The Mycobacterium tuberculosis rv3802c gene encodes an essential enzyme with thioesterase and phospholipase A activity. Overexpression of Rv3802 orthologs in Mycobacterium smegmatis and Corynebacterium glutamicum increases mycolate content and decreases glycerophospholipids. Although a role in modulating the lipid composition of the unique mycomembrane has been proposed, the true biological function of Rv3802 remains uncertain. In this study, we present the first M. tuberculosis Rv3802 X-ray crystal structure, solved to 1.7 Å resolution. On the basis of the binding of PEG molecules to Rv3802, we identified its lipid-binding site and the structural basis for phosphatidyl-based substrate binding and phospholipase A activity. We found that movement of the αβ-helix affords lipid binding and is required for catalytic turnover through covalent tethering. We gained insights into the mechanism of acyl hydrolysis by observing differing arrangements of PEG and water molecules within the active site. This study provides structural insights into biological function and facilitates future structure-based drug design toward Rv3802.

In 2015, 10.4 million incident cases of tuberculosis were reported with 1.4 million people succumbing to the disease (1). Lack of antibiotic adherence has led to an increase in multiple and totally drug-resistant strains of Mycobacterium tuberculosis, the causative agent of tuberculosis (2). As such, improved antibiotics and novel targets for antibiotic development are of high importance. Isoniazid and ethambutol are two first-line antibiotics and novel targets for antibiotic development (3). The M. tuberculosis cell wall remains an appealing target for antibiotic development (3, 4).

The M. tuberculosis cell envelope consists of four major components: an inner cytoplasmic membrane, a periplasmic region composed of the peptidoglycan and arabinogalactan, the outer membrane (mycomembrane), and the extracellular capsule (5). The hallmark fatty acids of mycobacteria are the mycolic acids (MA), which account for the majority of the mycomembrane (5). Secondary to MA, but of equal importance, are the essential phosphatidylinositol (PI)-based glycerophospholipids (6, 7). The major derivatives of PI are phosphatidylinositol mannoside (PIM) and the further glycosylated forms of PIM, lipoarabinomannan (LM) and mannose lipoarabinomannan (ManLM) (5). All three can be found within both the cytoplasmic membrane and mycomembrane, with the glycans of LM and ManLM extending into the extracellular capsule (5). Mycobacterial lipids and the subsequent modulation of the lipid composition have been associated with cell division, virulence, pathogenesis, and macrophage survival (5).

Residing within the gene cluster that encodes proteins responsible for MA and arabinogalactan biosynthesis is the rv3802c gene (8, 9). The biological function of rv3802c remains in question; however, the gene has been determined to be essential for M. tuberculosis viability (10–12). In addition to being required for general viability, the Mycobacterium avium ortholog of M. tuberculosis rv3802c has been shown to be associated with intestinal epithelium invasion (13). The rv3802c gene encodes an N-terminal translocation signal sequence that includes a predicted transmembrane region (14). Rv3802 was found only in cell wall extracts upon cellular lysis and maintained a molecular mass indicative of an uncleaved protein, suggesting that the enzyme remains anchored in the cytoplasmic membrane (14). Due to sequence similarities with cutinase-like enzymes, the rv3802c gene product was named CULP6 (cutinase-like protein 6), despite lacking cutinase activity (14). However, Rv3802 has been shown to possess both thioesterase and phospholipase A (PLA) activity utilizing residues Ser175-His299-Asp268 as the catalytic triad (8, 14). The structure of the Mycobacterium smegmatis ortholog, MSMEG_6394, has been solved, revealing that the enzyme possesses an α/β-hydrolase fold (12). The apo structure presented two intact disulfides and a small cavity leading to the identified active-site residues (12). M. tuberculosis Rv3802 shares a high level of sequence identity with a variety of Mycobacterium homologs (68.8% sequence identity between encoded M. tuberculosis Rv3802 and Corynebacterium glutamicum CULP6 (cutinase-like protein 6), despite lacking cutinase activity (14)). However, Rv3802 has been shown to possess both thioesterase and phospholipase A (PLA) activity utilizing residues Ser175-His299-Asp268 as the catalytic triad (8, 14). The structure of the Mycobacterium smegmatis ortholog, MSMEG_6394, has been solved, revealing that the enzyme possesses an α/β-hydrolase fold (12). The apo structure presented two intact disulfides and a small cavity leading to the identified active-site residues (12). M. tuberculosis Rv3802 shares a high level of sequence identity with a variety of Mycobacterium homologs (68.8% sequence identity between encoded M. tuberculosis Rv3802 and Corynebacterium glutamicum CULP6 (cutinase-like protein 6), despite lacking cutinase activity (14)). However, Rv3802 has been shown to possess both thioesterase and phospholipase A (PLA) activity utilizing residues Ser175-His299-Asp268 as the catalytic triad (8, 14). The structure of the Mycobacterium smegmatis ortholog, MSMEG_6394, has been solved, revealing that the enzyme possesses an α/β-hydrolase fold (12). The apo structure presented two intact disulfides and a small cavity leading to the identified active-site residues (12). M. tuberculosis Rv3802 shares a high level of sequence identity with a variety of Mycobacterium homologs (68.8% sequence identity between encoded M. tuberculosis Rv3802 and Corynebacterium glutamicum CULP6 (cutinase-like protein 6), despite lacking cutinase activity (14)). However, Rv3802 has been shown to possess both thioesterase and phospholipase A (PLA) activity utilizing residues Ser175-His299-Asp268 as the catalytic triad (8, 14). The structure of the Mycobacterium smegmatis ortholog, MSMEG_6394, has been solved, revealing that the enzyme possesses an α/β-hydrolase fold (12). The apo structure presented two intact disulfides and a small cavity leading to the identified active-site residues (12). M. tuberculosis Rv3802 shares a high level of sequence identity with a variety of Mycobacterium homologs (68.8% sequence identity between encoded M. tuberculosis Rv3802 and Corynebacterium glutamicum CULP6 (cutinase-like protein 6), despite lacking cutinase activity (14)).
**M. tuberculosis Rv3802 structure and lipid binding**

