Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Crystal structure and functional implications of LprF from *Mycobacterium tuberculosis* and *M. bovis*

Jin-Sik Kim,[a]‡ Li Jiao,[a]‡ Jeong-Il Oh,[b] Nam-Chul Ha[c]* and Yong-Hak Kim[d]*

[a]College of Pharmacy and Research Institute for Drug Development, Pusan National University, Busan, Republic of Korea, [b]Department of Microbiology, College of Natural Sciences, Pusan National University, Busan, Republic of Korea, [c]Department of Food and Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea, and [d]Department of Microbiology, Catholic University of Daegu School of Medicine, Daegu, Republic of Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: hanc210@snu.ac.kr, ykim@cu.ac.kr

Received 14 May 2014
Accepted 17 July 2014

PDB reference: LprF, 4qa8

The Gram-positive bacteria *Mycobacterium tuberculosis* and *M. bovis* are causative agents of tuberculosis in humans and cattle. The lipoprotein LprF is found in *M. tuberculosis* and *M. bovis* but not in the nonpathogenic *M. smegmatis*. To date, the role of LprF remains to be elucidated. In this study, the crystal structure of LprF has been determined at 1.1 Å resolution. The overall structure is similar to that of a homologue, LprG, with a central hydrophobic cavity that binds a triacylated glycolipid. LprF exhibited a central cavity structure similar to that of LprG, but with a smaller cavity that binds two alkyl chains. Consistently, subsequent mass-spectrometric analysis revealed that the bound ligand was a diacylated glycolipid, as found in the structure. Furthermore, an increased ratio of lipoarabinomannan to lipomannan in the mycobacterial cell wall was observed when *lprF* was introduced into *M. smegmatis*. These observations suggested that LprF transfers the diacylated glycolipid from the plasma membrane to the cell wall, which might be related to the pathogenesis of the bacteria.

1. Introduction

Mycobacterium tuberculosis (Mtb) is a pathogenic Gram-positive bacterium and is the causative agent of most cases of tuberculosis, a leading cause of death (Russell, 2001; Baena & Porcelli, 2009). Multiple drug-resistant and extremely drug-resistant strains are increasingly prevalent, which makes the development of new drugs essential (Bhowruth *et al.*, 2007). Mycobacteria have an unusual lipid-rich cell wall composed of a mycolyl–arabinogalactan–peptidoglycan complex (Daffé & Draper, 1998), in which the hydrophobic mycolic acid forms the outer membrane and is linked to the peptidoglycan layer *via* arabinogalactan (Hoffmann *et al.*, 2008). The mycolic acid layer functions as a hydrophobic mesh for intercalating additional glycolipids, including phosphatidyl-*myo*-inositol mannosides (PIMs), lipomannan (LM) and lipoarabinomannan (LAM). This tight mycobacterial cell wall provides a permeability barrier to various physical and chemical stresses such as antibiotics and chemotherapeutic agents, which is essential for virulence and growth within the host cells (Jarlier & Nikaido, 1994; Chatterjee, 1997; Minnikin *et al.*, 2002).

PIMs, LM and LAM are synthesized by sequential additions of mannoses and arabinoses to phosphatidylinositol, one of the major phospholipids in the mycobacterial cell membrane (Gilleron *et al.*, 2003). These glycolipids carry the lipid moieties that are responsible for anchoring to both the mycobacterial outer membrane and the plasma membrane

(Mishra *et al.*, 2011). Although the predominant forms of glycolipids are triacylated species, monoacylated, diacylated and tetraacylated forms are also found (Nigou *et al.*, 1999; Chatterjee *et al.*, 1992). The outer membrane-associated glycolipids contribute to resistance to bactericidal free radicals (Chan *et al.*, 1989) and modulate the host immune functions, including phagosome maturation and cytokine production (Briken *et al.*, 2004).

The structures and functions of three homologous lipoproteins, LppX, LprG and LpqW, from Mtb have recently been explored (Sulzenbacher *et al.*, 2006; Marland *et al.*, 2006; Drage *et al.*, 2010). These proteins are predicted to be triacylated at the cysteine residue of the N-terminus, resulting in anchoring to the outer leaflet of the plasma membrane through their acyl chains (Drage *et al.*, 2010; Sulzenbacher *et al.*, 2006; Marland *et al.*, 2006). These lipoproteins have been proposed to play a key role in the synthesis and transport of components of the mycobacterial outer membrane (Drage *et al.*, 2010; Sulzenbacher *et al.*, 2006; Marland *et al.*, 2006). LppX has been proposed to be involved in cell-wall biosynthesis by binding and transporting phthiocerol dimycocerosate (Sulzenbacher *et al.*, 2006), while LpqW has been shown to be essential in the synthesis of the cell-wall components PIM and LAM (Marland *et al.*, 2006). The crystal structure of LprG revealed that LprG has a central hydrophobic pocket that can accommodate a glycolipid with three acyl groups, which suggested a function for LprG in mycobacteria as a carrier of triacylated glycolipids during their trafficking and delivery to the mycobacterial cell wall (Drage *et al.*, 2010). It was further observed that the glycolipid-binding function of LprG facilitated the recognition of triacylated glycolipids by TLR2, thus enhancing the host immune response (Drage *et al.*, 2010).

M. bovis is a causative agent of tuberculosis, particularly in cattle but also in humans, with worldwide annual losses to agriculture of $3 billion (Garnier *et al.*, 2003). Most of the genes of Mtb and *M. bovis* are nearly identical, and the pathologies of the two bacteria are similar (Garnier *et al.*, 2003). The lipoprotein *lprF* was identified as a homologue of LprG in the same gene cluster as *lprG* and is found in the Mtb and *M. bovis* genomes but not in that of the nonpathogenic *M. smegmatis*. The amino-acid sequences of *lprF* of Mtb and *M. bovis* are identical. In this study, non-acylated LprF protein was produced in *Escherichia coli* and its crystal structure was solved using selenomethionyl-substituted LprF protein. Based on this structure further studies were performed, and the biochemical role of LprF is proposed with implications for its physiological roles in the pathogenesis of Mtb and *M. bovis*.

2. Materials and methods

2.1. Construction of plasmid for protein expression

A DNA fragment encoding the non-acylated form of *M. bovis* LprF (residues 40–261) was amplified from the genomic DNA of *M. bovis* using the polymerase chain reaction. The DNA fragment was inserted into the *Eco*RI and *Hin*dIII sites of the pPROEX-HTA vector (Invitrogen, USA). The A110Y mutation was introduced into the resulting plasmid using the QuikChange kit.

2.2. Protein expression and purification

The truncated form (residues 40–261) of the LprF protein was expressed in *E. coli* strain BL21 (DE3) in LB medium at 310 K until the OD_{600} reached 0.5 and was then induced by adding 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside and culturing at 303 K for 6 h. To obtain selenomethionyl-substituted LprF protein, *E. coli* strain B834 (DE3) was cultured in M9 medium supplemented with an amino-acid mixture containing L-(+)-selenomethionine at 310 K until the OD_{600} reached 0.5. Expression of the selenomethionyl LprF protein was induced by adding 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside and culturing at 303 K for 2 d. The cells were harvested by centrifugation after induction and were then stored at 193 K until use. To purify the protein, the cells were resuspended in lysis buffer consisting of 20 m*M* Tris pH 8.0, 150 m*M* NaCl, 2 m*M* β-mercaptoethanol. The resuspended lysate was disrupted by sonication and the cell debris was removed by centrifugation at 19 000*g* for 30 min. The supernatant was mixed with Ni–NTA affinity resin (Qiagen, Netherlands). After the slurry had been loaded into the column, unbound proteins were washed off with lysis buffer supplemented with 20 m*M* imidazole. The hexahistidine-tagged LprF protein was eluted with lysis buffer supplemented with 250 m*M* imidazole. Fractions containing the LprF protein were pooled and the hexahistidine tag was then cleaved by treatment with recombinant TEV protease. The LprF protein was loaded onto a HiTrap Q column (GE Healthcare, USA) and eluted from the column using a linear gradient of 0–1 *M* NaCl in 20 m*M* Tris pH 8.0. The LprF protein was then further purified using HiLoad Superdex 200 (GE Healthcare, USA) with lysis buffer. The purified protein was concentrated and stored at 4°C for use within a week or stored frozen at 193 K until use.

