MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine

Chih-Chia Su[1,Ψ], Philip A. Klenotic[1,Ψ], Jani Reddy Bolla[2,Ψ], Georgiana E. Purdy[3], Carol V. Robinson[2], and Edward W. Yu[1,*]

[1]Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA.

[2]Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, UK.

[3]Department of Molecular Microbiology and Immunology, Oregon Health and Science University, Portland, OR 97239, USA.

[Ψ]C.S., P.A.K. and J.R.B. contribute equally to this work.

[*]To whom correspondence should be addressed. Email: edward.w.yu@case.edu

Abstract

The cell envelope of *Mycobacterium tuberculosis* is notable for the abundance of mycolic acids (MAs), which are essential to mycobacterial viability, and other species-specific lipids. The mycobacterial cell envelope is extremely hydrophobic, contributes to virulence and antibiotic resistance. Yet, exactly how fatty acids and lipidic elements are transported across the cell envelope for cell wall biosynthesis is unclear. Mycobacterial membrane protein Large 3 (MmpL3) is essential and required for transport of trehalose monomycolates (TMMs), precursors of MA containing trehalose dimycolates (TDM) and mycolyl arabinogalactan peptidoglycan (mAGP), but the exact function of MmpL3 remains elusive. Here, we report a high-resolution crystal structure of *M. smegmatis* MmpL3, revealing a monomeric molecule that is structurally distinct from all known bacterial membrane proteins. A previously unknown MmpL3 ligand, phosphatidylethanolamine (PE), was discovered inside this transporter. We also show, via native mass spectrometry, that MmpL3 specifically binds both TMM and PE, but not TDM, in the micromolar range. These observations provide insight into the function of MmpL3 and suggest a possible role for this protein in shuttling a variety of lipids to strengthen the mycobacterial cell wall.

Introduction

Tuberculosis (TB) is the leading cause of mortality as a result of an infectious agent, exceeding both malaria and HIV (1, 2). In 2016, there were an estimated 10.4 million new cases of active TB and 1.7 million patients died from active TB infection (2). The causative agent, *Mycobacterium tuberculosis* (Mtb), has now infected more than one third of the world's population, and the emergence of multidrug resistant TB (MDR-TB) presents an increasingly difficult therapeutic challenge, so much so that MDR-TB is now the main cause of death due to antimicrobial resistance. Unfortunately, the lethality of TB combined with its multidrug-resistant capacity has now transformed this long-neglected tropical disease into a global health priority.

The unique architecture of the mycobacterial cell wall plays a predominant role in Mtb pathogenesis. This complex cell wall structure supports membrane stability and provides a barrier against the host environment and antibiotics, thereby contributing to intrinsic resistance of the organism. The outer membrane of Mtb is characterized by the abundance of very long chain mycolic acids (MAs). MA are transported across the inner membrane as trehalose monomycolates (TMMs), then are either covalently linked to the arabinogalactan-peptidoglycan layer as mycolyl arabinogalactan peptidoglycan (mAGP) or incorporated into trehalose dimycolates (TDM) which comprises the majority of the outer leaflet of this membrane. The outer leaflet also contains other non-covalently associated lipids, such as phthiocerol dimycocerosates and sulfolipids (3). This unique construction makes the outer membrane of Mtb very rigid and extremely impermeable to a wide range of compounds, including many antibiotics (3). Additionally, the surface-exposed lipids are immunomodulatory and play a crucial role in host-pathogen interactions (4-9).

TMM biosynthesis and its transport are essential for Mtb, as it is the precursor of both mAGP and TDM in the outer leaflet of the cell envelope. TMMs are synthesized in the cytoplasm via a highly conserved and well-characterized pathway (10), targeted by the first-line anti-TB drug isoniazid (11). However, exactly how these TMM is transported across the membrane from the cytosol to participate in mycobacterial cell wall remodeling is not well understood. Recent work demonstrated that mycobacterial membrane protein large (MmpL) transporters are critical for mycobacterial physiology and pathogenesis by shuttling fatty acids and lipid components to the mycobacterial cell wall. Mtb possesses 13 MmpL transporters (12). These membrane proteins share sequence similarity with the resistance-nodulation-cell division (RND) superfamily of transporters (13). Although MmpL3 is absolutely essential *in vitro*, there is strong evidence that MmpL4, MmpL5, MmpL7, MmpL8, MmpL10 and MmpL11 are all required for full virulence and the growth of mycobacteria in mouse lungs (14-18). Interestingly, of the 13 MmpLs, only MmpL3 is shown to export TMM (18, 19). Using a spheroplast system, a recent study strongly indicated that MmpL3 is a TMM flippase (20). This biochemical analysis is in line with genetic studies that determined that MmpL3 expression is necessary for Mtb survival and that depletion of MmpL3 in *M. smegmatis* results in the accumulation of TMMs concomitant with a reduction in levels of mAGP and TDM (18, 19). Thus, MmpL3 has now been shown to be essential for the biosynthesis of mAGP and TDM, as well as their incorporation into the mycobacterial cell wall, with both processes critical for mycobacterial replication and viability (21, 22). Collectively, studies from both Mtb and *M. smegmatis* have established the importance of MmpL3 in membrane biosynthesis and TB viability as well as being a target of several potent anti-TB agents.

As MmpL3 is absolutely essential for mycobacterial cell wall biogenesis, we decided to elucidate the molecular mechanism of lipid translocation across the membrane via this transporter. We here present a crystal structure of *M. smegmatis* MmpL3 to a resolution of 2.79 Å. Unexpectedly, a phosphatidylethanolamine (PE) molecule was found to be associated within a partially buried region of this protein. Using native mass spectrometry, we observed that MmpL3 is capable of binding PE and TMM, but not TDM, with dissociation constants within the micromolar range. Our work provides molecular insights into the mechanism of lipid transport via the MmpL3 membrane protein and suggests a possible role for MmpL3 of shuttling different lipids across the membrane for cell wall biogenesis.

Results

Overall structure of the *M. smegmatis* MmpL3 transporter

To obtain the structural information of MmpL3, we cloned the full-length *M. smegmatis* MmpL3 transporter, which contains 1,013 amino acids, into the *E. coli* expression vector pET15b, with a 6xHis tag at the C-terminus to generate pET15bΩ*mmpL3*. This MmpL3 protein was overproduced in *E. coli* BL21(DE3)*ΔacrB* cells and purified using Ni^{2+}-affinity and Superdex 200 size exclusion columns. Native mass spectrometry (native MS) shows that the MmpL3 transporter is monomeric in detergent solution (Fig. 1A). We also cloned and overproduced MmpL3 in *M. smegmatis* mc^2155 cells. Again, the native mass spectra indicate that this purified MmpL3 protein is a monomer in detergent solution (Fig. 1B).

Crystals of MmpL3 were obtained using vapor diffusion. The best crystal diffracted X-rays to a resolution of 3.30 Å (Table S1 and Fig. S1). However, these crystals were difficult to reproduce and most of the protein samples could not even be crystallized. We then harvested

several of these protein crystals and washed them with the crystallization precipitant solution. SDS-PAGE analysis indicated that the MmpL3 protein within these crystals were ~25 kDa smaller than the full-length protein. Since the C-terminal region (residues 733-1013) of MmpL3 contains a large number of prolines (52 proline residues), it is likely that this C-terminal sequence is intrinsically unstructured and unstable in solution.

