Structural insights on the substrate-binding proteins of the *Mycobacterium tuberculosis* mammalian-cell-entry (Mce) 1 and 4 complexes

Pooja Asthana[a,1], Dhirendra Singh[a,1], Jan Skov Pedersen[b], Mikko J. Hynönen[a], Ramita Sulu[a], Abhinandan V. Murthy[a,c], Mikko Laitaoja[d], Janne Jänis[d], Lee W. Riley[e], Rajaram Venkatesan[a,2]

a. Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland

b. Department of Chemistry and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark

c. Present address: Faculty of Medicine, University of Helsinki, Finland

d. Department of Chemistry, University of Eastern Finland, Joensuu, Finland

e. School of Public Health, University of California, Berkeley, CA, USA

1.These authors contributed equally

2. To whom correspondence may be addressed. Email: rajaram.venkatesan@oulu.fi

Key words:

Mycobacterium tuberculosis, mammalian-cell-entry proteins, Mce1, Mce4, substrate-binding proteins, ABC/lipid transporter, crystal structure, SAXS

Abstract

Tuberculosis (Tb), caused by *Mycobacterium tuberculosis* (*Mtb*), is responsible for more than a million deaths annually. In the latent phase of infection, *Mtb* uses lipids as the source of carbon and energy for its survival. The lipid molecules are transported across the cell wall via multiple transport systems. One such set of widely present and less-studied transporters is the Mammalian-cell-entry (Mce) complexes. Here, we report the properties of the substrate-binding proteins (SBPs; MceA-F) of the Mce1 and Mce4 complexes from *Mtb* which are responsible for the import of mycolic acid/fatty acids, and cholesterol respectively. MceA-F are composed of four domains namely, transmembrane, MCE, helical and tail domains. Our studies show that MceA-F are predominantly monomeric when purified individually and do not form homohexamers unlike the reported homologs (MlaD, PqiB and LetB) from other prokaryotes. The crystal structure of MCE domain of *Mtb* Mce4A ($MtMce4A_{39-140}$) determined at 2.9 Å shows the formation of an unexpected domain-swapped dimer in the crystals. Further, the purification and small-angle X-ray scattering (SAXS) analysis on MtMce1A, MtMce4A and their domains suggest that the helical domain requires hydrophobic interactions with the detergent molecules for its stability. Combining all the experimental data, we propose a heterohexameric arrangement of MtMceA-F SBPs, where the soluble MCE domain of the SBPs would remain in the periplasm with the helical domain extending to the lipid layer forming a hollow channel for the transport of lipids across the membranes. The tail domain would reach the cell surface assisting in lipid recognition and binding.

1. Introduction

Mycobacterium tuberculosis (*Mtb*) is a deadly intracellular pathogen causing the disease tuberculosis (Tb) and is responsible for more than a million deaths every year. Approximately, one fourth of the world population is latently infected by *Mtb* (1). *Mtb* is one of the very few bacteria which can use host lipids as the source of energy and carbon, and this property might be most relevant during the intra-phagosomal latent stage of infection (2). Interestingly, *Mtb* has nearly 200 lipid metabolizing proteins (3, 4). Therefore, lipid transport across the membrane plays a pivotal role in *Mtb* pathogenesis. One such set of lipid transporters found in *Mtb* is encoded by the, Mammalian-cell-entry (Mce) operons, Mce1, Mce2, Mce3, and Mce4 (4), comprising 10-14 genes each. These operons were named based on the initial observation that a DNA fragment (corresponding to Mce1A) from H37Ra when expressed in *E.coli* made them to invade HeLa cells (5). Similar to Mce1A, the expression of Mce3A and Mce4A in *E. coli* also provides the invasion ability to enter HeLa cells (6, 7). However, further studies have suggested that Mce operons encode for substrate-binding proteins (SBPs) (MceA-F) and permeases (YrbEA-B) to form an ABC transporter (8, 9). The ATPase for this ABC-transporter is proposed to be coded by the gene MceG (also known as Mkl) elsewhere (10). In addition, Mce1, Mce3 and Mce4 operons code for Mce-associated membrane proteins (Mam, also known as Mas) which probably stabilizes the Mce complexes (8). It is now well demonstrated that Mce1 is involved in the transport of mycolic acid/fatty acid and Mce4 imports cholesterol. While *Mtb* disrupted in Mce2 operon accumulates sulfolipid-1 at levels nearly 10 times that of wild type *Mtb* during stationary growth (2, 9, 11), the substrate specificity for Mce3 complex is still unknown. The role of protein encoded by Mce operons in the pathogenicity of *Mtb* has been well established in mouse models based on operon mutants (12, 13).

Recently, structures of homologs of Mce SBPs from *E.coli* (MlaD, PqiB, LetB) and *A. baumannii* (MlaD) were determined (14-16). Based on these homohexameric structures, two different mechanism of lipid transport has been reported. First, Mla complex-ferry transport mechanism, where the Mla operon carries a single Mce gene (MlaD) with a single MCE domain. In this case, the lipids are transferred to MlaD by a shuttle protein (MlaC) and the lipid molecule is then passed through the central hydrophobic pore of homohexameric MlaD. Second, the LetB and PqiB tunnel transport mechanism, where LetB forms a long stack of seven homohexameric MCE domains one above the other connecting the inner and outer membrane with central channel mediating lipid transport. Like LetB, PqiB also forms a central pore formed by three stacked Mce homohexamers with their long C-terminal helix forming a narrow channel for lipid transport.

In this study, we have recombinantly expressed and purified the SBPs MtMce1A-1F as well as MtMce4A-4F encoded in the Mce1 and Mce4 operons. In addition, based on sequence analysis and secondary structure prediction, various domains of MtMce1A and MtMce4A have been identified and a detailed characterization including small-angle X-ray scattering (SAXS) analysis has been performed. Furthermore, the crystal structure of the MCE domain of MtMce4A has been determined using Selenomethionine-single-wavelength anomalous dispersion (SeMet-SAD) phasing which forms an unexpected domain-swapped dimer in the crystals. This is the first reported structure of any mycobacterial protein encoded by the Mce operons. In addition, our studies show that these proteins are predominantly monomeric when expressed and purified individually. Moreover, the purification and SAXS studies of MtMce1A and MtMce4A domains further suggests that the helical domain

is mainly responsible for the interaction with the detergent micelles, which are required in the purification indicating that this region may form the hydrophobic pore for the transfer of lipid molecules in the complex. Based on these results, a possible model for the *Mtb* Mce complexes is provided.

2. Results

2.1 Mtb MceA-F SBPs have conserved four-domain architecture

A comparative sequence analysis and secondary structure prediction (*SI Appendix*, Fig. S1, S2) of *Mtb* Mce 1, 2, 3 and 4 A-F SBPs suggest that in spite of having very low sequence identity between them (~20% or less; *SI Appendix*, Table S1, S2), they have conserved domain architecture with each of them having four domains (Fig. 1*A* and 1*B*). The first domain is the N-terminal transmembrane (TM) domain (~30-40 amino acids) predicted to form a single transmembrane helix followed by the second domain with ~100 amino acids composed of mainly β-strands (usually 7 β-strands), now widely known as MCE domain. The third domain mainly consists of long helices (~200 amino acids) and is therefore referred here as 'helical domain'. The fourth domain is mainly predicted to be an unstructured domain and referred as 'tail domain'. Interestingly, the length of the tail domain varies between 6-260 amino acids between the various MtMceA-F SBPs, while the other domains are well conserved with respect to their topology and number of amino acids. For example, among the MtMce1A-1F proteins, the MtMce1E tail domain has 34 residues and MtMce1F has 218 residues. Additionally, the tail domains of MtMce1C, MtMce1D, MtMce4D and MtMce4F are proline-rich. Moreover, MtMce1E, MtMce2E, MtMce3E and MtMce4E act as probable lipoproteins also known as Lprk, LprL, LprM, and LprN respectively with a conserved 'lipobox' at the N-terminus (17).

2.2 MCE domain is the only soluble part of MtMce1A and MtMce4A

All the six SBPs encoded in the Mce1 and Mce4 operons (MtMce1A-1F and MtMce4A-4F) were recombinantly expressed in *E. coli* and purified (*SI Appendix*, Supplementary Results, Fig. S3) in the presence of detergents. Given that, all the substrate-binding proteins have been predicted to have similar secondary structures and domain architecture, further detailed domain-level characterization was performed for MtMce1A and MtMce4A. From secondary structure predictions (Section 3.1), the domain constructs of MtMce1A and MtMce4A categorized as MCE (MtMce1A$_{36-148}$, MtMce4A$_{39-140}$), MCE+Helical (MtMce1A$_{38-325}$, MtMce4A$_{39-320}$), Helical+Tail (MtMce1A$_{126-454}$, MtMce4A$_{121-400}$), and MCE+Helical+Tail domains (MtMce1A$_{38-454}$, MtMce4A$_{36-400}$) were successfully expressed in *E. coli*, and screened to evaluate their solubility in the presence and absence of detergents. Interestingly, the MCE domain of both MtMce1A and MtMce4A (MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$) were the only soluble constructs. Whereas, the MCE+Helical+Tail, MCE+Helical as well as Helical+Tail domains could be purified only in the presence of detergent even though the transmembrane domain has been deleted in all these constructs (Fig. 1*B* and *C*). Additionally, the extension of the soluble construct even with one (MtMce4A$_{39-154}$) or two helical domains (MtMce4A$_{39-190}$) resulted in insolubility, indicating that the helical domain requires detergent for its stability. The CD curves of MtMce1A and MtMce4A domains (Fig. 1*D* and *E*) indicated mixtures of α-helical and β-sheet content for all the MtMce1A and MtMce4A domains purified with DDM. Whereas, the soluble constructs (MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$) showed a typical β-sheet dominated spectra (*SI Appendix*, Table S3).