and MSMEG_6394) in addition to the Corynebacterium glutamicum ortholog NCgl2775 (11). Upon heat-induced cell stress, NCgl2775 was shown to play an essential role in the modulation of the mycomembrane composition (11). As a result, the total mycolate content increased, whereas glycerophospholipid content decreased (11). This elevated mycolate/glycerophospholipid ratio was also observed when both MSMEG_6394 and NCgl2775 were overexpressed under physiological conditions (11). Meniche et al. (11) therefore proposed naming this class of enzymes the "envelope lipids regulation factor," or the ElrF family.

Possessing lipase activity, Rv3802 was shown to be susceptible to *in vitro* inhibition by tetrahydrolipstatin (THL), a Food and Drug Administration-approved covalent lipase inhibitor (8). THL is reported to have a $K_{\text{app}}$ of high nanomolar affinity for Rv3802 (12). As such, a series of THL derivatives have been synthesized and tested, displaying improved *in vitro* and correlated *in vivo* activity against *M. tuberculosis* (15). It has been shown that up to 14 lipases undergo covalent modification by THL in *Mycobacterium bovis* (16). Although the *M. bovis* ortholog of Rv3802 was identified as being modified by THL, it was not indicated as one of the 14 statistically validated targets of the study (16).

Due to the essentiality and potential role of Rv3802 in cell wall modulation, we sought to further characterize the function and mechanism. We present the X-ray crystal structure of *M. tuberculosis* Rv3802 to 1.7 Å resolution. Due to binding of PEG molecules to both protein molecules within the asymmetric unit, we were able to characterize the lipid-binding site. Based on the location of the identified lipid-binding site and previous studies, further insights into a biological function for Rv3802 are gained as well as a basis for future structure-based drug design efforts.

**Results**

*M. tuberculosis Rv3802-PEG structure*

The *M. tuberculosis* Rv3802 crystal diffracted to 1.7 Å, having a space group of $P2_1 2_1 2_1$ with two molecules present in the asymmetric unit having a $C_a$ root mean square deviation of 0.14 Å (*X*-ray diffraction and model refinement statistics are given in Table 1). Molecule A was used for structural analyses if not otherwise noted. As predicted by homology to the *M. smegmatis* ortholog, *M. tuberculosis* Rv3802 was found to have an $\alpha/\beta$ hydrolase fold (Fig. 1) (12). Rv3802 has 10 helices (nine $\alpha$-helices and one $3_{10}$-helix) and a single, central $\beta$-sheet composed of five parallel $\beta$-strands (Fig. 1). The nucleophilic Ser$$^{175}$ is located on the aptly named nucleophile elbow connecting the $\beta$3-strand to the $\alpha$3-helix. The remaining catalytic triad residues, His$$^{299}$ and Asp$$^{268}$, are located on the $\alpha9$-helix and a connecting loop between the $\beta$3-strand and $\alpha8$-helix, respectively (Fig. 1). Two intact disulfides are present in the Rv3802 structure: Cys$^{72}$–Cys$^{164}$ and Cys$^{264}$–Cys$^{271}$ (Fig. 1). These conserved disulfides are also present in the MSMEG_6394 structure (12).

During the early stages of model refinement, two regions of linear difference density converging at the active site were observed. This non-protein electron density is interpreted as two molecules of PEG found bound to each of the two protein molecules in the asymmetric unit. Difference density maps for PEG molecules in protein molecule A can be found in Fig. 2a, and the PEG molecules in protein molecule B are shown in Fig. S1. PEG 3350 (average molecular mass is 3350 kDa or ~75 repeating monomer units) was present in the crystallization solution; however, the bound PEG polymers consist of only 5 ethylene glycol monomer units.

**Identification of lipid tail-binding site**

Due to the presence of PEG molecules within Rv3802 and PEG resembling the fatty acid chains of lipids, the lipid-binding site of Rv3802 and presumptive orthologs have been structurally identified (Fig. 2). The fatty acid portion of the lipid-binding site begins at the active site, with the terminal oxygen of PEG1 positioned within hydrogen bonding distance to both Ser$$^{175}$ and His$$^{299}$, 3.2 and 3.8 Å, respectively. The terminal oxygen of the PEG2 resides 6.1 Å away from the terminal oxygen of PEG1 and is positioned between the side chains of Tyr$$^{142}$ and Ile$$^{270}$. Both PEG molecules extend away from the active site in a linear fashion, flanked on both sides by mostly aliphatic amino acid side chains that are positioned on helices $\alpha6$ and $\alpha8$ as well as various loop regions (Fig. 2a). Without ligand present, the aliphatic side chains involved in PEG binding form a hydrophobic core of van der Waals interactions, as seen in the previously solved apo MSMEG_6394 structure (Fig. 2B) (12). Whereas Rv3802 and MSMEG_6394 have a relatively high level of sequence identity for all encoded residues, 68.8%, the residues that make up the identified fatty acid-binding region possess a higher level of conservation, 76.2% between *M. tuberculosis* and *M. smegmatis* specifically, as well as a higher level of conservation with other mycobacterial orthologs (Fig. 2C).