We then used native MS to determine the size of the full-length MmpL3 protein. Surprisingly, the data indicates that the protein sample contains three major fragments of 806, 776 and 763 amino acids, respectively (Fig. 1C). These three fragments are shorter than the full-length protein, which is composed of 1013 residues. A detailed analysis indicates that these three fragments contain residues 1-806, 1-776 and 1-763 of MmpL3. These protein fragments possess the full transmembrane region, but the C-terminal end was absent. Since it has been reported that this C-terminal domain is not essential for the function of MmpL3 (23), we decided to remove this C-terminal sequence and produce the truncated MmpL3 transporter, designated to be $MmpL3_{773}$ (residues 1-773), to improve crystal quality for structural determination. Our native MS data indicate that the $MmpL3_{773}$ protein also exists as a monomer, which presents as single protein fragment suitable for crystallization (Fig. 1D).

The best $MmpL3_{773}$ crystal diffracted X-ray to 2.79 Å resolution. We determined its structure using single anomalous dispersion (SAD) (Table S1, Fig. 2 and Fig. S2). We then used the structure of $MmpL3_{773}$ as the template to resolve the structure of full-length MmpL3, where we found that the C-terminal residues 753-1013 were missing. The conformation of this MmpL3 structure is identical to that of $MmpL3_{773}$. Superimposition of the two structures gives a root-mean-square-deviation (r.m.s.d.) of 0.6 Å (for 722 Cα atoms) (Fig. S3). Although MmpL3 belongs to a subclass of the RND superfamily, its three-dimensional topology is unique and very

different from the existing structures of RND transporters, including AcrB (24, 25), MexB (26), CusA (27, 28), MtrD (29), CmeB (30) and HpnN (31). Thus, the available structural information of RND proteins cannot be used to understand the function of this transporter.

The $MmpL3_{773}$ molecule consists of 12 transmembrane helices (TMs 1-12) and two periplasmic loops (loops 1 and 2), which create the periplasmic domain (Fig. 2A and B). Loop 1 is located between TMs 1 and 2, and loop 2 is found between TMs 7 and 8. These two loops contribute to generate subdomains PD1 and PD2 in the periplasm. Based on our structural information, the proline-rich C-terminal residues 733-1013 should form a cytoplasmic domain (CD) of MmpL3.

The N-terminal and C-terminal halves of $MmpL3_{773}$ are assembled in a twofold pseudo-symmetrical fashion. These two halves can be superimposed with a r.m.s.d. of 2.6 Å (for 294 Cα atoms) (Fig. S4). PD1 is composed of four α-helices and three β-strands. The majority of the PD1 amino acids come from loop 1. However, residues 437-448 of loop 2 also contribute to form helix α5 of PD1. PD2 constitutes three α-helices and three β-strands. Like PD1, the PD2 amino acids mainly arise from loop 2, but residues 49-60 of loop 1 participate in the formation of helix α1 of this periplasmic subdomain. The crossover of these two periplasmic loops allows for the two subdomains, PD1 and PD2, to be spatially adjoined within the periplasm. This structural feature is in good agreement with the finding that the two periplasmic loops closely interact with each other (32). There is a long flexible linker of 12 residues connecting the C-terminal end of TM1 and α1 of PD2. Similarly, a long, 15 residue flexible loop is found to link the C-terminal end of TM7 and α5 of PD1 together. The presence of these long linkers suggests that the periplasmic domains of MmpL3 are quite flexible in nature. The TMs are membrane embedded, but both TM2 and TM8 are significantly longer and protrude into the periplasmic region. These

two TMs directly tether PD1 and PD2, respectively, and form part of the periplasmic structure of the protein.

The MmpL3$_{773}$-phosphatidylethanolamine complex

Surprisingly, the MmpL3$_{773}$ molecule forms a channel-like cavity spanning the outer leaflet of the inner membrane up to the periplasmic domain (Fig. 2C). The beginning of this cavity is generated by a hydrophobic pocket created by TMs 7-10. This pocket opens to the outer leaflet of the inner membrane and periplasmic space. However, the majority of this cavity is found at the center of the periplasmic domain surrounded by the secondary structures of PD1 and PD2. This large space potentially constitutes a binding site for TMM.

Unexpectedly, two large extra electron densities were found within the structure of MmpL3$_{773}$. One of these densities was observed in the cavity surrounded by TMs 7-10. The shape of this extra density is compatible with a DDM detergent molecule (Fig. 3A). This is not surprising as we solubilized, purified and crystallized the MmpL3$_{773}$ transporter in solutions containing DDM detergent. Within 4.5 Å of the bound DDM, there are 13 residues, including L422, S423, L424, N524, A527, Q554, I557, F561, L564, P565, A568, L600 and I636 that provide hydrophobic and electrostatic interactions for anchoring this detergent molecule (Fig. 3A). Among these residues, the conserved leucine L564 has been observed to be important for the function of MmpL3 (33). Interestingly, within the vicinity of this binding site, we also find the conserved amino acids Q309, D555 and P630. These three residues are located at the beginning of the channel formed by the MmpL3 transporter. It has been shown that these three amino acids are important for the activity of the transporter in *M. smegmatis* (23).

The second extra density is observed within the large space between PD1 and PD2. The shape of this electron density resembles a large lipid molecule with two elongated hydrocarbon backbones (Fig. 3B). Presumably, this extra electron density belongs to a fortuitous ligand that we copurified and cocrystallized with the $MmpL3_{773}$ protein. Consistent with this observation, our MS data of purified $MmpL3_{773}$ indicate a charge state series corresponding to an adduct mass of ~700 Da (Fig. 1D). We then used liquid chromatography coupled with mass spectrometry (LC-MS) to identify this unknown bound ligand. The result indicates that this fortuitous ligand is phosphatidylethanolamine (16:0-17:1(9Z)) (PE) (Fig. S5). Because of the presence of these DDM and PE molecules, the conformation of our structure of MmpL3 should represent its ligand-bound form.

PE binds at the central cavity formed by PD1 and PD2 of the periplasmic domain. Two extended loops (residues 61-68, and residues 449-456) that both run across subdomains PD1 and PD2, are found to sandwich the bound PE. To secure the binding, residues 40-44 of the elongated loop connecting TM1 and PD2, as well as residues 426-429 that belong to the other elongated loop connecting TM7 and PD1 also participate to form the bottom of this large cavity. Several conserved amino acids are found to surround the wall of this central cavity. The binding of PE is extensive. Within 4.5 Å of the bound PE molecule, there are at least 32 amino acids involved in the binding. These residues are Q40, S41, F43, Y44, D64, T66, S67, V70, V109, T121, M125, F134, S136, D144, L171, L174, A175, Q421, I427, S428, E429, F445, F452, R453, T454, P456, R501, P502, A503, N504, Q517 and T549, which provide electrostatic and hydrophobic interactions to bind PE (Fig. 3B). Among them, the conserved residue Q40 has been reported to be crucial to the function of this transporter (23).

The proton-relay network

MmpL3 is a proton-motive-force (PMF)-dependent transporter that function via an antiport mechanism. Coupled with the movement of substrates towards the periplasm, protons have to flow into the cytoplasm to energize this translocation process. Within the transmembrane region, a hydrogen bond is formed between the conserved residues D256 and Y646. Likewise, the nearby conserved residues Y257 and D645 also form a hydrogen-bonded pair. It is found that S293, which is also conserved, is within 3.7 Å away from D645, interacts with this aspartate through electrostatic interaction (Fig. S6). Within the vicinity of these residues, the crystal structure suggests that the conserved charged residues E647 and K591 also form another hydrogen bond between them. Taken together, these data suggest that these residues may form an important network for proton transfer within MmpL3. This is in good agreement with previous studies that have emphasized the importance of many of these residues to the function of this transporter (23), perhaps by creating a proton-relay network for energy coupling.