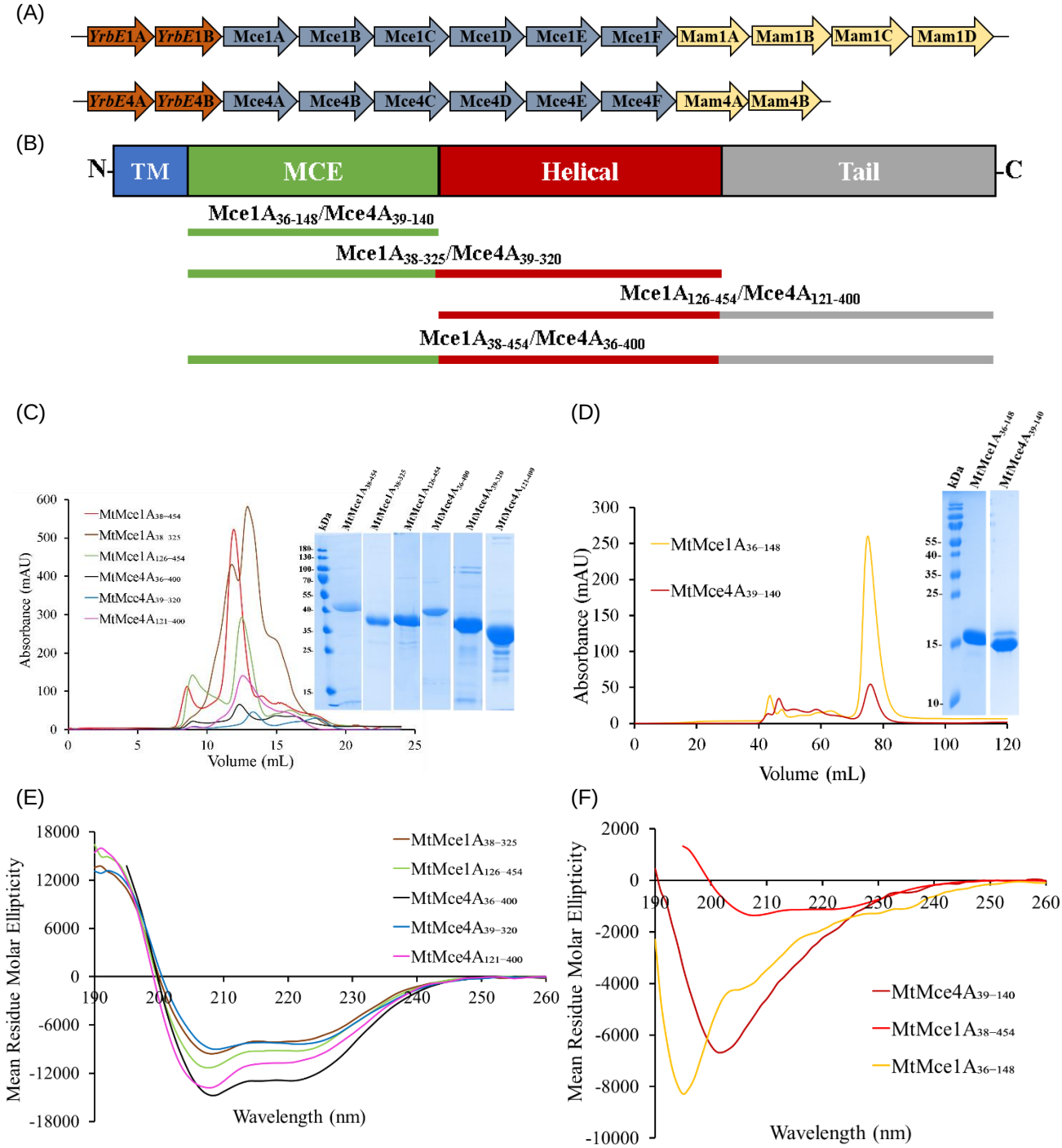


Fig.1: (A) Part of Mce1 and Mce4 operons of *Mtb* comprising permeases (YrbEA-B), SBPs (MceA-F) and Mam proteins. **(B)** Domain architecture of the individual Mce SBPs. The Mce SBPs are characterized by four domains named as transmembrane (TM), MCE, helical and tail domain. The constructs of MtMce1A and MtMce4A used in this study are mentioned below in the same color coding. **(C)** Size exclusion chromatography (SEC) elution profiles of MtMce1A$_{38-325}$, MtMce1A$_{126-454}$, MtMce1A$_{38-454}$ and MtMce4A$_{39-320}$, MtMce4A$_{121-400}$, and MtMce4A$_{36-400}$ on a 24 ml Superdex 200 10/300 column. The protein samples were analyzed on a 12 % SDS-PAGE (inset). **(D)** SEC elution profile of MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ on a 120 ml Superdex 75 HiLoad 16/600 column. The protein samples were analyzed on a 18% SDS-PAGE (inset) **(E)** The CD spectra of MtMce1A$_{38-325}$ (brown), MtMce1A$_{126-454}$ (green), MtMce4A$_{36-400}$ (black), MtMce4A$_{39-320}$ (blue) and MtMce4A$_{121-400}$ (pink). **(F)** The CD spectra of MtMce1A$_{38-454}$ (red), MtMce1A$_{36-148}$ (yellow) and MtMce4A$_{39-140}$ (maroon).

2.3 MtMce4A$_{39-140}$ crystallizes as a domain-swapped dimer

Structural studies were initiated on MtMce1A$_{38-454}$, MtMce4A$_{36-400}$ as well as soluble MCE domains MtMce1A$_{38-148}$ and MtMce4A$_{39-140}$. Despite extensive trials, only the MtMce4A$_{39-140}$ crystallized readily in several conditions in the space group $P6_5$. Given the low sequence identity of MtMce4A$_{39-140}$ with its homologous proteins (~15%), the structure of MtMce4A$_{39-140}$ was determined using SeMet-SAD to 2.9 Å resolution. Although the initial Mathew's coefficient calculations suggested the presence of 6-8 molecules in the asymmetric unit with a solvent content of about 43-57%, the solved structure showed that only four molecules are present in the asymmetric unit corresponding to a solvent content of about 71%. Interestingly, further refinement and model building of the structure revealed that the tetramer in the asymmetric unit is composed of two domain-swapped dimers (Fig. 2*A*). The domain-swapped dimer is formed by the extension of residues 107-141 from one molecule into the other molecule. The swapped region has two β-strands and an extended loop (Fig. 2*A*).

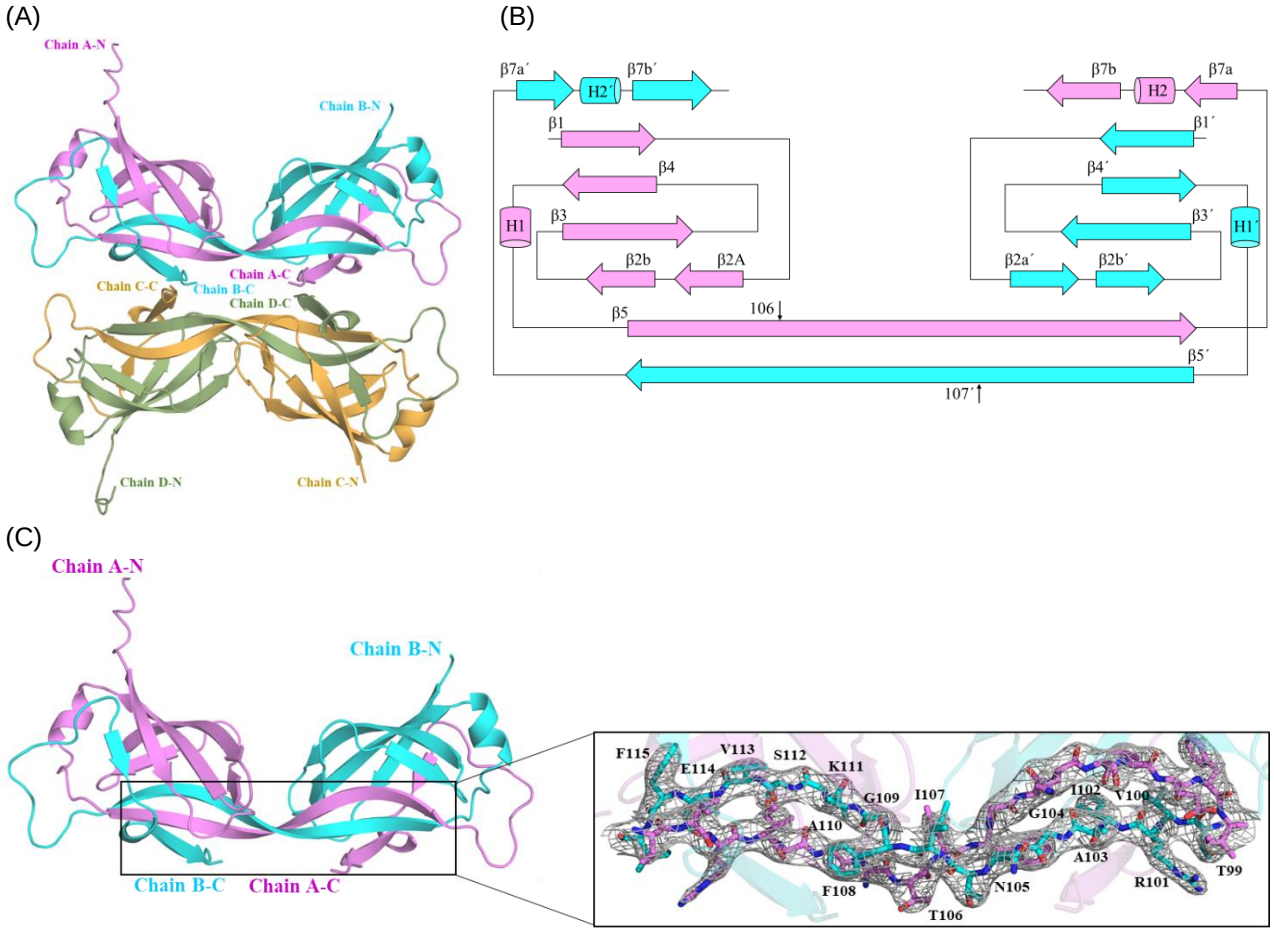


Fig.2: (A) Crystal structure of the MtMce4A$_{39-140}$ with 4 molecules in the asymmetric unit. **(B)** Topology of MtMce4A$_{39-140}$ domain-swapped dimer. The β-strands are shown as *arrows* and the helices as *cylinder*. The secondary structures of chain A and chain B are shown in *pink* and *cyan,* respectively. The secondary structure elements and residue numbers for chain B are indicated with prime (‘). The residues after the black vertical arrow are involved in the domain-swapping. **(C)** The domain-swapped dimer residues of β-5 and β-5´ are highlighted and shown in the inset. The 2Fo-Fc electron density maps contoured at 1.5 σ are shown in grey mesh. These residues are important for the arrangement of the domain-swapped dimer.

The topology diagram for the swapped dimer is shown in Fig. 2*B*. The residues involved in the formation of β-barrel are Thr40-Ser46 (β1), Leu52-Met54 (β2a), Lys59-Gly65 (β2b), Ile65-Ser74 (β3), Arg81-Asp87 (β4), Thr99-Thr106 (β5), Ile107-Ile116 (β5´) (considered as 6th β-strand), His131-Val132 (β7a´), Val137-Glu141

(β7b´). The residues from 107-141 are exchanged between the two monomers for completing the signature MCE-fold. The overall structure has visible electron density for all the residues corresponding to MtMce4A$_{39-140}$ except for the N-terminal residues 1-31, and C-terminal residues 143-146 corresponding to the vector region. The data collection and refinement statistics are mentioned in Table 1.