2.3. Crystallization of the selenomethionyl-substituted LprF protein

The native LprF protein and selenomethionyl-substituted LprF protein (residues 40–261) were concentrated to 15 mg ml^{-1} using an Amicon centrifugal filter (Millipore; 10 kDa cutoff). The initial screening for crystallization of the native LprF protein was performed using the sitting-drop vapour-diffusion method with conditions from the commercially available Crystal Screen HT (Hampton Research). After three months, crystals were obtained under the condition 0.2 *M* magnesium chloride, 0.1 *M* Tris–HCl pH 8.5, 30% PEG 4K at 295 K. Crystals of the native and the selenomethionyl-substituted LprF protein that were suitable for data collection were obtained in droplets consisting of 0.2 *M* magnesium chloride, 0.1 *M* Tris–HCl pH 7.8, 28% PEG 4K. The droplets were equilibrated by the hanging-drop vapour-diffusion method at 295 K for one month.

2.4. Data collection and structural determination

Native and selenomethionyl-substituted LprF crystals were flash-cooled in liquid nitrogen using the mother liquor as a cryoprotectant. A multiwavelength anomalous dispersion (MAD) data set was collected on beamline 7A at Pohang Accelerator Laboratory using a Quantum 270 CCD detector (ADSC). The diffraction data set was processed and scaled using the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals of LprF belonged to space group *P*1 (Table 1), indicating that the asymmetric unit contained one molecule. Phases were determined using the MAD data set for the selenomethionyl-substituted LprF crystals, and three selenium sites were found using *SOLVE* (Terwilliger & Berendzen, 1999). An initial model from *SOLVE*/*RESOLVE* (Terwilliger & Berendzen, 1999) was refined by *PHENIX* (Adams *et al.*, 2002) and built using *Coot* (Emsley & Cowtan, 2004). The structure of native LprF was determined by molecular replacement using the unrefined structure of selenomethionyl-substituted LprF as a search model (Adams *et al.*, 2002). The final $R_{\rm work}$ and $R_{\rm free}$ were 16.4 and 18.9%, respectively. The native and selenomethionyl-substituted LprF structures were nearly identical (r.m.s.d. of 0.122 Å between 150 C^{α} atoms).

Table 1
Data-collection and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native LprF	SeMet LprF		
Data collection				
X-ray source	Beamline 7A, Pohang Accelerator Laboratory			
Wavelength (Å)	0.9793	0.9789	0.9794	0.9721
Resolution limit (Å)	20–1.10 (1.12–1.10)	50–1.18 (1.20–1.18)		
Space group	*P*1	*P*1		
Unit-cell parameters (Å, °)	a = 31.5, b = 32.4, c = 51.7, α = 76.4, β = 80.8, γ = 71.2	a = 31.6, b = 32.5, c = 46.3, α = 89, β = 82.2, γ = 71		
Multiplicity	3.5 (2.4)	2.9 (1.9)	3.0 (1.9)	2.9 (1.8)
$R_{\rm merge}$ (%)	3.6 (15.6)	6.6 (24.5)	5.7 (22.7)	6.4 (24.7)
Completeness (%)	90.4 (82.0)	87.6 (53.4)	87.4 (50.8)	87.2 (50.1)
$\langle I/\sigma(I)\rangle$	47.8 (6.0)	30.3 (2.14)	30.9 (2.4)	30.2 (2.3)
Refinement				
Resolution (Å)	18.56–1.10	20–1.18		
R factor (%)	16.4	20.6		
$R_{\rm free}$† (%)	18.9	24.2		
Average B factor (Å^2)	23.0	22.1		
Wilson B factor (Å^2)	11.3	12.6		
R.m.s. deviation, bonds (Å)	0.006	0.008		
R.m.s. deviation, angles (°)	1.2	1.2		
Ramachandran plot				
Most favoured (%)	95.4	96.7		
Additionally allowed (%)	4.6	3.3		
Coordinate error (Å)	0.08	0.14		
PDB entry	4qa8			

† $R_{\rm free}$ was calculated with 5% of the data set.

2.5. Plasmid construction of LprF for expression in *M. smegmatis*

A DNA fragment encoding the full-length *lprF* gene and a 277 bp upstream sequence was amplified from the genomic DNA of Mtb using the polymerase chain reaction (PCR). We performed a two-step PCR. The first PCR, using the primers F-TCCGAGCCAGCGCGGCTCGGCGGA and R-TCCC-GCCGGGTTCGGGATGGTGACCGG, was performed to amplify the DNA region that includes the promoter and the *lprF* gene from the genomic DNA of Mtb. The second PCR was then performed using the DNA fragments from the first PCR as the template to add the restriction-enzyme sites and the hexahistidine tag at the C-terminus of LprF, using the primers F-CCCCGGATCCGAGCCAGCGCGGCTCGG and R-TTTAAGCTTTTAGTGATGGTGATGGTGATGTCCC-GCCGGGTTCGGGATGG. The resulting DNA fragments were digested with *Bam*HI and *Hind*III and ligated into the shuttle vector pNbv1 (Chion *et al.*, 2005), resulting in pNbv1-LprF. The DNA encodes a hexahistidine tag at the C-terminus. The A110Y mutation was introduced into the resulting plasmid using the QuikChange kit. For expression in *M. smegmatis* strain mc^2 155, the resulting plasmids were transformed by electroporation with a Micro Pulser (Bio-Rad, USA). The *M. smegmatis* strain was cultured in Middlebrook 7H9 medium (Sigma) supplemented with 0.02% Tween 80, 0.2% glucose and 50 µg ml^{-1} hygromycin at 310 K for two weeks.

2.6. Acid-fast staining and microscopy

Ziehl–Neelsen staining was performed to observe the acid fastness of the *M. smegmatis* strains (Shoub, 1923). Briefly, cells in the logarithmic growth phase were washed twice in PBS, air-dried on a glass slide and fixed by flaming. The slide was then treated sequentially with carbol–fuchsin solution, 3% HCl in ethanol and 0.3% methylene blue. Cell images were captured using a Carl Zeiss Axioskop (×1600) equipped with an MC80 exposure-time controller.

2.7. Extraction of LprF-bound lipid and TLC

The wild-type LprF protein and A110Y variant protein (1 mg each), purified from *E. coli* (as a truncated form), were subjected to lipid extraction by chloroform–methanol–water [10:10:1(*v*:*v*:*v*)] at room temperature or after 5 h heating at 353 K in chloroform–methanol–1 *N* HCl [10:10:1(*v*:*v*:*v*)]. The lipid extracts were dried *in vacuo* and dissolved in appropriate solvents for thin-layer chromatography (TLC) and mass spectrometry. TLC was performed using a pre-coated POLYGRAM SIL G/UV$_{254}$ TLC sheet (Macherey-Nagel, Düren, Germany) and developed in chloroform–methanol–formic acid–water [80:35:2:1.3(*v*:*v*:*v*:*v*)] and in chloroform–methanol–13 *M* ammonia–1 *M* ammonium acetate–water [180:140:9:9:23(*v*:*v*:*v*:*v*:*v*)] at ambient temperature. The TLC bands were scanned under UV light and visualized by

molybdenum blue staining solution to compare the R_f value with those of the phospholipids of *M. smegmatis*. The structure of PI was identified by subsequent mass-spectrometric analysis (see below).

2.8. Mass-spectrometric analysis

A truncated form and a mature form of the wild-type LprF protein overexpressed in the cytosol of *E. coli* or expressed in *M. smegmatis*, respectively, were used. The bound lipids were extracted with chloroform–methanol–water [10:10:1(*v*:*v*:*v*)]. The dried lipid was dissolved in 1 µl 50% methanol–0.1% trifluoroacetic acid and mixed with 1 µl gentisic acid matrix solution on stainless-steel plates and then crystallized under air. The lipid extracts were analyzed by tandem mass spectrometry using an AB4700 Proteomics analyzer (Applied Biosystems, Framingham, Massachusetts, USA) and a Thermo Velos Pro Mass instrument (Thermo Scientific). The AB4700 Proteomics analyzer was set to acquire negative-ion or positive-ion MS survey scans over the mass range 200–2000. Once the MS survey scans were completed, the data were processed to generate a list of monoisotopic precursor ions for the MS-MS scans with air as the collision gas at a pressure of 266 µPa. The MALDI TOF-TOF instrument was equipped with an Nd:YAG laser (355 nm, 3 ns pulse width, 200 Hz repetition rate) and was controlled by the Applied Biosystems *Explorer* v.1.1 software. The mass accuracy was calibrated to 50 p.p.m. in the full profile modes of MS and MS-MS. For the acquisition of electrospray/collision-induced MS^n data, LprF-bound lipids were extracted from wild-type LprF purified from *M. smegmatis*, the dried pellet was dissolved in 50% methanol–0.1% formic acid solution and a 1 µl sample was directly injected into a Velos Pro mass analyzer operated in negative mode with an exit voltage of −4 kV. A full-scan survey was performed between m/z 150 and 2000, and MS^n data for the three most intense ions from the preview survey scans were acquired in the ion trap with the following options: isolation width $\pm 0.8m/z$, collision energy 35%, dynamic exclusion duration 30 s. Tandem mass spectra were assigned to the fragment ions generated from the collision-induced dissociation of the precursor ions.

