FDWYQSVLS	
M.bov_EmbA	ADAAVI	ATLLWHV	IGATSSDDGYLLTVARVAPKA	GYVANYRYFGTTEAP	FDWYTSVLA	
M.mar_EmbA	TDAAVI	ATLLWHV	IGALSSDDGYNLTARIAPQA	GYVANYRYFGTTEAP	FDWYHELLA	
M.lep_EmbA	ADGGVI	ATLLWHV	IGATSSDDGYNLTARVSSSEA	GYLANYSRYFGATEAP	FDWYFTVLA	
M.tub_EmbC	LDTVVI	AVLVVWHF	VGANTSDDGYILTMARVSEHA	GYMANYSRWFGTEAP	PFGWYVDLLA	
M.smg_EmbC	LDGIVS	AMLVVWHF	VGANTADDDGYILTMARVSEHA	GYMANYSRWFGTEAP	PFGWYVDLLA	
M.bov_EmbC	LDTVVI	AVLVVWHF	VGANTSDDGYILTMARVSEHA	GYMANYSRWFGTEAP	PFGWYVDLLA	
M.mar_EmbC	LDGIVT	AVLVVWHF	VGANTSDDGYILTMARVSEHA	GYMANYSRWFGTEAP	PFGWYVDLLA	
M.lep_EmbC	LDGIVT	ITLAWWHF	VGANTSDDGYILTMARVSEHA	GYMANYSRWFGTEAP	PFGWYVDLLA

	340	350	360	370	380	390
M.tub_EmbB	LMTHVS	DASLWRR	FDLAAAGLVCWLLLSRE	VLPRI	...P	AVV
M.smg_EmbB	LMTKVS	DASLWRR	FDLAAAGLVCWLLLSRE	VLPRI	...P	AVV
M.bov_EmbB	LMTHVS	DASLWRR	FDLAAAGLVCWLLLSRE	VLPRI	...P	AVV
M.mar_EmbB	LMTHVS	DASLWRR	FDLAAAGLVCWLLLSRE	VLPRI	...P	AVV
M.lep_EmbB	LMTHVS	DASLWRR	FDLAAAGLVCWLLLSRE	VLPRI	...P	AVV
M.tub_EmbA	QLAAVS	TAGVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.smg_EmbA	HLAAS	TAGVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.bov_EmbA	QLAAVS	TAGVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.mar_EmbA	HLAAS	TAGVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.lep_EmbA	KLAAVS	TAGVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.tub_EmbC	LWAAVS	TASVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.smg_EmbC	LWAAVS	TASVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.bov_EmbC	LWAAVS	TASVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.mar_EmbC	LWAAVS	TASVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS
M.lep_EmbC	LWAAVS	TASVWWR	LPATLAGIACWLLVSRE	VLRRI	CGPGG	LAS

	400	410	420	430	440	450
M.tub_EmbB	MPFNNGLRPE	GI	I	ALGSLVTY	V	L
M.smg_EmbB	MPFNNGLRPE	GI	I	ALGSLVTY	V	L
M.bov_EmbB	MPFNNGLRPE	GI	I	ALGSLVTY	V	L
M.mar_EmbB	MPFNNGLRPE	GI	I	ALGSLVTY	V	L
M.lep_EmbB	MPFNNGLRPE	GI	I	ALGSLVTY	V	L
M.tub_EmbA	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.smg_EmbA	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.bov_EmbA	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.mar_EmbA	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.tub_EmbC	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.smg_EmbC	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.bov_EmbC	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.mar_EmbC	IPFNNGLRPE	PI	I	ALGSLVTY	V	L
M.lep_EmbC	IPFNNGLRPE	PI	I	ALGSLVTY	V	L

