Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A structural and functional investigation of a novel protein from *Mycobacterium smegmatis* implicated in mycobacterial macrophage survivability

Adam Shahine,[a,b] Dene Littler,[a,b] Rajini Brammananath,[b,c] Phooi Y. Chan,[b,c] Paul K. Crellin,[b,c] Ross L. Coppel,[b,c] Jamie Rossjohn[a,b,d]* and Travis Beddoe[a,b,e]*

[a]Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia, [b]Australian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia, [c]Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia, [d]Institute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, Wales, and [e]Department of Agriculture Sciences, La Trobe University, Bundoora, Victoria, Australia

Correspondence e-mail: jamie.rossjohn@monash.edu, t.beddoe@latrobe.edu.au

Received 11 December 2013
Accepted 13 May 2014

PDB reference:
MSMEG_5817, 4nss

The success of pathogenic mycobacterial species is owing in part to their ability to parasitize the generally inhospitable phagosomal environment of host macrophages, utilizing a variety of strategies to avoid their antimycobacterial capabilities and thereby enabling their survival. A recently identified gene target in *Mycobacterium smegmatis*, highly conserved within *Mycobacterium* spp. and denoted MSMEG_5817, has been found to be important for bacterial survival within host macrophages. To gain insight into its function, the crystal structure of MSMEG_5817 has been solved to 2.40 Å resolution. The structure reveals a high level of structural homology to the sterol carrier protein (SCP) family, suggesting a potential role of MSMEG_5817 in the binding and transportation of biologically relevant lipids required for bacterial survival. The lipid-binding capacity of MSMEG_5817 was confirmed by ELISA, revealing binding to a number of phospholipids with varying binding specificities compared with *Homo sapiens* SCP. A potential lipid-binding site was probed by alanine-scanning mutagenesis, revealing structurally relevant residues and a binding mechanism potentially differing from that of the SCPs.

1. Introduction

The genus *Mycobacterium* includes a number of deadly human pathogens, most notably *M. tuberculosis*, the causative agent of tuberculosis (TB). Despite extensive usage of the BCG vaccine and a series of drugs against TB, approximately one-third of the world's population is still infected, resulting in approximately 1.4 million deaths in 2011 (World Health Organization, 2012). The emergence of multiple drug-resistant (MDR) and extensively drug-resistant (XDR) strains of *M. tuberculosis* emphasizes the urgent need for the development of new antimycobacterials with novel modes of activity.

Pathogenic mycobacterial species have developed a multitude of mechanisms to parasitize the inhospitable environment of the macrophage, which is central to mycobacterial infection, latency and disease activation (Russell, 2001, 2011; Deretic *et al.*, 2006). Macrophages act as the main reservoir of infection for invading mycobacterial species, which contrasts with the accepted stance that these cells successfully act as our initial line of defence against bacterial infection. Invading mycobacteria are internalized by host macrophages through phagocytosis and are sorted into phagosomes, in which they are retained. At this stage, invading mycobacteria are able to arrest phagolysosome maturation, prolonging their survival (Armstrong & Hart, 1971; Vergne *et al.*, 2004). Invading mycobacteria are retained within the phagosome until the infected macrophage dies through necrosis or apoptosis.

Invading mycobacteria manipulate the host-signalling pathways required for correct vesicular trafficking and

therefore the infected phagosomes fail to mature and do not fuse with late endosomes for subsequent destruction by the lysosomes (Armstrong & Hart, 1971; Sturgill-Koszycki *et al.*, 1994; Deretic *et al.*, 2006). Additionally, invading mycobacteria have the ability to modulate the internal pH of the phagosomal lumen, retaining its neutral vacuolar pH of 6.2–6.4 and thus creating a relatively hospitable environment for survival (Via *et al.*, 1997). Elucidation of the mechanisms utilized by invading mycobacteria to corrupt host macrophage activities is vital for a greater insight into mycobacterial virulence and disease.

Invading mycobacteria utilize a wide variety of proteins and effector molecules to parasitize the macrophage phagosome, resulting in the absence of mature lysosomal hydrolases and incomplete luminal acidification (Via *et al.*, 1997). Invading mycobacteria have the ability to block phagolysosome biogenesis by modulating the function of macrophage small GTP-binding proteins (Rabs; Clemens *et al.*, 2000*a*). The Rab protein family, whose role is to control the identity of intracellular organelles and direct membrane trafficking *via* a molecular on–off switch mechanism (Clemens *et al.*, 2000*a*,*b*), plays a crucial role in macrophage phagolysosome formation. Through the wholesale conversion of Rab5 (early endocytosis) to Rab7 (late endosome), the phagosome matures to allow the correct generation of the phagolysosome (Fratti *et al.*, 2001; Pereira-Leal & Seabra, 2001; Kelley & Schorey, 2003; Pfeffer, 2005; Rink *et al.*, 2005). Normal Rab5 to Rab7 conversion is modulated by the Rab5 modulator EEA1 (early endosomal autoantigen 1) that associates with phosphatidylinositol 3-phosphate (PI3P) on organelle membranes, an interaction that is essential for proper membrane trafficking within the endosomal system (Sturgill-Koszycki *et al.*, 1994; Fratti *et al.*, 2003). Invading mycobacteria have the ability to block the conversion of Rab5 to Rab7 in a process known as the Rab conversion block (Clemens *et al.*, 2000*a*), utilizing two key mechanisms unique to mycobacteria. The lipoglycan lipoarabinomannan (LAM; Sturgill-Koszycki *et al.*, 1994; Fratti *et al.*, 2003; Hmama *et al.*, 2004; Kang *et al.*, 2005) and the enzyme SapM (a PI3P phosphatase; Saleh & Belisle, 2000; Vergne *et al.*, 2005) act to prevent the generation of PI3P within the macrophage and to break down excess PI3P, respectively, ensuring blocking of phagosome maturation (Deretic *et al.*, 2006). While we have a basic knowledge of mycobacterial phagosome maturation blocking, it is likely that additional mechanisms remain to be found. The identification and characterization of the mechanisms of macrophage survivability by pathogenic mycobacterial species will aid in further understanding mycobacterial pathogenesis, paving the way for the development of novel treatments.

The saprophytic *M. smegmatis* has been exploited as a model organism to investigate *M. tuberculosis* pathogenesis. As such, *M. smegmatis* has been used to investigate host–pathogen interactions in macrophages, as *M. smegmatis* has been shown to be able to mimic the growth and survival of *M. tuberculosis* within macrophages (Anes *et al.*, 2006). To investigate key factors in mycobacterial survival in macrophages, a random transposon mutant library of *M. smegmatis* was screened for reduced growth in macrophages (Billman-Jacobe *et al.*, 1999; Patterson *et al.*, 2000; Kovacevic *et al.*, 2006; Pelosi *et al.*, 2012). One such mutant, denoted Myco132, resulted in accelerated cell death of *M. smegmatis* in J774A.1 macrophages within the first 8 h (Pelosi *et al.*, 2012). The defective gene in Myco132 was identified as *MSMEG_5817*, which encodes a conserved hypothetical protein of 128 amino-acid residues with orthologues in a large variety of pathogenic and nonpathogenic mycobacteria (Pelosi *et al.*, 2012). However, MSMEG_5817 is most likely not required for normal growth, as Rv0807, the *M. tuberculosis* homologue of MSMEG_5817, has been shown to be non-essential (Sassetti *et al.*, 2003; Griffin *et al.*, 2011). Additionally, there was an up-regulation of NF-κB activation in macrophages infected with the *MSMEG_5817* mutant compared with those infected with wild-type *M. smegmatis* (Pelosi *et al.*, 2012). NF-κB activation plays key roles in driving host immune responses through the production of cytokines, suggesting that MSMEG_5817 may be involved in suppression of the NF-κB activation pathway to promote survivability. In addition, other studies have shown that Rv0807 is upregulated upon phosphate starvation owing to the loss of the two-component *senX3–regX3* regulatory system, suggesting that Rv0807 may play a role in phosphate metabolism (Parish *et al.*, 2003).

Upon conducting a *BLASTp* similarity search utilizing the nonredundant protein-sequence database, it was found that MSMEG_5817 shared high sequence identity with a vast number of orthologues within the Corynebacterineae sub-order, each of unknown function. Significant homology was also detected to a serine:pyruvate aminotransferase (SPT) from *Brevibacterium linens* BL2 (accession No. ZP_00381195). Upon comparing the SPT activities of wild-type *M. smegmatis* and the Myco132 mutant, it was found that the relative rate of pyruvate reduction was highly similar between the samples tested, indicating that MSMEG_5817 does not encode a functional SPT (Pelosi *et al.*, 2012).

We have previously cloned and expressed MSMEG_5817 as a recombinant protein from *Escherichia coli*, including the production of a selenomethionine derivative, and have performed purification, crystallization and preliminary X-ray diffraction studies to 2.40 Å resolution (Shahine *et al.*, 2013).

As a means of ascertaining its function, we set out to solve the crystal structure of MSMEG_5817. This structure would provide valuable insight into the structural characteristics of the protein, as well as to gauge a potential functional role through structural homology searches. Here, we describe the structural features of MSMEG_5817 and present preliminary evidence for its functional role.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant MSMEG_5817 protein

2.1.1. MSMEG_5817 and alanine-scanning mutagenesis mutants for *E. coli* expression. The gene encoding MSMEG_5817 was cloned and overexpressed in *E. coli*, and

both native and selenomethionine-derivative recombinant proteins were purified as described previously (Shahine *et al.*, 2013). *MSMEG_5817* mutants were generated from the parental pET-28c-*MSMEG_5817* plasmid DNA through site-directed mutagenesis. The mutants were PCR-amplified using *Pfu* DNA polymerase (Promega) in 10× *Pfu* buffer with specific sense and antisense primers. The PCR-amplified product was digested with *Dpn*I (BioLabs) for 2 h at 37°C before being transformed into chemically competent *E. coli* XL1 Blue cells. Successfully cloned transformants were confirmed by DNA sequencing. The expression and purification of recombinant MSMEG_5817 and its variants were conducted as described previously (Shahine *et al.*, 2013).