2.9. Analysis of mycobacterial lipid extracts

Glycolipids, phospholipids, trehalose dimycolates and mycolate methyl esters were analyzed according to previously described methods (Fukuda *et al.*, 2013). To detect LM and LAM obtained from the delipidated cell pellets by hot phenol extraction and protease K digestion, a 4–15% gradient SDS–PAGE gel was used for periodic acid Schiff staining (Doerner & White, 1990).

2.10. Adherence to hexadecane and the cellular water content

The surface hydrophobicity of 0.1× PBS-washed cells was measured by adherence to hexadecane droplets, as described elsewhere (Ofek *et al.*, 1983). To determine the cellular water content (%), the cells were filtrated on pre-weighed 0.2 µm glass-fibre filters, air-dried to evaporate intercellular water at room temperature and further dried under vacuum at 323 K. The cellular water content was calculated based on the difference in weight between the air-dried and vacuum-dried cells.

2.11. Antimicrobial susceptibility tests

The minimum inhibition concentrations (MICs) of antimycobacterial drugs were determined with *M. smegmatis* strains containing pNbv1, LprF and the A110Y variant. Broth dilutions of D-cycloserine (256 µg ml^{-1}), ethambutol (128 µg ml^{-1}), isoniazid (32 µg ml^{-1}), pyrazinamid (128 µg ml^{-1}), rifampin (128 µg ml^{-1}) and streptomycin (256 µg ml^{-1}) were performed in sterile 96-well plates containing 100 µl tryptic soy broth in each well. The cells were inoculated at ∼10^4 colony-forming units per well and cultivated at 180 rev min^{-1} and 310 K for 24–48 h before determination of the MIC values.

3. Results

3.1. Structural determination of LprF

We produced a truncated form of LprF that lacked the N-terminal residues 1–39 (the signal sequence and the cysteine in the lipobox motif) using an *E. coli* expression system. This LprF protein was identified to be a monomer in solution by gel filtration, as the protein was isolated as a single peak at ∼24 kDa (Supplementary Fig. S1[1]). Diamond plate-shaped crystals suitable for X-ray diffraction were obtained under the condition 0.2 *M* magnesium chloride, 0.1 *M* Tris–HCl pH 7.8, 28% PEG 4K at 295 K after one month. From the diffraction data, the crystals were found to belong to space group *P*1. The structure was solved using the multiwavelength anomalous dispersion (MAD) method with selenomethionyl-substituted crystals. We were able to trace most of the residues in the experimentally determined high-quality electron-density map. One protomer was contained in the asymmetric unit. The crystal structure of native LprF was refined to a resolution of 1.1 Å and further details of the structure determination and refinement are given in Table 1.

3.2. Overall structure and structural comparison with LprG and LppX

The overall LprF structure consists of a single domain in an α/β-fold with a β-sheet composed of 11 antiparallel strands and five α-helices on the opposite side of the β-sheet (Fig. 1*a*). The concave face of the β-sheet faces the α-helices, forming a large pocket (∼993 Å^3) between them (Fig. 1*b*). The hydrophobic pocket region can be divided into an entry portal and a central cavity (Fig. 1*b*). The entry portal to the cavity is partly open and is located near β3, β4 and β5, where the Pro88, Ala110, Val112, Phe125 and Ile149 residues are positioned. The central cavity, which is completely surrounded by these

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: WA5076).

residues except for the top region, is lined primarily with the side chains of hydrophobic residues (Leu162, Val77, Val79, Phe89, Val92, Val96, Tyr150, Ile155, Ile200, Val201, Pro202, Ile245 and Leu247; Fig. 1*c*). The characteristic hydrophobic environment in the pocket indicates that LprF can bind lipids in the cavity.

A *DALI* search for close structural homologues of LprF within the nonredundant set of protein structures from the PDB revealed that the lipoproteins LppX (30% sequence identity) and LprG (31% sequence identity) from Mtb share their overall fold with LprF (Holm & Rosenström, 2010). The structure of LppX is dominated by a large hydrophobic pocket suitable to accommodate a single phthiocerol dimycocerosate molecule (Sulzenbacher *et al.*, 2006; Fig. 2*a*). The crystal structure of Mtb LprG purified from *M. smegmatis* revealed a large hydrophobic pocket that accommodates the three alkyl chains of the ligand PIM (Fig. 2*b*). When compared with the structures of Mtb LprG and LppX, LprF exhibited a remarkable structural similarity to both proteins (Fig. 2*c*). LprG showed an r.m.s.d. of 0.872 Å between 145 C^{α} atoms, while LppX showed an r.m.s.d. of 1.376 Å between 109 C^{α} atoms. The hydrophobic pocket of LprG had a volume of 2680 $Å^3$ and the hydrophobic pocket of LppX had a volume of 2841 $Å^3$, and thus the pocket of LprF is approximately three times smaller than those of the others. The shape and the size of the hydrophobic pocket in LprF is relatively simple, short and narrow compared with the hydrophobic pockets of LppX and LprG. In particular, the cavity region is shallow in LprF (red boxes in Fig. 2*c*). This feature suggests that LprF binds a smaller lipid than the ligands found in LppX and LprG.

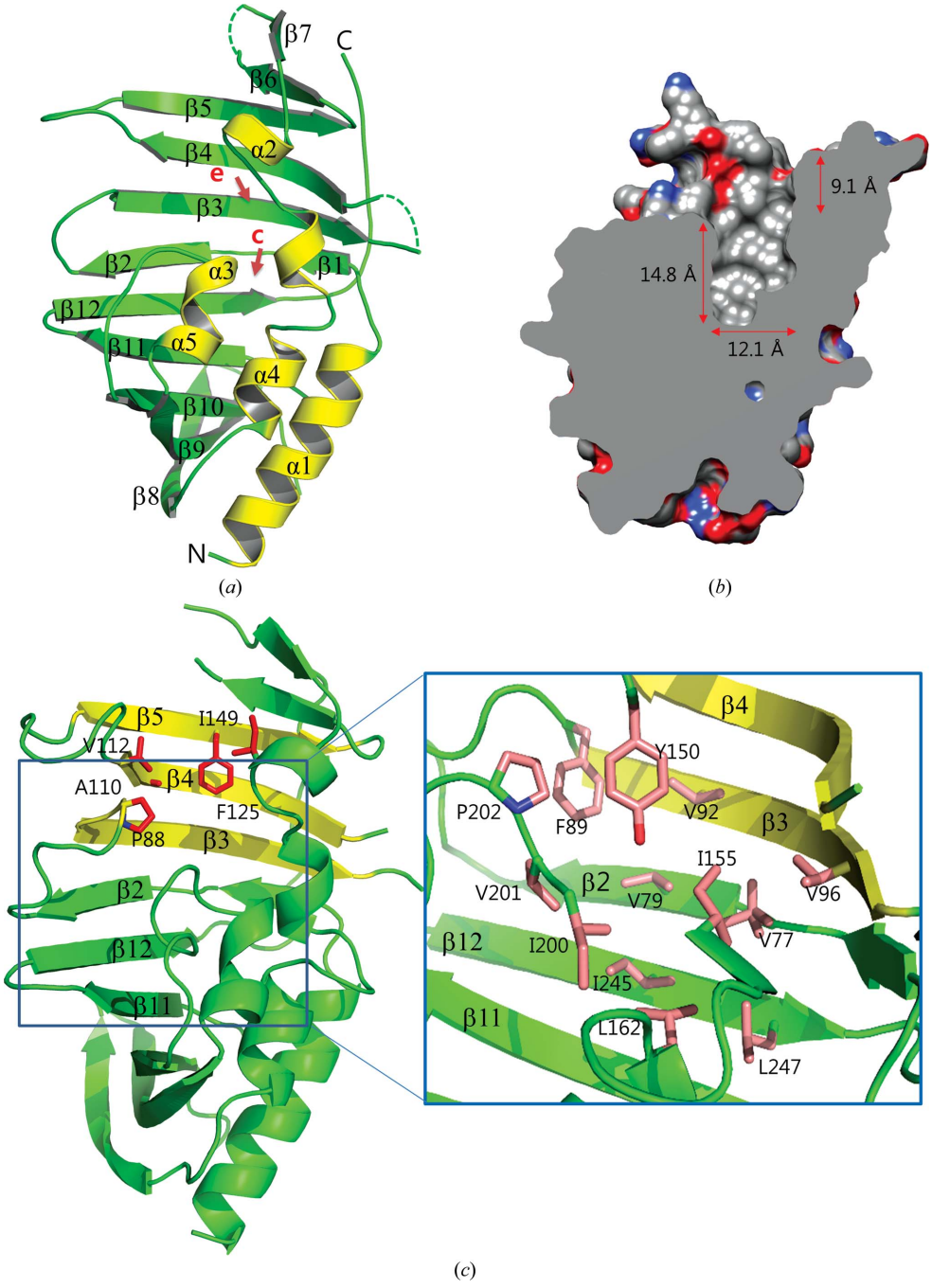

Figure 1
The crystal structure of LprF. (*a*) LprF ribbon structure. The α-helices are coloured yellow and β-sheets and loop regions are in green. The entry portal (e) and the hydrophobic cavity (c) are indicated by arrows. (*b*) LprF hydrophobic surface slab view clipped to the central pocket. The approximate dimensions of the central pocket are indicated. (*c*) The hydrophobic residues in the entry portal region (left) and in the cavity (right). The box indicates the cavity region shown in the right panel. For clarity, the β3 and β4 strands are coloured yellow.