Table 1: Statistics of MtMce4A$_{39-140}$ structure

Dataset	**SeMet-labelled**	**Native**
Crystallization		
Protein storage buffer	50 mM MOPS, 350 mM NaCl, 10% Glycerol, pH 7.0	50 mM MOPS, 350 mM NaCl, 10% Glycerol, 5 mM DTT, pH 7.0
Protein concentration	7.5 mg ml^{-1}	7.5 mg ml^{-1}
Crystallization conditions	100 mM MES, 700 mM ammonium sulfate, pH 6.0	100 mM Sodium HEPES, 100 mM $LiCl_2$, 20% PEG 400, pH 7.5
Cryo-protectant	25% Glycerol	20% Ethylene Glycol
Temperature	22 °C	22 °C
Data Collection:		
Beam line	BioMax	BioMax
Wavelength (Å)	0.968	0.953
Detector	Eiger 16M Hybrid-pixel	Eiger 16M Hybrid-pixel
Detector distance (mm)	357.46	276.71
Oscillation range (°)	0.1	0.1
Data Processing:		
Space group	$P6_5$	$P6_5$
a, b, c (Å)	134.0, 134.0, 105.5	131.2, 131.2, 105.5
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution range (Å)	48-3.6(3.9-3.6)*	47.8-2.9 (3.0-2.9) *
R_{pim} (%)	0.11 (1.45)	0.06 (0.79)
Multiplicity	12.4 (12.5)	15.4 (15.6)
Wilson B-factor (Å)	108.4	66.3
Solvent content	72.90	71.74
Total number of reflections	155214 (36878)	355222 (57858)
Number of unique reflections	12505 (2961)	23016 (3714)
CC (1/2)	99.6 (23.0)	99 (43.6)
<I>/σ<I>	7.1 (1.2)	11 (1.4)
Completeness (%)	99.8 (99.5)	100 (99.9)
Refinement statistics		
R-work	0.2165	0.1947
R-free	0.2466	0.2348
Number of atoms Protein Water	 3271 -	 3276 14
Average B-factor (Å^2) Protein Water	 158.89 -	 92.22 79.08
Ramachandran statistics Allowed (%) Favored (%) Outliers (%)	 6 93.8 0.2	 3 97 0
R.M.S deviations Bond lengths (Å) Bond angles (˚)	 0.002 0.560	 0.002 0.422
PDB id	7AI2	7AI3

*Values in parenthesis are for the highest resolution shell.

2.4 MtMce1A and *MtMce4A are predominantly monomeric in solution*

The domain-swapped dimeric nature of MtMce4A$_{39-140}$ was unexpected. To further verify their oligomeric states in solution, SEC-Multi angle light scattering (SEC-MALS) studies were conducted for all the purified MtMce1A and MtMce4A domains. Interestingly, these studies showed that MtMce1A and MtMce4A domains were predominantly monomeric in nature (Table 2 & *SI Appendix*, Fig. S4, S5). The MtMce1A and MtMce4A domains purified in DDM showed two peaks in the elution profile corresponding to protein-detergent complex (PDC) and empty detergent micelle. Whereas, MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ which are soluble and purified without DDM has a single scattering peak, corresponding to the monomeric molecular mass (*SI Appendix*, Fig. S4, S5). As the SEC-MALS analysis showed that both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ are monomeric in solution and MtMce4A$_{39-140}$ is a domain-swapped dimer in the crystal structure, the oligomeric states of MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ were also determined by native mass spectrometry (MS) at two different concentrations (5 µM and 50 µM) to determine if there is a concentration-dependent dimerization. These studies further confirmed that both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ are monomeric in solution at both the concentrations (*SI Appendix*, Fig. S6, S7).

Table 2: Molecular mass of MtMce1A and MtMce4A domains as calculated from the SEC-MALS experiment.

Protein name	Theoretical monomeric molecular mass (kDa)	Protein-DDM conjugate (SEC-MALS) (kDa)	Protein (SEC-MALS) (kDa)	Empty DDM micelle (SEC-MALS) (kDa)
MtMce1A$_{36-148}$	16.9	-	16.40	-
MtMce1A$_{38-325}$	36.0	103.0	39.0	66.0
MtMce1A$_{126-454}$	38.7	144.0	40.0	58.0
MtMce1A$_{38-454}$	48.3	159.0	58.0	68.0
MtMce4A$_{39-140}$	15.0	-	15.0	-
MtMce4A$_{39-320}$	34.4	102.5	34.9	67.5
MtMce4A$_{121-400}$	34.6	128.9	44.7	65.1
MtMce4A$_{36-400}$	43.5	142.0	55.4	70.1

Interestingly, comparison of the secondary structure content of MtMce4A$_{39-140}$ calculated from the CD spectrum with the crystal structure showed higher β sheet content (39.04%) in crystal than from the CD spectra (28.1%, *SI Appendix*, Table S4) indicating that the protein is more structured in the crystallization condition. Moreover, thermal melting analysis of MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ showed that they undergo heat-induced conformational changes (*SI Appendix*, Supplementary Results, Fig. S8 and S9). It is possible that in the purified conditions, both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ are in non-native conformations and MtMce4A$_{39-140}$ attains native conformation in the crystallization buffer. Therefore, the purified MtMce4A$_{39-140}$ was exchanged in crystallization buffer and analyzed by SEC-MALS. Surprisingly, SEC-MALS analysis showed only the presence of monomeric MtMce4A$_{39-140}$ also in the crystallization buffer (*SI Appendix*, Fig. S10). However, dimer

formation was observed when MtMce4A$_{39-140}$ was heated slowly, up to 50 °C in the crystallization buffer (0.7 M ammonium sulfate) (*SI Appendix*, Fig. S10). We speculate that the, incubation of this protein solution with the crystallization solution at 22 °C has probably facilitated the protein to attain a more compact and native state. The dimer could have selectively crystallized owing to better crystal contacts when compared to the monomer, indicating that the domain-swapped dimer is possibly a crystallization artifact.

2.5 Comparison of MtMce4A$_{39-140}$ structure with *E. coli* and *A. baumannii* homologs

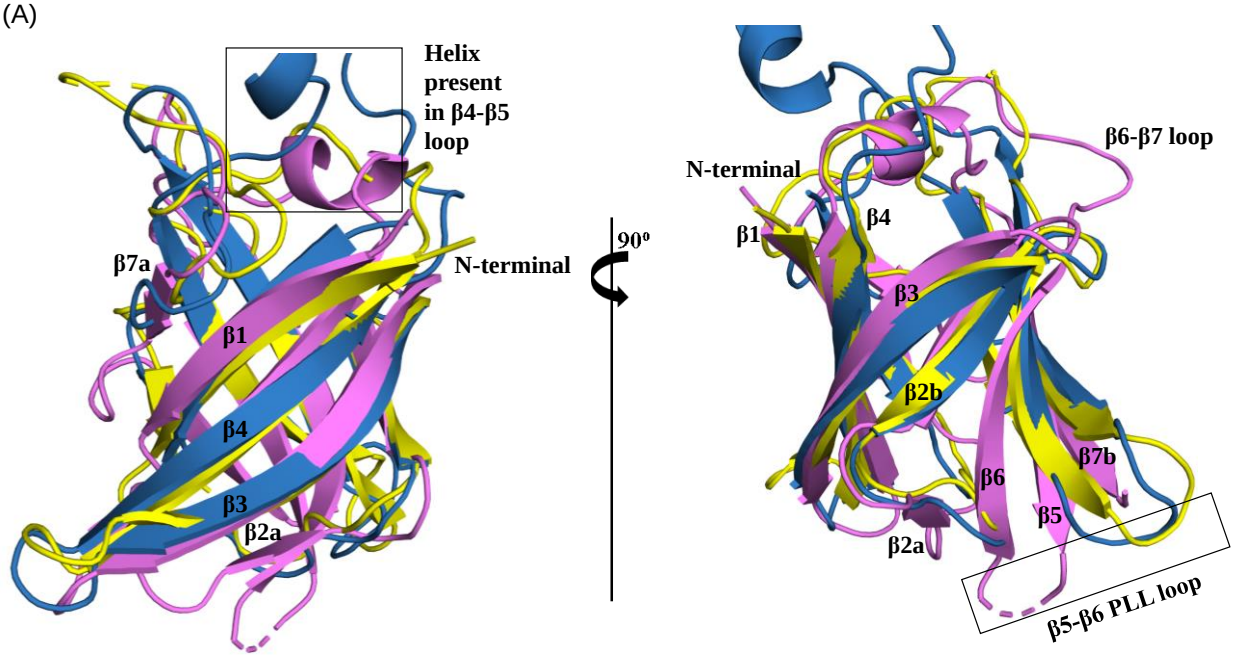
In spite of the initial non-native conformation of MtMce4A$_{39-140}$, the overall MCE fold with a seven-stranded β-barrel is conserved in the MtMce4A$_{39-140}$ crystal structure as also observed in the homologs from *E. coli* (EcMlaD, EcPqiB, and EcLetB) and *A. baumannii* MlaD (AbMlaD) (14-16, 18-20). As per our experimental data, the domain-swapped dimer formed by MtMce4A$_{39-140}$ is most likely a crystallization artifact. Therefore, we used the compact folded monomer for structural analysis and comparison. The compact monomer is formed by residues 32 to 106 from chain A and residues 107 to 145 from chain B.

Superposition of the MCE domain from MtMce4A on EcMlaD and AbMlaD yields a root mean square deviation (RMSD) of 1.70 Å and 2.62 Å, respectively (Fig. 3*A*). The overall topology of the protein is conserved with conformational differences mainly in the loop regions and a few other secondary structural elements. For example, the β2a-52 to 54 is present only in the MtMce4A$_{39-140}$ and not in EcMlaD and AbMlaD. The β4-β5 has an extra helix in MtMce4A$_{39-140}$ and AbMlaD (45 residue insertion) and this helix is absent in the EcMlaD. The β5 strand and the hydrophobic β5-β6 loop (also referred as Pore Lining Loop; PLL) involved in forming the hydrophobic central pore, has a different conformation in MtMce4A$_{39-140}$, which contrasts with the EcMlaD and AbMlaD. The β6-β7 loop in MtMce4A$_{39-140}$ is a proline-rich loop whereas it is lined with charged residues in EcMlaD and AbMlaD. In addition, the density of the β6-β7 loop is missing in EcMlaD crystal structure and is present in MtMce4A$_{39-140}$ and AbMlaD. The β7a-β7b is connected by a helix in MtMce4A$_{39-140}$, by a loop in EcMlaD and β7a is absent in AbMlaD. The homologous β7b strand is much smaller in EcMlaD and AbMlaD compared to Mce4A$_{39-140}$.