2.1.2. *Homo sapiens* sterol carrier protein (SCP). The expression vector pQTEV-SPC2 cloned into *E. coli* DH5α cells was purchased from Addgene (ID 31336). The sterol carrier protein from *H. sapiens* was expressed with an N-terminal hexahistidine tag with a TEV protease cleavage site. The pQTEV-SCP2 vector was transformed into *E. coli* BL21 (DE3) cells for recombinant protein expression, with successful transformants selected on LB–agar in the presence of 100 µg ml^{-1} ampicillin. Recombinant protein was expressed and purified in the same manner as described previously (Shahine *et al.*, 2013).

2.2. Crystallization, data collection and data processing of MSMEG_5817

Crystals of both native and selenomethionine-derivative MSMEG_5817 were obtained as described previously, with data collection and data processing conducted as described previously (Shahine *et al.*, 2013).

2.3. Structural determination of MSMEG_5817

The structure of MSMEG_5817 was determined as follows. Using anomalous differences, two selenium positions were identified and refined using *AutoSol* as part of the *PHENIX* program package using experimentally determined f' and f'' values (Terwilliger *et al.*, 2009) that were obtained previously (Shahine *et al.*, 2013) with default settings. The top solution resulted in a FOM value of 0.59 and an estimated CC × 100 value of 66.7 ± 12.9. Upon density modification the final R factor was 28.32%. The resultant density-modified Fourier map, based on the three selenomethionine-derivative MSMEG_5817 data sets, was used for automated model building with the *AutoBuild* program of the *PHENIX* program suite (Terwilliger *et al.*, 2008) with default settings. 210 of 256 residues were successfully built, with two molecules in the unit cell with an R_{work} of 24.68% and an R_{free} of 29.11%. The model built by *AutoBuild* was used as a molecular-replacement model for a 2.40 Å resolution native data set using *Phaser-MR* in the *PHENIX* program suite (Adams *et al.*, 2010), with the top solution solved with a TFZ value of 10.0. For cross-validation, a random set of 5% of the total reflections were kept aside from the refinement and used for the calculation of R_{free} (Brünger, 1992). The model was manually adjusted using the interactive graphics program *Coot* (Emsley *et al.*, 2010) and refined using *BUSTER* (Bricogne *et al.*, 2011) until no further improvement could be made, resulting in a final R_{work} of 18.92% and R_{free} of 22.52% (Table 1). The final round of *BUSTER* refinement was conducted with the `-TLSBasic` and `-autoncs` refinement settings. Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession code 4nss.

Table 1
Structural refinement and validation of MSMEG_5817.

Resolution range (Å)	40.42–2.40
Completeness (%)	98.00
R_{work}† (%)	18.92
R_{free}† (%)	22.52
R.m.s.d.s	
Bond lengths (Å)	0.010
Bond angles (°)	1.04
Ramachandran plot	
Favoured (%)	98.20
Allowed (%)	0.90
Outliers (%)	0.90
No. of modelled residues	230
No. of water molecules	71
Average B factors (Å^2)	
Protein atoms	61.05
Water molecules	60.93
PDB code	4nss

† $R_{work} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$ for all data excluding the 5% that comprise the R_{free} set used for cross-validation.

2.4. Structural analysis of MSMEG_5817

Refinement validation was conducted using *MolProbity* and the *POLYGON* tool in the *PHENIX* program suite (Adams *et al.*, 2010). All structural superpositions were achieved using the *SSM Superpose* feature of *Coot* (Emsley *et al.*, 2010). Secondary structure was confirmed using the *STRIDE* plugin in *PyMOL* (Zhu, 2011). Surface-area calculations were performed using the *APBS* plugin for *PyMOL* (Baker *et al.*, 2001). *H. sapiens* sterol carrier protein-2 (PDB entry 1ikt; Haapalainen *et al.*, 2001), *Aedes aegypti* sterol carrier protein-2 (PDB entry 1pz4; Dyer *et al.*, 2003) and *Oryctolagus cuniculus* sterol carrier protein-2 domain (PDB entry 1c44; Choinowski *et al.*, 2000) were used for all structural alignments.

2.5. Circular-dichroism spectra

Circular-dichroism spectra of MSMEG_5817 and variants were recorded on a Jasco J-815 circular-dichroism spectropolarimeter at 20°C with a data pitch of 0.1 nm and a scan speed of 100 nm min^{-1}. Briefly, all MSMEG_5817 samples were prepared at a concentration of 30 µ*M*, with the exception of the M109A and L121A mutants, which were prepared at a concentration of 10 µ*M*, in 50 m*M* sodium phosphate pH 6.0, 100 m*M* NaCl buffer. Far-ultraviolet circular-dichroism spectra (260–200 nm) were recorded with five data accumulations in a quartz cell with a path length of 0.1 cm. The resultant spectra were visualized using *Graphpad Prism*.

2.6. Thermal stability analysis *via* RT-PCR Thermofluor

Thermal stability assays of MSMEG_5817 and variants were performed using a real-time detection instrument

(Corbett Rotor-Gene 300). Briefly, all MSMEG_5817 samples were prepared at a concentration of 30 µ*M* in 50 m*M* sodium phosphate pH 6.0, 300 m*M* NaCl buffer. The fluorescent dye SYPRO Orange (LifeTechnologies) was added to each sample at a 10× concentration to enable monitoring of the protein-unfolding process. Samples were heated from 30 to 90°C at a rate of 1°C min^{-1} and the changes in fluorescent intensity were recorded accordingly with an excitation wavelength of 530 nm and an emission wavelength of 660 nm.

2.7. Phospholipid binding of MSMEG_5817 by chemiluminescent enzyme-linked immunoassay (ELISA)

The phospholipid-binding capabilities of MSMEG_5817 and variants were determined by immunoassay. Selected phospholipids (Avanti), fatty acid and cholesterol (Sigma) stocks were created by dissolving 1 mg in chloroform and freeze-drying under a stream of nitrogen gas before resuspension in 1:1 chloroform:methanol at a final concentration of 100 µ*M*. Briefly, for direct plating assays lipids were diluted 1:10 in 10 m*M* phosphate-buffered saline (PBS) to yield a final concentration of 10 µ*M* in 100 µl and were plated on Nunc-Immuno MicroWell Maxisorp 96-well solid plates overnight at 293 K. Excess phospholipids were washed in PBS + 1%(*v*/*v*) Tween 20 before blocking in PBS + 1%(*v*/*v*) Tween 20 + 3%(*w*/*v*) bovine serum album (BSA) for 2 h and washing again. MSMEG_5817 and variants were titrated between 10 and 100 µ*M* and allowed to bind for 1 h before washing. The presence of bound MSMEG_5817 protein was detected by HRP anti-hexahistidine antibody (R&D Systems) at a 1:2000 dilution. SureBlue TMB Microwell Peroxidase Substrate (KPL) was used to detect the presence of antibodies, with the reaction stopped by 0.1 *M* HCl. Absorbance was detected by FLUOstar Omega (BMG Labtech) at a wavelength of 450 nm. All experiments were conducted in triplicate, with data visualized and statistical *t*-tests conducted using *GraphPad Prism*.