3.3. LprF accommodates diacylated glycolipid

Most strikingly, we found a pair of chopstick-like electron densities in the hydrophobic pocket of LprF, indicating that LprF carries the diacylated lipid (Fig. 3*a*). To confirm whether the central pocket is responsible for the lipid binding as expected from the crystal structure, we designed a site-directed mutant of LprF in which the Ala110 residue located at the entry portal was substituted with a bulkier tyrosine residue

to narrow the cavity entrance (Figs. 3*b* and 3*c*). We produced the truncated form of the A110Y mutant protein in *E. coli* using the same procedure as for the wild-type LprF protein. Consistent with our expectation, we could not observe any extracted lipid moiety from the A110Y mutant protein by thin-layer chromatography (TLC), whereas the wild-type LprF protein showed a glycolipid band (Fig. 3*d*). These results demonstrate that the central cavity is the actual lipid-binding site.

3.4. Mass-spectrometric analysis of lipid extracts from LprF

To determine the structure of the lipid bound to the LprF protein, we eluted the lipid component from the TLC band and also extracted the lipid component from the purified LprF protein (residues 40–261) that was expressed in the *E. coli* cytosol and used for structural study. The structure of the ligand was determined using MALDI-TOF/TOF. As shown in Fig. 4(*a*), the ligand was 1,2-glycero-diacyl (C12:0/C15:1)-3-phospho-*myo*-inositol di-(L-β-D-heptose). The binding of phosphatidylinositol to LprF appears to be specific since only a trace amount of phosphatidylinositol is found in *E. coli* (Kozloff *et al.*, 1991). Our findings indicate that the ligand of LprF is related to glycolipids containing a diacylated phosphatidylinositol moiety, as this ligand shares the acylated phosphatidylinositol moiety with the ligands of the LprG protein as a core structure. Only the diacyl glycerol part was built into the central cavity of the structure, as no density map corresponding to the glycosyl groups was observed by X-ray crystallography (Fig. 3*a*). The interaction between LprF and the modelled ligand was mainly through van der Waals contacts between the hydrophobic side chains within the cavity (Figs. 1*b* and 1*a*).

The diacylated glycolipid in the crystal structure is derived from a mixture of lysates of *E. coli* as no fatty acid or lipid was added exogenously. To identify the ligands derived from mycobacteria, we produced the mature LprF protein in the membrane fraction by introducing the full-length gene including its own promoter region into *M. smegmatis* and extracted the lipid components from the protein. We had to express the protein in *M. smegmatis*, which grew reasonably, since it was hard to obtain and purify a sufficient amount of LprF protein from the slow-growing Mtb or *M. bovis*. The electrospray and collision-induced dissociation of the lipid generated three negative ion peaks at *m*/*z* 171, 363 and 393, although we failed to analyze the whole structure of the lipid owing to the complexity and the heterogeneity of the lipids extracted (Fig. 4*b*). The CID MS^n spectrum shows an inhomogeneous ion charge distribution around the *m*/*z* 170.83 ion peak, which is typically generated from a multiply charged ion of phosphoglyceride. Assignments of the data-dependent MS^n spectra of the *m*/*z* 363 and 393 peaks revealed that the diacylphosphoglycerol moiety contains myristate (C_{14}) and palmitate (C_{16}) groups. In agreement with the structure of the lipid-binding pocket of LprF (Fig. 3*a*), the overall structure of the mycobacterial lipid extract analyzed by electrospray and collision-induced dissociation suggests that LprF binds 1-myristoyl-2-palmitoyl-*sn*-glycero-3-phophate. When fatty-acid methyl-ester profiles of wild-type and mutant *S. smegmatis* strains used were analyzed, the percentage of myristic acid (C14:0; 2.92 $\pm$ 0.27%) was similar to the level of stearic acid (C18:0; 3.24 $\pm$ 0.75%), as shown in Supplementary Table S1. This means that the short myristyl chain is not a minor

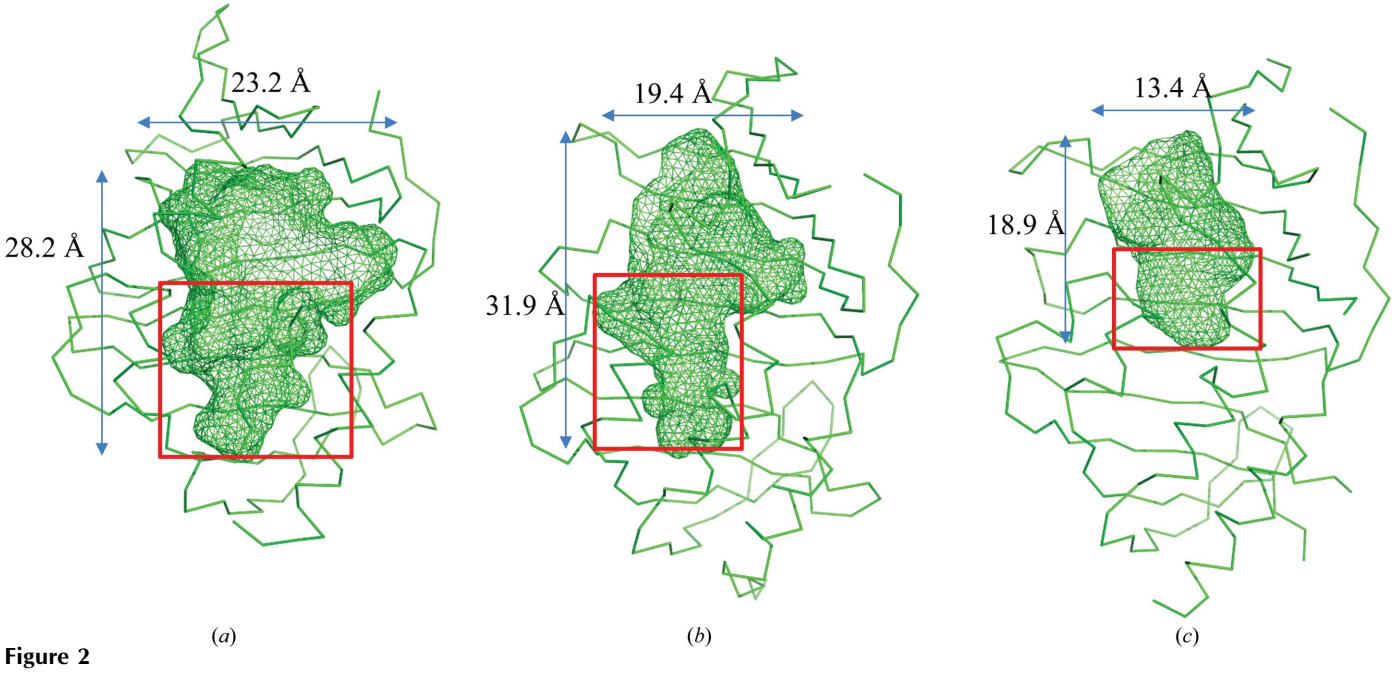

Figure 2
The different sizes of the hydrophobic pockets of LppX, LprG and LprF. The C^{α} traces of LppX, LprG and LprF are displayed in (*a*), (*b*) and (*c*), respectively. The hydrophobic pocket is shown as a mesh and the approximate sizes of the pockets are indicated. The red boxes indicate the cavity region, which is completely surrounded by amino-acid residues except at the top.

component of the mycobacterial lipid, although the most abundant fatty acids are palmitic acid (C16:0; 48.9 ± 3.5%) and oleic acid (C18:1; 12.3 ± 3.9%).