	460	470	480	490	500	510
M.tub_EmbB	VAGGRP	MLRILVRRHRLV	GT	P	LVSPMLAAG	TVILT
M.smg_EmbB	VAGGRP	MLRILVRRHRLV	GT	P	LVSPMLAAG	TVILT
M.bov_EmbB	VAGGRP	MLRILVRRHRLV	GT	P	LVSPMLAAG	TVILT
M.mar_EmbB	VAGGRP	MLRILVRRHRLV	GT	P	LVSPMLAAG	TVILT
M.lep_EmbB	VAGGRP	MLRILVRRHRLV	GT	P	LVSPMLAAG	TVILT
M.tub_EmbA	LTGARA	IAQIRRRRRATD	GL	LAP	LAVLAAAS	SLITV
M.smg_EmbA	LVGARA	IGRVVTTARRAGT	GL	LAP	LAVLAAAS	SLITV
M.bov_EmbA	LTGARA	IAQIRRRRRATD	GL	LAP	LAVLAAAS	SLITV
M.mar_EmbA	LTGARV	IAQIRRRRRATD	GL	LAP	LAVLAAAS	SLITV
M.tub_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.smg_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.bov_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.mar_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.lep_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV

	520	530	540	550	560	570
M.tub_EmbB	QAWYT	ENRYY	Y	L	I	P
M.smg_EmbB	QAWYT	ENRYY	Y	L	I	P
M.bov_EmbB	QAWYT	ENRYY	Y	L	I	P
M.mar_EmbB	QAWYT	ENRYY	Y	L	I	P
M.lep_EmbB	QAWYT	ENRYY	Y	L	I	P
M.tub_EmbA	IAWYQ	DFRYY	F	ETVES	N	EGSMSTR
M.smg_EmbA	IPWYQ	DFRYY	F	ETVES	N	EGSMSTR
M.bov_EmbA	IAWYQ	DFRYY	F	ETVES	N	EGSMSTR
M.mar_EmbA	IAWYQ	DFRYY	F	ETVES	N	EGSMSTR
M.tub_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.smg_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.bov_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.mar_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV
M.lep_EmbC	LVAIGP	LRTIILHRRSRRF	GV	LAP	LAVLAAAS	SLITV

	580	590	600	610	620	630
M.tub_EmbB	MGVIFG	TMFFLM	FTPTKVVHH	FGLFAA	VGAAMA	ALLTTVL
M.smg_EmbB	MGIIFA	TMFFLM	FTPTKWIHH	FGLFAA	VGAAMA	ALLTTVL
M.bov_EmbB	MGVIFG	TMFFLM	FTPTKVVHH	FGLFAA	VGAAMA	ALLTTVL
M.mar_EmbB	MGVIFG	TMFFLM	FTPTKVVHH	FGLFAA	VGAAMA	ALLTTVL
M.lep_EmbB	IGTILG	TMFFLM	FTPTKVVHH	FGLFAA	VGAAMA	ALLTTVL
M.tub_EmbA	IGTTAV	GLLLLT	FTPTKVAIQ	FCAFA	GLAGVL	GAVTAFT
M.smg_EmbA	CGSTAI	GLLLLT	FTPTKVAIQ	FCAFA	GLAGVL	GAVTAFT
M.mar_EmbA	IGTTAV	GLLLLT	FTPTKVAIQ	FCAFA	GLAGVL	GAVTAFT
M.lep_EmbA	IGTTAT	SLLLLT	FTPTKVAIQ	FCAFA	GLAGVL	GAVTAFT
M.tub_EmbC	IGITII	ISFLAM	FTPTKWTTH	FCVFA	GLAGSL	GALA
M.smg_EmbC	IGITII	ISFLAM	FTPTKWTTH	FCVFA	GLAGSL	GALA
M.bov_embC	IGITII	ISFLAM	FTPTKWTTH	FCVFA	GLAGSL	GALA
M.mar_embC	IGITII	ISFLAM	FTPTKWTTH	FCVFA	GLAGSL	GALA
M.lep_embC	IGITV	ISFLAM	FTPTKWTTH	FCVFA	GLAGSL	GALA

