Supporting Information for

Structure of the priming arabinosyltransferase AftA required for AG biosynthesis of *Mycobacterium tuberculosis*

Yicheng Gong[a], Chuancun Wei[a], Jun Wang[a,b], Nengjiang Mu[a], Qinhong Lu[a], Chengyao Wu[a], Ning Yan[a], Huifang Yang[a], Yao Zhao[a,] Xiuna Yang[a], Sudagar S. Gurcha[c], Natacha Veerapen[c], Sarah M. Batt[c], Zhiqiang Hao[d], Lintai Da[d], Gurdyal S. Besra[c,1], Zihe Rao[a,e], Lu Zhang[a,1]

[a] Shanghai Institute for Advanced Immunochemical Studies, School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.

[b] School of Life Sciences, Tianjin University, Tianjin 300072, China.

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

[d] Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.

[e] Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China.

[1]To whom correspondence may be addressed.

Email: g.besra@bham.ac.uk (G.S.B.) and zhanglu1@shanghaitech.edu.cn (L. Z.)

This file includes:

Figures S1-S10

Tables S1-S2

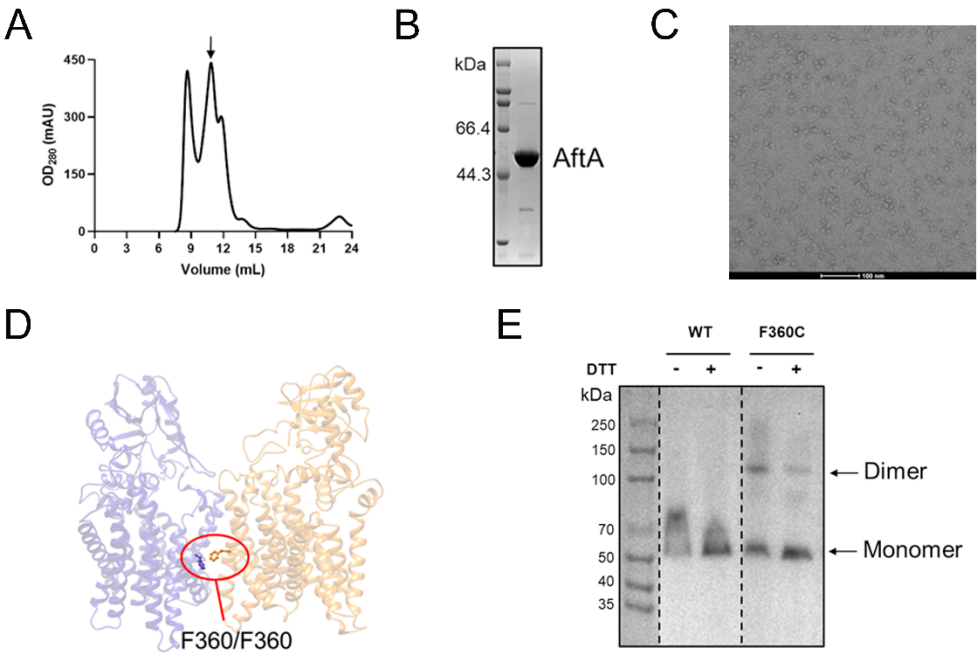


Fig. S1. Characterization of *Mtb* AftA.

(A) Size-exclusion chromatography on a Superdex 200 gel filtration column (GE healthcare) for *Mtb* AftA 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 band about 50 kDa corresponds to AftA protein used for EM study.

(C) Negative staining of AftA protein.

(D) Dimer interface with the a Phe360 pair chosen for cysteine mutation indicated with red circles.

(E) Western blot analysis of dimerization on cell membrane. Cell membranes containing 10×His-AftA, 10×His-AftA(F360C), were treated with and without 10 mM DTT. For F360C without DTT, the stronger monomeric AftA signal than the dimeric signal may be due to the dynamic equilibrium between monomeric and dimeric state, or the non-fully-oxidized environment of the membrane where F360C was embedded.