As the ligands from *E. coli* and mycobacteria share a common structural core, our results are consistent with the ligand structure from the crystal structure and the mass-spectrometric analysis using LprF expressed in *E. coli*. Furthermore, our results indicate that LprF carries a diacylated glycolipid that contains a diacylated glycerophosphate or a diacylated phosphatidylinositol moiety, which is analogous to LprG, which carries triacylated PIM, LM and LAM (Cao & Williams, 2010; Fischer *et al.*, 2004).

3.5. LprF affects the ratio of LAM/LM in the mycobacterial cell wall

Combined with the proposed function of LprG in translocating the glycolipids to the cell wall (Drage *et al.*, 2010), it was hypothesized that LprF, with its smaller pocket, is involved in the translocation of diacylated glycolipids. The diacylated glycolipids may be formed as a co-product or a reaction intermediate in the process of the biosynthesis of triacylated glycolipids (Guerin *et al.*, 2010). To test the hypothesis, we constructed an *M. smegmatis* strain transformed with a low-copy plasmid pNbv1 containing the full-length *M. bovis lprF* gene, including its own promoter region, and analyzed the LAM/LM components and lipid extracts by SDS–

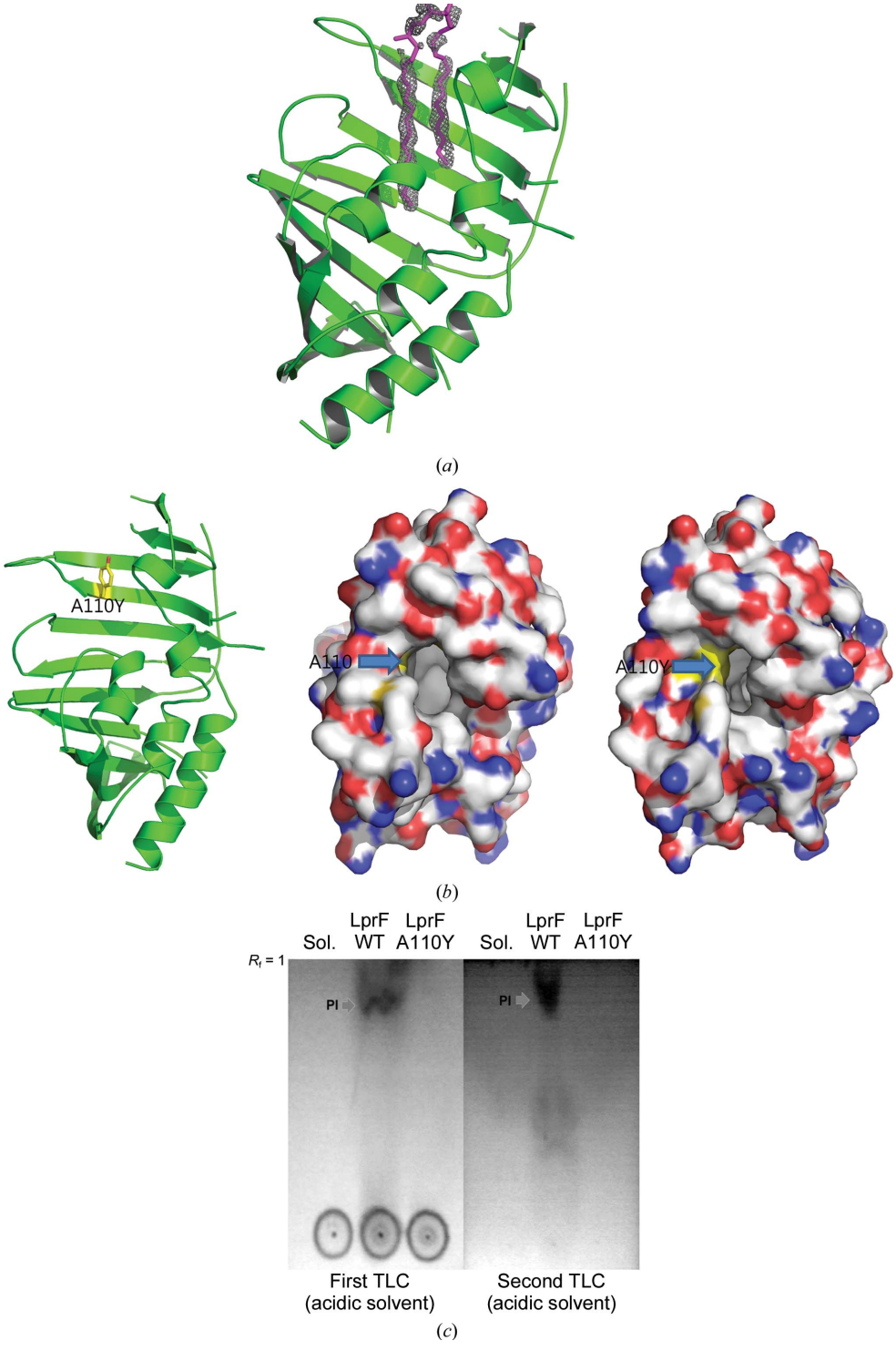

Figure 3
Diacylated glycolipid binding of LprF. (*a*) A chopstick-like electron density is shown in the hydrophobic pocket region of the LprF ribbon structure. The electron density of the $2F_o - F_c$ map is shown at the 1σ level, using the model structure without the ligand. The diacyl (C12:0/C15:1)-glycero-3-phosphate moiety is modelled in the electron-density map. (*b*) The modelled structure of the LprF A110Y variant. The altered residue is shown in yellow sticks. (*c*) A surface representation of wild-type LprF (left) and the modelled A110Y variant (right) in the top-view orientation. The arrow indicates the altered residue; C atoms are coloured grey, O atoms red and N atoms blue. Ala110 or Tyr110 is coloured yellow. Note that the central pocket is largely lined with hydrophobic residues. (*d*) Thin-layer chromatographic analysis of LprF-bound lipid. 1 mg each of wild-type LprF and A110Y variant proteins expressed in *E. coli* as a truncated form (residues 40–261) were used for lipid extraction, and concentrated spots on a TLC plate were developed using acidic (left) and basic (right) solvent systems as described in §2. From the right panel, the TLC band visualized using molybdenum blue staining solution was tentatively assigned to phosphatidylinositol (PI) by comparison of the R_f value (0.81) with those of the phospholipids of *M. smegmatis* (Fig. 5*c*).

PAGE and thin-layer chromatography. Compared with the previous results from wild-type *M. smegmatis* (Fukuda *et al.*, 2013), the pNbv1 vector insertion resulted in no significant difference in the lipid composition, although the band positions (R_f) differed between the two TLC experiments. As shown in Fig. 5(*a*), the ratio of LAM to LM resolved by