| Table 1 | X-ray diffraction data and structure refinement statistics. Values in parentheses are for the outer shell. |
|---------|----------------------------------|
| **Data collection** | **Rv3802-PEG** |
| Space group | P 21 21 21 |
| $a$, $b$, $c$ (Å) | 57.19, 91.30, 101.39 |
| $\alpha$, $\beta$, $\gamma$ (degrees) | 90, 90, 90 |
| Resolution (Å) | 44.32–1.72 |
| $R_{merge}$ | 0.09 (0.51) |
| $I/\sigma I$ | 23.23 (6.08) |
| Completeness (%) | 99.63 (96.24) |
| Redundancy | 14.8 (14.8) |
| **Refinement** | |
| Resolution (Å) | 44.32–1.72 |
| No. of reflections | 56,631 |
| $R_{work}/R_{free}$ | 0.1582/0.1830 |
| No. of atoms | |
| Protein | 3921 |
| Ligand/ion | 64 |
| Water | 648 |
| $B$-factors (Å$^2$) | 17.7 |
| Protein | 33.1 |
| Ligand/ion | 28.0 |
| Water | |
| Root mean square deviations | |
| Bond lengths (Å) | 0.007 |
| Bond angles (degrees) | 1.06 |
| Ramachandran (%) | 98.0 |
| Favored | |
| Outliers | 0.4 |
M. tuberculosis Rv3802 structure and lipid binding

When α8-helix is in the open conformation, a sizable substrate-binding site is apparent (Fig. 4A). However, in the closed state visualized in the M. smegmatis ortholog, only a small cavity leading to the nucleophilic serine is observed (12). Therefore, for a large substrate possessing hydrophobic, aliphatic character to bind, α8-helix must move to expose the substrate-binding site. It is expected that initial substrate binding positions the scissile ester/thioester bond in proximity to Ser175, affording bifurcation of the active site into two distinct areas that accommodate either the hydrophobic tail or the hydrophilic headgroup. With the reasonable expectation that the PEG molecules bind within the lipid tail-binding site, the lipid head-binding site becomes apparent (Fig. 4B). The headgroup-binding site is unchanged from that observed in the closed state of M. smegmatis ortholog, being the observed small cavity described previously (12). This site is composed of the hydrophobic side chains of Trp84, Leu102, Phe174, Ala292, and Ala300; the negatively charged residues or carboxyl backbones of Thr83, Glu85, and Ala300; and the positively charged side chain of Lys100. These residues are fully conserved between Rv3802 and the M. smegmatis ortholog aside from Lys100 and Ala300; these two residues are Leu and Asn in M. smegmatis, respectively.

Perpendicular to and intersecting the substrate-binding site is a solvent channel leading to the active site (Fig. 4A). In both protein molecules within the asymmetric unit, ordered water molecules are found alongside His299, opposite of the lipid headgroup-binding site. In molecule A, the terminal PEG1 oxygen is 3.2 Å from the Ser175 hydroxyl, whereas the Ser175 hydroxyl is 3.0 Å from the ε nitrogen of His299 (Fig. 5A). Additionally, a water molecule is found 4.2 Å away from His299. In molecule B, the terminal PEG1 oxygen is pointing toward the plausible oxyanion hole, 3.4 Å equidistant to both backbone amides of Thr83 and Asn176. A water molecule is now found within hydrogen-bonding distance, 2.9 Å, to the ε nitrogen of His299 (Fig. 5B). It should be noted that, although not depicted in the figures, solvent molecules are positioned in a similar fashion in the lipid head-binding sites of both protein molecules A and B.

Influence of α8 movement on hydrolase activity

To investigate the effects that α8 movement has on catalysis and to validate the structural observations, M. tuberculosis Rv3802 was mutated to limit α8 movement. In the closed state, Asn133 and Glu289 of the M. smegmatis ortholog are ~5 Å apart; therefore, we mutated the corresponding residues in M. tuberculosis Rv3802, Asn132 and Asn289, to cysteine residues to form an engineered disulfide. The resulting N132C/N288C mutant would therefore have limited α8 helical movement due to covalent tethering. The Rv3802 N132C/N288C variant exhibited lower expression levels but was purified in an identical fashion to WT Rv3802. Both WT and N132C/N288C Rv3802 run as single monomeric bands on a non-reducing SDS-polyacrylamide gel, indicating the lack of intermolecular disulfide bonds (Fig. S2A). Additionally, the amount of free thiols present in...
Figure 3. Superposition of *M. tuberculosis* Rv3802 (gray) and ortholog MSMEG_6394 (purple; PDB code 3AJA); residue numbering is given as Rv3802/MSMEG_6394 (12).