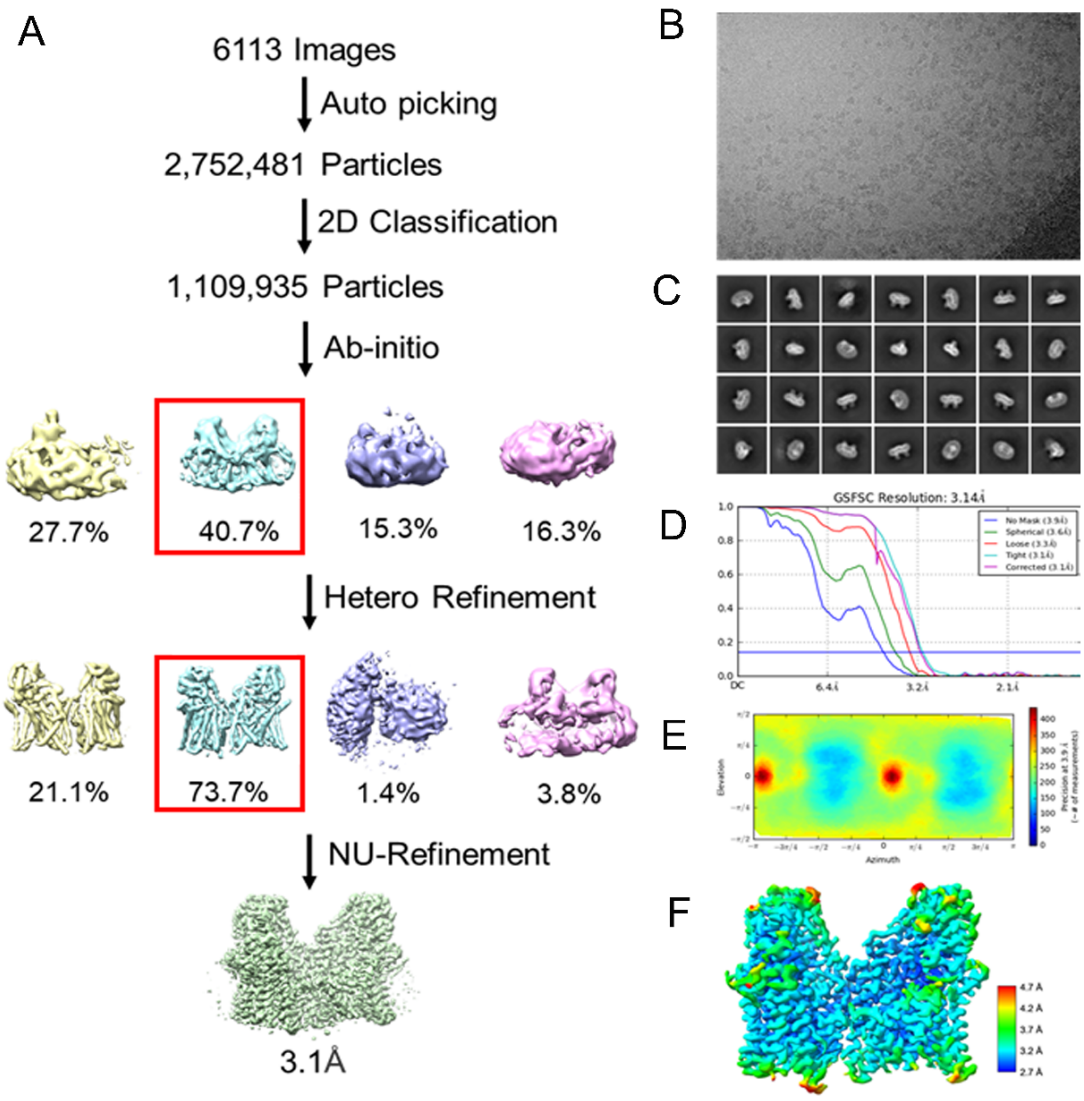


Fig. S2. Cryo-EM data processing and validation of *Mtb* AftA.

(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.