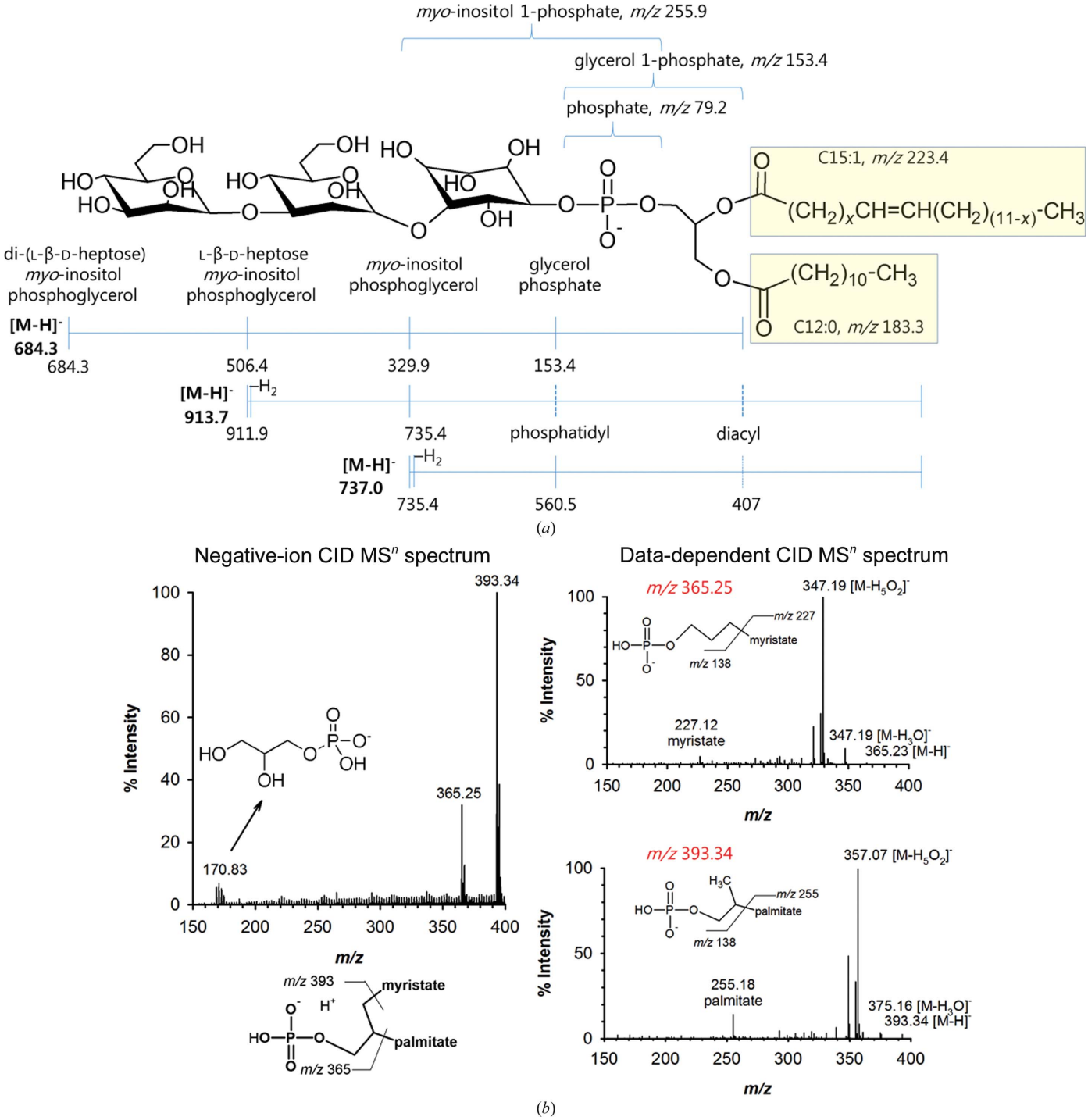

Figure 4
Ligand structures analyzed by mass spectrometry. (*a*) Overall structure of 1,2-glycero-diacyl (C12:0/C15:1)-3-phospho-*myo*-inositol di-(L-*β*-D-heptose) assigned based on the MALDI TOF-TOF mass spectrometric results given in Supplementary Figs. S2(*a*)–S2(*c*). The mass-spectrometric analysis of lipid extracted from a truncated form of LprF overproduced in *E. coli* suggests that LprF binds to a phosphophatidylinositol ligand. (*b*) Electrospray/collision-induced dissociation of lipid extract from the mature LprF protein purified from the membrane fraction of *M. smegmatis*. The left panel shows the generation of *m*/*z* 170.83, 365.25 and 393.24 ions by negative-ion CID of the lipid extract, and the inset shows the chemical structure of phosphoglyceride identified by the generation of an inhomogeneous ion-charge distribution of multiply charged ions around the *m*/*z* 170.83 ion peak. In the right panels, data-dependent MS^n spectra of *m*/*z* 365.25 and 393.34 ions are assigned to the fragment ions generated from monoacyl deoxy-glycero-3-phosphate ions. The CID MS^n spectra suggest that the mature LprF binds to 1-myristoyl-2-palmitoyl-*sn*-glycero-3-phophate1, as shown below.

Table 2
Water content and hexadecane adsorption of *M. smegmatis* cells containing LprF, LprF A110Y and empty pNbv1 vector.

Results from three independent cell cultures are reported as the mean ± standard deviation. The percentage adherence of cells to droplets of hexadecane at the initial OD_{600} of 0.5 was obtained from the difference in percentage turbidity between the initial and minimum values estimated by a three-parameter exponential decay equation fit to the time-course curves in Fig. 6(*c*). Statistically significant differences ($p < 0.001$; marked with an asterisk) are shown by *F*-tests on the time scales of hexadecane adsorption.

Strain	pNbv1	LprF	A110Y
Water content of the cell (%)	78.9 ± 1.2	79.5 ± 0.8	80.9 ± 0.6
Adherence to hexadecane (%)	4.4 ± 0.1	0.8 ± 0.2*	7.1 ± <0.1*

SDS–PAGE was significantly raised from LprF-expressing *M. smegmatis* when compared with cells containing the empty plasmid and the A110Y variant plasmid. However, the compositions of other lipid extracts, including glycolipids (PIMs), phospholipids, glycolipoproteins, trehalose dimycolate and mycolic acid derivatives, were not significantly changed by the expression of LprF (Figs. 5*b*–5*e*). Our findings suggest that LprF may be involved in the preferential translocation of diacylated LAM to the outer cell wall compared with diacylated LM (Kang *et al.*, 2005). An alternative explanation is that the diacylated LAM is more abundant than the diacylated LM. However, it cannot be excluded that other diacylated PIMs or LM are also recognized by LprF, and further study is needed to elucidate the mechanism.

3.6. LprF increases the resistance of *M. smegmatis* to ethambutol

To determine whether LprF affects the cell-wall integrity, we measured the sensitivity of each *M. smegmatis* stain to membrane-permeable compounds such as crystal violet,