PE binding to MmpL3

As our structure of MmpL3 contained a bound PE, we wanted to further characterize this MmpL3-PE interaction. To achieve this, we performed a lipid binding experiment using native MS. A solution containing 5 μM $MmpL3_{773}$, 200 mM ammonium acetate (pH 8.0) and 0.05% (w/v) lauryldimethylamine N-oxide (LDAO) was used for the experiment. Equimolar amounts of $MmpL3_{773}$ and PE were mixed together and then the mass spectra under our optimized conditions were recorded (Fig. 4A). The experimental result depicts a charge state series for a second higher mass species at ~15% intensity of the peak corresponding to $MmpL3_{773}$. The 700 Da mass difference is consistent with the binding of one molecule PE to MmpL3 (Fig. 4A).

Subsequently, we incubated the $MmpL3_{773}$ protein with increasing concentrations of PE. At concentrations above 20 μM, we observed an increase in the intensity of the one PE-bound species, as well as an additional charge state series that corresponds binding of a second molecule of PE to $MmpL3_{773}$. At 80 μM of PE, the predominant species is the lipid-bound form of $MmpL3_{773}$. Based on our mass spectral data, we extracted and plotted the relative intensity of PE-bound forms as a function of lipid concentration. The data suggests a dissociation constant (K_d) of 19.5 ± 6.3 μM for the first PE binding (Fig. 4B).

TMM binding to MmpL3

Since there is strong evidence that MmpL3 is a TMM transporter, we decided to perform TMM binding experiments to probe the affinity of MmpL3 to TMM. Similar to the binding experiments described above, equimolar concentrations of $MmpL3_{773}$ and TMM were mixed together to acquire mass spectra under the conditions similar to those for PE binding. As expected, the data suggest a charge state series for a second higher mass species at ~25% intensity of the peak corresponding to $MmpL3_{773}$. The mass difference of 1,425 Da is consistent with the binding of one molecule of TMM to $MmpL3_{773}$, indicating that MmpL3 is capable of recognizing TMM (Fig. 4C).

To further define the affinity of TMM for MmpL3, we incubated the protein with increasing concentrations of TMM. Like the case of PE binding, at concentrations higher than 20 μM TMM, we observed evolution of an additional charge state series, corresponding to binding of a second molecule of TMM to $MmpL3_{773}$. Again, we extracted and plotted the relative intensity of TMM-bound forms with respect to TMM concentration, suggesting a K_d of 3.7 ± 1.3 μM for the first TMM binding (Fig. 4D).

MmpL3 does not bind TDM

As MmpL3 is able to specially bind TMM, we decided to elucidate if this transporter can bind TDM as well. We purified glycolipids from *M. smegmatis* mc^2155 cells. MS analysis indicates that the two main species of this glycolipid sample are TDM and TMM (Fig. 5). We then added this sample to the purified $MmpL3_{773}$ membrane protein and then recorded the mass spectra. In addition to the $MmpL3_{773}$ and $MmpL3_{773}$-PE peaks, the MS data show a new charge state series corresponding to an adduct mass of ~1,428 Da, which is the mass of TMM. Surprisingly, our data show that only short chain TMM lipids (~1,428 Da) bind specifically to $MmpL3_{773}$ even though different longer chain lengths of TMM lipids are present in the solution (red inlet in Fig. 5). The result is consistent with the binding data that $MmpL3_{773}$ selectively interacts with TMM. Of specific interest, the spectra did not present any other adduct peaks that correspond to the TDM bound species, demonstrating that $MmpL3_{773}$ does not bind TDM. Collectively, our data indicate that TMM, but not TDM, is preferentially bound by the MmpL3 transporter (Fig. 5).

Discussion

The translocation of TMM across the cytoplasmic membrane is an essential step in mycobacterial cell wall biogenesis. This process is absolutely required for the biosyntheses of mAGPs and TDMs that subsequently allow the cell envelope of Mtb to form a very rigid and an extremely impermeable layer to a wide range of antimicrobial compounds. However, exactly how TMMs are shuttled across the membrane for cell wall biosynthesis remains largely

unknown, which hampers progress in understanding the molecular mechanisms of Mtb cell wall biogenesis.

Our crystal structure of *M. smegmatis* MmpL3 indicates that this transporter functions as a monomer, which is able to take up TMM from the outer leaflet of the cytoplasmic membrane via a channel constituted by the transporter. A cleft surrounded by TMs 7-10 forms the entrance of this channel, which spans the outer leaflet of the cytoplasmic membrane and up to the central cavity of the periplasmic domain between PD1 and PD2. The TMM molecule could be shuttled through the channel to reach the periplasmic domain of the protein. This bound lipid could then be released to the inner leaflet of the outer membrane, where biosyntheses of mAGP and TDM take place (Fig. 6). To facilitate this process, proton transfer via the proton-relay network mediates the energy needed for substrate translocation. Our X-ray crystallographic and native mass spectrometric data suggest that MmpL3 is capable of specifically recognizing PE lipids, positing a possibility that this membrane protein is able to deliver TMM and other outer membrane lipids to facilitate cell wall remodeling.

Methods

Expression and purification of MmpL3 and MmpL3$_{773}$

The *M. smegmatis* MmpL3 protein was cloned into the pET15bΩ*mmpL3* expression vector in frame with a 6xHis tag at the C-terminus. This tagged MmpL3 protein was overproduced in *Escherichia coli* BL21(DE3)*ΔacrB* cells, which harbor a deletion in the chromosomal *acrB* gene. Cells were grown in 12 L of LB medium with 100 μg/ml ampicillin at 37°C. When the $OD_{600\ nm}$ reached 0.5, the culture was treated with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to induce *mmpL3* expression. Cells were then harvested within 3 h

of induction. The collected bacteria were resuspended in low salt buffer (100 mM sodium phosphate (pH 7.2), 10 % glycerol, 1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM phenylmethanesulfonyl fluoride (PMSF)), and then disrupted with a French pressure cell. The membrane fraction was collected and washed twice with high salt buffer (20 mM sodium phosphate (pH 7.2), 2 M KCl, 10 % glycerol, 1 mM EDTA and 1 mM PMSF), and once with 20 mM HEPES-NaOH buffer (pH 7.5) containing 1 mM PMSF as described previously (27). The membrane protein was then solubilized in 2% (w/v) n-dodecyl-β-D-maltoside (DDM). Insoluble material was removed by ultracentrifugation at 100,000 x g. The extracted protein was then purified with a Ni^{2+}-affinity column. The purified protein was dialyzed against 20 mM Na-HEPES (pH 7.5) and concentrated to 20 mg/ml in a buffer containing 20 mM Na-HEPES (pH 7.5) and 0.05% DDM. A final purification step was performed using a Superdex 200 size exclusion column loaded with buffer solution containing 20 mM Na-HEPES (pH 7.5) and 0.05% DDM. The purity of the MmpL3 protein (>95%) was judged using SDS-PAGE stained with Coomassie Brilliant Blue. The purified protein was then concentrated a second time to 20 mg/ml in a buffer containing 20 mM Na-HEPES (pH 7.5) and 0.05% DDM.

The expression and purification procedures for the $MmpL3_{773}$ protein are identical to those of MmpL3. For 6xHis selenomethionyl-substituted (SeMet)-$MmpL3_{773}$ protein expression, a 10 ml LB broth overnight culture was transferred into 120 ml of LB broth containing 100 μg/ml ampicillin and grown at 37°C. When the $OD_{600\ nm}$ reached 1.2, cells were harvested by centrifugation at 6000 g for 10 min, and then washed two times with 50 ml of M9 minimal salts solution. The cells were re-suspended in 120 ml of M9 media and then transferred into a 12 L pre-warmed M9 solution containing 100 μg/ml ampicillin. The cell culture was incubated at 37°C with shaking. When the $OD_{600\ nm}$ reached 0.4, 100 mg/L of lysine, phenylalanine and threonine, 50

mg/L isoleucine, leucine and valine, and 60 mg/L of L-selenomethionine were added. The culture was induced with 0.2 mM IPTG after 15 min. Cells were then harvested within 4 h after induction. The procedures for purifying SeMet-MmpL3$_{773}$ were identical to those of the native protein.