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

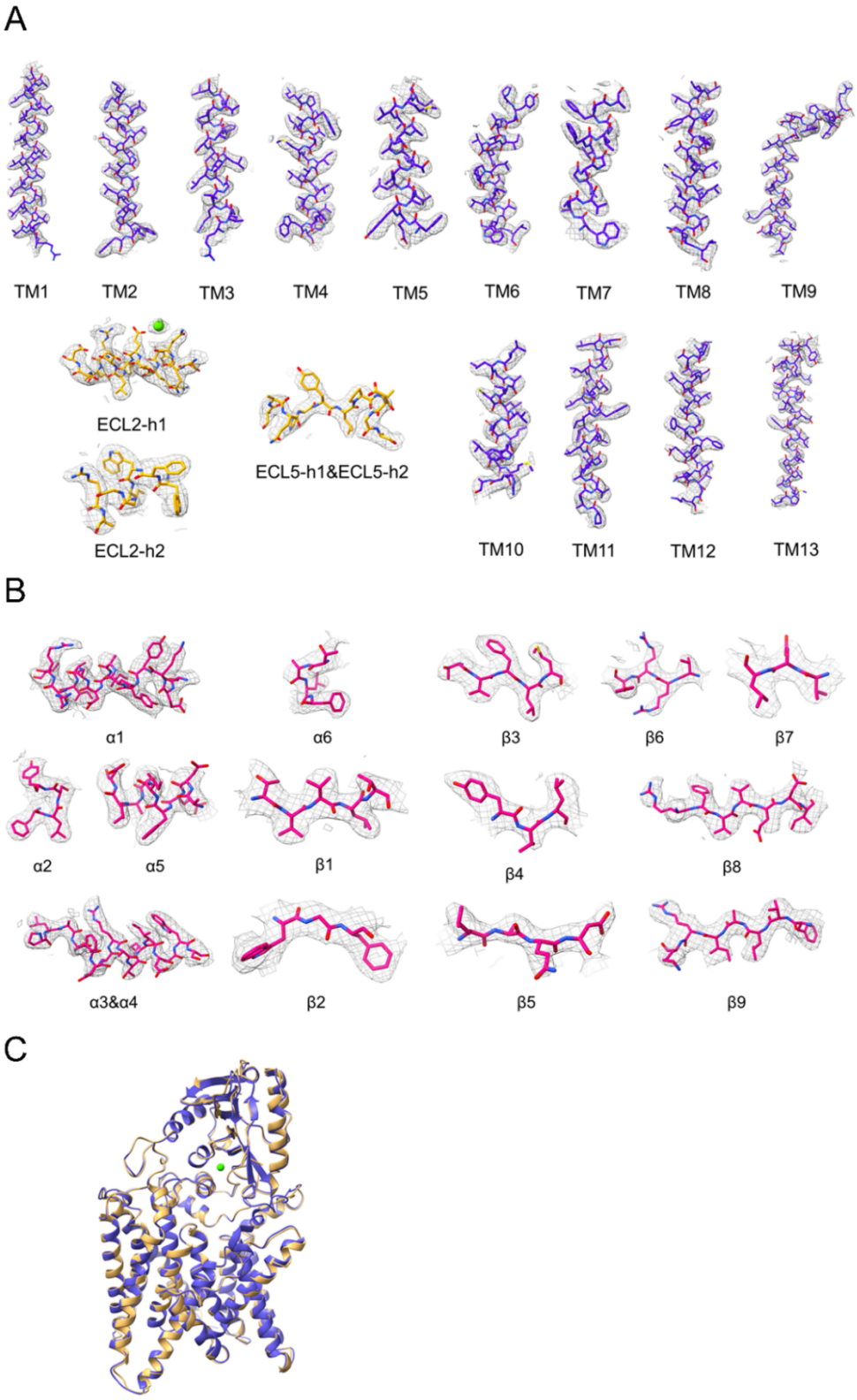


Fig. S3. Example regions of cryo-EM maps of AftA.
(A) The cryo-EM map (threshold 0.28) of TM1-13 and ECL2&ECL5.

(B) The cryo-EM map (threshold 0.28) of CTD including α1-6 and β1-9.

(C) Alignment of the two AftA protomers.

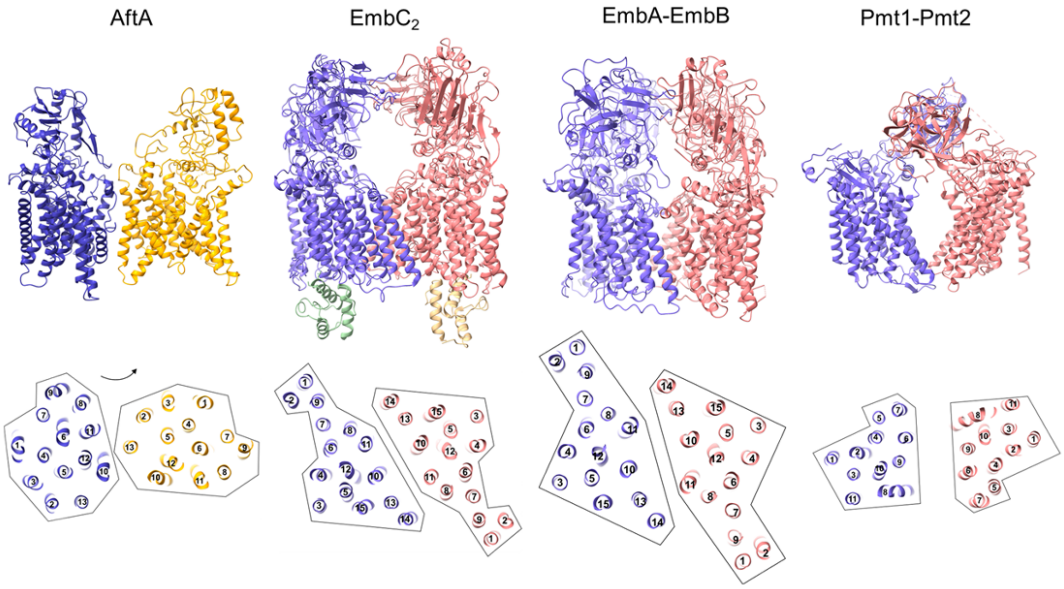


Fig. S4. Comparison of dimer assembly pattern of AftA with other GT-C enzymes. (up) Cartoon representations of AftA, Pmt1-Pmt2 (PDB code 6P25), EmbA-EmbB (PDB code 7BVF), EmbC-EmbC (PDB code 7BVE). Each structure is colored by chain; (down) Arrangement of the TM helices for the structures is shown as a slice through the TM domain. Two AftA monomers show an asymmetric rotation while other dimers are arranged in a two-fold ($EmbC_2$) or a pseudo-two-fold (EmbA-EmbB, Pmt1-Pmt2) symmetry.

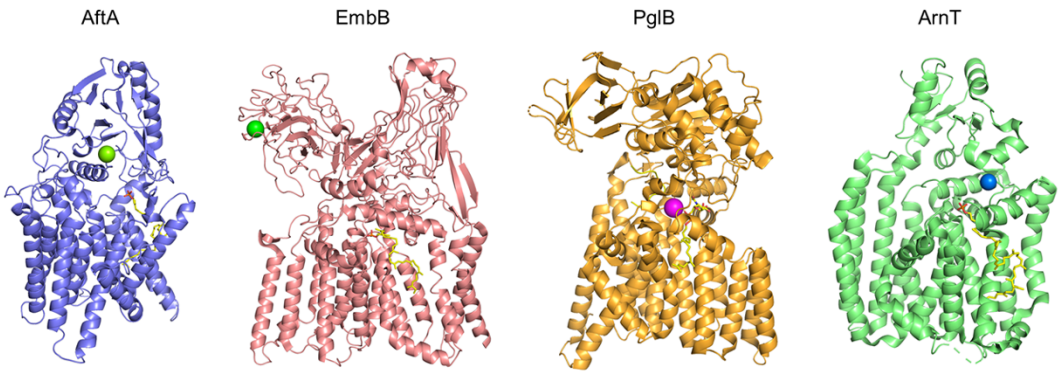


Fig. S5. the metal ion binding mode in AftA is different from three representative GT-C enzymes.

Cartoon representations of AftA, EmbB(PDB code 7BVF), PglB(PDB code 5OGL), ArnT (PDB code 5F15). Ca^{2+}, Mn^{2+} and Zn^{2+} are represented by green, magenta and blue spheres.