A, the α8-helix and surrounding residues are found to be structurally dynamic upon substrate binding. B, the N-terminal of α8-helix moves ~9 Å between the open and closed states; the region shown spans from residues 265/266 to 302/303. C, despite dynamic movement near the active site, catalytic and surrounding residues are positioned similarly between the open and closed states. The side chain of Asn176 in Rv3802 was modeled in the most likely conformation as assessed by inspection of nearby hydrogen bond donors and acceptors, resulting in the discrepancy with the previously solved MSMEG_6394 Asn177.

Figure 2. The identified lipid tail (fatty acid)-binding site is highly conserved and hydrophobic. A, $F_o - F_c$ likelihood weighted omit map contoured to 3σ for salmon-colored PEG molecules. Amino acid side chains within 4 Å of the PEG molecules are shown in cyan, and active-site residues are orange, with other non-carbon atoms in CPK color. These residues comprise the lipid tail-binding site. B, without ligand present, the amino acid side chains of the lipid tail-binding site form a wall of hydrophobic interactions, as seen in the apo MSMEG_6394 structure (PDB code 3AJA) (12). C, sequence alignment of the fatty acid-binding region from Mycobacteria. Residues of differing sequence are shaded, and starred residues are shown in A and B, whereas a red underline corresponds to secondary structure location.
M. tuberculosis Rv3802 structure and lipid binding

The residual hydrolase activity of the N132C/N288C mutant toward 4MH indicated that N132C/N288C mutant retained sufficient dynamics or space near the active site for 4MH to bind, despite the hindered helical movement. Whereas 4MH has a 7-carbon alkyl chain, the substrate is significantly shorter than phosphatidyl-based substrates that could occupy the entire lipid tail-binding site. Therefore, PLA activity of WT and N132C/N288C mutant was tested using PI as a substrate. Following 1 h of incubation with PI, hydrolyzed product indicative of PLA activity was observed for WT (Fig. 6C, lane 1). However, no PLA activity was observed for the N132C/N288C mutant (Fig. 6C, lane 2).

Discussion

M. tuberculosis Rv3802 structure and PLA classification

Given the structural and enzymatic attributes of Rv3802, the Rv3802 family of enzymes most likely falls within the PLA2 enzyme class (18). Rv3802 was previously determined to have PLA activity toward phosphatidyl species (phosphatidylserine, phosphatidycholine, and phosphatidylethanolamine) (8). In contrast to thioesters, the ester moieties of phospholipids are generally stable in aqueous solutions, so the enzymatic activity observed in vitro is strongly suggestive of biological activity. The two general classes of PLAs differ with regard to specificity for either the sn-1 acyl chain (PLA1) or sn-2 acyl chain (PLA2) of the glycerol phosphate moiety of phosphatidyl species (17). For unknown reasons, PLA2 enzymes possess numerous disulfides, whereas the PLA1 enzymes typically lack disulfides (18). M. tuberculosis Rv3802 was found to possess two disulfides (Fig. 1) (8). These two disulfides are present in the M. smegmatis ortholog structure and were found to be conserved across all Rv3802 orthologs (12).

The first conserved disulfide, Cys72-Cys164, covalently links the N-terminal peptides to the base of α2-helix (Fig. 1). This is of potential structural importance, as the 41 residues preceding Val75, Cys72 included, are predicted to be unstructured until the predicted N-terminal transmembrane region (Ile12-Ile24) (secondary structure and transmembrane prediction in Fig. S4). Therefore, this disulfide may help retain proper protein fold of the structured catalytic hydrolase domain while being immediately preceded by a large, unstructured domain. The second conserved disulfide, Cys264-Cys271, induces a kink in a loop to appropriately position the catalytically important Asp268 residue. This second disulfide therefore ensures the proper positioning of Asp268 relative to His299, maintaining the catalytic triad despite movement of the nearby α7-helix upon lipid binding.

Mechanism of catalysis

PLA2 enzymes catalyze fatty acid hydrolysis through two general mechanisms: a calcium-dependent His/Asp mechanism or a calcium-independent mechanism (18). Given the identified catalytic triad and the lack of calcium-dependent enzyme activity, Rv3802 is a calcium-independent PLA2 enzyme. Flanked by aromatic side chains (Trp84, Phe174) and displaying little movement of catalytic residues between open and closed enzymatic states, the active site of Rv3802 resembles that of a typical hydrolase enzyme (Fig. 3C) (19). These features...
observed in the Rv3802 active site are known to help promote substrate hydrolysis over transfer (19). These structural attributes help explain why Rv3802 has no reported transerase activity in the presence of acyl acceptors (8).

Based on the enzymatic activity, structure, and catalytic triad of Rv3802, catalysis most likely proceeds through a Ping-Pong reaction mechanism. The reaction can be broken down into two half-reactions. The first half-reaction proceeds through nucleophilic attack on the carbon of the ester/thioester carbonyl by Ser^{179}, which is deprotonated by His^{299}. The terminal oxygen of PEG1 would therefore mimic the carbon of the carbonyl moiety of substrate before nucleophilic attack, depicted in Fig. 5A. Following attack, the plausible oxyanion hole consisting of backbone amide nitrogen atoms of Thr^{83} and Asn^{176} stabilizes the tetrahedral intermediate. Tetrahedral intermediate collapses yield the acyl-enzyme intermediate and the release of the now deacylated lipid at the end of the first half-reaction. The acyl-enzyme intermediate form of the enzyme is approximated by the terminal oxygen of PEG1 from protein molecule B (Fig. 5B). The second half-reaction would therefore proceed through the activation of a water molecule by His^{299} and subsequent nucleophilic attack on the acyl-enzyme intermediate. The nucleophilic water molecule is observed within hydrogen bonding distance to His^{299} in protein molecule B and is depicted in Fig. 5B. Tetrahedral collapse following this second nucleophilic attack results in the release of a free fatty acid and a regenerated enzyme.