Expression of MmpL3 in *M. smegmatis*

The *M. smegmatis* MmpL3 protein that contains a 6xHis tag at the C-terminus was overproduced in *M. smegmatis* mc^{2}155 cells using the vector pSD26KΩ*mmpL3*. Cells were grown in 6 L of 7H9 expression medium with 25 μg/ml kanamycin at 37^{o}C. When the $OD_{600\ nm}$ reached 0.5, the culture was induced with 2% 0.2% acetamide and grow for another 12 h. The procedures for purifying MmpL3 expressed in *M. smegmatis* were identical to those for MmpL3 expressed in *E. coli* cells.

Crystallization of *M. smegmatis* MmpL3 and MmpL3$_{773}$

Crystals of the MmpL3 or MmpL3$_{773}$ protein were obtained using sitting-drop vapor diffusion. The crystals were grown at room temperature in 24-well plates with the following procedures. A 2 μl protein solution containing 20 mg/ml MmpL3 or MmpL3$_{773}$ in 20 mM Na-HEPES (pH 7.5) and 0.05% (w/v) DDM was mixed with a 2 μl of reservoir solution containing 25% PEG 400, 0.1 M sodium acetate (pH 5.4) and 0.05 M magnesium acetate. The resultant mixture was equilibrated against 500 μl of the reservoir solution at 25^{o}C. The crystallization conditions for SeMet-MmpL3$_{773}$ were the same as those for crystallization of the MmpL3$_{773}$ crystals. Crystals of MmpL3, MmpL3$_{773}$ and SeMet-MmpL3$_{773}$ grew to a full size in the drops within a month. Typically, the dimensions of the crystals were 0.2 mm x 0.2 mm x 0.2 mm. Cryoprotection of these crystals was achieved by raising the PEG 400 concentration to 30%.

Data collection, structural determination and refinement

All diffraction data were collected at 100K at beamline 24ID-C located at the Advanced Photon Source, using a Pilatus 6M detector (Dectris Ltd., Switzerland). Diffraction data were processed using DENZO and scaled using SCALEPACK.(34) Crystals of MmpL3, MmpL3$_{773}$ and SeMet-MmpL3$_{773}$ belong to space group $P2_12_12_1$ (Table S1).

The MmpL3$_{773}$ protein contains 25 methionines (excluding the N-terminal methionine). Within the asymmetric unit of SeMet-MmpL3$_{773}$, all of the 25 selenium sites were identified using SHELXC and SHELXD (35) as implemented in the HKL2MAP package (36). Single anomalous dispersion (SAD) was employed to obtain experimental phases using the program Autosol (37) in PHENIX (38). The resulting phases were then subjected to density modification and NCS averaging using the program PARROT (39) using the native structure factor amplitudes. The SeMet sites were also used to trace the molecule by anomalous difference Fourier maps, where we could ascertain the proper registry of SeMet residues. The initial model of MmpL3$_{773}$ was constructed manually using program Coot (40). Then the model was refined using PHENIX (38), leaving 5% of reflections in the Free-R set. Iterations of refinement were performed using PHENIX (38). Model buildings were done using Coot (40), which led to the current model (Table S1).

The crystal structures of full-length MmpL3 was determined by molecular replacement (MR), utilizing the final structure of MmpL3$_{773}$ as a search model. The procedures for structural refinement and model building were the same as those of MmpL3$_{773}$ (Table S1).

Native mass spectrometry

Prior to mass spectrometry analysis, protein solutions were buffer exchanged into 200 mM ammonium acetate pH 8.0, with 2× CMC of detergent of interest using a Biospin-6 (BioRad) column and introduced directly into the mass spectrometer using gold-coated capillary needles prepared in-house (41). Data were collected on a modified QExactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) optimized for analysis of high mass complexes, using methods previously described for membrane proteins (41). The instrument parameters were as follows: capillary voltage 1.2 kV, S-lens RF 100%, quadrupole selection from 2,000 to 20,000 m/z range, collisional activation in the HCD cell 100-200 V, argon UHV pressure 1.12×10^{-9} mbar, temperature 60 °C, resolution of the instrument at 17,500 at m/z = 200 (a transient time of 64 ms) and ion transfer optics (injection flatapole, inter-flatapole lens, bent flatapole, transfer multipole: 8, 7, 6, 4 V respectively). The noise level was set at 3 rather than the default value of 4.64. No in source dissociation was applied. Where required baseline subtraction was performed to get a better-quality mass spectrum.

Lipid analysis

Lipidomic analysis was performed on $MmpL3_{773}$ purified in DDM in order to gain insights on the lipids that co-purify with these proteins. Co-purified lipids have been identified in a similar manner that has been described previously (41). Briefly, protein was digested with trypsin overnight at 37 °C, lyophilized and re-dissolved in 68% solution A (ACN:H2O 60:40, 10 mM ammonium formate and 0.1% formic acid) and 32% solution B (IPA:ACN 90:10, 10 mM ammonium formate and 0.1% formic acid) for analysis by reverse phase liquid chromatography tandem MS (RP LC-MS/MS). LC-MS/MS was performed using a Dionex UltiMate 3000 RSLC Nano system coupled to an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher

Scientific). The peptide/lipid mixture was loaded onto a C18 column (Acclaim PepMap 100, C18, 75 µm × 15 cm; Thermo Scientific) at a flow rate of 300 nl min-1. After 10 min solvent B was ramped to 65% over 1 min, then 80% over 6 min, before being held at 80% for 10 min, then ramped to 99% over 6 min and held for 7 min. Typical MS conditions were spray voltage of 1.6 kV and capillary temperature of 275 °C. The LTQ-Orbitrap XL was set up in negative ion mode and in data-dependent acquisition mode to perform five MS/MS scans per MS scan. Survey full-scan MS spectra were acquired on the Orbitrap (m/z 400–2,000) with a resolution of 60,000. Collision-induced dissociation (CID) fragmentation in the linear ion trap was performed for the five most intense ions at an automatic gain control target of 30,000 and a normalized collision energy of 38% at an activation of q = 0.25 and an activation time of 30 ms.

Binding of PE and TMM by MmpL3$_{773}$

PE and TMM binding experiments were performed with the protein in 200 mM ammonium acetate supplemented with 0.05% (w/v) LDAO. In order to obtain the binding constant for the interaction between MmpL3$_{773}$ and PE or TMM, PE or TMM was added in increasing amounts while keeping the protein concentration constant. Peak intensities were extracted and the binding affinities were determined as described previously (41). The optimized mass spectrometry conditions for PE or TMM binding are same as described above in the naïve mass spectrometry section except that the HCD cell activation energy is 200V.

Accession Codes

Atomic coordinates and structure factors have been deposited at the RCSB Protein Data Bank with accession codes 6N3T (MmpL3$_{773}$) and 6N40 (MmpL3).