	640	650	660	670	680	690
M.tub_EmbB	LLALCW	ATTNGW	VVSSYG	VPPNSAM	BKIDGIT	VSTIFFA
M.smg_EmbB	VLAFCF	ASTNGW	VVSSYGR	APFNNNS	VSKVGGV	QISAIFF
M.bov_EmbB	LLALCW	ATTNGW	VVSSYG	VPPNSAM	BKIDGIT	VSTIFFA
M.mar_EmbB	TLALCW	ATTNGW	VVSSYG	VPPNSAM	BKIDGIT	VSTIFFA
M.lep_EmbB	VMTLCA	ATTNGW	VVSSYG	VPPNSAM	BKIDGIT	VSTIFFA
M.tub_EmbA	VLAWAT	SINGWF	VVGNVGV	PPYDIO	QVIA	SHPVIT
M.smg_EmbA	VLAWAT	SINGWF	VVGNVGV	PPYDIO	QVIA	SHPVIT
M.bov_EmbA	VLAWAT	SINGWF	VVGNVGV	PPYDIO	QVIA	SHPVIT
M.mar_EmbA	VLAWAT	SINGWF	VVGNVGV	PPYDIO	QVIA	SHPVIT
M.tub_EmbC	VLALSF	AVNGW	VVSSNRC	VPWNSNF	FEKWS	LSLTALL
M.smg_EmbC	VTALSF	AVNGW	VVSSNRC	VPWNSNF	FEKWS	LSLTALL
M.bov_embC	VLALSF	AVNGW	VVSSNRC	VPWNSNF	FEKWS	LSLTALL
M.mar_embC	VMALSF	AVNGW	VVSSNRC	VPWNSNF	FEKWS	LSLTALL
M.lep_embC	VVALSF	AVNGW	VVSSNRC	VPWNSNF	FEKWS	LSLTALL

	700	710	720	730	740	
M.tub_EmbB	E....G	RLIRALT	..APV	PIVAGF	MAAVF	VASMV
M.smg_EmbB		RVVDR	LTAA	PIVAGF	MVVVM	MAAIG
M.bov_EmbB	E....G	RLIRALT	..APV	PIVAGF	MAAVF	VASMV
M.mar_EmbB	E....G	RLIRALT	..APV	PIVAGF	MAAVF	VASMV
M.lep_EmbB	E....G	RLIRALT	..APV	PIVAGF	MAAVF	VASMV
M.tub_EmbA	HT...E	VKDNRNR	RLAST	PLL	VAVIM	VAGEV
M.smg_EmbA	HT...E	VKDNRNR	RLAST	PLL	VAVIM	VAGEV
M.bov_EmbA	HT...E	VKDNRNR	RLAST	PLL	VAVIM	VAGEV
M.mar_EmbA	HT...E	VKDNRNR	RLAST	PLL	VAVIM	VAGEV
M.lep_EmbA	HT...E	VKDNRNR	RLAST	PLL	VAVIM	VAGEV
M.tub_EmbC	RRTAR	PTFRAR	LIVQS	PLA	IATW	LVLF
M.smg_EmbC		PDPR	QRWQL	LLVA	PLA	IATW
M.bov_embC	RRTAR	PTFRAR	LIVQS	PLA	IATW	LVLF
M.mar_embC		PDPR	QRWQL	LLVA	PLA	IATW
M.lep_embC		SAKIT	RFVRI	DIRV	QSLA	ITW

	750	760	770	780	790	800
M.tub_EmbB	.VGG.	GGLADD	VLEP	DTNAG	FMKPL	DGD
M.smg_EmbB	.AGG.	GGLADD	VLEP	DTNAG	FMKPL	DGD
M.bov_EmbB	.VGG.	GGLADD	VLEP	DTNAG	FMKPL	DGD
M.mar_EmbB	.TGG.	GGLADD	VLEP	DTNAG	FMKPL	DGD
M.lep_EmbB	.TGG.	GGLADD	VLEP	DTNAG	FMKPL	DGD
M.tub_EmbA	LS..S	CAMADD	VLEP	PDNAG	MLQ	VP
M.smg_EmbA	GDS..	CAMADD	VLEP	PDNAG	MLQ	VP
M.bov_EmbA	LS..S	CAMADD	VLEP	PDNAG	MLQ	VP
M.mar_EmbA	LS..S	CAMADD	VLEP	PDNAG	MLQ	VP
M.lep_EmbA	LAPSV	CAMADD	VLEP	PDNAG	MLQ	VP
M.tub_EmbC	.AGKT	GGLAED	VLEP	LDNAG	MLP	VTAP
M.smg_EmbC	.TGKT	GGLAED	VLEP	LDNAG	MLP	VTAP
M.bov_embC	.AGKT	GGLAED	VLEP	LDNAG	MLP	VTAP
M.mar_embC	.TGKS	GGLAED	VLEP	LDNAG	MLP	VTAP
M.lep_embC	.TGQT	GGLAED	VLEP	LDNAG	MLP	VTAP