**Lipid binding**

Based on the structural similarity of PEG to the short-chain fatty acids of various lipids, it is therefore logical that a similar repositioning of the α-helix must occur for binding of natural substrate and subsequent catalysis to occur. The Rv3802 N132C/N288C mutant, with limited α-helix movement, displayed a lower ability to bind 4MH, resulting in a significant loss in activity compared with WT enzyme (Fig. 6B). However, when the larger PI substrate was used, no PLA activity was observed for the N132C/N288C mutant (Fig. 6C). This loss in activity as a result of hindered helical movement highlights the importance of protein dynamics with respect to substrate binding. Given the large variety of possible structural forms of α-helix and adjacent loops, it is difficult to fully assess how this dynamic interplay affects substrate recognition.

*Initial substrate recognition may proceed through binding of the lipid head, as this site does not change between open and closed states of the enzyme. The MD simulation of apo Rv3802 suggests that the α-helix quickly adopts the closed conformation and remains in that state without substrate present when in the aqueous phase (Fig. S3). Therefore, substrate binding may drive helical opening in a solvated environment. However, given the uncertainty of the true biological function of Rv3802, it is difficult to fully assess the mechanism of lipid binding based on the Rv3802-PEG structural observations and mutant kinetics. PLA₁ enzymes have been found to bind only the surface-exposed lipid headgroup, to reside partially within the membrane itself, and to bind the entire lipid or extract the lipid from the membrane for hydrolysis (18, 20).

When α-helix is in the open state, a vast hydrophobic lipid-binding site is apparent (Fig. 2). In agreement with typical lipid-binding sites, the identified Rv3802 lipid-binding site is lined with aliphatic amino acid side chains (Fig. 2) (8, 21). The Rv3802 hydrophobic binding channel is relatively linear, extending away from the active site, therefore resulting in a site capable of binding two alkyl chains in a parallel linear arrangement similar to the arrangement of the two observed PEG molecules. Therefore, the binding of PEG molecules observed within Rv3802 provides a structural basis for the binding of the two acyl chains of the phosphatidyl substrate.

**Structure-based insights into biological function**

Meniche et al. (11) determined that Rv3802 orthologs NCgl2775 and MSMEG_6394 play an essential role in modu-
lating lipid composition. Specifically, they noted the increase of the mycolic acid/glycerophospholipid ratio, suggesting that this family of enzymes plays a direct role in decreasing glycerophospholipid content within the mycomembrane (11). It should be noted that NCgl2775 is not essential under physiological growth, whereas the \textit{M. tuberculosis} orthologs are essential (11, 12). This differing essentiality of Rv3802 orthologs may be attributed to the differing compositions of the outer membranes between mycobacteria and corynebacteria (22). For example, the inner layer of the mycomembrane of corynebacteria has a much higher composition of cardiolipin relative to other glycerophospholipids, when compared with mycobacteria (22).

PI species represent a major portion of the glycerophospholipids found within the plasma membrane and mycomembrane layers in both \textit{M. tuberculosis} and \textit{M. smegmatis} (23). Previously, Rv3802 was shown \textit{in vitro} to hydrolyze a fatty acid from PIM2 upon treatment of \textit{M. smegmatis} cell wall extracts with the enzyme (8). PIMs play an important role in the permeability of the \textit{M. tuberculosis} cell wall, affect proper cell division, and influence pathogenesis (24–27). Interestingly, when a conditional knockout of MSMEG_6394 was cultured, cell morphology was jagged and elongated compared with WT (12). PIM2 has four acyl substituents, two on the phosphoglycerol, one on the inositol, and the fourth on the mannose moiety. The observed lipid-binding site of Rv3802 is significantly larger than the palmitoyl-binding pocket of other hydrolases known to hydrolyze esters of PIM (28). Specifically, the enzyme PatA resides within the PIM biosynthetic pathway and is responsible for the transfer of a palmitoyl group to the 6-position on the mannose sugar (28). Based on the larger Rv3802 lipid-binding site compared with PatA, a sensible biological function for Rv3802 is the hydrolysis of one or both of the fatty acids from the phosphoglycerol component of PIM or other PI species. PIM has been shown to be an antigen of CD1d-restricted T cells; however, PIM binding to CD1d is abolished upon the hydrolysis of the diacyl phosphoglycerol moiety following treatment with a PLA$_2$ enzyme (27). Whereas Rv3802 has the enzymatic ability to facilitate this chemistry, further in-depth studies are required to truly validate the biological context of such action.

Using the Rv3802-PEG structure as a basis, PI was modeled into the open form of the enzyme (Fig. 7). This model provides a reasonable structural basis for binding of phosphatidyl-based substrates to Rv3802 (8). Structurally, PI can be easily accommodated within the lipid-binding site of Rv3802 in a manner that promotes catalysis. As modeled, the \textit{sn}-2 acyl chain would be subject to hydrolysis. Although little positive charge is present at this site to counter the negatively charged phosphodiester moiety, stabilization of interactions with the phosphate moiety may be afforded through solvent or by backbone amide nitrogen atoms of the highly dynamic loop upon restructuring as a function of substrate binding (Ala$^{92}$–Glu$^{96}$). Due to the proximity of this loop to the lipid head-binding site, this region may also be relevant to sugar binding of larger PI species, such as PIM.