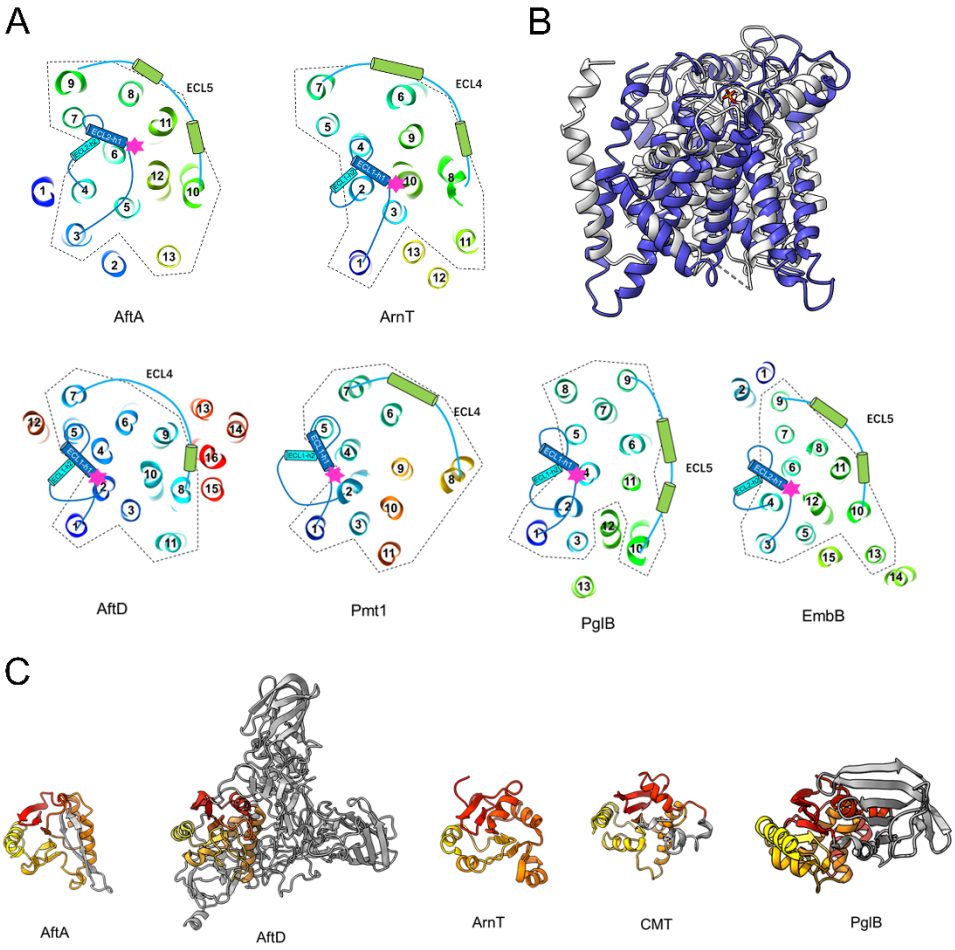
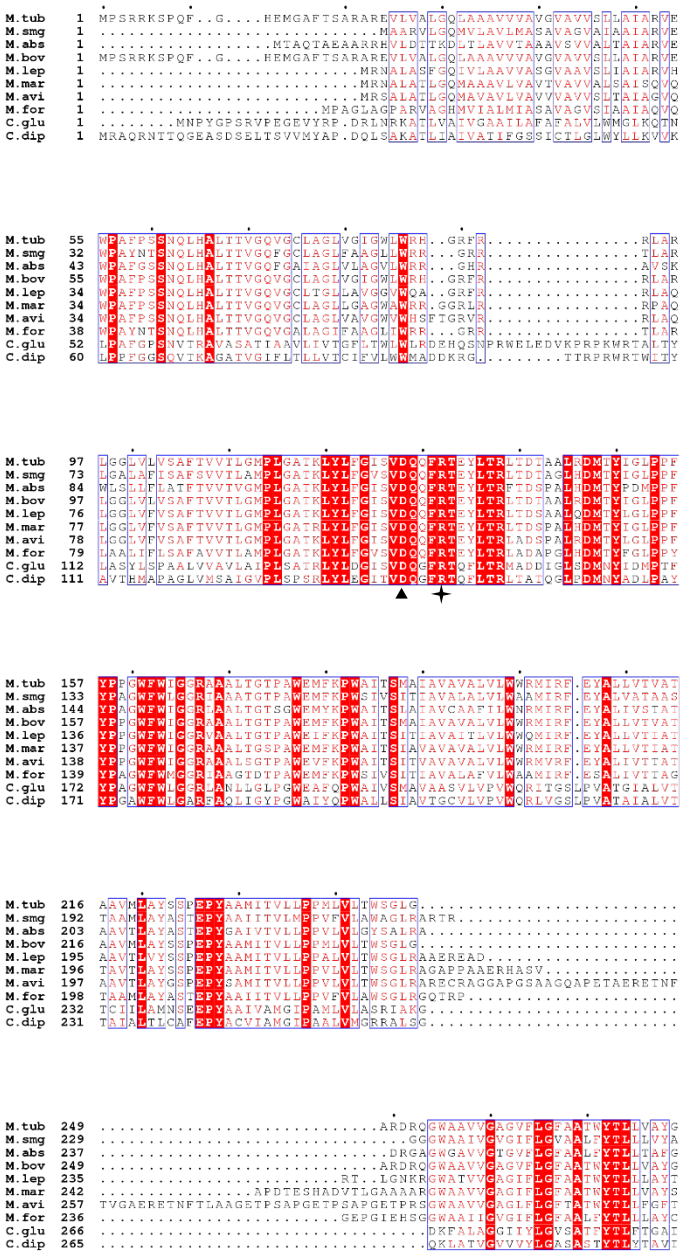


Fig. S6. Conservation of the GT-C core in both TMD and CTD.