	810	820	830	840
M.tub_EmbB	AIVMK	ENQ	PGTDY	DWDAPT
M.smg_EmbB	AIRLNL	NQ	PGTDY	DWNRIK
M.bov_EmbB	AIVMK	ENQ	PGTDY	DWDAPT
M.mar_EmbB	AMVMK	ENQ	PGTDY	DWDQV
M.lep_EmbB	AIRIT	ENQ	PGTDY	DWEAPT
M.tub_EmbA	PVVSK	PGLV	NSDASP	NKPNA
M.smg_EmbA	PVAAN	P	TPNSD	GPVDK
M.bov_EmbA	PVVSK	PGLV	NSDASP	NKPNA
M.mar_EmbA	PVVSK	PGLV	NSDASP	NKPNA
M.lep_EmbA	PVVSK	PGLV	NSDASP	NKPNT
M.tub_EmbC	PVMER	P	DRSFLN	.DDGLIT
M.smg_EmbC	PVMEQ	P	TDNFAD	SDSGV
M.bov_embC	PVMER	P	DRSFLN	.DDGLIT
M.mar_embC	PVMER	P	DRSFVN	.DDGLVT
M.lep_embC	PVMEP	P	DRSFVK	.ENGMTT

850 860 870 880 890 900

M.tub_EmbB TGAQQQSTLVSAYWLPKPDD...GHPLVVVTAAGKIAGNSVLHGYPGTVVVEYAMP
M.smg_EmbB TEAQQESRLSSAYWLPARDETERAAHPLVVVTAAGTITGESVANGLTGTQVVDLEYATR
M.bov_EmbB TGAQQQSTLVSAYWLPKPDD...GHPLVVVTAAGKIAGNSVLHGYPGTVVVEYAMP
M.mar_EmbB TGSQQQSKLVSAYWLPKPDD...GHPLVVVTAAGKIAGNSVLHGYPGTVVVEYAMP
M.lep_EmbB TGAQQQSRLLVSAYWLPKPDD...RHPLVVVTAAGKITGNSVLHGHTYGTVVVEYADP
M.tub_EmbA EN.NLAATAISAYWOLP...PRSPDRPLVVVSAAGAIWYKEDGDFIYGOSLKLQWGV
M.smg_EmbA EN.KLAATAISAYWOLP...PRTPDRPLVIVAAAGAIWYKEDGDFIYGOSLKLQWGV
M.bov_EmbA EN.NLAATAISAYWOLP...PRSPDRPLVVVSAAGAIWYKEDGDFIYGOSLKLQWGV
M.mar_EmbA EN.SLAATAISAYWOLPGDWKANIAGPLVVVTAAGAIWYKEDGDFIYGOSLKLQWGV
M.lep_EmbA EN.TLAASATISAYWOLPLHWKESIADPLVVVSAAGAIWYKEDGDFIYGOSLKLQWGV
M.tub_EmbC AGVQVPAMLRSGWYRLPTNEQRDR..APLLVVVTAAGRFD.....SREVRVQWATD
M.smg_EmbC SGTQQPAMLRSGWYRLPDRDQAG...PLLVVSAAGRFD.....QGEVVEVQWATD
M.bov_EmbC AGVQVPAMLRSGWYRLPTNEQRDR..APLLVVVTAAGRFD.....SREVRVQWATD
M.mar_EmbC SGIQVPAMLRSGWYRLPADQRKK..TPLLVVSAAGRFD.....PREVVEVQWATD
M.lep_EmbC SGIQVVARLRSGWYRLPARDKAG...PLLVVSAAGRFD.....HEVVKVQWATD