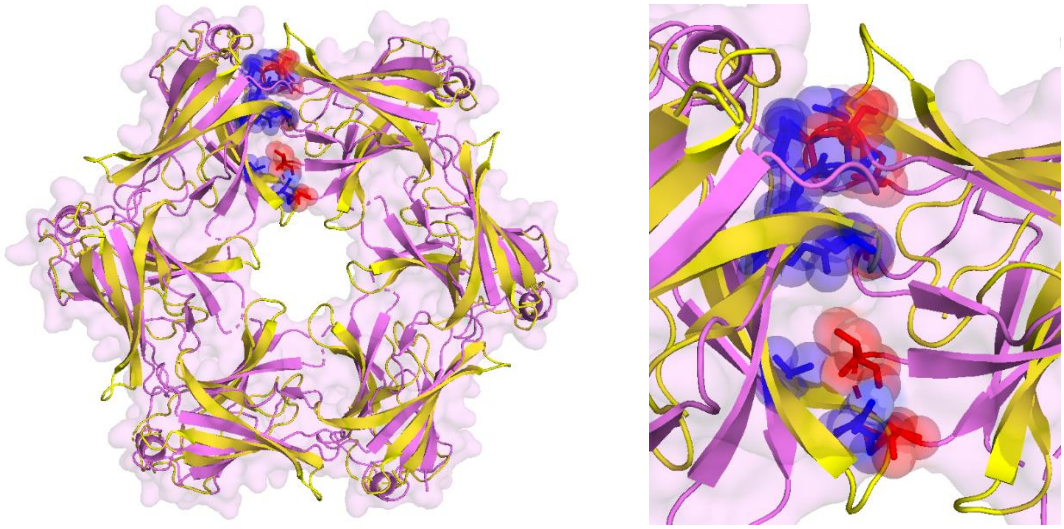
Similarly, MtMce4A$_{39-140}$ was superposed with MCE domains of EcPqiB1-3 and EcLetB1-7 monomers (*SI Appendix*, Fig. S11). The superposition showed that the β-barrel fold is conserved, and the observed differences are mainly in the loop regions. For example, the β2a-52 to 54 is unique in MtMce4A$_{39-140}$ and is absent throughout in EcPqiB1-3 and EcLetB1-7. The β3-β4 loop conformation present on the exterior surface varies amongst MtMce4A$_{39-140,}$ EcPqiB1-3 and EcLetB1-7. Noteworthy, the length of β3-β4 loop remains constant (4 residues) in all the MCE domains except the EcPqiB3, which has 18 residues in the loop. The PLL (β5-β6 loop) comprising the hydrophobic channel is much longer (16-27 residues) in EcPqiB1-2 and EcLetB1-7 when compared to the MtMce4A$_{39-140}$ (5 residues). We found that the PLL in EcPqiB3 has only 7 residues and it is the only MCE domain amongst EcPqiB and EcLetB, which share this feature with MtMce4A$_{39-140}$. Interestingly, the conformation of all the PLL varies throughout the MCE domains of EcPqiB1-3 and EcLetB1-7 (*SI Appendix*, Fig. S11). Furthermore, the β6-β7 loop in MtMce4A$_{39-140}$ has a different conformation as compared to EcPqiB1-3 and EcLetB1-7. Amongst the available Mce SBP structures and MtMce1/4A-F,

MtMce4A$_{39-140}$ has maximum number of proline residues in the β6-β7 loop. The role of this proline rich loop is still not understood.

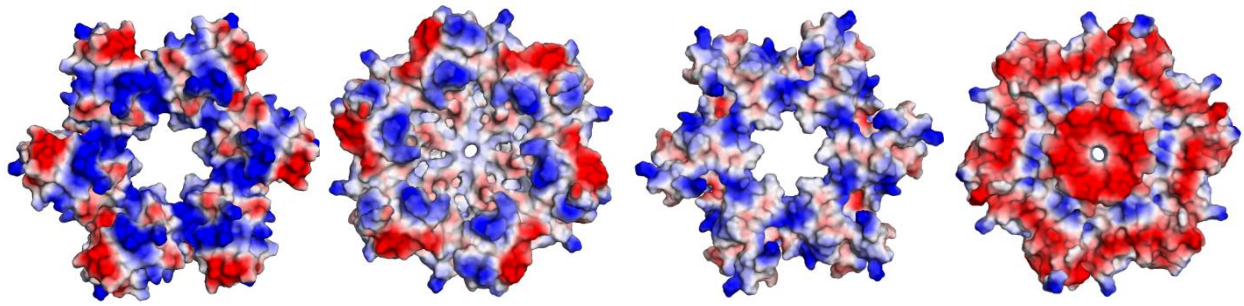
(A)



(B)



(C) (D)



MtMce4A$_{39-140}$ **EcMlaD** **MtMce4A$_{39-140}$** **EcMlaD**

Inner membrane facing **Periplasmic facing**

Fig 3: (A) Structural superposition of MCE domains of *Mtb* (MtMce4A$_{39-140}$;pink), *E.coli* (EcMlaD; yellow) and *A.baumannii* (AbMlaD; blue). **(B)** Cartoon representation of the hypothetical homohexamer of MtMce4A$_{39-140}$ (*pink*) generated based on EcMlaD homohexamer (*yellow*). The residues from two monomers (chains A and B) involved in the steric clashes are shown as sphere and sticks in blue and red colors. These clashes are between β2- β3 loop residues K61,Y62, R63 of chain A and β3- β4loop residues S76, G77, Q79 of chain B; between β5 strand residue A103 of chain A with β5- β6 loop residue I107 of chain B; between β6 strand residue E114 of chain A with β5- β6 loop residue A50 of chain B and between β7 strand residue L140 of chain A with β5- β6 loop residue T106 and I107 of chain B. **(C)** Electrostatic potential surface (inner membrane facing side) of the hypothetical homohexamer of MtMce4A$_{39-140}$ (left) and EcMlaD homohexamer (right). The inner membrane facing sides are comparable with each other with high positive charges. **(D)** Electrostatic potential surfaces (periplasmic facing side) of the hypothetical homohexamer of MtMce4A$_{39-140}$ (left) and EcMlaD homohexamer (right). The periplasmic side of MtMce4A$_{39-140}$ has more positively charged residues, which is in contrast to the highly negative surface of EcMlaD.

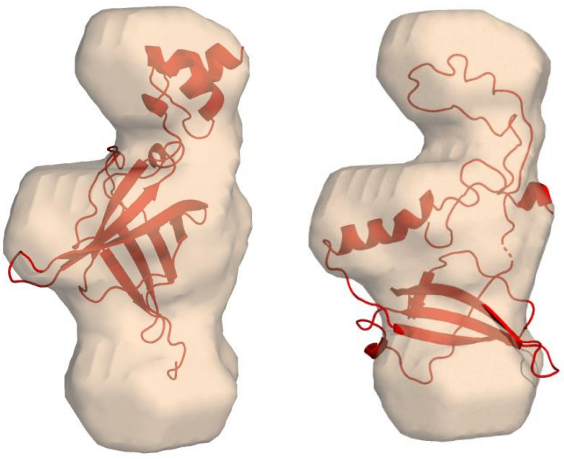
The monomeric nature of MtMce4A$_{39-140}$ is contrasting to the other Mce proteins (MlaD, PqiB, LetB) from *E. coli* and *A. baumannii,* which forms a homohexamer (14-16, 18-20). Based on the EcMlaD, we generated a hypothetical homohexamer of MtMce4A$_{39-140}$ (Fig. 3*B)*. Interestingly, the domain interface of the hypothetical homohexamer of MtMce4A$_{39-140}$ has multiple steric clashes, which will preclude the formation of homohexamers in MtMce4A (clashes between chain A and B are shown in Fig. 3*B*). These clashes are absent in EcMlaD, AbMlaD*,* EcPqiB1-3, and EcLetB1-7, where homohexamers have been formed.

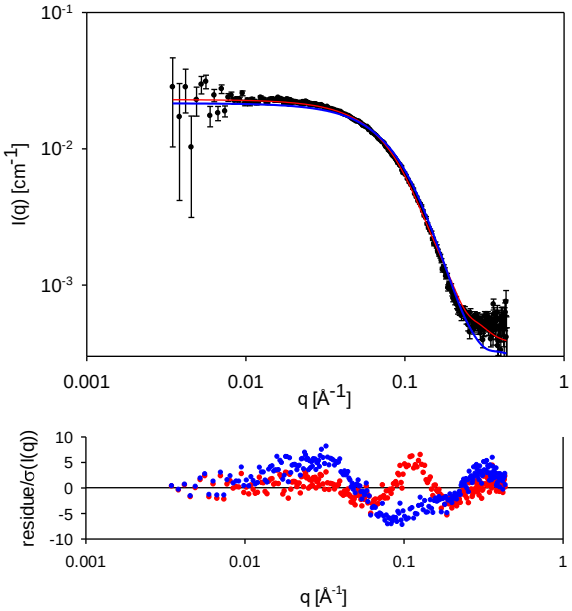
2.6 Elongated conformation of MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ in solution

SAXS experiments were performed to gain information on the shape and size of the purified MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$. Measured intensities, *I*(*q*), are displayed as a function of the modulus, *q*, of the scattering vector. Structural parameters calculated from the scattering intensities are given in the Table S5. The radius of gyration (R_g) and maximum interatomic distances (D_{max}) were determined to be 21.6 Å and 70 Å for MtMce1A$_{36-148}$, and 21.7 Å and 80 Å for MtMce4A$_{39-140,}$ respectively (*SI Appendix*,S12). Interestingly, the determined D_{max} of both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ is much higher than the maximum diameter of homologous EcMlaD (35 Å) pointing towards an elongated structure for both the proteins. Further, *ab initio* molecular shape reconstructed by DAMMIN indicate that both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ attains an elongated shape in the purified conditions (Fig. 4*A* and *B*). From SEC-MALS and SAXS, we know that the proteins exist as monomers in solution. Therefore the *ab initio* shape of MtMce4A$_{39-140}$ was fitted with the compact monomer (made from residues 32 to 106 from chain A and residues 107 to 145 from chain B) and elongated (similar to the monomer in a domain-swapped dimer) monomeric variation of MtMce4A$_{39-140}$. The agreement in terms of reduced chi-squared (χ^2) of compact and elongated model calculated against the experimental SAXS data was 10.0 and 2.0, respectively (Fig 4B). Similarly, in case of MtMce1A$_{36-148}$, a template-based model (obtained from Robetta) was used for fitting the SAXS data and the χ^2 for compact and elongated model were 14.0 and 11.0, respectively (Fig 4A), also here favoring slightly the elongated model. Further, the domain-swapped region of MtMce1A$_{36-148}$ was optimized by rigid-body refinement and this improved the χ^2 to 4.2. In summary, the elongated models fit relatively better than the compact model in both cases, suggesting that both MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$ attains an elongated shape in the purified conditions.