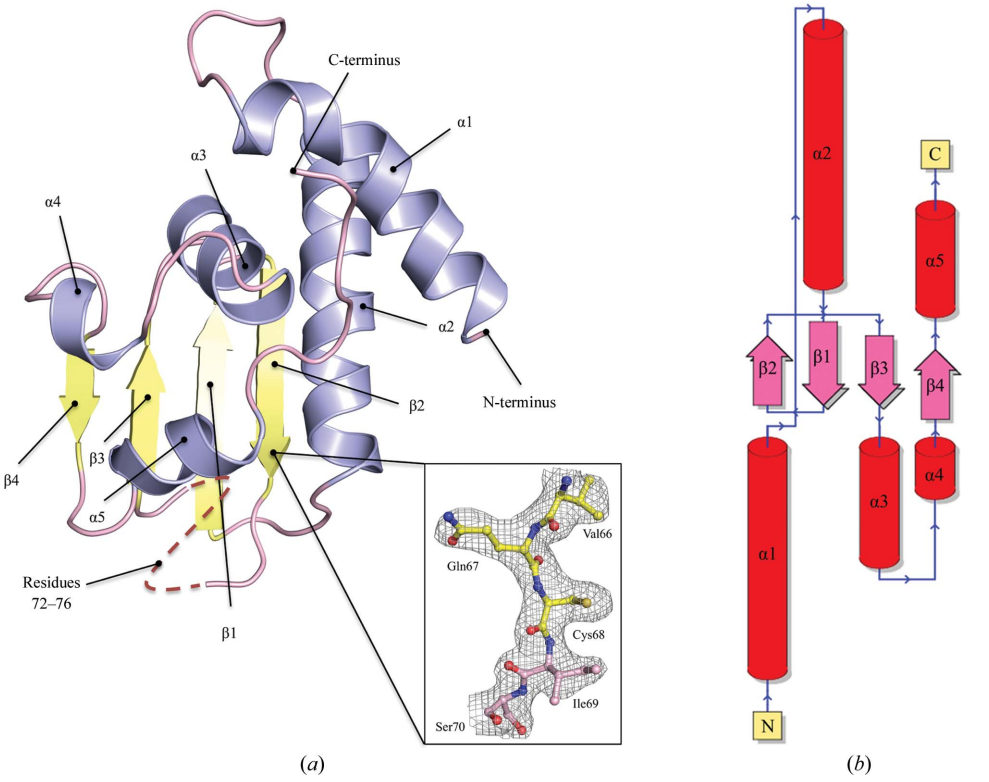

Figure 1
Overall crystal structure of MSMEG_5817. (*a*) Cartoon representation of the 2.4 Å resolution MSMEG_5817 crystal structure. The structure is colour-coded by secondary-structure element, with α-helices, β-strands and loops coloured blue, yellow and pink, respectively. The structure comprises a single antiparallel β-sheet consisting of four β-strands, β1 (residues 55–60), β2 (residues 63–68), β3 (residues 81–84) and β4 (residues 108–110), with a classical β-bulge occurring between the β2 strand and the β3 strand. Two β-hairpin structures are present, with the first spanning between the β3 strand and the β4 strand and the second occurring through the β1 strand sandwiched between the β3 strand and the β4 strand. The latter contributes to a ψ-loop spanning residues 64–85. Residues 72–76 missing from the crystal structure contribute to the formation of the ψ-loop. The β-sheet is stacked by a layer of five α-helices: α1 (residues 11–24), α2 (residues 33–50), α3 (residues 87–95), α4 (residues 87–95) and α5 (residues 100–104). A highly hydrophobic cavity is formed between the β-sheet and the α3, α4 and α5 helices. The quality of the final density is visualized (inset) around residues 66–70, centred on Ile69, which are visualized in ball-and-stick representation and colour-coded based on their secondary structure. The σ_A-weighted $(2m|F_o| - D|F_c|)$ electron-density maps are contoured at 1.0σ. (*b*) Topology diagram of the MSMEG_5817 structure colour-coded by secondary-structure element, with α-helices, β-strands and loops coloured red, pink and blue, respectively.

3. Results and discussion

3.1. Structural determination of MSMEG_5817

To gain insight into the function of MSMEG_5817 and its orthologues, the crystal structure of the MSMEG_5817 protein was determined. As no significant structural homologue was available, the crystal structure of MSMEG_5817 was determined *via* multiple-wavelength anomalous dispersion (MAD) using data collected at three experimentally determined wavelengths. Initial experimental phases were obtained with *AutoSol* as part of the *PHENIX* program package (Terwilliger *et al.*, 2009), using experimentally determined f' and f'' values. The highly symmetrical primitive tetragonal $P4_32_12$ space group, with unit-cell parameters a = 50.5, b = 50.5, c = 205.57 Å, could accommodate two polypeptide molecules in the asymmetric unit (molecules *A* and *B*), corresponding to a Matthews coefficient of 2.2 Å^3 Da^{-1} at 15 kDa per molecule (Matthews, 1968).

The structure of MSMEG_5817 was determined to a resolution of

2.40 Å, with final R_{work} and R_{free} statistics of 18.92 and 22.52%, respectively. The refinement statistics are summarized in Table 1. The construct utilized in crystallization included residues 1–128 of MSMEG_5817 with ten vector-derived residues and an N-terminal hexahistidine purification tag. Of this construct, residues 9–128 in molecule *A* and residues 8–128 in molecule *B* were modelled from the electron density. Electron density is missing for one segment, consisting of residues 72–76 of molecule *A* and 72–77 of molecule *B*, which could not be traced in the electron-density map and therefore was not included in the final model. Molecules *A* and *B* can be superimposed with a core root-mean-square deviation (r.m.s.d.) value of 0.40 Å. There was no evidence of higher order oligomerization within the crystal lattice, which coincides with the purification of MSMEG_5817 as a monomer by size-exclusion chromatography (Shahine *et al.*, 2013). The final model of MSMEG_5817 consisted of 230 amino-acid residues and 71 water molecules. A Ramachandran plot analysis of the final model using *MolProbity* (Chen *et al.*, 2010) showed that 98.20% of the residues are in the most favoured regions, while 0.90% of the residues are in the allowed region and 0.90% of the residues are in the outlier region. The outlier residues are Ile69 in both molecules *A* and *B*.

3.2. Crystal structure of MSMEG_5817

The 2.40 Å resolution MSMEG_5817 crystal structure (Fig. 1*a*) reveals a single β-sheet comprised of four antiparallel β-strands surrounded by five α-helices, adopting an $\alpha1$–$\alpha2$–$\beta1$–$\beta2$–$\beta3$–$\alpha3$–$\alpha4$–$\beta4$–$\alpha5$ arrangement (Fig. 1*b*). The overall tertiary structure of MSMEG_5817 exhibits a Rossmann-like fold α/β two-layer sandwich (Rao & Rossmann, 1973; García *et al.*, 2000; Sillitoe *et al.*, 2013). A tightly packed cavity exists at the interface between the α-helical layer and the β-sheet, which may provide an interface for ligand binding. The four antiparallel β-strands form the first layer of the protein structure, arranged in a −1 2 1 topology. A single antiparallel classic β-bulge occurs between Val60 on the $\beta3$ strand and Phe63 and Val64 on the $\beta2$ strand across a β-turn. This type IV hydrogen-bonded β-turn is comprised of residues Val60-Pro61-Pro62-Phe63-Val64, with an i to i + 3 C^{α} distance of 5.8 Å. A large β-hairpin structure is formed between $\beta3$ and $\beta4$ with a hairpin class of 19:21, which includes $\alpha3$ and $\alpha4$ that form part of the structural architecture over the β-sheet. A secondary β-hairpin structure is present in the form of a ψ-loop, in which the $\beta1$ strand is sandwiched in between the $\beta2$ and $\beta3$ strands, forming hydrogen bonds between the two strands. The ψ-loop, which is generally considered to be rare in protein structures (Hutchinson & Thornton, 1996), results in a 12-residue loop including an extended conserved residue patch spanning residues 74–83. The architecture of the ψ-loop is stabilized by hydrogen-bond formation between the C^{α} atom of Ile69 with Ser55 on the $\beta1$ strand and Ala51 at the C-terminal end of $\alpha2$, with bond distances of 2.7 and 2.9 Å, respectively. Owing to the rarity of ψ-loops within protein structures, the presence of Ile69 as a Ramachandran outlier is explained. The Cys68 residue is situation at the C-terminal end of $\beta2$ before the ψ-loop and faces into the core of the protein.

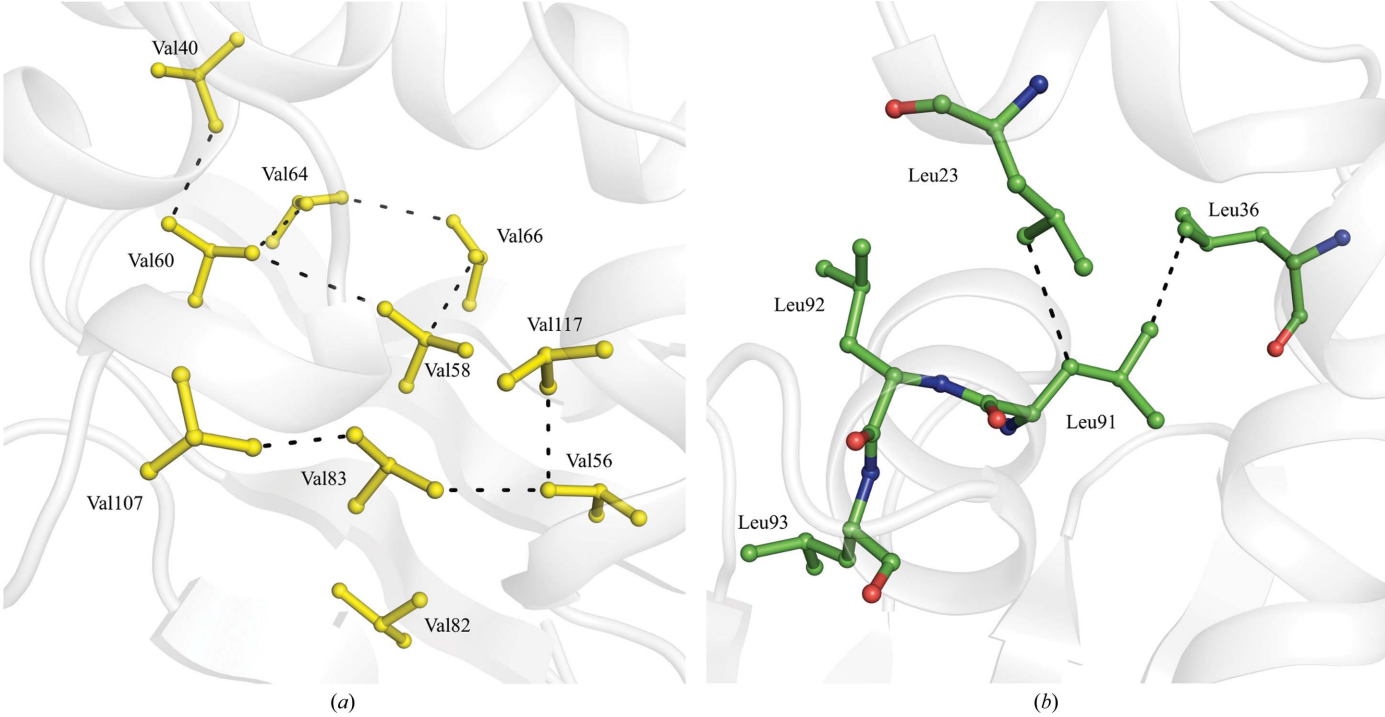

Figure 2
Hydrophobic residue content. (*a*) Stick representation of valine residue side chains comprising the hydrophobic core of MSMEG_5817 (yellow). Black dashed lines represent cross-valine hydrophobic interactions that stabilize the hydrophobic cavity of the protein, with bond distances ranging between 3.7 and 4.3 Å. (*b*) Coordination of the conserved Leu91 residue positioned on the $\alpha3$ helix. Hydrophobic interactions between Leu91 and Leu23 on the $\alpha1$ helix and Leu36 on the $\alpha2$ helix are denoted by black dashed lines, with bond distances of 3.7 and 3.6 Å, respectively.