910 920 930 940 950 960

M.tub_EmbB GPGALVPAGRMVDPDLGYEQPKAWRNLRFAKAMPADAVAVRVVAEDLSLTPEDWIAVT
M.smg_EmbB GPDGTLVPAGRVTPYDVG..PTPSWRNLRYPARSEIPDADAVAVRVVAEDLSLTPEDWIAVT
M.bov_EmbB GPGALVPAGRMVDPDLGYEQPKAWRNLRFAKAMPADAVAVRVVAEDLSLTPEDWIAVT
M.mar_EmbB GPGALVPAGRMVDPDLGYEQPKAWRNLRFAKAMPADAVAVRVVAEDLSLTPEDWIAVT
M.lep_EmbB GPNGLVPAAGRLVDPDLGYEQPKAWRNLRFAKAMPADAVAVRVVAEDLSLTPEDWIAVT
M.tub_EmbA GPDGRIQPLGVVFPIDIG..PQPAWRNLRFPFLAWAPPANVARIVADDPNLSPEQWFAET
M.smg_EmbA RPDGTIYQALSEVQPIDIF..QQKAWRNLRFPFLAWAPPANVARIVADDPNLSPEQWFAET
M.bov_EmbA GPDGRIQPLGVVFPIDIG..PQPAWRNLRFPFLAWAPPANVARIVADDPNLSPEQWFAET
M.mar_EmbA RPDGTTQPLGVVFPIDIG..PQPAWRNLRFPFLAWAPPANVARIVADDPNLSPEQWFAET
M.lep_EmbA RPDGIIQPLQVMPIDIG..PQPAWRNLRFPFLAWAPPANVARIVADDPNLSPEQWFAET
M.tub_EmbC EQAAAGHHGSGMEFADVG..AAPAWRNLRAPLSAIPSTATQVRLVADDDLPAPQHWIAET
M.smg_EmbC EQAAANPEPGSGITFDVG..AAPAWRNLRAPLSAIPSTATQVRLVADDDLPAPQHWIAET
M.bov_EmbC EQAAAGHHGSGMEFADVG..AAPAWRNLRAPLSAIPSTATQVRLVADDDLPAPQHWIAET
M.mar_EmbC EQAAAGKHGSGMEFADVG..AAPAWRNLRAPLSAIPSTATQVRLVADDDLPAPQHWIAET
M.lep_EmbC SGAASGQPGSAFQFSQDVG..ASPWRNLRAPLSAIPSTATQVRLVADDDLPAPQHWIAET

970 980 990 1000 1010 1020

M.tub_EmbB PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.smg_EmbB PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.bov_EmbB PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.mar_EmbB PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.lep_EmbB PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.tub_EmbA PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.smg_EmbA PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.bov_EmbA PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.mar_EmbA PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.lep_EmbA PPRVPELRSIQEYVGSSTQPVLLDWAVGLAFPCQOEMLHANGTAETPKFRITPDYSAKKLD
M.tub_EmbC PPRIPRVRTIQNVYGAADPVFIDWLVLGLAFPCQOEFQGHQYGVDETPKWRITPDRFGAEAN
M.smg_EmbC PPRIPRVRTIQNVYGAADPVFIDWLVLGLAFPCQOEFQGHQYGVDETPKWRITPDRFGAEAN
M.bov_EmbC PPRIPRVRTIQNVYGAADPVFIDWLVLGLAFPCQOEFQGHQYGVDETPKWRITPDRFGAEAN
M.mar_EmbC PPRIPRVRTIQNVYGAADPVFIDWLVLGLAFPCQOEFQGHQYGVDETPKWRITPDRFGAEAN
M.lep_EmbC PPRIPRVRTIQNVYGAADPVFIDWLVLGLAFPCQOEFQGHQYGVDETPKWRITPDRFGAEAN