Supplementary information

Supplementary Materials

Supplementary Figs. 1 to 6

Supplementary Table 1

References

1. Wallis RS*, et al.* (2016) Tuberculosis--advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers. *Lancet Infect Dis* 16(4):e34-46.
2. WHO (2017) Global tuberculosis report 2017.
3. Brennan PJ & Nikaido H (1995) The envelope of mycobacteria. *Annu Rev Biochem* 64:29-63.
4. Camacho LR*, et al.* (2001) Analysis of the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* 276(23):19845-19854.
5. Gilmore SA*, et al.* (2012) Sulfolipid-1 biosynthesis restricts Mycobacterium tuberculosis growth in human macrophages. *ACS Chem Biol* 7(5):863-870.
6. Perez RL*, et al.* (2000) Cytokine message and protein expression during lung granuloma formation and resolution induced by the mycobacterial cord factor trehalose-6,6'-dimycolate. *J Interferon Cytokine Res* 20(9):795-804.
7. Bekierkunst A*, et al.* (1969) Granuloma formation induced in mice by chemically defined mycobacterial fractions. *J Bacteriol* 100(1):95-102.
8. Indrigo J, Hunter RL, Jr., & Actor JK (2003) Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. *Microbiology* 149(Pt 8):2049-2059.
9. Geisel RE, Sakamoto K, Russell DG, & Rhoades ER (2005) In vivo activity of released cell wall lipids of Mycobacterium bovis bacillus Calmette-Guerin is due principally to trehalose mycolates. *J Immunol* 174(8):5007-5015.
10. Takayama K, Wang C, & Besra GS (2005) Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. *Clin Microbiol Rev* 18(1):81-101.
11. Banerjee A*, et al.* (1994) inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science* 263(5144):227-230.
12. Cole ST*, et al.* (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393(6685):537-544.
13. Tseng TT*, et al.* (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol* 1(1):107-125.
14. Lamichhane G, Tyagi S, & Bishai WR (2005) Designer arrays for defined mutant analysis to detect genes essential for survival of Mycobacterium tuberculosis in mouse lungs. *Infect Immun* 73(4):2533-2540.
15. Domenech P, Reed MB, & Barry CE, 3rd (2005) Contribution of the Mycobacterium tuberculosis MmpL protein family to virulence and drug resistance. *Infect Immun* 73(6):3492-3501.
16. Domenech P*, et al.* (2004) The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. *J Biol Chem* 279(20):21257-21265.
17. Cox JS, Chen B, McNeil M, & Jacobs WR, Jr. (1999) Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature* 402(6757):79-83.
18. Grzegorzewicz AE*, et al.* (2012) Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane. *Nat Chem Biol* 8(4):334-341.

19. Varela C*, et al.* (2012) MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem Biol* 19(4):498-506.
20. Xu Z, Meshcheryakov VA, Poce G, & Chng SS (2017) MmpL3 is the flippase for mycolic acids in mycobacteria. *Proc Natl Acad Sci U S A* 114(30):7993-7998.
21. Belisle JT*, et al.* (1997) Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. *Science* 276(5317):1420-1422.
22. Bhatt A, Kremer L, Dai AZ, Sacchettini JC, & Jacobs WR, Jr. (2005) Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *J Bacteriol* 187(22):7596-7606.
23. Belardinelli JM*, et al.* (2016) Structure-Function Profile of MmpL3, the Essential Mycolic Acid Transporter from Mycobacterium tuberculosis. *ACS Infect Dis* 2(10):702-713.
24. Murakami S, Nakashima R, Yamashita E, & Yamaguchi A (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* 419(6907):587-593.
25. Murakami S, Nakashima R, Yamashita E, Matsumoto T, & Yamaguchi A (2006) Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* 443(7108):173-179.
26. Sennhauser G, Bukowska MA, Briand C, & Grutter MG (2009) Crystal structure of the multidrug exporter MexB from Pseudomonas aeruginosa. *J Mol Biol* 389(1):134-145.
27. Long F*, et al.* (2010) Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature* 467(7314):484-488.
28. Su CC*, et al.* (2011) Crystal structure of the CusBA heavy-metal efflux complex of Escherichia coli. *Nature* 470(7335):558-562.
29. Bolla JR*, et al.* (2014) Crystal structure of the Neisseria gonorrhoeae MtrD inner membrane multidrug efflux pump. *PLoS One* 9(6):e97903.
30. Su CC*, et al.* (2017) Structures and transport dynamics of a Campylobacter jejuni multidrug efflux pump. *Nat Commun* 8(1):171.
31. Kumar N*, et al.* (2017) Crystal structures of the Burkholderia multivorans hopanoid transporter HpnN. *Proc Natl Acad Sci U S A* 114(25):6557-6562.
32. Chim N*, et al.* (2015) The Structure and Interactions of Periplasmic Domains of Crucial MmpL Membrane Proteins from Mycobacterium tuberculosis. *Chem Biol* 22(8):1098-1107.
33. Li W*, et al.* (2014) Novel insights into the mechanism of inhibition of MmpL3, a target of multiple pharmacophores in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 58(11):6413-6423.
34. Otwinowski Z & Minor W (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* 276:307-326.
35. Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58(Pt 10 Pt 2):1772-1779.
36. Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* 37.
37. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582-601.

38. Adams PD*, et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58(Pt 11):1948-1954.
39. Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):470-478.
40. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
41. Bolla JR*, et al.* (2018) Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ. *Nat Chem* 10(3):363-371.

Legends of Figures

Fig. 1. Mass spectra of purified MmpL3 proteins. (A) Mass spectrum of full-length MmpL3 expressed in *E. coli* indicates that the protein exists as a monomer. Calculated and observed masses, including the C-terminal 6×His tag, are 110,222 Da and 110,246 Da. (B) Mass spectrum of full-length MmpL3 expressed in *M. smegmatis*. In addition to the monomeric protein in solution, the presence of degraded protein bands indicates that the purified protein is unstable. The two observed masses are 110,371 Da and 91,990 Da. (C) Mass spectrum of a five-day old sample of full-length MmpL3 expressed and purified from *E. coli*. The spectrum depicts three major degraded species in the solution sample. The observed masses of these species are 87,998 Da, 84,842 Da and 83,359 Da, corresponding to residues 1-806, 1-776 and 1-763 of the protein, respectively. (D) Mass spectrum of the $MmpL3_{773}$ protein. The spectrum indicates that this protein exists as a monomer in solution. Observed and calculated masses in this case are 85,950 Da and 85,925 Da. The mass observed from the adduct peaks is 86,660 Da, which corresponds to the $MmpL3_{773}$-PE complex.

Fig. 2. Structure of the *M. smegmatis* MmpL3 transporter. (A) Secondary structural topology of the $MmpL3_{773}$ monomer. The topology was constructed based on the crystal structure of $MmpL3_{773}$. The transmembrane helices (TMs) are colored blue. The periplasmic loops 1 and 2 and in red and green, respectively. The cytoplasmic domain (CD) of MmpL3 is colored brown. (B) Ribbon diagram of a monomer of $MmpL3_{773}$ viewed in the membrane plane. The TMs, periplasmic loops 1 and 2, and CD are colored slate, red, green and orange, respectively. The bound DDM is in yellow spheres and the bound PE is in pink spheres. The periplasmic loops 1 and 2 cross over each other to form the periplasmic domains 1 and 2 (PD1 and PD2). (C) The

$MmpL3_{773}$ monomer forms a channel spanning the outer leaflet of the inner membrane and up to the periplasmic domain. The orientation of this MmpL3 molecule has been rotated by 60° clockwise, with respect to the vertical pseudo C2 symmetry axis of $MmpL3_{773}$, when compared with the orientation of (B). The channel (colored gray) was calculated using the program CAVER (http://loschmidt.chemi.muni.cz/caver). The TMs, periplasmic loops 1 and 2, and CD are colored the same as (B).

Fig. 3. The DDM and PE binding sites of the MmpL3-PE complex. (A) The F_o - F_c electron density map of bound DDM in MmpL3. The bound DDM is shown as a stick model (yellow, carbon; red, oxygen). The F_o - F_c map is contoured at 3σ. Residues involved in DDM binding are in green sticks. (B) The F_o - F_c electron density map of bound PE in MmpL3. The bound PE is shown as a stick model (magenta, carbon; red, oxygen; blue, nitrogen). The F_o - F_c map is contoured at 3σ. Residues involved in PE binding are in green sticks.