MSMEG_5817 is comprised of a high percentage of hydrophobic residues, predominantly alanine, valine and leucine, which make up 39.1% of its amino-acid sequence. The valine residues, which contribute 11.7% of the amino-acid sequence, are predominantly positioned on the $\beta1$–$\beta3$ strands, with side chains facing the interior of the structure (Fig. 2*a*). The valine residues act to contribute to the overall hydrophobicity of the protein core. Interestingly enough, Val82, located on the $\beta3$ strand, is the only valine residue with a solvent-exposed side chain within the structure (Fig. 2*a*). The 11 leucine residues, comprising 8.6% of the total amino-acid sequence, contribute to stabilizing the α-helical architecture surrounding the β-sheet (Fig. 2*b*). These helix–helix interactions are primarily driven by a stretch of three leucine residues, Leu89, Leu90 and Leu91, positioned in the $\alpha3$ helix. Leu91 acts to stabilize helix–helix interactions between the $\alpha3$ helix and the $\alpha1$ and $\alpha2$ helices through hydrophobic interactions with neighbouring leucine residues (Fig. 2*b*). Additional hydrophobic interactions are made between Leu91 and Leu23 in the $\alpha1$ helix and Leu36 in the $\alpha2$ helix, with bond lengths of 3.7 and 3.6 Å, respectively (Fig. 2*b*).

Despite the highly hydrophobic nature of the MSMEG_5817 structure, the recombinant protein was successfully purified as a soluble protein. This is aided by the

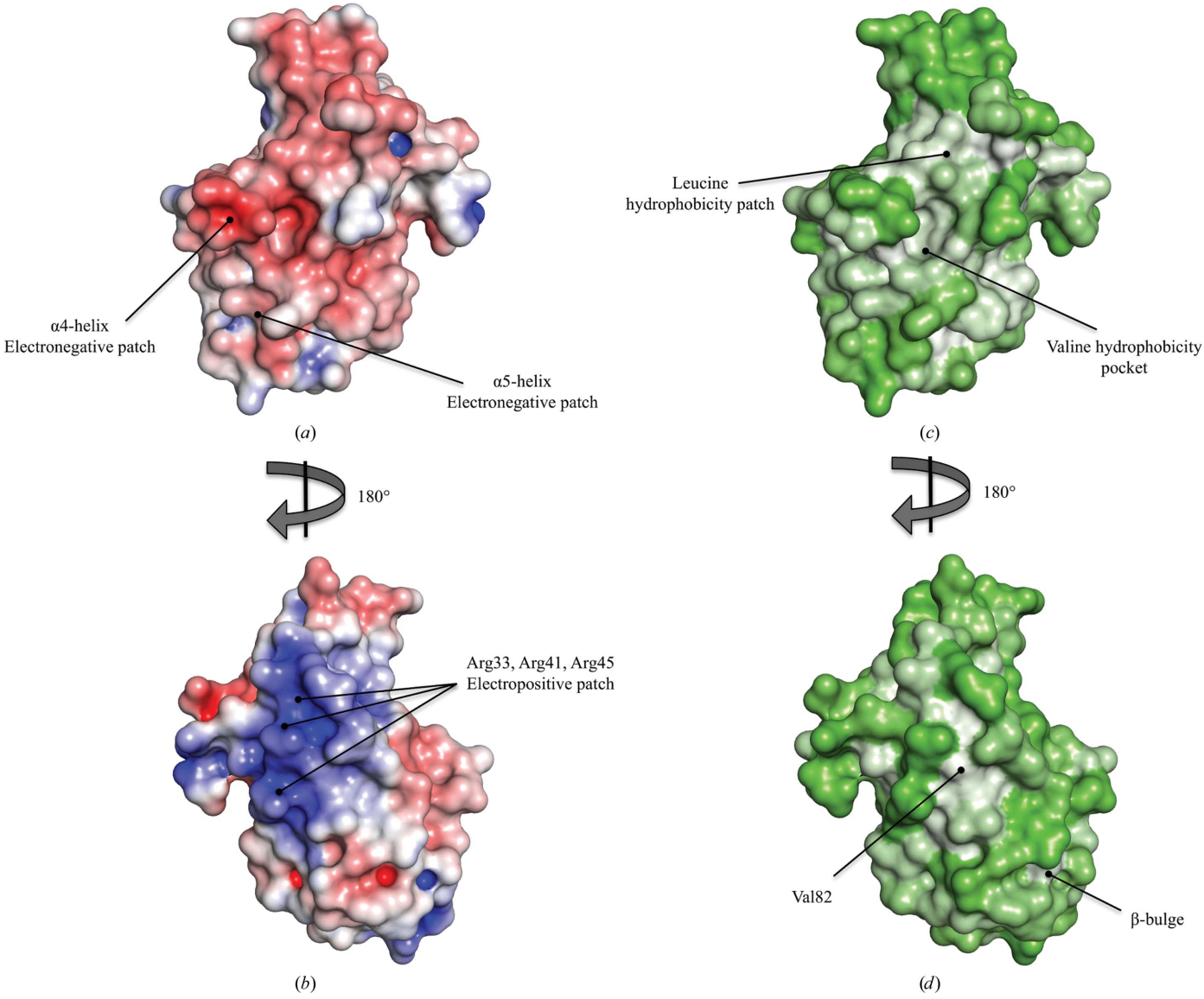

Figure 3
Surface electrostatic and hydrophobic properties of the MSMEG_5817 crystal structure. (*a*) Front and (*b*) rear views of the solvent-accessible surface representation coloured by electrostatic potential as calculated by *APBS* (Baker *et al.*, 2001). The potential contours are shown on a scale from $+5.0k_BT\,e^{-1}$ (blue) to $-5.0k_BT\,e^{-1}$ (red); white indicates no change. (*a*) The positions of two electronegative patches involving residues Asp99 and Asp102 on the $\alpha4$ helix, as well as residues Glu116 and His119 on the $\alpha5$ helix. (*b*) The position of an extended electropositive patch involving residues Arg33, Arg41 and Arg45. (*c*) Front and (*d*) rear views of a surface representation coloured by hydrophobicity according to the normalized consensus hydrophobicity scale (Eisenberg *et al.*, 1984). Colour representation with a colour gradient of most hydrophobic to least hydrophobic ranging from white to green, respectively. Indicated are the positions of (*c*) the leucine hydrophobic patch and valine hydrophobic pocket and (*d*) the solvent-exposed Val82 residue and the classical β-bulge.

high percentage of solvent-exposed arginine residues, which make up 9.4% of the amino-acid sequence and are positioned throughout the surface of the protein. The localization of three arginine residues, Arg33, Arg41 and Arg45, situated in the α2 helix contributes to a highly electrostatically positive region facing the rear of the structure (Fig. 3*b*).

In addition, there is an extended negative electrostatic patch spanning between the α-helical layer across the β-sheet, comprised of α3, α4 and α5. This primarily involves Asp99 and Asp102 in the α4 helix and Glu116 and His119 in the α5 helix (Fig. 3*a*). In each case, the residues are poorly conserved throughout the orthologues analysed, save for Glu116, which is substituted by an aspartic acid in *M. leprae*. In each case, the side chains of these negatively charged residues face away from the core of the structure and, if this is indeed a ligand-binding cavity, would not be involved in direct ligand contacts.

A differing trend was observed upon comparing electrostatic potential with hydrophobic content. The highest localization of hydrophobicity is situated within the cavity sandwiched between the α-helical layer and the β-sheet (Fig. 3*c*). This region of hydrophobicity primarily involves the valine residues spanning the β-sheet, combined with the leucine residues linking the α3 helix to the α1 and α2 helices, as described previously. The rear of the protein structure reveals a small hydrophobic patch indicative of the anti-parallel β-bulge spanning between the β2 and β3 strands (Fig. 3*d*). This localization of hydrophobicity at the cavity interface, coupled with the lack of electrostatically charged residues facing inwards, may suggest that if this is indeed a binding cavity then MSMEG_5817 may be involved in facilitating the binding of nonpolar ligands.