The essential nature of Rv3802 to \textit{M. tuberculosis} viability makes it an alluring new target for future drug development. Derivatives of THL exhibit enhanced \textit{in vitro} inhibition against Rv3802 and related \textit{in vivo} bioactivity (15). However, THL has been shown to target numerous \textit{M. tuberculosis} lipases and directly affect MA content (16). Therefore, inhibition of \textit{M. tuberculosis} by THL proceeds through a shotgun, polypharmacology strategy that has little specificity toward Rv3802. The identification of the lipid-binding site of Rv3802 highlights an extremely intriguing and potentially promising site for drug development. The presented structure and subsequent analysis can be utilized for designing inhibitors specific to Rv3802.

**Experimental procedures**

**Molecular cloning**

An \textit{Escherichia coli} codon-optimized, synthetic gene encoding the periplasmic domain of \textit{M. tuberculosis} Rv3802c was purchased from Integrated DNA Technologies. The bases of 5’-CGCTGTTCAGGGACCT-3’ and 5’-GCGTCCGGATCC-3’ were added to the 5’- and 3’-ends of the gene, respectively. The resulting gene was PCR-amplified using primers with sequences identical to those added to the gene and inserted into a pET32-derived plasmid linearized with PshAI using a Gibson Assembly$^\text{TM}$ Master Mix. The protein expression construct was verified by restriction digestion. The protein expression construct was amplified in \textit{E. coli} and purified.

An in vitro expression construct was also amplified in \textit{E. coli} and purified. For the purification of the in vitro expression construct, an in vitro transcription and translation were performed using the TNT coupled translation system (Promega). The resulting protein was purified using a nickel-affinity column.

**Figure 7. PI modeled into the lipid-binding site of Rv3802.** The fatty acid chains are easily accommodated within the lipid-binding site, allowing the \textit{sn}-2 ester to be positioned by the nucleophilic Ser$^{175}$, placing the inositol moiety within the proposed sugar-binding site. FA1, tuberculosis acid; FA2, palmitic acid, \textit{sn}-2 position of glycerol.
M. tuberculosis Rv3802 structure and lipid binding

were then amplified together to produce a full-length mutant gene and inserted into an expression plasmid in an identical fashion to WT. Both WT and mutant plasmids were sequenced to confirm the desired DNA sequence.

Protein expression and purification

Protein expression and purification were adapted and modified from a methodology published previously (14). Both WT and N132C/N288C mutant protein were expressed and purified in an identical fashion. Chemically competent T7 Express E. coli cells (New England Biolabs) were transformed with plasmid encoding WT or mutant protein. Subsequent cultures were grown at 37 °C in lysogeny broth medium containing carbencillin. Protein expression was induced for 3 h at 37 °C, following the addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside once culture reached an A600 of 0.6. Induced cells were pelleted and resuspended in a 50 mM Tris, pH 7.5, 300 mM NaCl buffer and placed at −80 °C for storage.

Induced cells were thawed in a warm bath, followed by the addition of DNase I and lysozyme, and allowed to incubate on ice for half an hour. Resuspended cells were then sonicated, and after addition of DNase I and lysozyme, and allowed to incubate on ice for half an hour. Resuspended cells were then sonicated, and the pellet was resuspended in 50 mM Tris, pH 7.5, 300 mM NaCl buffer and placed at −80 °C for storage.

Resuspended cells were pelleted and resuspended in a 50 mM Tris, pH 7.5, 3 mM NaCl buffer. Resuspended cells were again pelleted at 18,515 × g for 40 min at 4 °C. Supernatant was discarded, and the pellet was resuspended in 50 mM Tris, pH 7.5, 3 mM NaCl buffer and placed at −80 °C for storage.

The resulting pelleted inclusion bodies containing Rv3802 were solubilized in 50 mM Tris, pH 7.5, and 4 mM guanidine HCl (GdnHCl) and pelleted at 18,515 × g for 40 min at 4 °C. The supernatant containing soluble Rv3802c was loaded onto an equilibrated 5-ml HiTrap Talon crude cobalt column (GE Healthcare) and was washed for 15 column volumes with 50 mM Tris, pH 7.5, 4 mM GdnHCl. Protein was eluted over a 20-column volume gradient with increasing imidazole concentration. Rv3802 was refolded through the removal of GdnHCl as a result of extensive dialysis using a 50 mM Tris, pH 7.5, 300 mM NaCl buffer. Protein must be at a low concentration (~0.25 mg/ml) within the dialysis bag for refolding, or significant precipitation is observed. Refolded protein was centrifuged at 18,515 × g for 1 h at 4 °C to remove any residual insoluble protein. Soluble, pure protein was dialyzed into either 100 mM HEPES, pH 7.5, for crystallization or 50 mM NaPO4, pH 7.5, for cysteine quantification, 4MH, and PLA assays.