(A) MtMce1A$_{36\text{-}148}$

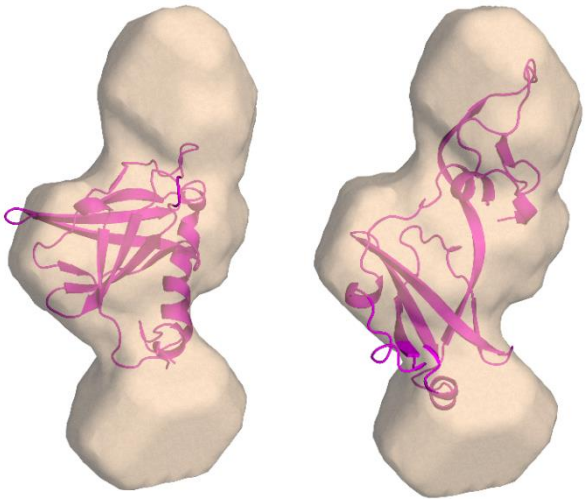
Compact Elongated

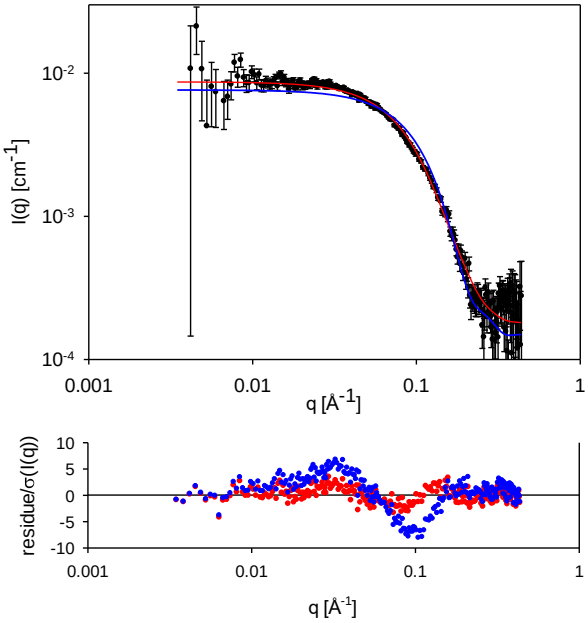




(B) MtMce4A$_{39\text{-}140}$

Compact Elongated





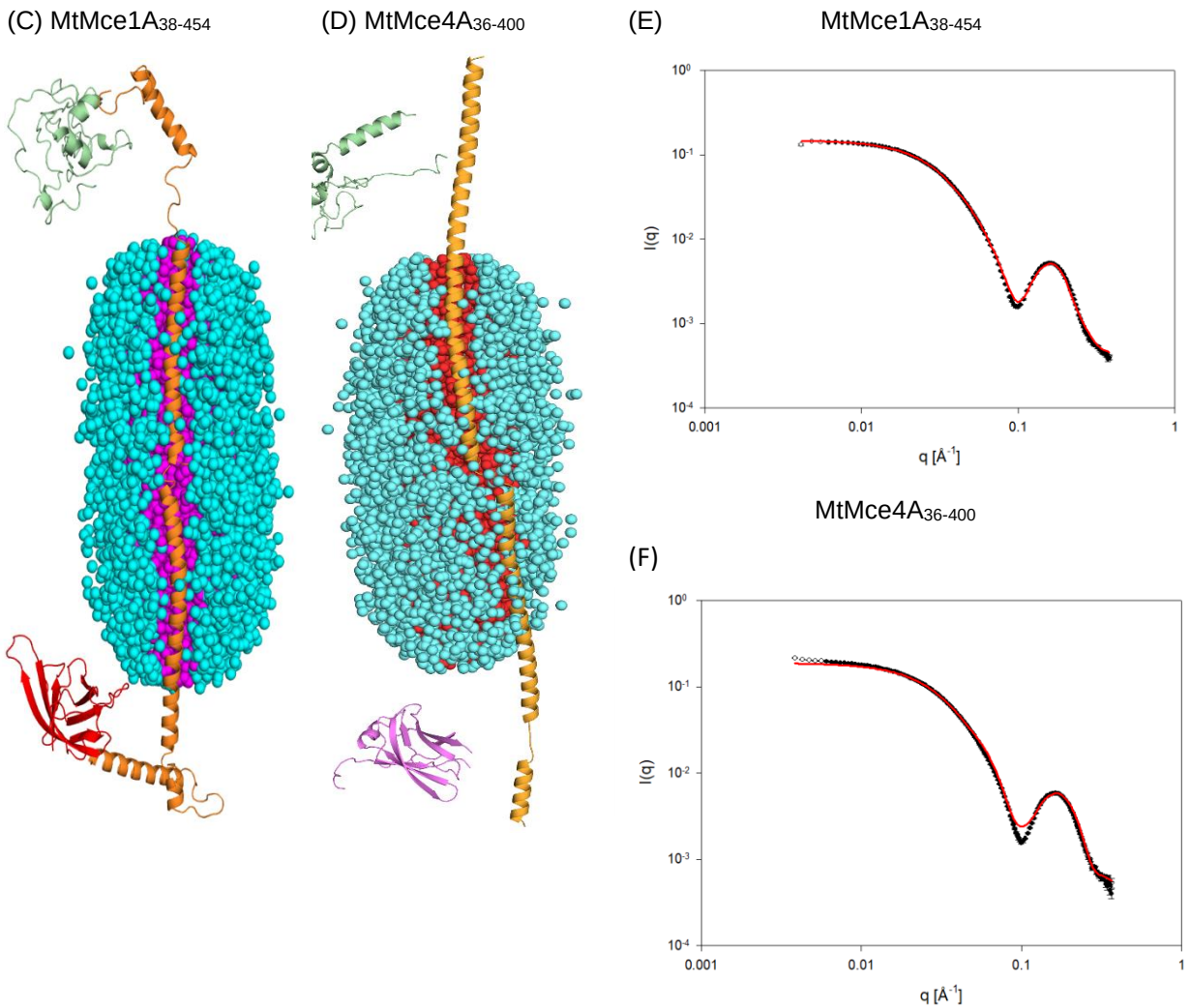


Fig. 4: (A) The *ab initio* shape generated by DAMMIN for MtMce1A$_{36-148}$ superposed on the elongated (left) and compact (right) monomeric models of MtMce1A$_{36-148}$. The corresponding fits of the experimental SAXS data (black) with the elongated (red) and compact (blue) monomers are shown below. **(B)** The *ab initio* shape generated by DAMMIN for MtMce4A$_{39-140}$ superposed on the elongated (left) and compact (right) monomeric models of MtMce4A$_{39-140}$. The corresponding fits of the experimental SAXS data (black) with the elongated (red) and compact (blue) monomers are shown below. **(C)** The model of MtMce1A$_{38-454}$ and **(D)** MtMce4A$_{36-400}$ interacting with the core of DDM micelle.The DDM molecules are represented as spheres and the protein as cartoon. **(E)** The fit (red line) of experimental SAXS data (*black dots*) with the proposed models of MtMce1A$_{38-454}$ and **(F)** MtMce4A$_{36-400}$.

2.7 Helical domain of MtMce1A and MtMce4A interacts with DDM core

In addition to the soluble MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$, SAXS measurements were also performed for the other MtMce1A and MtMce4A domain constructs in SEC-inline mode as they have been purified in the presence of DDM (*SI Appendix*, Table S6 and S7). The elution profile has two peaks, one for the PDCs and one for the empty micelles. The SAXS scattering data of the PDCs display a minima at a scattering vector modulus of 0.1 Å^{-1} followed by a broad bump, which reconfirms the presence of nearly intact detergent micelle together with the protein. To further analyze the SAXS data, models of MtMce1A and MtMce4A were generated and optimized as described in the methods section. In both MtMce1A and MtMce4A models, the helix turns back at residues E248 and D215, respectively, to form a coiled-coil structure, where the coiled-coil helices are held together by hydrophobic interactions and which brings the tail domain closer to the MCE domain. These models are referred as ‘‘coiled-coil models’’. A second variation of this model was also generated by opening

the helical domain to form an extended helix keeping the tail domain far away from MCE domain. This model is referred as “extended helical model” (Fig. 4*C* and *D*).

With our experimental data, it is clear that the MCE domain is soluble and the presence of helical domain required detergent for the purification. Therefore, the detergent micelle has to interact with the helical domain. Although the helical domain has a high number of hydrophobic amino acids, it also has polar residues which precludes the possibility of the helix completely inserted into the core of the micelle like a typical transmembrane protein. Furthermore, calculations of the SAXS intensity with the helix inserted into the core, show that SAXS intensity for these models smears out the minimum, so that it is not as deep as in the data. Therefore, a core-shell model of the detergent micelle was used where the helical domain is interacting with the surface of the core (dodecyl chains) of the micelle (21). On testing multiple micelle shapes, the best fit was obtained when using a superellipsoid shape with the long axis along the helical domain which maximizes the interaction of the helix with the core of the micelle. The micelle size (aggregation number) was initially estimated from SEC-MALS and SAXS scattering analysis (21) to be in the range of 125-200. However, in cases where the fits were not satisfactory, it was further varied in a reasonable range for obtaining the best fit to the SAXS data.

With these assumptions, both the coiled-coil as well as extended helical models for each of the MtMce1A and MtMce4A constructs were optimized (10 independent runs) together with the micelle with appropriate aggregation number to fit the SAXS data. For MtMce1A$_{38\text{-}325,}$ as well as MtMce4A$_{39\text{-}320}$ (MCE+Helical domains), both, the coiled-coil and extended, models showed convincing fits with χ^2 values ranging from 3-20. In case of MtMce1A$_{38\text{-}325,}$ the best fit to the data are very good with the low *q*, minimum and broad bump being very well reproduced (*SI Appendix*, Fig. S13). For MtMce4A$_{39\text{-}320}$, the best fit is very good in the whole range for the extended model, however, with a tendency that the model structure is slightly too large and the minimum in the model curve being slightly higher than the data (*SI Appendix*, Fig. S14). In contrast, for the best fit with the coiled-coil structure, the size of the model seems slightly too small and there are more systematic deviations around the minimum and bump (*SI Appendix*, Fig. S14). From the χ^2 values and these observations, it can be concluded that the extended model is best in terms of reproducing the data.