3.3. Comparison of MSMEG_5817 to structural homologues

Owing to the lack of conserved domains, and the absence of characterized proteins with amino-acid sequence identity, the function of MSMEG_5817 is presently unknown. In order to gain functional insight based on the crystal structure of MSMEG_5817, a *DALI* search of the Protein Data Bank was conducted (Holm & Rosenström, 2010). The closest structural homologues identified from the search results were a number of sterol carrier proteins (SCPs), the function of which is to

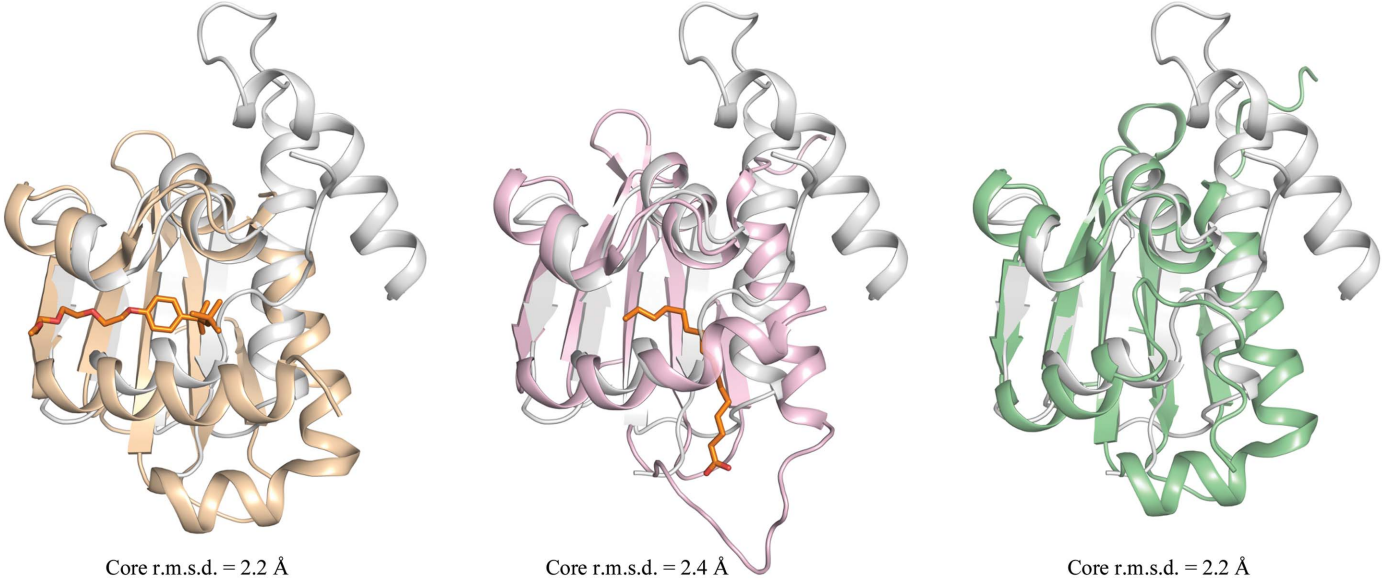

Figure 4
Cartoon structure superposition of MSMEG_5817 and sterol carrier proteins (SCPs). MSMEG_5817 (grey) is overlayed with (*a*) *H. sapiens* SCP (hSCP; wheat), (*b*) *A. aegypti* SCP (AaSCP; magenta) and (*c*) *O. cuniculus* SCP (OcSCP; green), with core r.m.s.d. values of 2.2, 2.4 and 2.2 Å, respectively. The presence of bound Triton X-100 (*a*) and palmitic acid (*b*) in the SCP structures is shown in orange stick representation. (*d*) Sequence alignment between MSMEG_5817 and identified SCPs. The alignment was prepared with *ClustalW*2 and visualized using *ESPript* v.2.2. Strict sequence-identical residues are denoted with a red background, while similar residues are visualized in red text with a white background. Sequence similarities in groups are denoted by blue boxes. Gaps are represented by dots.

bind and transport sterols (Avdulov *et al.*, 1999) and lipids (Schroeder *et al.*, 1995) across cellular membranes (Kernstock & Girotti, 2007). Three SCPs were identified from *H. sapiens* (PDB entry 1ikt; Haapalainen *et al.*, 2001; Szyperski *et al.*, 1993), *A. aegypti* (PDB entry 1pz4; Dyer *et al.*, 2003) and *O. cuniculus* (PDB entry 1c44; Choinowski *et al.*, 2000), with core r.m.s.d. values of 2.2, 2.4 and 2.2 Å, respectively (Figs. 4*a*, 4*b* and 4*c*). The SCPs exhibit a Rossman-like fold α/β two-layer sandwich, structurally similar to MSMEG_5817. The crystal structures of the *H. sapiens* SCP (hSCP) and *A. aegypti* SCP (AaSCP) were solved in the presence of a lipid bound within the binding groove: Triton X-100 (Haapalainen *et al.*, 2001) and palmitic acid (Dyer *et al.*, 2003), respectively. The sequence identities between MSMEG_5817 and the SCPs are very low, with AaSCP having the highest percentage identity at 15.7% (Fig. 4*d*). Despite this, the secondary-structure elements overlay tightly between MSMEG_5817 and the SCPs, with the overall protein architecture being predominantly retained.

While the similarities between the tertiary structures of MSMEG_5817 and the SCPs are apparent, there are some subtle differences in their secondary-structure compositions. The SCPs exhibit a larger binding cleft arranged in a −1 −1 3X −1 topology comprised of a five-stranded antiparallel β-sheet with a slight curvature. In comparison, the MSMEG_5817 β-sheet is arranged in a −1 2 1 topology comprised of a flat four-stranded antiparallel β-sheet (Figs. 4*a*, 4*b* and 4*c*) The larger β-sheets present in the SCPs accommodate a larger lipid-binding interface. In addition, the MSMEG_5817 structure contains a N-terminal helix that is not conserved in other SCPs (Figs. 4*a*, 4*b* and 4*c*) that results in r.m.s.d.s that are larger than normal between conserved proteins. However, the N-terminal helix is conserved across other Corynebacterineae orthologues (Fig. 5), suggesting a conserved function within this family.

The previously determined crystal structures of hSCP and AaSCP both contain a bound lipid within their respective binding cavities. Upon comparison, the residues involved in contacts between the SCPs and the respective lipid differ between the two proteins. In the case of the AaSCP, the basal α-helix unravels to accommodate binding against the polar carboxylic acid head group. The residues binding the lipid tail are predominantly conserved with respect to hSCP, principally comprising large aliphatic residues situated on the α-helices surrounding the β-sheet. This supports the function of the SCPs as nonspecific lipid-binding proteins exhibiting a large binding cavity accommodating a broad range of lipids both in functionality and size.

In comparison, the positions of the lipid-binding residues are poorly conserved between MSMEG_5817 and the SCPs.

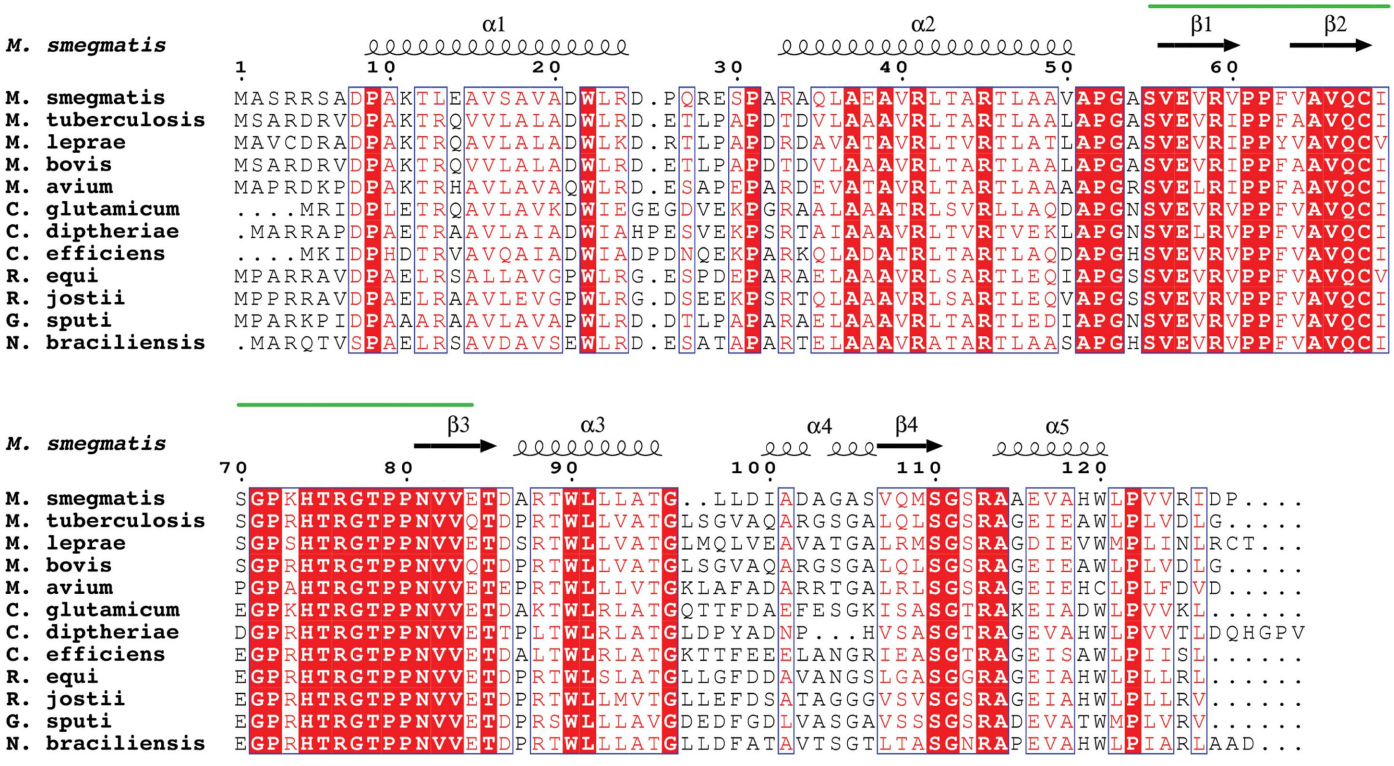

Figure 5
Amino-acid sequence alignment of members of the MSMEG_5817 protein family. This list includes the following: *Mycobacterium smegmatis* (YP_890045.1), *M. tuberculosis* (NP_215322.1), *M. leprae* (NP_302448.1), *M. bovis* (NP_854488.1), *M. avium* (WP_003875797.1), *Corynebacterium glutamicum* (NP_601783.1), *C. diphtheriae* (YP_005138847.1), *C. efficiens* (WP_006769165), *Rhodococcus equi* (YP_004005639.1), *R. jostii* (YP_704781.1), *Gordonia sputi* (WP_005204186.1) and *Nocardia brasiliensis* (YP_006805579.1). The alignment was prepared with *ClustalW*2 and was visualized using *ESPript* v.2.2. The secondary-structure elements correspond to the structure of MSMEG_5817. Strict sequence-identical residues are denoted with a red background, while similar residues are visualized in red text with a white background. Sequence similarities in groups are denoted by blue boxes. The green line indicates the position of the ψ-loop.