Crystallization and data collection

Purified Rv3802 was concentrated to 4.5 mg/ml for crystallization. Screening of Rv3802 against the index screen (Hampton Research) using the hanging drop method resulted in the formation of a single orthorhombic crystal. The crystal formed in a 1:1 ratio of well solution (0.1 M BisTris, pH 6.5, 25% (w/v) PEG 3350) to protein, requiring 1 month to grow with incubation at 16 °C. The resulting crystal was looped and flash-cooled in liquid nitrogen. Extensive efforts to crystallize the N132C/N288C mutant have been exerted; to date, diffraction quality crystals have not been obtained. X-ray diffraction data were collected using synchrotron radiation at the Advanced Photon Source, LS-CAT beamline F, at Argonne National Laboratory.

Structure determination and refinement

Diffraction data were indexed and scaled in a P212121 space group using HKL2000 (29). The phase solution with two molecules in the asymmetric unit came from molecular replacement (PHASR-MR) using a model generated with SWISS-MODEL that was based on the previously solved MSMEG_6394 structure (PDB entry 3AJA) (30, 31). Atoms lacking 2Fo − Fc density were deleted, and the resulting model was subjected to rigid body and simulated annealing refinements (PHENIX Refine) (30). Deleted atoms were built back into corresponding Fo − Fc and 2Fo − Fc difference density maps using COOT (32). The progressing model was subjected to rounds of xyz coordinate, real-space, occupancy, and individual B-factor refinements (PHENIX-Refine) with manual modeling in COOT (30, 32). The four molecules of PEG were added using LigandFit and subjected to model refinement (33). Model refinement was complete after a final Rwork/Rfree of 0.1582/0.1830 was achieved. The 0.4% Ramachandran outlier is due to Val237 and is warranted, given the strong electron density for all atoms of the residue in both molecules within the asymmetric unit. MolProbity was utilized to validate the structure (34). Rv3802 model coordinates and structure factors were deposited in the PDB with accession code 5W95.

Cysteine quantification

To ensure proper disulfide formation of the engineered N132C/N288C mutant, thiols present in WT and mutant were quantified and compared using fluorescent labeling with a non-reversible thiol-reactive probe. Respective enzymes were compared in non-reduced and reduced states. A 5 mM concentration of respective enzyme in 50 mM sodium phosphate buffer, pH 7.5, was incubated with 0.3 mM tris(2-carboxyethyl)phosphine (reduced) or equivalent volume buffer (non-reduced) for 30 min at 37 °C (200-μl reaction volumes). Samples were titrated with 240 μM monochlorobimane (100 mM stock, DMSO), mixed, and incubated at 37 °C for 30 min. A standard curve to quantify free thiols was generated by serial dilution of l-cysteine in the presence of 0.3 mM tris(2-carboxyethyl)phosphine, reduction for a half hour at 37 °C, and modification with 240 μM monochlorobimane for 30 min at 37 °C (Fig. S2D). Fluorescent reads were conducted in triplicates using λex = 390 nm and λem = 490 nm on a Synergy H4 plate reader (Biotek).

In vitro hydrolase assays

A fluorescence-based assay using 4-methylumbelliferyl heptanoate (4MH) (Sigma-Aldrich) was developed for Rv3802, which affords a sensitive method for activity determination and is amenable to drug screening, with a Z′ score of 0.84 (Z′ determination and methods shown in Fig. S5). In the presence of enzyme, the heptyl chain is hydrolyzed from 4MH, producing the now fluorescent 4-methylumbelliferone molecule and heptanoic acid. All assays were performed in triplicate in a 50 mM sodium phosphate, pH 7.5, buffer. For Michaelis–Menten kinetics, 50 mM WT Rv3802c and 75 mM N132C/N288C mutant were used with varying concentrations of serial diluted 4MH (20 mM DMSO stock). All fluorescent reads were conducted at 37 °C using λex = 360 nm and λem = 450 nm on a Synergy H4 plate reader (Biotek). Background fluorescence as a conse-
sequence of 4MH hydrolysis was subtracted, and relative fluorescence units were converted to concentration of standard curve (Fig. S6). Reaction rates and Michaelis–Menten kinetic parameters were determined using PRISM version 7.

PLA activity

PLA activity was assessed using a modified version of a previously established Rv3802-PIM hydrolysis experiment (8). Briefly, soy PI (Avanti® Polar Lipids Inc.) was resuspended in 50 mM sodium phosphate, pH 7.5, forming a 2 mg/ml suspension of PI. Reactions consisted of 300 μg of PI suspension, 50 μg of respective enzyme or an equivalent volume of buffer and were brought to a final volume of 400 μl with 50 mM sodium phosphate, pH 7.5, buffer. Reactions were incubated for 1 h at 37 °C, being mixed every 10 min, and quenched with 200 μl of 0.3 CHCl₃/CH₂OH/HCl. The lower organic phase was extracted, spotted onto a silica TLC plate, and resolved using a mobile phase of 80:20:2 CHCl₃/CH₂OH/NH₄OH. Inositol-containing species were visualized by treating the TLC plate with 5% H₂SO₄ in methanol and charring.

Structural and bioinformatic analysis

X-ray crystal structure alignments, measurements, and subsequent analysis were performed using PyMOL (35). Sequence alignments were performed with ClustalW, and figures were generated through the use of T-COFFEE and BOXSHADE (36, 37). Ipred4 was used for secondary structure prediction (38). Transmembrane regions were predicted using the TMHMM server (39).