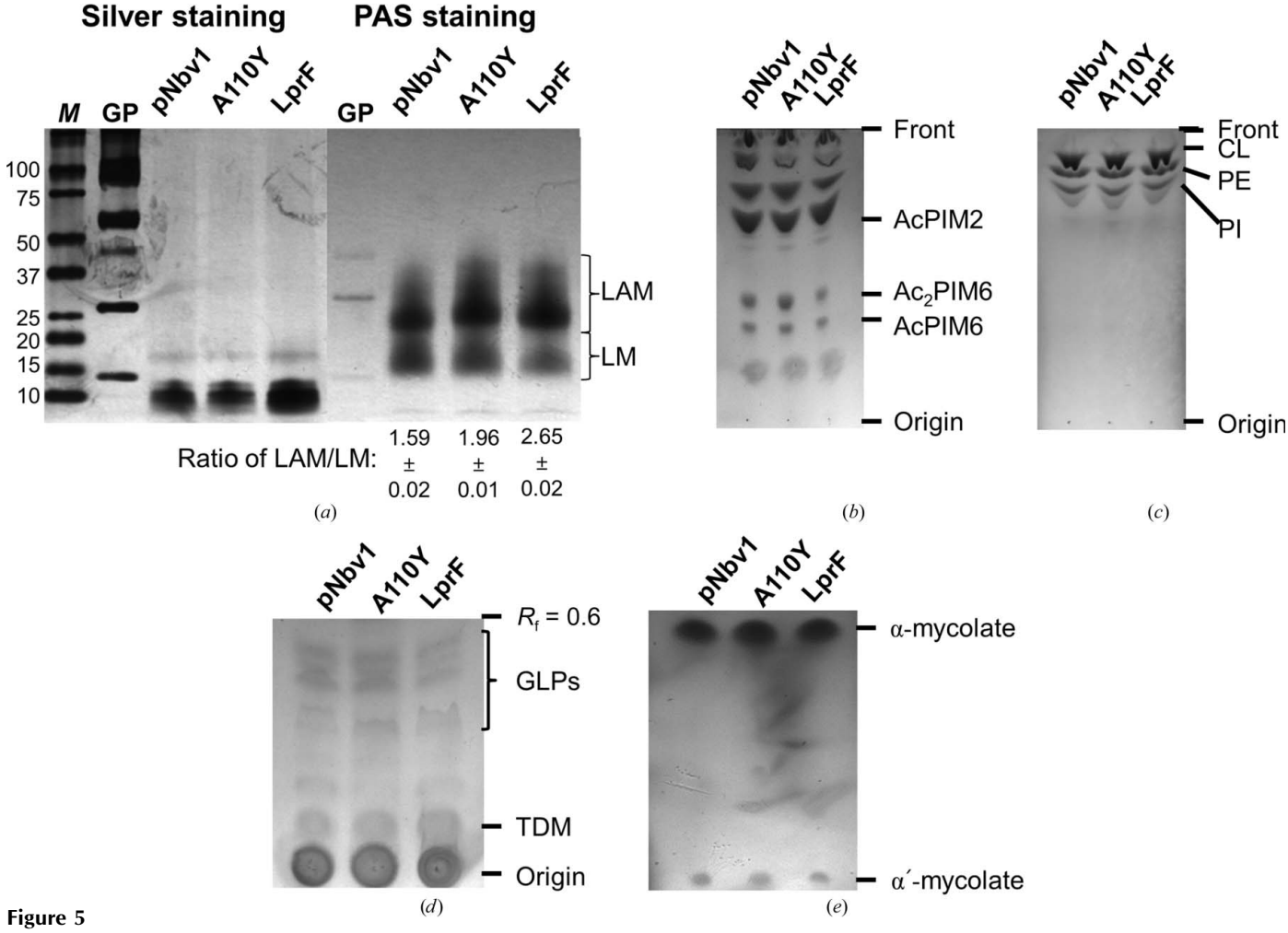

Figure 5
Effects of LprF on the composition of mycobacterial lipids. (*a*) Lipomannan (LM) and lipoarabinomannan (LAM) separated by SDS–PAGE. Bands are visualized by silver staining (left) and periodic acid Schiff (PAS) staining (right), and the relative intensity of LAM to LM calculated by the *ImageQuant* v.5.2 program is shown below the PAS-stained gel. Lane GP, CandyCane glycoprotein molecular-weight standards (Molecular Probes, Invitrogen). (*b*) Glycolipids in silica gel developed by chloroform–methanol–13 *M* ammonia–1 *M* ammonium acetate–water (180:140:9:9:23) as a solvent system and visualized by orcinol staining. $AcPIM_2$ and $AcPIM_6$ are triacylated phosphatidylinositiol species carrying two and six mannoses, respectively. (*c*) Phospholipids in silica gel developed using the same solvent as the glycoproteins in (*b*) and visualized by molybdenum blue staining. CL, cardiolipin; PE, phosphatidylethanolamine; PI, phosphatidylinositol. (*d*) Glycopeptidolipids (GLPs) and trehalose dimycolate in silica gel developed by chloroform–methanol (9:1) and visualized by orcinol staining. (*e*) Methyl esters of α- and α′-mycolic acids extracted from the peptidoglycan–arabinogalactan core. Mycolic acid methyl esters were separated in silica gel by hexane–ethyl ether (4:1) and visualized by chromic acid staining.

Table 3
Minimum inhibitory concentrations (mg l^{-1}) of antimycobacterial drugs.

Strain	pNbv1	LprF	A110Y
D-Cycloserine	>256	>256	>256
Ethambutol	1	2-4	1
Isoniazid	8	8	8
Pyrazinamide	>256	>256	>256
Rifampin	64	64	64
Streptomycin	4	4	4

malachite green and SDS (Fukuda *et al.*, 2013). These lipophilic compounds are toxic to mycobacteria and have been used to test the permeability of the membrane. As shown in Fig. 6(*a*), the introduction of LprF into *M. smegmatis* did not affect the sensitivity of the mycobacterium to membrane-permeable compounds. Moreover, each strain containing wild-type, A110Y mutant LprF or empty pNbv1 plasmid displayed similar cell morphology and acid-fast staining, indicating little or no effect of LprF on the mycolic acid layer (Fig. 6*b*). Taken together, our results indicate that the cell-wall integrity was not significantly changed by LprF.

To assess whether LprF influences the physicochemical properties of the mycolate layer, we measured the hydrophobicity of the cell surface. The expression of wild-type full-length LprF significantly decreased the adherence to hexadecane droplets, indicating a decrease in the hydrophobicity of the cell surface, but did not affect the amount of water contained in the mycobacterial cells (Table 2 and Fig. 6*c*). However, the A110Y mutation in LprF resulted in an increase in the adherence of cells to the hydrophobic hexadecane in an inverse relationship compared with wild-type LprF. These results suggest that LprF might influence the cell-surface hydrophobicity of the mycobacteria without affecting the cellular water content. We next measured whether LprF affects the susceptibility of antimycobacterial drugs using the *M. smegmatis* strains. It was found that LprF rendered *M. smegmatis* resistant to low levels of ethambutol (EMB) compared wth the other strains containing pNbv1 and the A110Y variant (Table 3). Since EMB resistance of mycobacteria has been implicated in the glycosyl composition of LAM (Khoo *et al.*, 2001), the change in the composition of LAM by LprF could explain the increased resistance to EMB.

4. Discussion

In this study, we determined the crystal structure of LprF at high resolution, revealing a central hydrophobic pocket that is smaller than the pockets of the homologous proteins LprG and LppX. A glycolipid containing two acyl chains was found in the crystal structure of LprF. The introduction of a single point mutation in this pocket blocked the glycolipid-binding function of LprF, confirming the binding ability of the pocket. We subsequently found that the function of LprF is related to the change in the cell-wall property, presumably by the translocation of the diacylated glycolipid from the plasma membrane to the mycobacterial outer membrane. The lipid-extract analysis further suggested that LprF increases the LAM:LM ratio of the mycobacterial cell wall, indicating a preferential transfer of LAM by LprF. Furthermore, we observed that LprF increased the resistance to EMB when the gene was introduced into *M. smegmatis*, which might be related to the altered LAM composition.

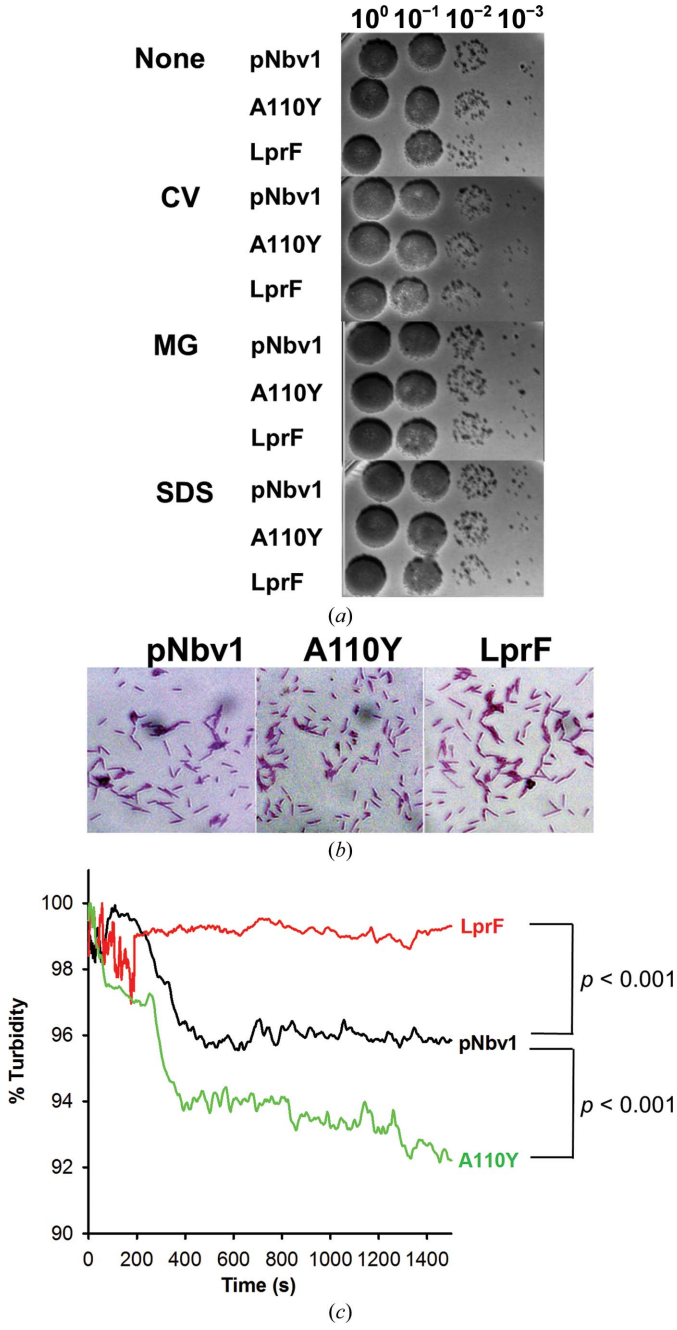

Figure 6
Cell-wall integrity and hydrophobicity. (*a*) Sensitivity of cells containing wild-type LprF, A110Y mutant LprF and empty pNbv1 vector to the membrane-permeable chemicals crystal violet (CV), malachite green (MG) and SDS. (*b*) Acid-fast staining of cells containing wild-type LprF, A110Y mutant LprF and empty pNbv1 vector. (*c*) Time-course curves of hexadecane adsorption of stationary-phase *M. smegmatis* cells containing wild-type LprF, A110Y mutant LprF and empty pNbv1 vector. Statistically significant differences (*F*-tests; $p < 0.001$) between strains are shown in the time course of hexadecane adsorption.