(A) 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. From left to right are AftA, ArnT(PDB code 5F15), AftD(PDB code 6W98), Pmt1(PDB code 6P25), PglB(PDB code 5OGL), EmbB(PDB code 7BVF). The catalytically essential Asp is shown as a magenta asterisk at the beginning of ECL1(2 in AftA and EmbB)-h1.

(B) Superimpose the TMD core (TM3-13) of AftA on the TMD core (TM1-11) of ArnT which has a donor UndP around its GT-C core, the phospholipid moiety of the sugar donor lies within the proposed donor cavity. AftA is shown as blue ribbon, while ArnT is gray.

(C) Structures of the luminal dome of representative GT-C members are aligned to the CTD of AftA. The structurally conserved core is highlighted in red, orange and yellow. From left to right are AftA, AftD(PDB code 6W98), ArnT(PDB code 5F15), CMT(PDB code 7ZLH), PglB(PDB code 5OGL).


M.tub 1 MPSRRKSPQF..G....HEMGAFTSARAREVLVALGQLAAAVVVAVGVAVVSLLAIARVE
M.smg 1 .....MAARVLGQMVLAVLMASAVAGVAIAAIARVE
M.abs 1 .....MTAQTAEAARREVLDITKDLTLAVVTAAAVSVVALTAIARVE
M.bov 1 MPSRRKSPQF..G....HEMGAFTSARAREVLVALGQLAAAVVVAVGVAVVSLLAIARVE
M.lep 1 .....MRNALASFGQIVLAAVVASGVAAVSLIAIARVH
M.mar 1 .....MRNALAILGQMAAAVLVAVTVAVVALSAISQVQ
M.avi 1 .....MRSALAILGQMAVAVLVAVVVAVVSLTAIAGVQ
M.for 1 .....MPAGLAGPARVAGHMVIALMIASAVAGVSIAAIAQVQ
C.glu 1 .....MNPYGPSRVPEGEVYRP.DRLNRKATLVAIVGAAILAFAFALVLWMGLKQTN
C.dip 1 MRAQRNTTQGEASDSELTSVVMYAP.DQLSAKATLLAIVATIFGSSICTLGLWYLLKVVK
M.tub 55 WPAFPSSNQLHALTTVGQVGCLAGLVGIGWLWRH..GRFR.....RLAR
M.smg 32 WPAYNTSNQLHALTTVGQFGCLAGLFAAGLLWRR...GRR.....TLAR
M.abs 43 WPAFGSSNQLHALTTVGQFGAIAGLVLAGVLWRR...GHR.....AVSK
M.bov 55 WPAFPSSNQLHALTTVGQVGCLAGLVGIGWLWRH..GRFR.....RLAR
M.lep 34 WPAFPSSNQLHALITVGQVGCLTGLLAVGGVWQA..GRFR.....RLAQ
M.mar 34 WPAFPSSNQLHALITVGQVGCLAGLLGAGWAWRR.GGRLR.....RPAQ
M.avi 34 WPAFPSSNQLHALTTVGQVGCLAGLVAVGWVWHS FTGRVR.....RLAQ
M.for 38 MPAYNTSNQLHALTTVGQVGALAGITFAAGLTWRR...GRR.....TLAR
C.glu 52 LPAFGFSNVTRAVASATIAAVLTVTGFLTWLWLRDRHQSNPRWELEDVKPRPKWRTALTY
C.dip 60 LPPFGGSQVTKAGATVGIFLTLLVTCIFVLWWMADDKRG.....TTRPRWRTWITY
M.tub 97 LGGLVLVSAFTVVTLGMPLGATRLYLFGISVDQCFRTEYLTRLTDTAALRDMTYIGLPPF
M.smg 73 LGALAFISAFSVVTLAMPLGATRLYLFGVSVDQCFRTEYLTRLTDTAGLHDMTYIGLPPF
M.abs 84 WLSLLFLATFTVVTVGMPLGATRLYLFGISVDQCFRTEYLTRFTDSPALHDMTYPDMPPF