In case of MtMce1A$_{38\text{-}454}$ (MCE+Helical+Tail domains), the extended helical model (Fig. 4C) has a better fit with a χ^2 range of 15-24 compared to the coiled-coil model (χ^2 =22-50). The extended model fits the data well in the full *q* range with a small deviation around the minimum, where the model curve is not quite low enough (Fig. 4E). The coiled-coil model is too small with some deviations at low *q*, and the optimization compensates partly by displacing the MCE domain away from the Helical+Tail domain, leading to some disconnectivity of the structure (*SI Appendix*, Fig. S15). In case of MtMce4A$_{36\text{-}400}$, the data are not fitted well at the high *q* values for both models, although, the low *q* data fits better in the extended model, favoring the extended helical model (Fig. 4D). Similar to the MtMce1A$_{38\text{-}454}$ coiled-coil model, the MCE domain and the Helical+Tail domain gets disconnected in the MtMce4A$_{36\text{-}400}$ coiled-coil model favoring the extended model (Fig. 4F and *SI Appendix*, S15).

The Helical+Tail domain fits for MtMce4A$_{121-400}$ extended and coiled-coil models have similar χ^2 values and are in the range of 4.6-8.0. The extended model fits very well in the high and low *q* range and the coiled-coil has deviations at the high *q* data (*SI Appendix*, Fig. S16). Both these models have less deep minima with respect to the data. The MtMce1A$_{126-454}$ extended model fitting has a χ^2 range of 40-75, whereas, the coiled-coil model fit shows a χ^2 between 114-182. Similar, to the MtMce4A$_{121-400}$ models, the minima is also less deeper in MtMce1A$_{126-454}$ models. It could also be concluded that the extended models fit better in case of Helical+Tail domain (*SI Appendix*, Fig. S17). Additionally, we have to accept that the tail domain is unstructured with greater uncertainty in the structure prediction. This could also be a reason for poorer SAXS fits for all the constructs with the tail domain. The counting statistics of the data for the samples varies somewhat and therefore χ^2 values also varies, it is observed that the χ^2 values are often higher for data with good counting statistics. Therefore, we decided also to calculate R factors and weighted R factors as used in crystallography. R factors are dominated by the high intensities at low q, whereas weighted R factors are a normalized measure of (χ^2)^0.5. The determined values are both in the range 1-5 % (*SI Appendix*, Table S6, S7). They reveal that the deviation between data and fits are more similar than the χ^2 values suggest.

Considering the low resolution information in the SAXS data as well as the possible errors in the generated MtMce1A and MtMce4A models, which are partly based on structural predictions, our analysis gives the best possible explanation for the observed SAXS data in a qualitative and in a semi-quantitative manner. Among the coiled-coil and extended helical models, the fit for the extended helical models are relatively better overall. It is possible that the extended helical model is more favorable because the DDM micelle can interact with a large area of the helix in the extended helical model as compared to the smaller area in the coiled-coil model.

3. Discussion

We have classified the *Mtb* Mce1A-1F and Mce4A-4F SBPs into four different domains based on the secondary structure prediction, which shows the presence of a unique tail domain in SBPs from *Mtb*. Further characterization shows that the full length as well as all the domains of MtMce1A and MtMce4A remain as monomers in solution when purified individually. Moreover, only the MCE domain is soluble whereas the MCE+Helical and the Helical+Tail domains require detergents for their solubility. Further, SAXS analysis of MtMce1A, MtMce4A and their domains suggests that the individual protein may adopt the 'extended helical' model where the MCE and tail domains are far away from each other. Also, structural analysis of MtMce4A$_{39-140}$ suggests that the homohexamer cannot be formed at least in Mce4A due to multiple steric clashes. It is possible that in case of *Mtb*, the six MceA-F SBPs can interact with each other to form heterohexamers rather than homohexamers. Taken together, on the basis of all these data, we propose a model for the *Mtb* Mce-complex, where the MceA-F SBPs would exist as an "extended helical heterohexamer" (Fig. 5). In this model, the MCE domain from each SBP will form a ring like structure and the long helical domain will form a hollow channel. Inferring from the SAXS data, we speculate that the space occupied by the shell of the DDM micelle on the helical domain of MtMce1A and MtMce4A would be replaced by the interacting MceB-F SBPs. Additionally, the core of the DDM micelle could correspond to the hollow hydrophobic channel formed by the six interacting MceA-F SBPs in the physiological condition. The length of *Mtb* periplasmic space is ~140 Å

(22) and the calculated D_{max} from SAXS for the MCE+Helical+Tail domains is approximately 200 Å. Therefore, we propose that the MCE domain would be present in the periplasm, the helical domain would extend from the periplasm to the mycolic acid layer and the tail domain would extend to the surface of the cell wall. The tail domain reaching the surface of the cell wall could play a role in the substrate recognition, binding and transport of lipids into the 'channel'. Although the individual tail domains of MceA-F SBPs are predicted to be unstructured, it is possible that they acquire some structure upon interaction with each other. In the inner membrane side, the MCE domain heterohexamer would be connected to the YrbEAB permeases, which will form the transmembrane channel. The Mkl (MceG) interacting with the YrbEAB in the cytoplasmic side of the inner membrane would act as the driver of the transport in an ATP-dependent manner. Additionally, the Mam proteins and LucA would act as stabilizing proteins (23, 24) to form the Mce complex. Therefore, the Mce complex would form a direct connection between the cytoplasm and cell surface facilitating the transport of lipids across the cell wall (Fig. 5).

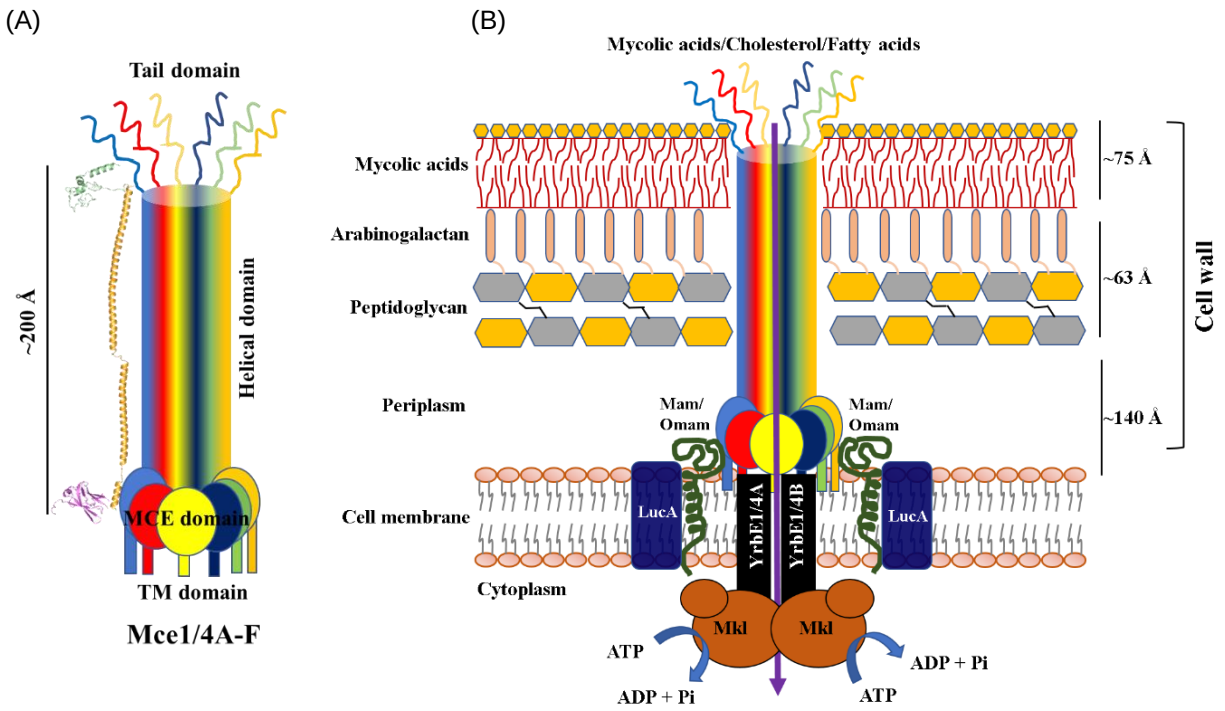


Fig. 5: (A) Proposed heterohexameric arrangement of *Mtb* Mce1/4A-F SBPs. The model for MtMce4A is represented in cartoon. **(B)** Schematic representation of Mce1/4 complex inside the cell wall of *Mtb*. The Mce1/4A-F SBPs would be embedded in the outer lipid layer, where the soluble N-terminus would be predicted to remain in the periplasmic space. The proposed permeases YrbE1/4AB would encompass the inner cell membrane connecting the Mce1/4A-F with Mkl. The Mkl would act as an ATPase and would convert the energy of ATP for the transport of lipid molecules across the membranes. The Mam/Omam and LucA would act as accessory proteins, involved in stabilizing the entire complex.

The central pore of all the reported Mce SBP hexamers are comprised of highly hydrophobic residues also known as PLL, which allows the transport of small hydrophobic lipid molecules across the membranes. The variation in the length of PLL depends on the transport mechanism followed by the particular Mce complex. For example, EcMlaD and AbMlaD has a smaller PLL (6 residues) because they follow a ferry-based transport mechanism. Whereas, the PLL is longer in EcPqiB and EcLetB (17-27 residues) in agreement with that they follow a tunnel-based lipid transport mechanism. It has been reported that the PLL of EcMlaD, AbMlaD,

EcPqiB$_{1-3}$, and EcLeTB$_{1-7}$ follows a pattern of ''ϕxxϕϕ'', where 'ϕ' denotes hydrophobic amino acid and 'x' represent any amino acid (16). This pattern has been followed in MtMce1A ($_{112}$ATTVF$_{116}$) and MtMce4A ($_{104}$GNTIF$_{108}$), but it does not align with other Mce SBPs from *Mtb* (*SI Appendix*, Fig. S18). Instead the other *Mtb* Mce SBPs follow a pattern of ''xxxϕϕ''in the PLL. This conserved "duo" of consecutive hydrophobic residues in the central pore of MtMce4A-4F SBPs is indicating towards formation of a heterohexameric channel. In addition, the helical domain of the MtbMceA-F SBPs also has high number of hydrophobic residues although a clear 'motif' is not observed.