1030 1040 1050 1060 1070 1080

M.tub_EmbB TDIWEDGNTGCLLGITDILLRAHVMAITYLSRDWARDWGSLR.KFDTLV..DAPPQLELGL
M.smg_EmbB TDIWQDGINCLLGITDILLRAHVMAITYLSRDWARDWGSLR.KFDTLV..DAPPQLELGL
M.bov_EmbB TDIWEDGNTGCLLGITDILLRAHVMAITYLSRDWARDWGSLR.KFDTLV..DAPPQLELGL
M.mar_EmbB TDIWEDGNTGCLLGITDILLRAHVMAITYLSRDWARDWGSLR.KFDTLV..DAPPQLELGL
M.lep_EmbB TDIWEDGNTGCLLGITDILLRAHVMAITYLSRDWARDWGSLR.KFDTLV..DAPPQLELGL
M.tub_EmbA SNLWQSSSTGCPFLFTQALRLTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.smg_EmbA SNLWQSSSTGCPFLFTQALRLTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.bov_EmbA SNLWQSSSTGCPFLFTQALRLTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.mar_EmbA SNLWQSSSTGCPFLFTQALRLTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.lep_EmbA SNLWQSSSTGCPFLFTQALRLTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.tub_EmbC SPVMDH.NGCGPLGITETLMRATTVAICYLKDDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.smg_EmbC SPVMDH.NGCGPLGITETLMRATTVAICYLKDDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.bov_EmbC SPVMDH.NGCGPLGITETLMRATTVAICYLKDDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.mar_EmbC SPVMDH.NGCGPLGITETLMRATTVAICYLKDDWYRDWGSVEQYHRLVPADQAPDAVVEEG
M.lep_EmbC SPVMDH.NGCGPLGITETLMRATTVAICYLKDDWYRDWGSVEQYHRLVPADQAPDAVVEEG

1090

M.tub_EmbB TATRSGLWSFGKIRIRGP
M.smg_EmbB SQTHSGLWSFGKIRIRGP
M.bov_EmbB TATRSGLWSFGKIRIRGP
M.mar_EmbB TATRSGLWSFGKIRIRGP
M.lep_EmbB TATRSGLWSFGKIRIRGP
M.tub_EmbA VITVPGWGRGPIRALP
M.smg_EmbA STRVFGWGRGPIRALP
M.bov_EmbA VITVPGWGRGPIRALP
M.mar_EmbA VITVPGWGRGPIRALP
M.lep_EmbA VITVPGWGRGPIRALP
M.tub_EmbC TVTRSGWWSFAPLRRLG.
M.smg_EmbC TATRSGLWSFAPLRRLG.
M.bov_EmbC TVTRSGWWSFAPLRRLG.
M.mar_EmbC TATRSGLWSFAPLRRLG.
M.lep_EmbC TATRSGLWSFAPLRRLG.

Fig. S13. Sequence alignment.

(A) Sequence alignment of Emb proteins from *Mycobacterium tuberculosis*,
 1350 *Mycobacterium smegmatis*, *Mycobacterium bovis*, *Mycobacterium marinum* and
Mycobacterium leprae. Ethambutol binding sites, di-arabinose binding sites, AcpM
 binding sites and clinical ethambutol resistant mutant hotspots are highlight as red ▲,
 ●, +, +, respectively below the aligned sequences. CL1 and Loop₇₆₆₋₈₀₆ in *Msm*
 EmbA are labeled as red bars.

1355 (B) Sequence alignment of Emb catalytic site Asp in DDx motif with catalytic sites of
 two GT-C family members, *Cupriavidus metallidurans* ArnT and *Campylobacter lari*
 PglB.

Fig. S14. Comparison of substrates binding on Emb proteins.

(A) Superposition of the active site region of *Msm* EmbA (pink) in complex with DPA and *Msm* EmbB (purple) in complex with di-saccharide. The D-site arabinose group of the di-saccharide in *Msm* EmbB overlaps with the arabinose moiety of DPA in *Msm* EmbA.

(B) Superposition between *Msm* EmbC (orange) and *Msm* EmbA (magenta) shows the phosphate (ball-and-stick) in EmbC binds to a similar position to the phosphate groups (sticks) of DPA in EmbA. The distance between two phosphorous atoms is 2.2 Å upon superposition between two overall structures.

(C) Superposition of *Msm* EmbA, *Msm* EmbB and *Msm* EmbC on periplasmic region. In *Msm* EmbA the characteristic Loop₇₆₆₋₈₀₆ (cyan) inserts into the active site, whilst

1375 in *Msm* EmbB, the corresponding Loop₇₇₆₋₈₀₆ is shorter and folded on the PC domain,
and in *Msm* EmbC the corresponding fragment (residue range 780-810) is flexible and
missing (dashed line). The di-arabinose groups in *Msm* EmbB and *Msm* EmbC bind in
similar manner.