M.bov 97 LGGLVLVSAFTVVTLGMPLGATRLYLFGISVDQCFRTEYLTRFTDTAALRDMTYIGLPPF
M.lep 76 LGGLVFVSAFTVVTLGMPLGATRLYLFGISVDQCFRTEYLTRLTDSAALQDMTYLGLPPF
M.mar 77 LGGLVFVSAFTVVTLGMPLGATRLYLFGISVDQCFRTEYLTRLTDSPALHDMTYTGLPPF
M.avi 78 LGGLVFVSAFTVVTLGMPLGATKLYLFGISVDQCFRTEYLTRLADSPALRDMTYLGLPPF
M.for 79 LAALIFLSAFAVVTLAMPLGATKLYLFGVSVDQCFRTEYLTRLADAPGLHDMTYFGLPPY
C.glu 112 LASYLSPAALVVAVLAIPLSATRLYLDGISVDQGFRTQFLTRMADDIGLSDMNYIDMPTF
C.dip 111 AVTHMAPAGLVMSAIGVPLSPSRLYLEGLTVDQGFRTQFLTRLTAIQGLPDMNYADLPAY
M.tub 157 YPPGWFWIGPRAAALTGTPAWEMFKPWAITSMAIAVAVALVIWWRMIRF.EYALLVTVAT
M.smg 133 YPAGWFWIGPRFAATGTPAWEMFKPWSIVSTTIAVALALVIWAAMIRF.EYALVATAAS
M.abs 144 YPAGWFWIGPRFAALTGTSGWEMYKPWAITSLATAVCAAFIIWNNRMIRF.EYALIVSTAT
M.bov 157 YPPGWFWIGPRAAALTGTPAWEMFKPWAITSMAIAVAVALVIWWRMIRF.EYALLVTVAT
M.lep 136 YPPGWFWIGPRVAALTGTPAWEIFKPWAITSITIAVAITLVIWWQMIRF.EYALLVTIAT
M.mar 137 YPPGWFWIGPRAAALTGSPAWEMFKPWAITSLAVAVAVALVIWWRMIRF.EYALLVTIAT
M.avi 138 YPPGWFWIGPRAAALSGTPAWEVFKPWAITSITVAVAVALVIWWRMVRF.EYALIVTTAT
M.for 139 YPAGWFWMGPRFAAGTDTPAWEMFKPWSIVSITIAVALAFVIWAAMIRF.ESALIVTTAG
C.glu 172 YPAGWFWLGPRFANLLGLPGWEAFQPWAIVSMAVAASVLVPVWQRITGSLPVATGIALVT
C.dip 171 YPGAWFWLGARFAQLIGYPGWAIYQPWALLSLAVTGCVLVPVWQRLVGSLPVATAIALVT
M.tub 216 AAVMLAYSSPEPYAAMITVLLPPMLVLTWSGLG.....
M.smg 192 TAAMLAYASTEPYAAIITVLMPPVFVLAWAGLRARTR.....
M.abs 203 AAVTLAYASTEPYGAIVTVLLPPVLVLGYSALRA.....
M.bov 216 AAVMLAYSSPEPYAAMITVLLPPMLVLTWSGLG.....
M.lep 195 AAVTLVYSSPEPYAAMITVLLPPALVLTWSGLRAAEREAD.....
M.mar 196 TAVTLAYSSPEPYAAMITVLLPPVLVLTWSGLRAGAPPAAERHASV.....
M.avi 197 AAVTLAYGSPEPYSAMITVLLPPVLVLTWSGLRARECRAGGAPGSAAGQAPETAERETNF
M.for 198 TAAMLAYASTEPYAAIITVLLPPVFVLAWSGLRGQTRP.....
C.glu 232 TCIILAMNSEEPYAAIVAMGIPAMLVLASRIAKG.....
C.dip 231 TAIALTLCAFEPYACVIAMGIPAALVMGRRALSG.....
M.tub 249 .....ARDRQGWAAVVGAGVPLGFAATNYTLLVAYG
M.smg 229 .....GGGWAAIVGVGIPLGVAALPYTLLVYA
M.abs 237 .....DRGAGWGAVVGTGVPLGFAALPYTLLTAFG
M.bov 249 .....ARDRQGWAAVVGAGVPLGFAATNYTLLVAYG
M.lep 235 .....RT..LGNKRGWATVVGAGIPLGFAATNYTLLLAYT
M.mar 242 .....APDTESHADVTLGAAAARGWAAVVGAGIPLGFAATNYTLLVAYS
M.avi 257 TVGAERETNFTLAAGETPSAPGETPSAPGETPRSGWAAVVGAGIPLGFTATNYTLLFGFT
M.for 236 .....GEPGIEHSGGWAAIIGVGIPLGFAALPYTPLLAYC
C.glu 266 .....DKFALAGGIIYLGVSATPYTPFTGAI
C.dip 265 .....QKLATVGVVVYLGASASTYTPWTAVT