The inner membrane-facing surface of homohexamers of EcMlaD, AbMlaD and 1st ring of EcPqiB and EcLetB has predominantly positively charged residues, which would interact with the negatively charged outer leaflet of the plasma membrane. The hypothetical MtMce4A$_{39-140}$ homohexamer also share the same surface electrostatics on the inner membrane facing side (Fig. 4*C*). The periplasmic side of the EcMlaD, AbMlaD homohexamer is lined with negatively charged residues, which are needed for the hydrophilic interactions. In contrast, the periplasmic side of MtMce4A$_{39-140}$ is positively charged and is more comparable to the 3rd ring of EcPqiB (Fig. 4*C*). EcPqiB$_3$ and MtMce4A$_{39-140}$, both are connected to a long helix and we predict that this helix is shielding the interaction of MCE domain with the periplasm. With the current knowledge of *Mtb* Mce SBPs and its comparison with homologous Mce SBPs, we propose that the lipid transport mechanism in *Mtb* Mce-complex would be tunnel-based as observed in EcPqiB and EcLetB. We believe that the main difference may be the involvement of the tail domain in substrate recognition and shuttling of the lipids from the outer surface of the *Mtb* whereas, the shuttling in EcPqiB and EcLetB probably involves other outer membrane proteins. Indeed, a higher resolution structure of MtMceA-F SBPs or the entire Mce complex is needed to further understand the structural arrangement as well as the lipid transport mechanism of the mycobacterial Mce complexes.

4. Materials and Methods

4.1 Biochemicals

The genomic DNA of *Mtb H37Rv* was purchased from ATCC. *Phusion* DNA polymerase and restriction enzymes used for cloning were purchased from Thermo Scientific (Massachusetts, USA) and New England BioLabs. The Ni-NTA chromatography resin was obtained from Qiagen (Hilden, Germany).

4.2 Cloning, expression and purification of MtMce1A-1F and MtMce4A-4F

Individual MtMce1A-1F and MtMce4A-4F genes were PCR amplified using *Mtb H37Rv* genomic DNA as the template with specific primers (*SI Appendix*, Table S8, S9). Each amplicon was cloned into pETM11 vector (EMBL) using restriction based cloning method resulting in an N-terminal His$_6$-Tag followed by a TEV protease site, *MceA-F* gene and C-terminal His$_6$ Tag. For protein expression, the plasmid was transformed into *E. coli* BL21-RIPL competent cells. Overnight cultures were grown at 30 ºC until O.D$_{600}$ reached 0.6, and the expression of the protein was induced with 0.5 mM IPTG (isopropyl β-D-thiogalactopyranoside) at 16 ºC overnight. Cells were harvested by centrifugation at 4000 g. The bacterial pellet was resupended in the desired lysis buffer with a suitable detergent (*SI Appendix*, Table S10). Cells were lysed using sonicator (30% amplitude, 3 min) and the lysate was centrifuged at 15000 g at 4 ºC for 30 min. The supernatant was then

filtered (0.45 µm; Millipore), and proteins were allowed to bind to Ni^{2+}-NTA matrix for 1 hour. The beads were washed, and bound proteins were eluted from the Ni-NTA column using 400 mM of Imidazole in the desired elution buffer (*SI Appendix*, Table S10). At this step the concentration of the detergent was reduced to 5 mM. The eluted protein was analyzed on a 12% or 18% SDS-PAGE and was concentrated (spin concentrator, molecular cut-off: 30 kDa, Millipore) and injected on a SEC column (Superdex 200 10/300 or Superdex 75 HiLoad 16/600 column, GE-Healthcare). The identity of the eluted protein was further confirmed with peptide mass finger printing analysis.

4.3 Expression and purification of MtMce1A and MtMce4A domains

Based on the secondary structure analysis, MtMce1A and MtMce4A domain constructs were generated. They were cloned in pETM11 using restriction free cloning methods: The constructs were named according to the secondary structural features: 1) MCE domain (MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$), 2) MCE+Helical+Tail domain (MtMce1A$_{38-454}$ and MtMce4A$_{36-400}$), 3) Helical+Tail domain (MtMce1A$_{126-454}$ and MtMce4A$_{121-400}$) 4) MCE+Helical domain (MtMce1A$_{38-325}$ and MtMce4A$_{39-320}$) and 5) Tail domain (MtMce4A$_{321-400}$). The expression and purification protocols were similar to the corresponding full-length proteins. Only the MCE domain constructs (MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$), were soluble without the use of detergents and different buffer was used for lysis and elution (*SI Appendix*, Table S10).

For Selenomethionine (SeMet) labelled MtMce4A$_{39-140}$, the construct was transformed to auxotrophic strain of *E. coli* (B834) and grown according to the protocol from Molecular Dimension (25). The expression and purification protocols were similar to the native MtMce4A$_{39-140}$. SeMet incorporation was confirmed by electrospray ionization liquid chromatography-mass spectrometry (ESI LC-MS) which showed 100% incorporation of SeMet into the protein.

4.4 SEC-MALS of Mce1A and Mce4A domains

SEC-MALS analysis of purified MtMce1A and MtMce4A domains was carried out using SEC coupled the miniDAWN™ TREOS (*Wyatt Technologies)*. Approximately 5-6 mg ml^{-1} of purified protein was loaded on the pre-equilibrated Superdex 200 10/300 using an autosampler and was run at a rate of 0.4 ml min^{-1} at 4 °C in a Shimadzu HPLC/FPLC system. The samples were then passed through a refractive index (RI) detector and subsequently through the MALS detector. The cumulative data collected for UV, MALS and RI, were analyzed using the *ASTRA* software (*Wyatt technologies*). The proteins were bound to detergent; therefore, protein-conjugate analysis method was incorporated for analysis. The detergent was considered as a modifier and the recommended dn/dc values of 0.1473 ml g^{-1} for n-Dodecyl β-D-maltoside (DDM) were used for the protein-conjugate analysis. The soluble MtMce1A$_{36-148}$ and MtMce4A$_{39-140}$, analysis was done without using the protein conjugate template.

For understanding the effect of heat and higher ionic strength on the oligomeric state of the protein, purified MtMce4A$_{39-140}$ was subjected to buffer exchange (0.1 M MES [2-(N-morpholino)ethanesulfonic acid], 0.7 M ammonium sulfate, pH 6.0) using a 10 kDa molecular mass cut-off amicon concentrator. Mce4A$_{39-140}$ was heated to 50 °C in a thermocycler, with initial one min incubation at 20 °C followed by 0.8 °C increase per min

up to 50 °C and a final incubation at 50 °C for one min. The heated protein was then centrifuged at 10000 g for 5 mins and the supernatant was injected to Superdex 200 10/300 column pre-equilibrated with a buffer containing 0.1 M MES and 0.7 M ammonium sulfate at pH 6.0 coupled to a MALS detector. The results were processed using both UV and RI signals as the source of concentration to calculate the molecular mass.

4.5 SAXS of MtMce1A and MtMce4A domains

Data collection: SAXS data for the purified Mce1A and Mce4A domains were collected on the B21 beamline of Diamond Light Source (DLS), UK. Data was collected based on the standard protocols for inline-SEC SAXS and batch mode measurement using a Pilatus 2M two-dimensional detector at a sample-detector distance of 4.014 m and at a wavelength of λ = 0.99 Å. Inline SEC-SAXS measurements were collected for domains purified in the presence of detergent DDM (MtMce1A$_{36-325}$, MtMce1A$_{121-454}$, MtMce1A$_{38-454}$, MtMce4A$_{39-320}$, MtMce4A$_{121-400}$, and MtMce4A$_{36-400}$) at an initial concentration of 5 mg ml^{-1}. Batch mode measurements were collected for MtMce1A$_{38-148}$, and MtMce4A$_{39-140}$ at 2 mg ml^{-1} and 1 mg ml^{-1} respectively, with bovine serum albumin (BSA) as a control. For each batch mode concentration, 25 frames were collected.

Data processing: Data processing and analysis was done using the ScÅtter and ATSAS software packages (26). The 2D data were averaged to give a 1D data set of intensity, $I(q)$, vs q, where q is the modulus of the scattering vector. The scattering of the buffer was subtracted from the protein scattering using the ScÅtter program. The data were rebinned using home-written software to be approximately equidistantly spaced on a logarithmic q scale. The radius of gyration (R_g), forward scattering I (0) and maximum particle distance (D_{max}) were calculated using PRIMUS. The hydrated volume V_p was computed from the Porod invariant. The molecular weight was calculated based on two methods: Volume of correlation (27) and SAXSMoW (28). *Ab initio* shape was generated using DAMMIN (29). For MtMce4A$_{39-140}$, crystal structure was used for superposition to the *ab initio* shape and the missing residues were generated using the template based modelling server "Robetta" (30, 31). Whereas, for MtMce1A$_{36-148}$, the entire model was generated with Robetta using the MtMce4A$_{39-140}$ structure as the template. The models were evaluated against the experimental data using a home-written program (32, 33). The helical domain for MtMce1A$_{38-325}$ and MtMce4A$_{39-320}$ were generated using *iTasser*. A summary of the data collection and analysis parameters are summarized in supplementary Table 3 and 4.

Detergent and protein model fitting for MtMce1A$_{38-325}$ and MtMce4A$_{39-320}$: The SAXS data for the complexes of DDM and the various constructs were also analyzed using home-written software. The program is based on the methods described in (21, 32-34). The DDM micellar structure is represented by Monte Carlo points in a tri-axial core-shell structure with super-ellipsoidal shape with shape parameter t = 3 (35), and the protein is represented by the atoms in the PDB structures. When the protein overlaps with the core-shell structure, the corresponding Monte Carlo points were removed. The volume of the core was estimated from the number of points and the point density, and the aggregation number was calculated by dividing the core volume by the volume of a C12 chain (353 Å^3). The shell contains both DDM headgroups and solvating buffer, and the thickness of the shell was fixed at 10 Å. In practice, the aggregation number was kept fixed and the length of the long axis and of one of the short axes optimized, whereas the length of the third axis was calculated from

these two and the aggregation number. The Monte Carlo points were assigned an excess scattering length corresponding to the electron densities of C12 tails and heads for points in the core and in the shell, respectively, taking into account the glycerol content of the buffer. Similarly, the excess scattering length of the atoms of the protein was adjusted taking into account the glycerol. The scattering of a hydration layer was added to the protein in the places where it is not in contact with the micelle. The protein structure was divided into three domains, namely MCE, Helical and Tail domains, to allow rigid-body refinement. The domains (MCE+Helical, Helical+Tail, and MCE+Helical+Tail, respectively for the three constructs) were connected by soft restraints as described in Vilstrup et al (33). The algorithm for generating the micelle including the estimates of the excess scattering length were checked by fitting a data frame from pure micelle from the elution profile, and it gave a satisfactory fit.