(D) Di-saccharide binding in the active site of *Msm* EmbB. Contacting residues are
1380 shown as sticks. Cryo-EM map density of ligands (di-arabinose and phosphate)
(threshold 0.3) is shown as yellow meshes.

Fig. S15. Ethambutol binding and structural mapping of ethambutol resistance mutation sites.

(A) Mapping of the top 16 clinical ethambutol (cyan and red spheres) resistance mutation sites on *Mtb* EmbB. The regions these mutations belong to are highlighted in different colors.

1390 (B) Statistics of the regions on *Mtb* EmbB where most frequent mutations occur in all the 1,814 ethambutol resistant sites.

(C) (left) Mapping of the ethambutol resistance mutation sites of *Mtb* EmbC on the *Msm* EmbC structure. Ethambutol is shown as sticks representation in light blue. ethambutol mutation sites are shown as pink spheres. The most frequently occurring
1395 mutation site, Met306 of *Mtb* EmbB (corresponding to Met286 of *Msm* EmbC), is highlighted in magenta. The dashed circle shows that the mutation sites (within the EH1-EH2 region) are clustered inside or around the active site. (right) Table lists ethambutol mutants in *Mtb* EmbC and corresponding residues to *Msm* EmbC (22-28). The highlighted sites in the table represent those mapped inside the dashed circle on
1400 the EmbC structure.

(D) The ethambutol binding site of *Msm* EmbC. Ethambutol, P_i (likely from DP) and surrounding residues (within 4 Å distance) are shown as sticks. Densities of ethambutol and P_i from the cryo-EM map (threshold 0.2) are shown as meshes.

(E) Structural superposition of ethambutol bound *Msm* EmbC and di-arabinose bound
1405 *Msm* EmbC₂ shows ethambutol occupies both D-site and A₀-site arabinose groups of Ara₂. Thus, ethambutol binding will block both donor and acceptor substrate binding near the catalytic residue.

(F) Superposition of ethambutol binding site amongst *Mtb* EmbB, *Msm* EmbB and *Msm* EmbC.

1410

Table S1. Statistics of the cryo-EM structures presented in this study.

	<i>Mtb</i> EmbA-EmbB-AcpM ₂ +ethambutol	<i>Msm</i> EmbA-EmbB-AcpM ₂ +ethambutol	<i>Msm</i> EmbA-EmbB-AcpM ₂ +di-arabinose	<i>Msm</i> EmbC ₂ -AcpM ₂ +ethambutol
EMDB ID	30218	30216	30219	30217
PDB ID	7BVF	7BVC	7BVG	7BVE
Cryo-electron microscopy data collection				
Microscope	FEI Titan Krios			
Voltage (keV)	300			
Camera	Gatan K3-Summit			
Automation software	SerialEM			
Normal / Calibrated magnification	29,000 / 60,976			
Exposure rate (e ⁻ / (pixel ² ·s))	16.8			
Exposure time (s)	2.0			
Number of frames collected	40			
Defocus range (μm)	0.8-2.5			1.5-2.5
Pixel size (Å/pixel)	0.82			
3D reconstruction				
Number of movies	7,612	5,100	2,608	3,711
Symmetry imposed	C1			C2
Initial Particle Number	2,186,192	1,855,947	408,623	707,538
Final particle images (No.)	521,803	227,206	209,894	217,550
Resolution (unmasked, Å)	3.5	3.7	3.7	3.4
Resolution (masked, Å)	2.97	2.90	3.10	2.81
Sharpening B-factor (Å ²)	-113.6	-96.7	-106.7	-104.4
Local resolution range (Å)	1.8-7.5	1.8-7.4	1.8-8.0	1.8-7.2
Coordinate and B-factor refinement				
Model composition				
Atoms (non-H)	16,934	17,206	17,247	17,144
Protein residues	2,221	2,231	2,234	2,236
Ligands	5	8	8	8
Mean B-factor protein atoms (Å ²)	32.17	61.94	56.70	79.47
Mean B-factor Ligand atoms (Å ²)	41.87	61.27	51.56	86.76
Rmsd bonds (Å)	0.005	0.003	0.004	0.010
Rmsd bond angles (°)	0.778	0.609	0.763	1.412