Fig. 4. Determination of dissociation constants for the bindings of PE and TMM to $MmpL3_{773}$. (A) Mass spectra recorded for solutions of $MmpL3_{773}$ with increasing concentrations of PE. At 5 μM PE a charge state series is observed (light orange), corresponding to bound PE, which increases in intensity as the PE concentration is increased to 80 μM. A second PE binding peak (gray) emerges at concentrations above 20 μM. (B) Plot of relative fractional intensity of lipid binding peaks over the total peak intensity versus PE concentration (see Experimental Methods), yielding a curve for the first binding event and linear-like fit for the second, consistent with non-specific PE binding. Each data point and standard deviation are calculated from the average of five observed charge states in three independent experiments. Error bars represent standard

deviations (n = 3). (C) Mass spectra recorded for solutions of MmpL3$_{773}$ with increasing concentrations of TMM. At 2.5 μM TMM a charge state series is observed (orange), corresponding to bound TMM, which increases in intensity as the TMM concentration is increased to 40 μM. A second TMM binding peak (purple) emerges at concentrations above 20 μM. (D) Plot of the relative fractional intensity of lipid binding peaks over total peak intensity versus TMM concentration (see Experimental Methods), yielding a curve for the first binding event and linear-like fit for the second, consistent with non-specific TMM binding. Each data point and standard deviation are calculated from the average of five observed charge states in three independent experiments. Error bars represent standard deviations (n = 3).

Fig. 5. Analysis of different components of purified *M. smegmatis* glycolipids and their bindings to MmpL3$_{773}$. Mass spectrum of MmpL3$_{773}$ with purified *M. smegmatis* glycolipids shows a preferential binding to TMM (orange charge state series) but not to TDM. Purified lipid fraction has both TMM (red inlet) and TDM (blue inlet) at the lower m/z region.

Fig. 6. Proposed mechanism for TMM translocation via MmpL3. This schematic diagram indicates that the MmpL3 transporter is capable of picking up a TMM molecule from the outer leaflet of the cytoplasmic membrane. This TMM molecule will pass through the channel formed by MmpL3 and arrive the periplasmic lipid binding site. The TMM moiety will then be exported to the inner leaflet of the outer membrane for the biosyntheses of TDMs and mAGPs.

Fig. 1

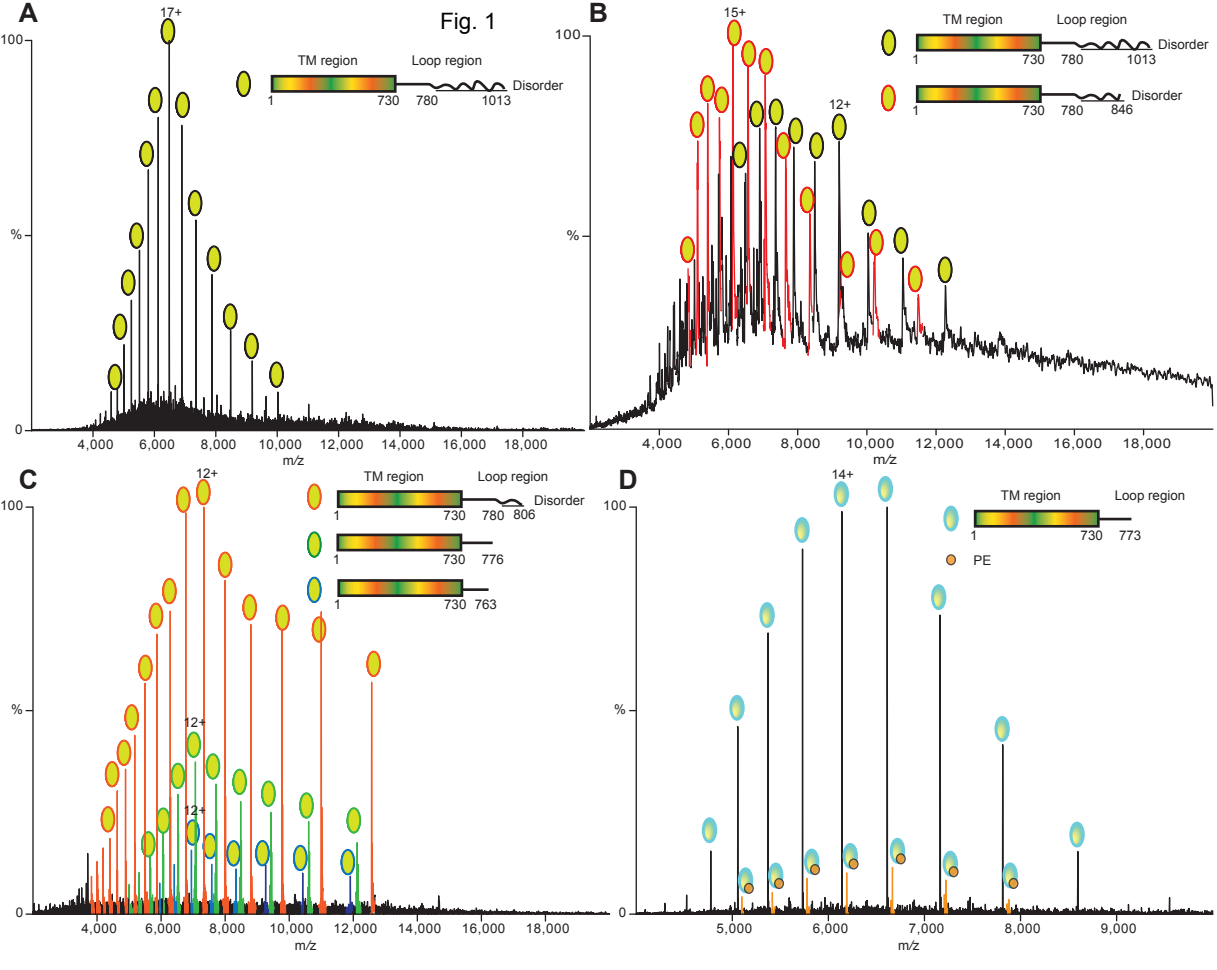

A
17+
100
%
0
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
m/z
TM region
Loop region
Disorder
1
730
780
1013
B
15+
12+
100
%
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
m/z
TM region
Loop region
Disorder
1
730
780
1013
Disorder
1
730
780
846
C
12+
12+
12+
100
%
0
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
m/z
TM region
Loop region
Disorder
1
730
780
806
1
730
776
1
730
763
D
14+
100
%
0
5,000
6,000
7,000
8,000
9,000
m/z
TM region
Loop region
1
730
773
PE