The only conserved residue present within the binding groove of MSMEG_5817, assuming a similar binding mechanism to the SCPs, is Leu121. The Leu121 side chain faces the binding groove situated on the α5 helix and may be involved in direct contact with a bound ligand.

3.4. Sequence similarity of MSMEG_5817 to other Corynebacterineae orthologues

In order to gain insight into the functional role of MSMEG_5817, we compared the sequence of MSMEG_5817 with those of its orthologues from the closely related mycobacterial species *M. tuberculosis* and *M. leprae*, as well as those from other representatives of the suborder Corynebacterineae (Fig. 5). The sequences are highly conserved, with a minimum of 53% pairwise sequence conservation. A *BLASTp* search was unable to detect any putative conserved domains. The longest stretch of conserved residues is located from residues 74 to 83. A high degree of variation is observed in residues 97–105 that represent the α4 helix. The highly conserved valine residues Val56, Val66, Val82 and Val83 in MSMEG_5817 maintain the potential hydrophobic binding pocket. The conservation of these valine residues across all orthologues, as well as the presence of additional hydrophobic residues replacing the other valine positions in MSMEG_5817, suggests that all orthologues may bind nonpolar ligands.

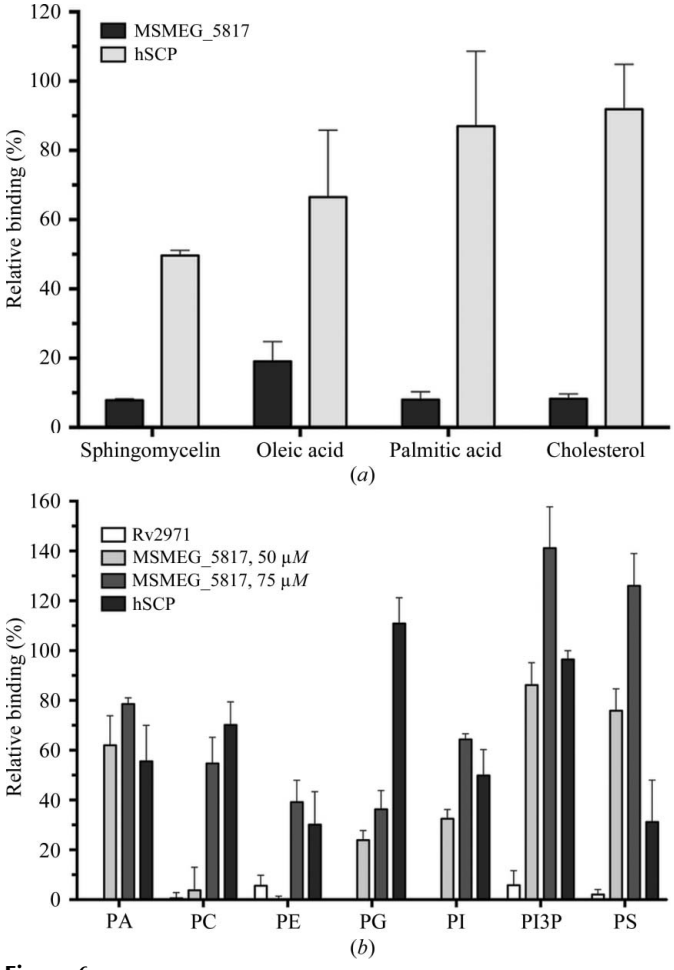

Figure 6
Comparative lipid-binding capabilities of MSMEG_5817 and *H. sapiens* SCP (hSCP) as determined by ELISA. (*a*) The lipid-binding capabilities of MSMEG_5817 were assessed against lipids previously determined to be bound by hSCP with high affinity (Schroeder *et al.*, 2007). Data were normalized against the results of hSCP binding to cholesterol as a positive control. Limited to no binding was observed for MSMEG_5817, indicating a difference in lipid specificity compared with hSCP. (*b*) A range of biologically relevant phospholipids were selected to probe the lipid-binding capabilities of MSMEG_5817. Data were normalized against hSCP as a positive control and the *M. tuberculosis* aldo–keto reductase Rv2971 as a negative control. The results indicate binding to a number of phospholipids in a dose-dependent response manner. The ELISA experiment was conducted as described in §2.7. Data are represented as relative percentage binding against the controls. Results were visualized using *GraphPad Prism*, with error bars representing the mean percentage ± SEM of triplicate experiments.

3.5. Phospholipid-binding capabilities of MSMEG_5817 by ELISA

Owing to the conservation of the hydrophobic binding pocket and the structural similarities of MSMEG_5817 to the SCP class of proteins, we hypothesized that MSMEG_5817 and its orthologues may be involved in the binding of biologically relevant lipids. SCPs have previously been characterized to bind to cholesterol, a range of phospholipids and saturated and unsaturated fatty acids of varying lengths (van Amerongen *et al.*, 1989; Wirtz & Gadella, 1990; Schroeder *et al.*, 1995; Avdulov *et al.*, 1999; Kernstock & Girotti, 2007).

The lipid-binding capabilities of MSMEG_5817 were probed by immobilized lipid ELISA (enzyme-linked immunoassay). hSCP has previously been observed to bind tightly to cholesterol, with a reported affinity of 4.0 n*M* (Stolowich *et al.*, 1999). The lipid-binding capabilities of MSMEG_5817 was initially probed against cholesterol, as well as a number of highly prevalent biological fatty acids, including palmitic acid, which was present in the crystal structure of AaSCP (Dyer *et al.*, 2003). Recombinant hSCP was found to bind oleic acid, palmitic acid and cholesterol, as observed previously (Fig. 6*a*; Dyer *et al.*, 2003; Kernstock & Girotti, 2007). In comparison, MSMEG_5817 was found to not bind any of the initially tested lipids. In addition, no growth defects were observed upon growing MSMEG_5817 mutant on minimal medium where oleic acid, palmitic acid and cholesterol were the only carbon sources (data not shown). This observation is in keeping with a recent study looking at genes that are essential for cholesterol catabolism in *M. tuberculosis*, with Rv0807, the *M. tuberculosis* homologue of MSMEG_5817, being shown not to be required for growth that is dependent on cholesterol as sole carbon source (Griffin *et al.*, 2011).

Further experiments exploring the potential lipid-binding capabilities of MSMEG_5817 were performed using commercially available lipid strips (Echelon). It was shown that MSMEG_5817 bound to a range of biologically relevant phospholipids; however, no binding was observed to cholesterol (data not shown). It has previously been reported that hSCP is capable of binding phosphatidylcholine (PC) as well as a range of phosphatidylinositols with varying specificities (Gadella & Wirtz, 1991; Avdulov *et al.*, 1999; Zhou *et al.*, 2004).

Verification of phospholipid binding by MSMEG_5817 was also performed using immobilized lipid ELISA. MSMEG_5817 was found to bind a number of phospholipids with a dose-dependent response, with the strongest binding observed to phosphatidic acid (PA), phosphatidylserine (PS) and phosphatidylinositol 3-phosphate (PI3P) (Fig. 6*b*). Weak