Model-to-map scores				
CC (mask)	0.89	0.86	0.86	0.83
CC (volume)	0.84	0.83	0.84	0.79
CC (box)	0.75	0.68	0.71	0.74
Mean CC for ligands	0.81	0.74	0.76	0.59
Validation				
MolProbity score	1.93	1.78	1.97	1.77
Clash score	5.70	5.21	8.48	5.31
Rotamer outliers (%)	0.00	0.00	0.00	0.28
C β outliers (%)	0.00	0.00	0.00	0.24
CaBLAM outliers (%)	7.07	5.14	5.67	4.43
EMRinger score	3.48	3.17	3.18	4.58
Ramachandran plot				
Favored (%)	87.12	91.28	91.47	92.18
Allowed (%)	12.83	8.67	8.48	7.60
Outliers (%)	0.05	0.04	0.04	0.22

Table S2. Data collection and refinement statistics of crystal structure of *Msm*

1415 **EmbC₂-AcpM₂.**

	<i>Msm</i> EmbC ₂ -AcpM ₂ (Tb derivative)	<i>Msm</i> EmbC ₂ -AcpM ₂ (di-arabinose)
PDB code	7BVH	
Data collection		
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	1.6491	1.0000
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	121.99, 176.48, 207.31	121.08, 176.33, 207.77
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	49.02-4.60 (4.43-4.20) ^a	49.60-3.30 (3.48-3.30)
No. of unique reflections	33,063 (4,727)	67,591 (9,775)
Completeness (%)	99.9 (100.0)	99.9 (100.0)
<i>R</i> _{merge} (%) ^b	0.118 (1.576)	0.114 (5.176)
<i>R</i> _{pim} (%) ^c	0.029 (0.391)	0.027 (1.184)
<i>CC</i> _{1/2} ^d	0.999 (0.903)	0.998 (0.417)
Redundancy	33.3 (33.1)	19.9 (19.9)
Mean <i>I</i> /σ(<i>I</i>)	18.0 (2.6)	12.2 (0.7)
Wilson B factors (Å ²)	190	147
Phasing Statistics		
No. of heavy-atom sites	6	
Figure of merit	0.3	
BAYES-CC score	41	
Refinement		
Resolution (Å)		49.67-3.30
No. of reflections used		67,568
<i>R</i> _{work} / <i>R</i> _{free} (%) ^e		23.2 / 26.5
No. of atoms		
Protein		17,088
Ligand		228
<i>B</i> factor (Å ²)		
Protein		78
Ligand		84
R.m.s deviations		
Bond lengths (Å)		0.010
Bond angles (°)		1.210

Ramachandran plot (%)	
Favored	91.6
Allowed	7.0
Outliers	1.5

^a Values in parentheses are for highest-resolution shell.

^b $R_{\text{merge}} = \sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean intensity of the observations of I_{ih} of reflection h .

^c R_{pim} : precision-indicating (multiplicity-weighted) R_{merge} .

^d $CC_{1/2}$: percentage of correlation between intensities from random half-datasets.

1420 ^e $R_{\text{work}} = \sum_h |F_o - F_c| / \sum_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h . R_{free} is mathematically equivalent to R_{work} , but was measured over 5% of the data.

Table S3. The binding affinity analyzed by MST assay.

1425

Protein	Ligand	Binding affinity (K_d)
<i>Msm</i> EmbC ₂ (WT)	DPA	$3.0 \pm 1.1 \mu\text{M}$
<i>Msm</i> EmbC ₂ (R383A)	DPA	No binding up to 1.25 mM
<i>Msm</i> EmbC ₂ (T570S)	DPA	No binding up to 1.25 mM
<i>Msm</i> EmbC ₂ (H574A)	DPA	$122.0 \pm 44.0 \mu\text{M}$
<i>Msm</i> EmbC ₂ (H575A)	DPA	$137.0 \pm 63.0 \mu\text{M}$
<i>Mtb</i> EmbA-EmbB	Ethambutol	$0.42 \pm 0.12 \mu\text{M}$
<i>Msm</i> EmbA-EmbB	Ethambutol	$0.31 \pm 0.08 \mu\text{M}$
<i>Msm</i> EmbC ₂	Ethambutol	$11.1 \pm 1.2 \mu\text{M}$