The SAXS data for all constructs have a deep minimum around q = 0.1 $Å^{-1}$, followed by a pronounced secondary maximum. This behavior is qualitatively very similar to that of pure DDM micelles, and the first tests revealed that such a q dependence could not be obtained when the protein penetrates significantly into the core of the micelles. Further tests showed that reasonable agreement with the SAXS data was obtained when the helix of the protein was along the long axis of the DDM micelle. Therefore, starting structures with this position were used in the optimizations. Additionally, a soft restraint that keeps the helix in contact with the micelle was introduced. The structure was optimized by random searches, initially with large amplitudes, which were gradually decreased during the optimization (33) or each structure, 10 independent runs were performed, each with 4000 cycles of optimization. The structure with the best agreement with the SAXS data in terms of reduced chi-squared, χ^2, was selected as the resulting structure. Initially the aggregation numbers were estimated from the SEC-MALS results, however, in some cases, this did not give good fits to the SAXS data. Therefore, the aggregation number was varied in a reasonable range for these cases.

4.6 Crystallization, data collection and structure determination of MtMce4A$_{39-140}$

Purified MtMce4A$_{39-140}$ and SeMet labelled-MtMce4A$_{39-140}$ were concentrated up to 7.5 mg ml^{-1} and used for all the crystallization experiments. Crystallization was performed using sitting drop vapor diffusion method at three different drop ratios (100:150, 150:150, 150:100 nl; protein: reservoir) at 22 °C. Crystals were observed in all three-drop ratios. Crystals were transferred to reservoir solution containing 25% of the appropriate cryoprotectant and flash-frozen in liquid nitrogen.

The data for both, the native MtMce4A$_{39-140}$ and SeMet-MtMce4A$_{39-140}$ crystals were collected at Biomax beamline, MaxIV (Lund, Sweden). The data processing and scaling was done using XDS (36) and AIMLESS (37), respectively, which suggested that the space group was P6$_1$ or P6$_5$.The SeMet-MtMce4A$_{39-140}$ structure was solved using SeMet-SAD phasing using the CRANK2 (38) pipeline. During hand-determination, the space group P6$_5$ was chosen based on figure of merit (FOM). The model obtained from CRANK2 pipeline was further build manually and the resulting R-free/R-work was 0.37 and 0.34 (with total 20 heavy atom sites used).

This refined model generated from SeMet-MtMce4A$_{39-140}$ was used as the initial model for expert mode molecular replacement (Expert-MR) of native MtMce4A$_{39-140}$ (39). The output from Expert-MR was used as an

initial model for Autobuilding in Phenix (40). Further model building was done manually. Final refinement steps gave an approximate R-free and R-work of 0.23 and 0.19 respectively. This final model of native-MtMce4A$_{39-140}$ was then again used to refine the SeMet-MtMce4A$_{39-140}$ data, which gave a final R-free and R-work of 0.24 and 0.21 respectively.

4.7 Circular Dichroism (CD) spectroscopy with MtMce1A and MtMce4A domains

The MtMce1A and MtMce4A domains were diluted in water to obtain a lower buffer and salt concentration. The protein concentration used for CD measurements (Chirascan CD spectrophotometer, Applied Photophysics, Surrey, U.K.) was 0.05 mg ml^{-1}. The secondary structure analysis of MtMce1A and MtMce4A domains purified with DDM and without DDM using CDNN and BestSel, respectively (41, 42) softwares. For the determination of the thermal melting temperature (T_m), the sample was heated at a rate of 1 °C per min from 22 °C to 92 °C. The melting curves were calculated by comparing the spectra from 190 to 280 nm with the global fit analysis protocol as implemented in the Global3 software.

Author Contributions:

P.A. and D.S. performed the work on Mce4A-4F and Mce1A-1F proteins, respectively. P.A. and R.V. determined the crystal structure of Mce4A domain. J.S.P. performed the SAXS analysis and simulations for Mce1A and Mce4A domains. M.J.H., R.S. and A.V.M. assisted in the cloning and/or purification of Mce1C & Mce1E, Mce1B and Mce1D, respectively. M.L. and J.J. conducted the native mass spectrometry experiments. L.W.R contributed in the discussion, manuscript writing and co-supervision. P.A., D.S., and R.V. designed the experiments, analyzed the data and wrote the manuscript with comments from all authors. R.V. conceptualized and supervised the work and acquired funding.

Acknowledgments

This work was supported by funding from Jane and Aatos Erkko foundation, Sigrid Juselius Foundation, I4Future doctoral program (MSCA-COFUND by Horizon 2020 European Union, grant agreement number 713606) and Biocenter Oulu. R.V. also acknowledges Academy of Finland for the Academy Research Fellowship (319194). MaxIV, DLS, ESRF and iNEXT (funded by Horizon 2020 European Union, grant agreement number 653706, and 730872) are acknowledged for their support in X-ray and SAXS data collection. We acknowledge the support of Biocenter Oulu (proteomics and protein analysis, structure biology and DNA sequencing core facilities). We acknowledge the support of Dr. Hongmin Tu for CD analysis.

Database Depositions

The X-ray structure for SeMet-MtMce4A$_{39-140}$ and MtMce4A$_{39-140}$ has been deposited in the PDB with the accession codes 7AI2 and 7AI3, respectively.

References:

1. Organisation WH (2018) Global tuberculosis report WHO Press, Geneva, Switzerland,.
2. Pandey AK & Sassetti CM (2008) Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* 105(11):4376-4380.

3. Camus JC, Pryor MJ, Medigue C, & Cole ST (2002) Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology (Reading, England)* 148(Pt 10):2967-2973.
4. Cole ST*, et al.* (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393(6685):537-544.
5. Arruda S, Bomfim G, Knights R, Huima-Byron T, & Riley LW (1993) Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. *Science* 261(5127):1454-1457.
6. El-Shazly S, Ahmad S, Mustafa AS, Al-Attiyah R, & Krajci D (2007) Internalization by HeLa cells of latex beads coated with mammalian cell entry (Mce) proteins encoded by the mce3 operon of Mycobacterium tuberculosis. *Journal of Medical Microbiology* 56(9):1145-1151.
7. Saini NK*, et al.* (2008) Characterization of Mce4A protein of Mycobacterium tuberculosis: role in invasion and survival. *BMC microbiology* 8:200.
8. Perkowski EF*, et al.* (2016) An orphaned Mce-associated membrane protein of Mycobacterium tuberculosis is a virulence factor that stabilizes Mce transporters. *Mol Microbiol* 100(1):90-107.
9. Casali N & Riley LW (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* 8:60.
10. Joshi SM*, et al.* (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc Natl Acad Sci U S A* 103(31):11760-11765.
11. Marjanovic O, Iavarone AT, & Riley LW (2011) Sulfolipid accumulation in Mycobacterium tuberculosis disrupted in the mce2 operon. *J Microbiol* 49(3):441-447.
12. Sassetti CM & Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 100(22):12989-12994.
13. Shimono N*, et al.* (2003) Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon. *Proc Natl Acad Sci U S A* 100(26):15918-15923.
14. Ekiert DC*, et al.* (Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. *Cell* 169(2):273-285.e217.
15. Kamischke C*, et al.* (2019) The Acinetobacter baumannii Mla system and glycerophospholipid transport to the outer membrane. *eLife* 8:e40171.
16. Isom GL*, et al.* (2020) LetB Structure Reveals a Tunnel for Lipid Transport across the Bacterial Envelope. *Cell* 181(3):653-664.e619.
17. Sutcliffe IC & Harrington DJ (2004) Lipoproteins of Mycobacterium tuberculosis: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiology Reviews* 28(5):645-659.
18. Liu C, Ma J, Wang J, Wang H, & Zhang L (2020) Cryo-EM Structure of a Bacterial Lipid Transporter YebT. *Journal of molecular biology* 432(4):1008-1019.
19. Mann D*, et al.* (2020) Structural basis for lipid transport by the MLA complex. *bioRxiv*:2020.2005.2030.125013.
20. Coudray N*, et al.* (2020) Structure of MlaFEDB lipid transporter reveals an ABC exporter fold and two bound phospholipids. *bioRxiv*:2020.2006.2002.129247.
21. Døvling Kaspersen J*, et al.* (2014) Low-resolution structures of OmpA·DDM protein-detergent complexes. *Chembiochem : a European journal of chemical biology* 15(14):2113-2124.

22. Vincent AT, *et al.* (2018) The Mycobacterial Cell Envelope: A Relict From the Past or the Result of Recent Evolution? *Frontiers in Microbiology* 9(2341).
23. García-Fernández J, Papavinasasundaram K, Galán B, Sassetti CM, & García JL (2017) Unravelling the pleiotropic role of the MceGATPase in Mycobacterium smegmatis. *Environmental Microbiology* 19(7):2564-2576.
24. Nazarova EV, *et al.* (2017) Rv3723/LucA coordinates fatty acid and cholesterol uptake in Mycobacterium tuberculosis. *eLife* 6:e26969.
25. Ramakrishnan V, Finch JT, Graziano V, Lee PL, & Sweet RM (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362(6417):219-223.
26. Franke D, *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography* 50(Pt 4):1212-1225.
27. Rambo RP & Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496(7446):477-481.
28. Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, & Polikarpov I (2019) SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Science* 28(2):454-463.
29. Svergun DI (1999) Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal* 76(6):2879-2886.
30. Raman S, *et al.* (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 77 Suppl 9(0 9):89-99.
31. Song Y, *et al.* (2013) High-resolution comparative modeling with RosettaCM. *Structure (London, England : 1993)* 21(10):1735-1742.
32. Steiner EM, *et al.* (2018) The structure of the N-terminal module of the cell wall hydrolase RipA and its role in regulating catalytic activity. *Proteins* 86(9):912-923.
33. Vilstrup J, *et al.* (2020) Crystal and solution structures of fragments of the human leucocyte common antigen-related protein. *Acta Crystallogr D Struct Biol* 76(Pt 5):406-417.
34. Calcutta A, *et al.* (2012) Mapping of unfolding states of integral helical membrane proteins by GPS-NMR and scattering techniques: TFE-induced unfolding of KcsA in DDM surfactant. *Biochimica et biophysica acta* 1818(9):2290-2301.
35. Maric S, Lind TK, Lyngsø J, Cárdenas M, & Pedersen JS (2017) Modeling Small-Angle X-ray Scattering Data for Low-Density Lipoproteins: Insights into the Fatty Core Packing and Phase Transition. *ACS Nano* 11(1):1080-1090.
36. Kabsch W (2010) XDS. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):125-132.
37. Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography* 69(Pt 7):1204-1214.
38. Skubák P & Pannu NS (2013) Automatic protein structure solution from weak X-ray data. *Nature Communications* 4(1):2777.
39. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of applied crystallography* 40(Pt 4):658-674.

40. Liebschner D, *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 75(Pt 10):861-877.
41. Micsonai A, *et al.* (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic acids research* 46(W1):W315-W322.
42. Micsonai A, *et al.* (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences* 112(24):E3095-E3103.