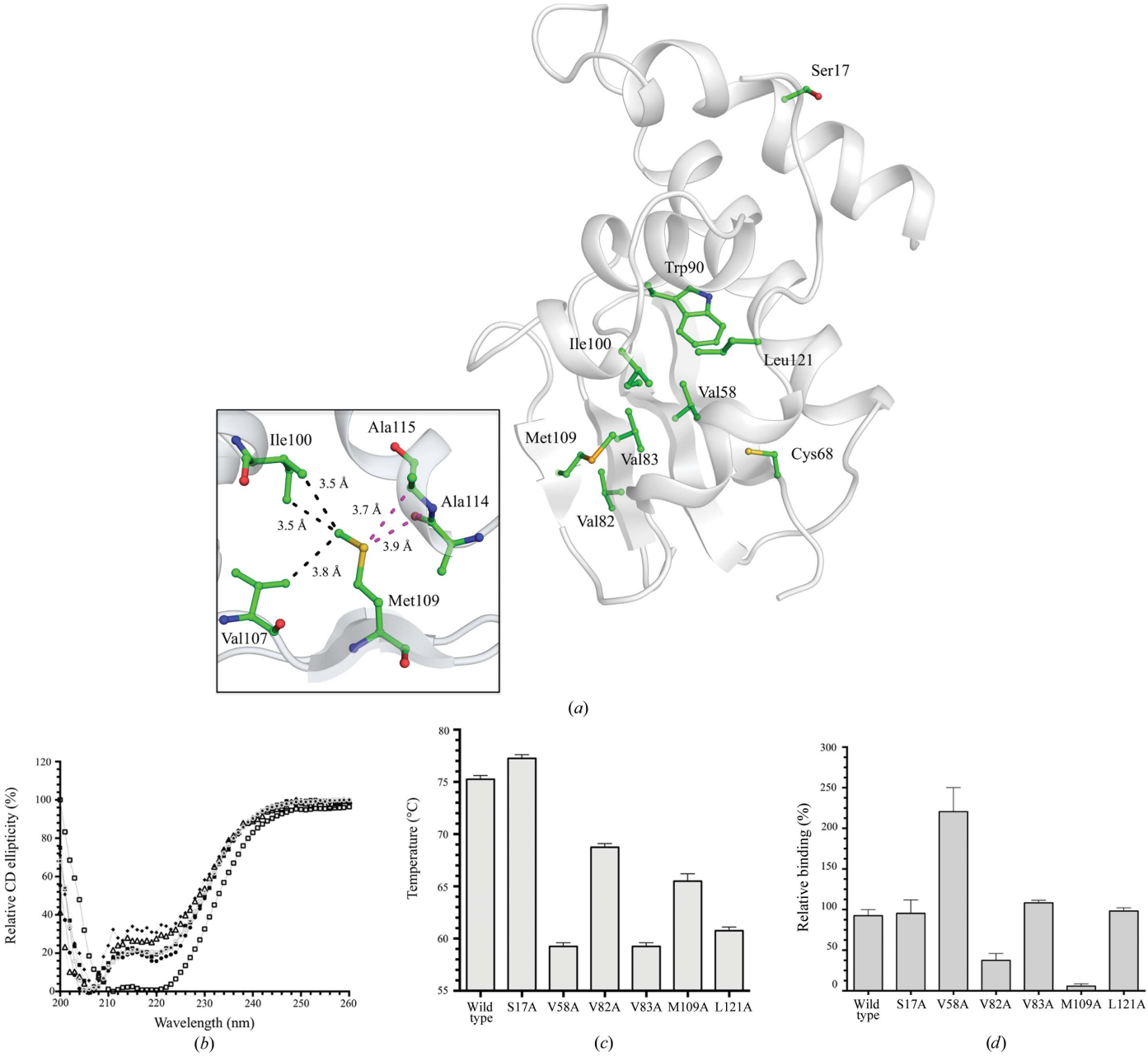

Figure 7
Mutant analysis of MSMEG_5817 by alanine-scanning mutagenesis. (*a*) Stick representation of the positions of the residues mutated to alanine. Residues for mutagenesis studies were selected based on sequence identity between MSMEG_5817 orthologues and conservation with SCP ligand-binding residues. Inset: stick representation of amino-acid residue contacts with Met109 positioned in the β4 strand. The C^{ε} atom of Met109 forms two hydrophobic interactions with Ile100 with bond distances of 3.5 Å, as well as a single hydrophobic interaction with Val107 with a bond distance of 3.8 Å, represented by the black dashed line. In addition, the S^{δ} atom of Met109 forms van der Waals interactions with the main chains of Ala114 and Ala115 in the α5 helix, with bond distances of 3.7 and 3.9 Å, represented by magenta dashed lines. (*b*) Circular-dichroism spectra of MSMEG_5817 and mutants. Far-UV CD spectra were recorded of wild-type MSMEG_5817 (filled circles) and the S17A (filled squares), V58A (triangles), V82A (filled diamonds), V83A (circles), M109A (squares) and L121A (diamonds) mutants between wavelengths of 200 and 260 nm. Spectra were visualized using *GraphPad Prism*. (*c*) Thermal stability assay of MSMEG_5817 mutants compared with the wild type, as measured by Thermofluor. The greatest reduction in thermal stability was observed for the V58A and V63A mutants. Results are represented as the mean temperature ± SEM in triplicate experiments. (*d*) Lipid-binding capabilities of MSMEG_5817 mutants by ELISA. Binding capabilities to PI3P of MSMEG_5817 mutants compared with wild-type MSMEG_5817 as measured by ELISA as described in §2.7. The greatest reduction in binding capabilities was observed for the V82A and M109A mutants. Results were normalized against the wild type (100%) and a PBS blank (0%). Results are represented as mean percentages ± SEM in triplicate experiments. All graphs were visualized using *GraphPad Prism*.

binding was observed with phosphatidylglycerol (PG) and phosphatidylinositol (PI) and minimal or no binding was observed with phosphatidylcholine (PC) and phosphatidylethanolamine (PE). Stronger binding was observed with PI3P compared with PI, indicating substrate specificity for the additional phosphate on the PI3P head group. No binding was observed for the saturated and unsaturated fatty acids tested, which contribute to phospholipid fatty-acid tails, indicating substrate specificity for the larger phospholipid head groups of PI3P and PS. In contrast, recombinant hSCP was found to bind to a range of the phospholipids, with the highest binding to PG, while the aldo–keto reductase (Rv2971) from *M. tuberculosis* showed no binding to phospholipids (Fig. 6*b*). These results indicate a difference in substrate specifies between MSMEG_5817 and hSCP, with MSMEG_5817 binding preferentially to phospholipids with predominantly electronegative head groups.

3.6. Probing the putative MSMEG_5817 binding pocket

In the absence of a crystal structure of MSMEG_5817 with a bound ligand, the potential ligand-binding residues were probed by utilizing alanine-scanning mutagenesis. The lipid ELISA indicated differences in binding mechanisms and substrate specificities between MSMEG_5817 and hSCP. The mutated residues were selected based on sequence identity within the MSMEG_5817 orthologue family and the SCP family, along with positional conservation of residues involved in ligand contacts within the SCPs (Fig. 7*a*).

The mutants were expressed and purified as soluble protein, with the exception of W90A and C68A, which were expressed as insoluble protein, indicating essentiality in maintaining the structural integrity of MSMEG_5817. No expression of I100A was detected by Western blotting. Each of the mutant proteins was purified by size-exclusion chromatography, revealing a strong peak relating to monomeric protein. In the case of M109A and L121A, a high degree of soluble aggregate was present and was successfully removed by size-exclusion chromatography.

In order to ascertain the effects of the mutants on the overall secondary-structure composition of MSMEG_5817, circular dichroism (CD) in the far-UV range was conducted. The CD spectra measured between wavelengths of 260 and 200 nm indicated conservation of α-helical content between wild-type MSMEG_5817 and each of the purified mutants, with the exception of M109A (Fig. 7*b*). The changes within the M109A mutant may indicate destabilization of the β4 sheet, resulting in a stronger α-helical representative trace, with a prominent dip at 220 nm observed in the CD spectra. Met109 forms hydrophobic interactions with Ile100 present on the α4 helix, as well as van der Waals interactions between the S atom and the C^{α} atom of residue Ala115 in the α5 helix (inset in Fig. 7*a*), which is conserved amongst the orthologues (Fig. 5). This could be an indication as to the lack of expression of I100A, as well as the high degree of soluble aggregate observed for M109A. However, Met109 is poorly conserved between orthologues and is only present in *M. leprae* (Fig. 5). Within mycobacterial species a leucine residue is predominantly present at position 109, with smaller residues present amongst the broader Corynebacteria orthologues.

The structural integrity of each of the mutants was analysed by RT-PCR Thermofluor using the fluorescent dye SYPRO Orange to determine thermal stability (T_m) values (Ericsson *et al.*, 2006). A significant drop in the T_m value was observed for each of the mutants, with the exception of S17A, which acted as the control mutant. The greatest decrease in the T_m value was observed for the internal valine mutants V58A and V83A, with a decrease of $\sim$15°C from $\sim$70°C for wild-type MSMEG_5817 to $\sim$54°C for the mutants (Fig. 7*c*). These results indicate that the internal valine residues tested are vital for maintaining the hydrophobic integrity of the protein core, stabilizing the overall protein fold (Fig. 4*a*).

To gain insight into the effects of the mutant binding capabilities, an ELISA was conducted using PI3P as the strongest binding lipid. The results of the ELISA indicated no statistical differences in binding capabilities between the wild-type and the internal V83A ($p < 0.19$) and L121A ($p < 0.57$) mutants (Fig. 7*d*). The V58A ($p < 0.05$) mutation resulted in a significant increase in lipid binding, suggesting that this mutation affects the environment in the potential binding pocket, thus increasing the affinity for the phospholipid PI3P.

The binding of PI3P by wild-type MSMEG_5817 and the control S17A mutant was the same. There was a complete loss of PI3P binding by the M109A mutant ($p < 0.001$), which may be owing to alteration of the overall secondary-structure composition of MSMEG_5817 or to direct residue binding. Interestingly enough, there was a 60% reduction in binding of PI3P by the solvent-exposed V82A mutant ($p < 0.05$), which may indicate differences from the binding mechanism shown by hSCP, which has a preference for cholesterol over phospholipids.