Fig. 2

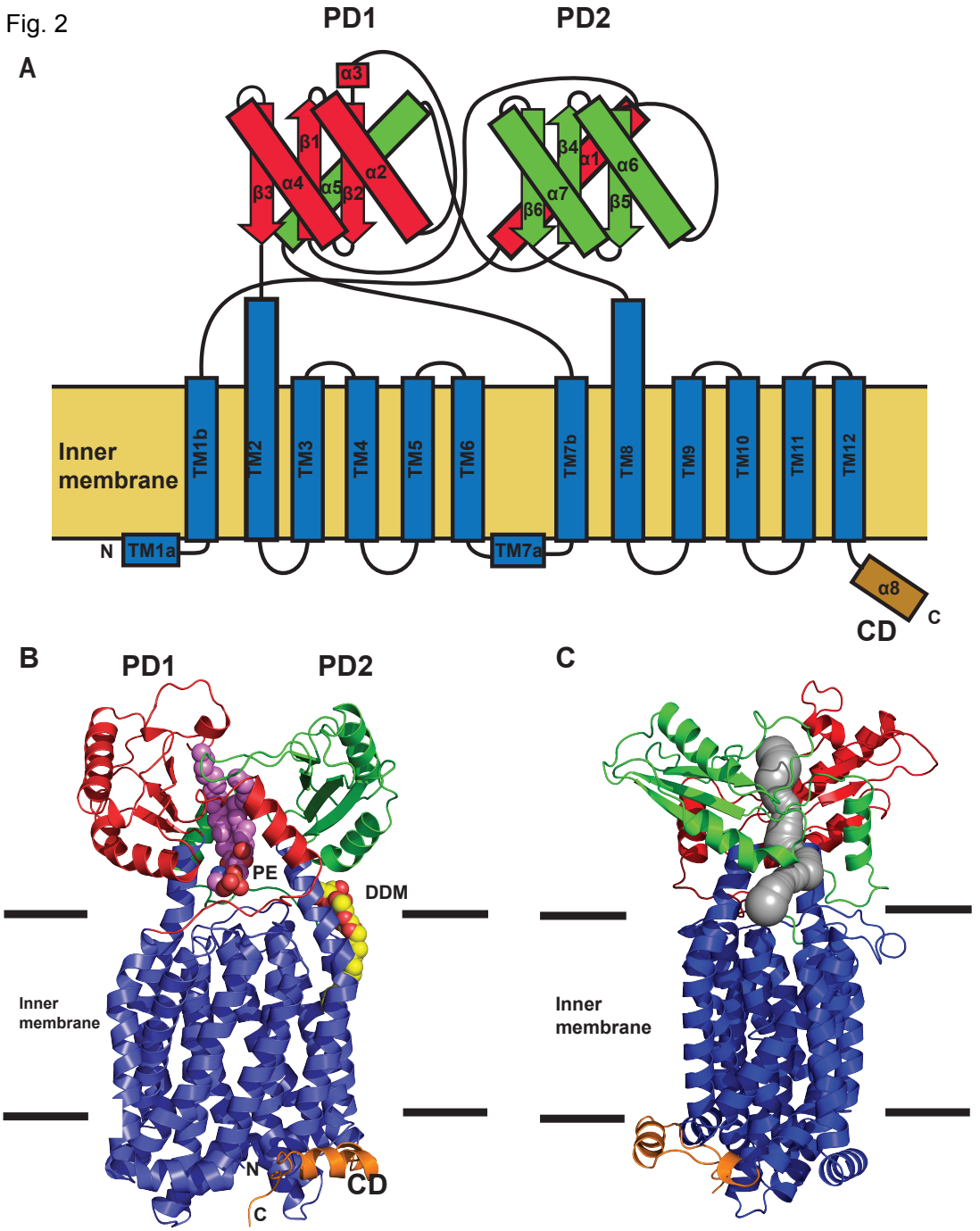
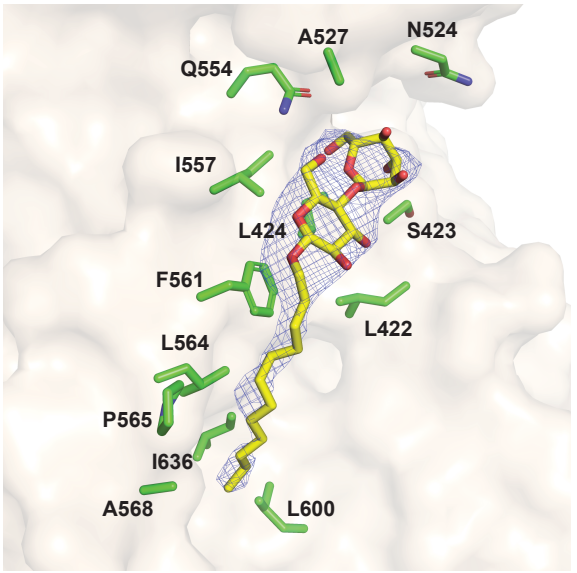

A
PD1
PD2
α3
β1
α4
α5
β3
β2
α2
β4
α1
α6
α7
β6
β5
Inner membrane
TM1b
TM2
TM3
TM4
TM5
TM6
TM7b
TM8
TM9
TM10
TM11
TM12
N
TM1a
TM7a
α8
CD
C
B
PD1
PD2
PE
DDM
Inner membrane
N
C
CD
C

Fig. 3

A


A527
N524
Q554
I557
L424
S423
F561
L422
L564
P565
I636
A568
L600

B

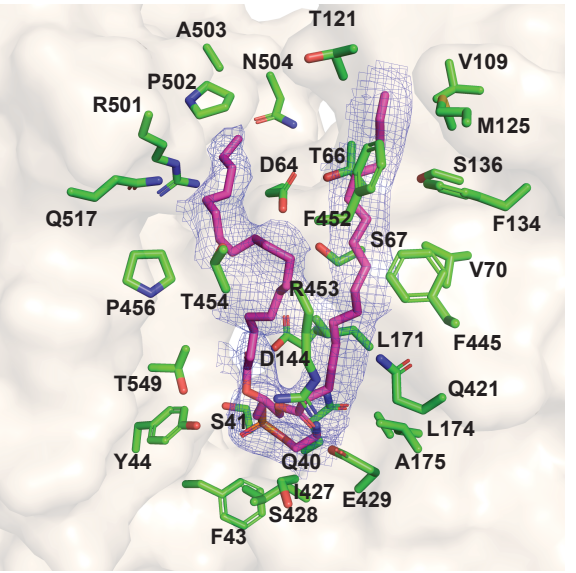

T121
A503
N504
V109
P502
R501
M125
T66
D64
S136
Q517
F452
F134
S67
V70
P456
T454
R453
L171
F445
D144
T549
Q421
S41
L174
Y44
Q40
A175
I427
E429
S428
F43