LprG is known to be important for the virulence of mycobacteria (Farrow & Rubin, 2008; Drage *et al.*, 2010). As this glycolipid-carrier function of LprG overlaps with the function of LprF, it is implicated in mycobacterial physiology and host–pathogen interactions. Unlike LprG, LprF is found in Mtb and *M. bovis* but is not found in the nonpathogenic *M. smegmatis*. The cell-wall compositions of Mtb (or *M. bovis*) and *M. smegmatis* are very similar but are not identical (Mishra *et al.*, 2011). Since the variation in the glycolipids in the cell-wall composition could determine the pathogenesis of the mycobacteria, the glycolipid-carrier function and the change in the LAM:LM ratio by LprF may be more significantly associated than LprG with the pathogenesis of Mtb and *M. bovis*. The glycosyl groups of LAM or LM may constitute an important structural entity engaged in receptor binding and subsequent immunopathogenesis (Khoo *et al.*, 2001), because the glycosyl groups of the glycolipids are exposed and can be directly recognized by the host proteins.

What might the difference in the functions of LprG and LprF be in terms of the pathogenesis of the mycobacteria? The structural and functional difference between the two proteins lies in the number of acyl chains bound in the hydrophobic pocket of each protein. The mycobacterial glycolipids are suggested to regulate the host immune response *via* TLR2 (Chatterjee, 1997; Brennan, 2003). In the host immune system, the diacylated moiety and the triacyl moiety are recognized by different toll-like receptors (TLRs). The TLR2–TLR1 heterodimer complex recognizes triacylated lipopeptide and activates the immune response (Morr *et al.*, 2002), whereas the TLR2–TLR6 heterodimer recognizes diacylated lipopeptide (Jin *et al.*, 2007). The binding pocket of TLR2 binds only a diacyl chain, and the third acyl chain of the lipoprotein binds to TLR1. In contrast, TLR6 does not have a prominent acyl-binding pocket (Jin *et al.*, 2007). It has been reported that when TLR2 (most likely in the TLR2–TLR1 heterodimer complex) recognizes the triacylated glycolipid carried by LprG, the recognition of the glycolipid is facilitated by the lipid-carrier function of LprG, leading to activation of the immune response of the host (Drage *et al.*, 2010). Considering the roles of lipoproteins in the activation of the immune response, there are two possible ways for LprF to carry a diacylated glycolipid. The diacylated glycolipid recognized by the TLR2–TLR6 heterodimer may mimic the diacylated lipoprotein that activates the immune response. In contrast, the diacylated glycolipid may antagonize signalling other than TLR2–TLR6-mediated signalling by binding to TLR2. To elicit the function of LprF in the pathogenesis of Mtb and *M. bovis* and host cell-surface receptors, further investigations are needed.

In conclusion, this study revealed the hydrophobic pocket of LprF, which carries a diacylated glycolipid. Lipid analysis suggested that LprF functions in mycobacteria as a carrier of diacylated glycolipids (probably LAM) from the plasma membrane to the outer cell membrane, increasing the ratio of LAM to LM. This result provides us with a new insight into the unique role of LprF in the pathogenic Mtb and *M. bovis*, although further study is required to understand the function of LprF in the context of the host–pathogen interaction.

The authors declare no competing financial interests. This study involved the use of beamline 7A at Pohang Accelerator Laboratory (Pohang, Republic of Korea). This study was supported by a grant from the Ministry of Knowledge Economy (MKE), Korea Institute for Advancement of Technology (KIAT) through the Inter-ER Cooperation Projects (R0000577) to N-CH and a grant from the National Research Foundation of Korea (Grant No. 2013R1A1A2061369) to Y-HK.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Baena, A. & Porcelli, S. A. (2009). *Tissue Antigens*, **74**, 189–204.
Bhowruth, V., Dover, L. G. & Besra, G. S. (2007). *Prog. Med. Chem.* **45**, 169–203.
Brennan, P. J. (2003). *Tuberculosis*, **83**, 91–97.
Briken, V., Porcelli, S. A., Besra, G. S. & Kremer, L. (2004). *Mol. Microbiol.* **53**, 391–403.
Cao, B. & Williams, S. J. (2010). *Nat. Prod. Rep.* **27**, 919–947.
Chan, J., Fujiwara, T., Brennan, P., McNeil, M., Turco, S. J., Sibille, J. C., Snapper, M., Aisen, P. & Bloom, B. R. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2453–2457.
Chan Kwo Chion, C. K., Askew, S. E. & Leak, D. J. (2005). *Appl. Environ. Microbiol.* **71**, 1909–1914.
Chatterjee, D. (1997). *Curr. Opin. Chem. Biol.* **1**, 579–588.
Chatterjee, D., Hunter, S. W., McNeil, M. & Brennan, P. J. (1992). *J. Biol. Chem.* **267**, 6228–6233.
Daffé, M. & Draper, P. (1998). *Adv. Microb. Physiol.* **39**, 131–203.
Doerner, K. C. & White, B. A. (1990). *Anal. Biochem.* **187**, 147–150.
Drage, M. G., Tsai, H.-C., Pecora, N. D., Cheng, T.-Y., Arida, A. R., Shukla, S., Rojas, R. E., Seshadri, C., Moody, D. B., Boom, W. H., Sacchettini, J. C. & Harding, C. V. (2010). *Nature Struct. Mol. Biol.* **17**, 1088–1095.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Farrow, M. F. & Rubin, E. J. (2008). *J. Bacteriol.* **190**, 1783–1791.
Fischer, K., Scotet, E., Niemeyer, M., Koebernick, H., Zerrahn, J., Maillet, S., Hurwitz, R., Kursar, M., Bonneville, M., Kaufmann, S. H. & Schaible, U. E. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 10685–10690.
Fukuda, T., Matsumura, T., Ato, M., Hamasaki, M., Nishiuchi, Y., Murakami, Y., Maeda, Y., Yoshimori, T., Matsumoto, S., Kobayashi, K., Kinoshita, T. & Morita, Y. S. (2013). *mBio*, **4**, e00472-12.
Garnier, T. *et al.* (2003). *Proc. Natl Acad. Sci. USA*, **100**, 7877–7882.
Gilleron, M., Quesniaux, V. F. & Puzo, G. (2003). *J. Biol. Chem.* **278**, 29880–29889.
Guerin, M. E., Korduláková, J., Alzari, P. M., Brennan, P. J. & Jackson, M. (2010). *J. Biol. Chem.* **285**, 33577–33583.
Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. M. & Engelhardt, H. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 3963–3967.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Jarlier, V. & Nikaido, H. (1994). *FEMS Microbiol. Lett.* **123**, 11–18.
Jin, M. S., Kim, S. E., Heo, J. Y., Lee, M. E., Kim, H. M., Paik, S.-G., Lee, H. & Lee, J.-O. (2007). *Cell*, **130**, 1071–1082.
Kang, P. B., Azad, A. K., Torrelles, J. B., Kaufman, T. M., Beharka, A., Tibesar, E., DesJardin, L. E. & Schlesinger, L. S. (2005). *J. Exp. Med.* **202**, 987–999.
Khoo, K.-H., Tang, J.-B. & Chatterjee, D. (2001). *J. Biol. Chem.* **276**, 3863–3871.

Kozloff, L. M., Turner, M. A., Arellano, F. & Lute, M. (1991). *J. Bacteriol.* **173**, 2053–2060.
Marland, Z., Beddoe, T., Zaker-Tabrizi, L., Lucet, I. S., Brammananth, R., Whisstock, J. C., Wilce, M. C., Coppel, R. L., Crellin, P. K. & Rossjohn, J. (2006). *J. Mol. Biol.* **359**, 983–997.
Minnikin, D. E., Kremer, L., Dover, L. G. & Besra, G. S. (2002). *Chem. Biol.* **9**, 545–553.
Mishra, A. K., Driessen, N. N., Appelmelk, B. J. & Besra, G. S. (2011). *FEMS Microbiol. Rev.* **35**, 1126–1157.
Morr, M., Takeuchi, O., Akira, S., Simon, M. M. & Mühlradt, P. F. (2002). *Eur. J. Immunol.* **32**, 3337–3347.
Nigou, J., Gilleron, M. & Puzo, G. (1999). *Biochem. J.* **337**, 453–460.
Ofek, I., Whitnack, E. & Beachey, E. H. (1983). *J. Bacteriol.* **154**, 139–145.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Russell, D. G. (2001). *Nature Rev. Mol. Cell Biol.* **2**, 569–577.
Shoub, H. L. (1923). *J. Bacteriol.* **8**, 121–126.
Sulzenbacher, G., Canaan, S., Bordat, Y., Neyrolles, O., Stadthagen, G., Roig-Zamboni, V., Rauzier, J., Maurin, D., Laval, F., Daffé, M., Cambillau, C., Gicquel, B., Bourne, Y. & Jackson, M. (2006). *EMBO J.* **25**, 1436–1444.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.