4. Conclusions

Vital to the success of mycobacterial pathogenesis is the ability of mycobacteria to parasitize the host macrophage phagosome, creating a hospitable environment for survival and replication. The key to controlling mycobacterial infection is to identify the novel biochemical pathways utilized for this process for the purposes of rational drug design. The recently identified gene product of *MSMEG_5817* from *M. smegmatis* has been found to be vital for mycobacterial survival within host macrophages (Pelosi *et al.*, 2012). The *MSMEG_5817* gene encodes a protein of 128 amino-acid residues of unknown function that is highly conserved within the Corynebacterineae suborder. In the present study, we have determined the crystal structure of MSMEG_5817. The crystal structure revealed a Rossmann-like fold α/β two-layer sandwich forming a highly hydrophobic interface cavity and with high structural homology to the SCP family. Coupled together, this supports the hypothesis that MSMEG_5817 may be involved in the interaction of apolar ligands through its hydrophobic cavity.

Investigation of the lipid-binding capabilities of MSMEG_5817 revealed binding to a number of phospholipids, with the highest specificities for those with predominantly negatively charged head groups, namely PA, PS and PI3P. This specificity was in direct contrast to the lipid-binding capabilities of hSCP. Alanine-scanning mutagenesis of the hydrophobic cavity of MSMEG_5817 demonstrated that the conserved Val82 residue plays an important role in ligand binding, with the ligand-binding interface potentially located at the base of the structure, centred on Val82 and the ψ-loop, in a similar position to the AaSCP binding site. The result for the M109A mutant was surprising owing to changes in secondary structure and the complete loss of ligand-binding capability despite its poor conservation within the MSMEG_5817 family. This suggests that the potential interface has delicate architecture.

While binding to a number of phospholipids was observed, further studies are required to identify the natural ligand and identify the true function of the MSMEG_5817 family. Owing to the binding of MSMEG_5817 to PI3P, we speculate that it may have a potential role in processing host-cell PI3P, thereby contributing to the block in phagolysosomal maturation; however, confirmatory studies are required. Nevertheless, the structural and functional characterization presented in this report has paved the way for the further characterization of this family of proteins and their functional role within the Corynebacterineae suborder of bacteria.

We thank the staff of the Australian Synchrotron and Monash Macromolecular Crystallization Facility for assistance with crystallization and X-ray data collection. This work was supported by the Australian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics and the National Health and Medical Research Council of Australia. JR is an NHMRC Australia Fellow and TB is a Pfizer Australian Research Fellow.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Amerongen, A. van, Demel, R. A., Westerman, J. & Wirtz, K. W. A. (1989). *Biochim. Biophys. Acta*, **1004**, 36–43.

Anes, E., Peyron, P., Staali, L., Jordao, L., Gutierrez, M. G., Kress, H., Hagedorn, M., Maridonneau-Parini, I., Skinner, M. A., Wildeman, A. G., Kalamidas, S. A., Kuehnel, M. & Griffiths, G. (2006). *Cell. Microbiol.* **8**, 939–960.

Armstrong, J. A. & Hart, P. D. (1971). *J. Exp. Med.* **134**, 713–740.

Avdulov, N. A., Chochina, S. V., Igbavboa, U., Warden, C. S., Schroeder, F. & Wood, W. G. (1999). *Biochim. Biophys. Acta*, **1437**, 37–45.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Billman-Jacobe, H., McConville, M. J., Haites, R. E., Kovacevic, S. & Coppel, R. L. (1999). *Mol. Microbiol.* **33**, 1244–1253.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P. S. A., Smart, O. S., Vonrhein, C. & Womack, T. O. (2011). *BUSTER*. Cambridge: Global Phasing Ltd.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Choinowski, T., Hauser, H. & Piontek, K. (2000). *Biochemistry*, **39**, 1897–1902.

Clemens, D. L., Lee, B.-Y. & Horwitz, M. A. (2000*a*). *Infect. Immun.* **68**, 2671–2684.

Clemens, D. L., Lee, B.-Y. & Horwitz, M. A. (2000*b*). *Infect. Immun.* **68**, 5154–5166.

Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., Davis, A., de Haro, S., Naylor, J., Lee, H.-H. & Vergne, I. (2006). *Cell. Microbiol.* **8**, 719–727.

Dyer, D. H., Lovell, S., Thoden, J. B., Holden, H. M., Rayment, I. & Lan, Q. (2003). *J. Biol. Chem.* **278**, 39085–39091.

Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 140–144.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.

Fratti, R. A., Backer, J. M., Gruenberg, J., Corvera, S. & Deretic, V. (2001). *J. Cell Biol.* **154**, 631–644.

Fratti, R. A., Chua, J., Vergne, I. & Deretic, V. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 5437–5442.

Gadella, T. W. J. Jr & Wirtz, K. W. A. (1991). *Biochim. Biophys. Acta*, **1070**, 237–245.

García, F. L., Szyperski, T., Dyer, J. H., Choinowski, T., Seedorf, U., Hauser, H. & Wüthrich, K. (2000). *J. Mol. Biol.* **295**, 595–603.

Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J. & Sassetti, C. M. (2011). *PLoS Pathog.* **7**, e1002251.

Haapalainen, A. M., Van Aalten, D. M. F., Meriläinen, G., Jalonen, J. E., Pirilä, P., Wierenga, R. K., Hiltunen, J. K. & Glumoff, T. (2001). *J. Mol. Biol.* **313**, 1127–1138.

Hmama, Z., Sendide, K., Talal, A., Garcia, R., Dobos, K. & Reiner, N. E. (2004). *J. Cell Sci.* **117**, 2131–2140.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.

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.

Kelley, V. A. & Schorey, J. S. (2003). *Mol. Biol. Cell*, **14**, 3366–3377.

Kernstock, R. M. & Girotti, A. W. (2007). *Anal. Biochem.* **365**, 111–121.

Kovacevic, S., Anderson, D., Morita, Y. S., Patterson, J., Haites, R., McMillan, B. N. I., Coppel, R., McConville, M. J. & Billman-Jacobe, H. (2006). *J. Biol. Chem.* **281**, 9011–9017.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Parish, T., Smith, D. A., Roberts, G., Betts, J. & Stoker, N. G. (2003). *Microbiology*, **149**, 1423–1435.

Patterson, J. H., McConville, M. J., Haites, R. E., Coppel, R. L. & Billman-Jacobe, H. (2000). *J. Biol. Chem.* **275**, 24900–24906.

Pelosi, A., Smith, D., Brammananth, R., Topolska, A., Billman-Jacobe, H., Nagley, P., Crellin, P. K. & Coppel, R. L. (2012). *PLoS One*, **7**, e31788.

Pereira-Leal, J. B. & Seabra, M. C. (2001). *J. Mol. Biol.* **313**, 889–901.

Pfeffer, S. R. (2005). *J. Biol. Chem.* **280**, 15485–15488.

Rao, S. T. & Rossmann, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.

Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. (2005). *Cell*, **122**, 735–749.

Russell, D. G. (2001). *Nature Rev. Mol. Cell Biol.* **2**, 569–577.

Russell, D. G. (2011). *Immunol. Rev.* **240**, 252–268.

Saleh, M. T. & Belisle, J. T. (2000). *J. Bacteriol.* **182**, 6850–6853.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.

Schroeder, F., Atshaves, B. P., McIntosh, A. L., Gallegos, A. M., Storey, S. M., Parr, R. D., Jefferson, J. R., Ball, J. M. & Kier, A. B. (2007). *Biochim. Biophys. Acta*, **1771**, 700–718.

Schroeder, F., Myers-Payne, S. C., Billheimer, J. T. & Wood, W. G. (1995). *Biochemistry*, **34**, 11919–11927.

Shahine, A. E., Chan, P. Y., Littler, D., Vivian, J., Brammananth, R., Crellin, P. K., Coppel, R. L., Rossjohn, J. & Beddoe, T. (2013). *Acta Cryst.* F**69**, 566–569.
Sillitoe, I., Cuff, A. L., Dessailly, B. H., Dawson, N. L., Furnham, N., Lee, D., Lees, J. G., Lewis, T. E., Studer, R. A., Rentzsch, R., Yeats, C., Thornton, J. M. & Orengo, C. A. (2013). *Nucleic Acids Res.* **41**, D490–D498.
Stolowich, N., Frolov, A., Petrescu, A. D., Scott, A. I., Billheimer, J. T. & Schroeder, F. (1999). *J. Biol. Chem.* **274**, 35425–35433.
Sturgill-Koszycki, S., Schlesinger, P. H., Chakraborty, P., Haddix, P. L., Collins, H. L., Fok, A. K., Allen, R. D., Gluck, S. L., Heuser, J. & Russell, D. G. (1994). *Science*, **263**, 678–681.
Szyperski, T., Scheek, S., Johansson, J., Assmann, G., Seedorf, U. & Wüthrich, K. (1993). *FEBS Lett.* **335**, 18–26.
Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* D**65**, 582–601.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Vergne, I., Chua, J., Lee, H.-H., Lucas, M., Belisle, J. & Deretic, V. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 4033–4038.
Vergne, I., Chua, J., Singh, S. B. & Deretic, V. (2004). *Annu. Rev. Cell Dev. Biol.* **20**, 367–394.
Via, L. E., Deretic, D., Ulmer, R. J., Hibler, N. S., Huber, L. A. & Deretic, V. (1997). *J. Biol. Chem.* **272**, 13326–13331.
Wirtz, K. W. A. & Gadella, T. W. J. Jr (1990). *Experientia*, **46**, 592–599.
World Health Organization (2012). *Global Tuberculosis Report*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/en/.
Zhou, M., Parr, R. D., Petrescu, A. D., Payne, H. R., Atshaves, B. P., Kier, A. B., Ball, J. M. & Schroeder, F. (2004). *Biochemistry*, **43**, 7288–7306.
Zhu, H. (2011). *DSSP & Stride Plugin for PyMOL*. Biotechnology Center (BIOTEC), TU Dresden, Germany.