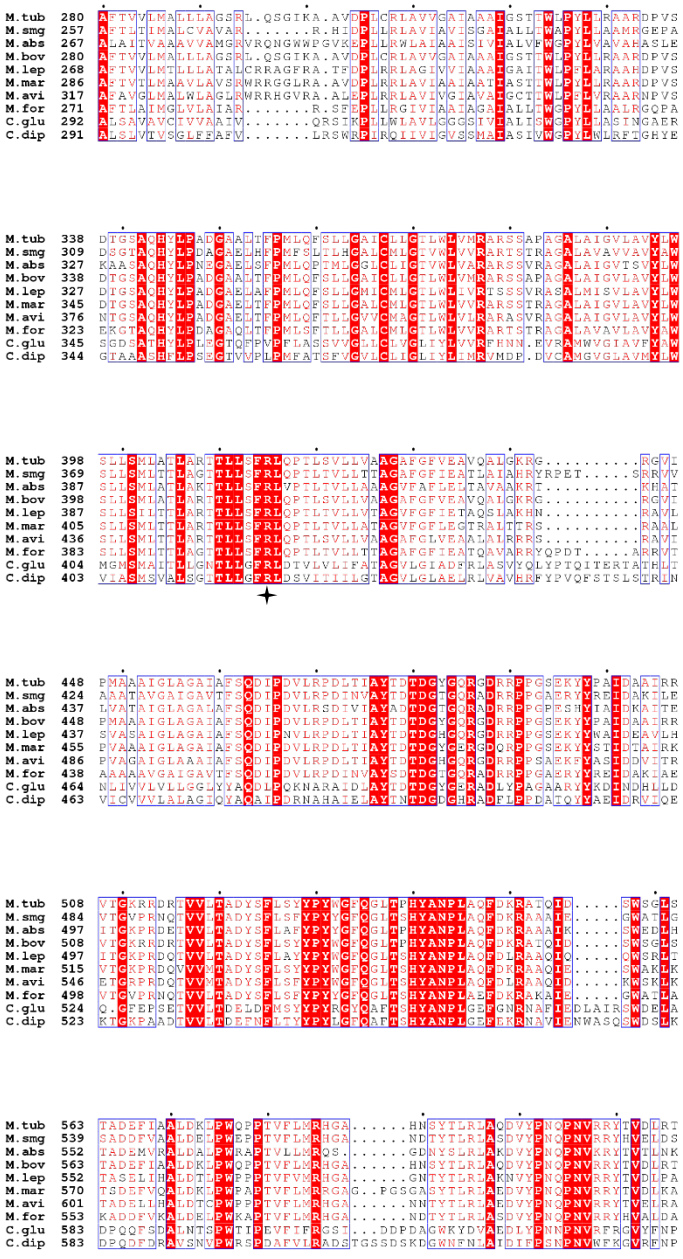


Fig. S7. Sequence alignment of AftA within the Corynebacterianeae.
Sequence alignment of AftA from *Mycobacterium tuberculosis*, *Mycobacterium smegmatis*, *Mycobacteroides abscessus*, *Mycobacterium bovis*, *Mycobacterium leprae*, *Mycobacterium marinum*, *Mycobacterium avium*, *Mycolicibacterium fortuitum*, *Corynebacterium glutamicum* and *Corynebacterium diphtheriae*. Catalytic site Asp and conserved amino acids mentioned in the article are highlight as ▲ and ✦.

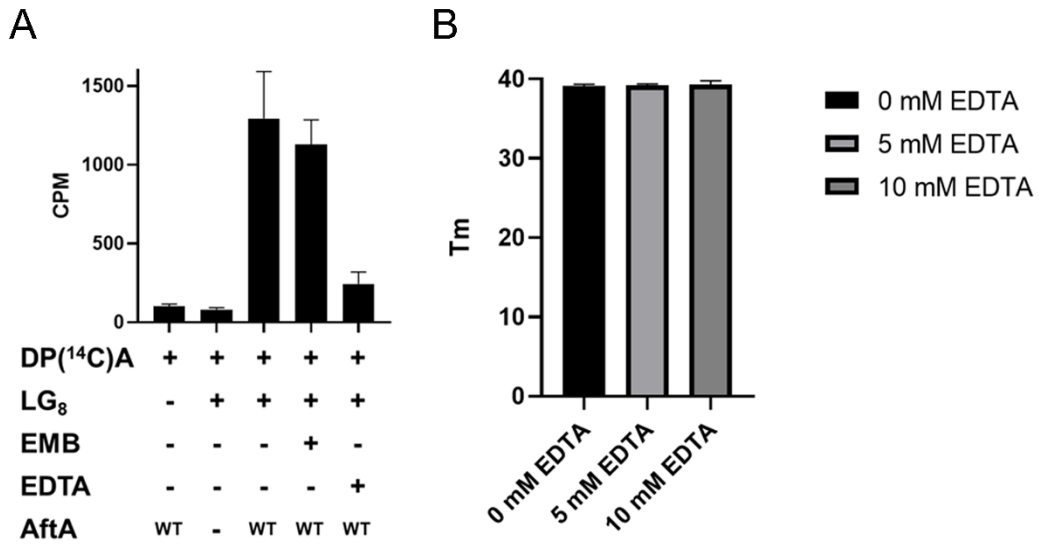


Fig. S8. Effect of EDTA and ethambutol on the activity and thermostability of AftA.
(A) Enzymatic activity of WT *Mtb* AftA when adding EDTA and ethambutol.
(B) Thermostability of AftA protein in EDTA at different concentrations.

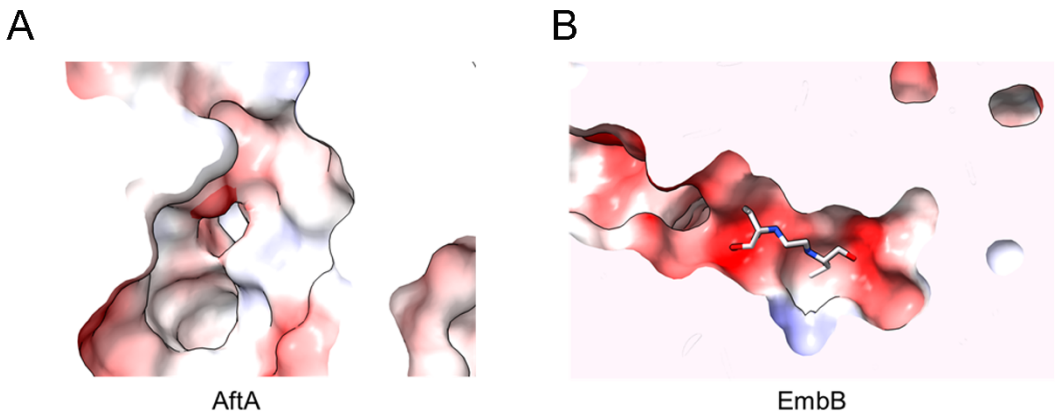


Fig. S9. Comparison of active sites of *Mtb* AftA and *Mtb* EmbB.
(A) Section of AftA active site, shown as electrostatic surface.
(B) Section of EmB active site bound with ethambutol, shown as electrostatic surface.