Fig. 4

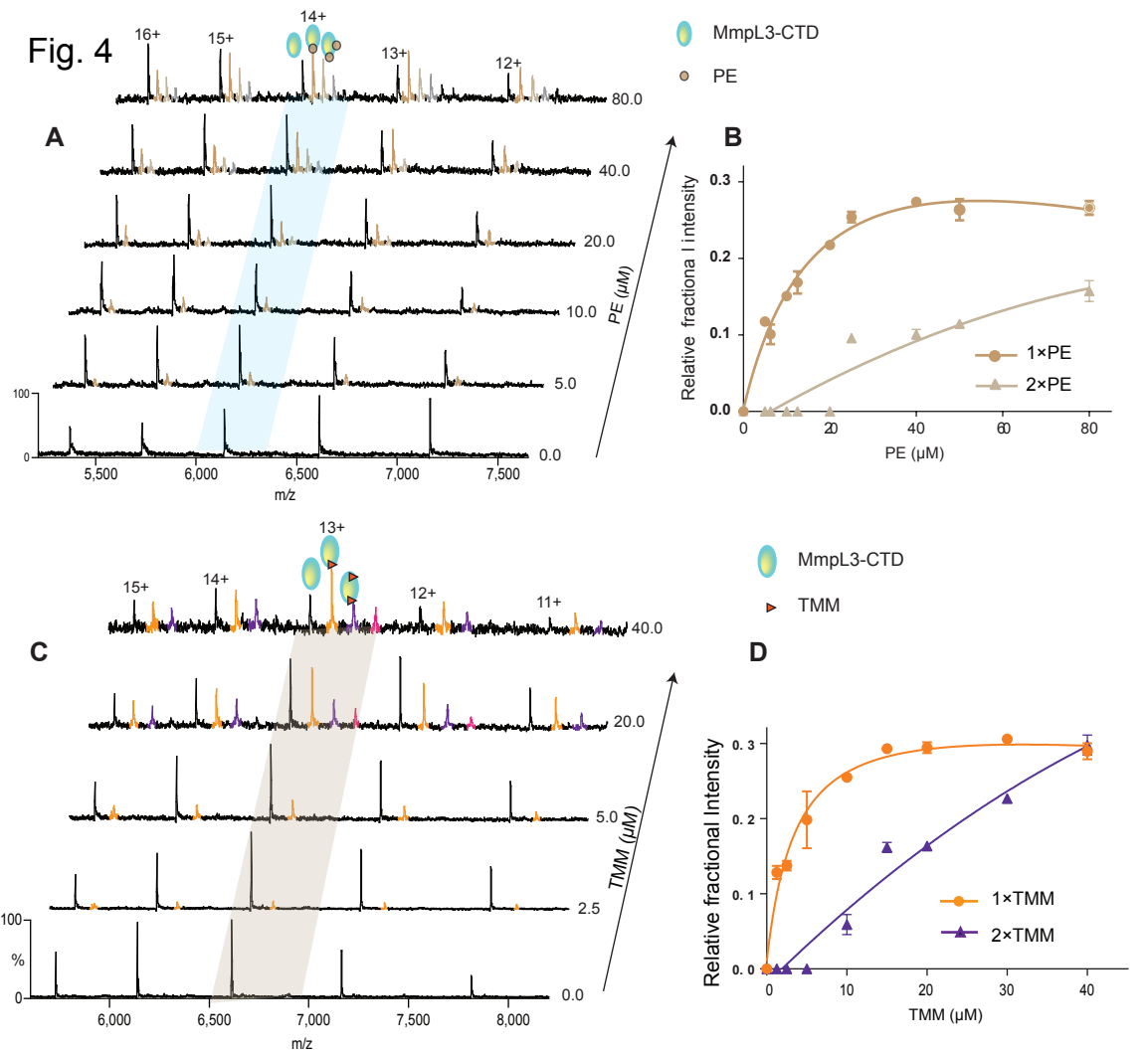

16+
15+
14+
13+
12+
80.0
40.0
20.0
10.0
5.0
0.0
PE (μM)
A
100
0
5,500
6,000
6,500
7,000
7,500
m/z
MmpL3-CTD
PE
B
Relative fractiona l intensity
0.3
0.2
0.1
0.0
0
20
40
60
80
PE (μM)
1×PE
2×PE
13+
15+
14+
12+
11+
40.0
C
20.0
5.0
2.5
0.0
TMM (μM)
100
%
0
6,000
6,500
7,000
7,500
8,000
m/z
MmpL3-CTD
TMM
D
Relative fractional Intensity
0.3
0.2
0.1
0.0
0
10
20
30
40
TMM (μM)
1×TMM
2×TMM

Fig. 5

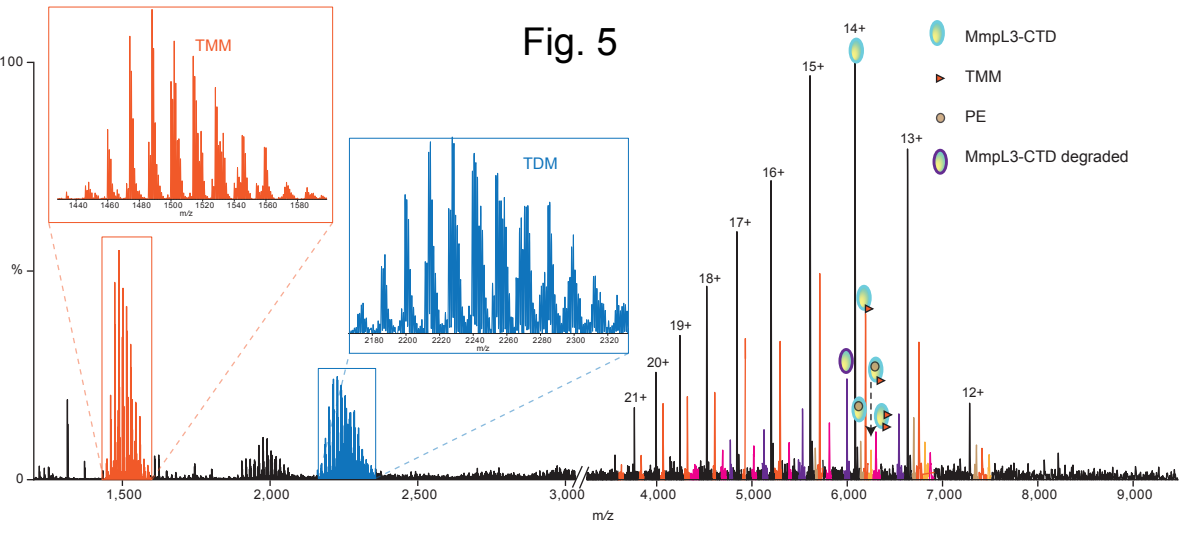
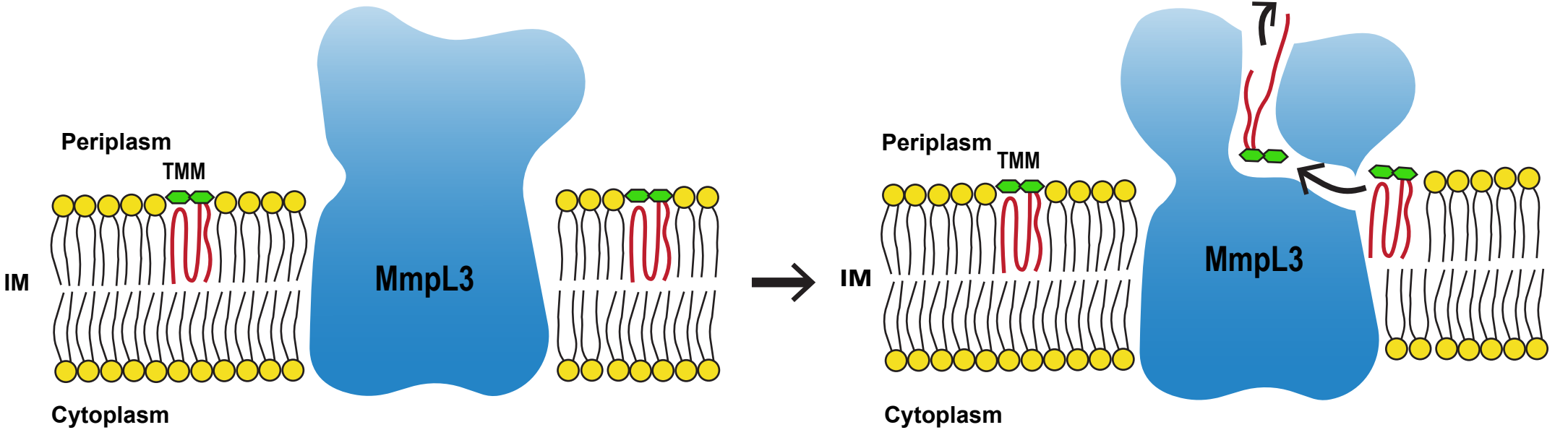

TMM
1440 1460 1480 1500 1520 1540 1560 1580
m/z
TDM
2180 2200 2220 2240 2260 2280 2300 2320
m/z
100
%
0
14+
15+
13+
16+
17+
18+
19+
20+
21+
12+
MmpL3-CTD
TMM
PE
MmpL3-CTD degraded
1,500
2,000
2,500
3,000
4,000
5,000
6,000
7,000
8,000
9,000
m/z

Fig. 6


Periplasm
TMM
IM
MmpL3
Cytoplasm
Periplasm
TMM
IM
MmpL3
Cytoplasm