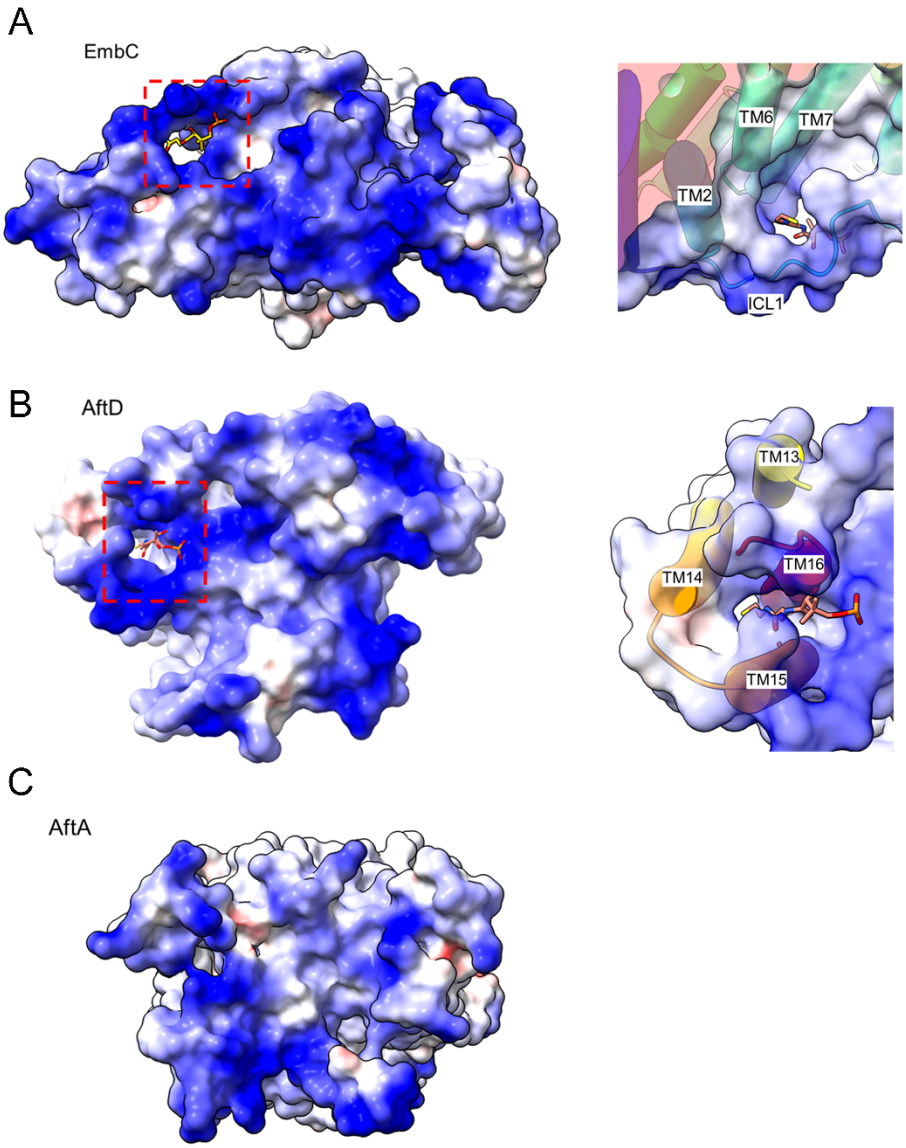


Fig. S10. AcpM binding sites.

(A) Left, cytoplasmic surface of EmbC, shown as electrostatic surface. Right, the 4′-Ppant binding tunnel formed by TM2, TM6, TM7 and ICL1 of EmbC.

(B) Left, cytoplasmic surface of AftD, shown as electrostatic surface. Right, the 4′-Ppant binding tunnel formed by TM13-16 of AftD.

(C) Cytoplasmic surface of AftA, shown as electrostatic surface.

Table S1. Cryo-EM data collection and modeling statistics.

AftA apo	
EMDB: 35410	
PDB ID: 8IF8	
Cryo-electron microscopy data collection	
Microscope	FEI Titan Krios
Voltage (keV)	300
Camera	Gatan K3-Summit
Automation software	SerialEM
Normal / Calibrated magnification	105,000 / 59,000
Exposure rate (e^- / (pixel^2·s))	16.8
Total dose (e^-/Å^2)	60
Exposure time (s)	2.4
Number of frames collected	40
Defocus range (μm)	1.2-2.0
Pixel size (Å/pixel)	0.832
3D reconstruction	
Number of movies	6113
Symmetry imposed	C1
Initial Particle Number	2,752,481
Final particle images (No.)	147,896
Resolution (unmasked, Å)	3.9
Resolution (masked, Å)	3.14
Sharpening B-factor (Å^2)	-103.9
Local resolution range (Å)	2.7-4.7
Coordinate and B-factor refinement	
Model composition	
Atoms (non-H)	9307
Protein residues	1218
Ligands	2
Mean B-factor protein atoms (Å^2)	54.31
Mean B-factor Ligand atoms (Å^2)	56.81
Rmsd bonds (Å)	0.006
Rmsd bond angles (°)	0.745
Model-to-map scores	
CC (mask)	0.88
CC (volume)	0.85
CC (box)	0.70
Mean CC for ligands	0.82

Validation	
MolProbity score	1.85
Clash score	6.79
Rotamer outliers (%)	0
Cβ outliers (%)	0
CaBLAMoutliers (%)	3.58
Ramachandran plot	
Favored (%)	92.23
Allowed (%)	7.69
Outliers (%)	0.08

Table S2. Summary of model building

Protopmer Name	**Chain**	**Total residues/ Range built**	**Unmodelled residues**	**% atomic model**	**Ligands**
AftA Protomer 1	B	643/ 20-639	1-19, 640-643	96.4%	Ca^{2+}
AftA Protomer 2	C	643/ 21-339, 351-598, 609-639	1-20, 340-350, 599-608	93.0%	Ca^{2+}