Author contributions—C.M.G., C.M.S., and S.D. formal analysis; C.M.G., C.M.S., and S.D. visualization; C.M.G. and S.D. writing-original draft; C.M.G., C.M.S., and S.D. writing-review and editing; D.R.R. conceptualization; D.R.R. supervision; D.R.R. funding acquisition; D.R.R. validation; D.R.R. project administration.

Acknowledgments—This research used resources of the Advanced Photon Source, a United States Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357. This work used resources of the Compute and Data Environment for Science (CADES) at ORNL, which is managed by UT-Battelle, LLC for the DOE under Contract DE-AC05-00OR22725.

References

1. World Health Organization (2016) 2016 Global Tuberculosis Report, World Health Organization, Geneva.
2. Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D., and Ravignione, M. (2016) Tuberculosis. Nat. Rev. Dis. Primers 2, 16076 CrossRef Medline
3. Mitchison, D., and Davies, G. (2012) The chemotherapy of tuberculosis: past, present and future. Int. J. Tuberc. Lung Dis. 16, 724–732 CrossRef Medline
4. Jackson, M., McNeil, M. R., and Brennan, P. J. (2013) Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. Future Microbiol. 8, 855–875 CrossRef Medline
5. Jackson M. (2014) The mycobacterial cell envelope—lipids. Cold Spring Harb. Perspect. Med. 4, a021105 CrossRef Medline
6. Kaur, D., Guerin, M. E., Skovierová, H., Brennan, P. J., and Jackson, M. (2009) Chapter 2: Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis. Adv. Appl. Microbiol. 69, 23–78 CrossRef Medline
7. Jackson, M., Crick, D. C., and Brennan, P. I. (2000) Phosphatidylinositol is an essential phospholipid of mycobacteria. J. Biol. Chem. 275, 30092–30099 CrossRef Medline
8. Parker, S. K., Barkley, R. M., Rino, J. G., and Vasil, M. L. (2009) Mycobacterium tuberculosis Rv3802c encodes a phospholipase/thioesterase and is inhibited by the antymycobacterial agent tetrahydrolipstatin. PLoS One 4, e4281 CrossRef Medline
9. Marrakchi, H., Lanéelle, M. A., and Daffé, M. (2014) Mycolic acids: structures, biosynthesis, and beyond. Chem. Biol. 21, 67–85 CrossRef Medline
10. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis, Mol. Microbiol. 46, 77–84 CrossRef Medline
11. Meniche, X., Labarre, C., de Sousa-d’Auria, C., Huc, E., Laval, F., Tropis, M., Bayan, N., Portevin, D., Guihlout, C., Daffé, M., and Houssin, C. (2009) Identification of a stress-induced factor of Corynebacterineae that is involved in the regulation of the outer membrane lipid composition. J. Bacteriol. 191, 7323–7332 CrossRef Medline
12. Credlin, P. K., Vivian, J. P., Scoble, J., Chow, F. M., West, N. P., Brammanthan, R., Proellochs, N. I., Shahine, A., Le Nours, J., Wilce, M. C., Britton, W. J., Coppel, R. L., Rossjohn, J., and Beddoe, T. (2010) Tetrahydrolipstatin inhibition, functional analyses, and three-dimensional structure of a lipase essential for mycobacterial viability. J. Biol. Chem. 285, 30050–30060 CrossRef Medline
13. Mittler, E., Daroogehe, K., Mehta, P. K., Cirillo, S. L., Cirillo, J. D., and Bermudez, L. E. (2005) Identification of Mycobacterium avium genes that affect invasion of the intestinal epithelium. Infect. Immun. 73, 4214–4221 CrossRef Medline
14. West, N. P., Chow, F. M., Randall, E. J., Wu, J., Chen, J., Ribeiro, J. M., and Britton, W. J. (2009) Cutinase-like proteins of Mycobacterium tuberculosis: characterization of their variable enzymatic functions and active site identification. FASEB J. 23, 1694–1704 CrossRef Medline
15. West, N. P., Cergol, K. M., Xue, M., Randall, E. J., Britton, W. J., and Payne, R. J. (2011) Inhibitors of an essential mycobacterial cell wall lipase (Rv3802c) as tuberculosis drug leads. Chem. Commun. 47, 5166–5168 CrossRef Medline
16. Ravindran, M. S., Rao, S. P., Cheng, X., Shukla, A., Cazenave-Gassiot, A., Yao, S. Q., and Wenk, M. R. (2014) Targeting lipid esterases in mycobacteria grown under different physiological conditions using activity-based profiling with tetrahydrolipstatin (THL). Mol. Cell Proteomics 13, 435–448 CrossRef Medline
17. Dennis, E. A. (2015) Introduction to thematic review series: phospholipases: central role in lipid signaling and disease. J. Lipid Res. 56, 1245–1247 CrossRef Medline
18. Dennis, E. A., Cao, J., Hsu, Y. H., Magrioti, V., and Kokotos, G. (2011) Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. Chem. Rev. 111, 6130–6185 CrossRef Medline
19. Jochens, H., Hessler, M., Stiba, K., Padhi, S. K., Kazlauskas, R. J., and Bornscheuer, U. T. (2011) Protein engineering of α/β-hydrolase fold enzymes. ChemBioChem 12, 1508–1517 CrossRef Medline
20. Mouchlis, V. D., Bucher, D., McCammon, J. A., and Dennis, E. A. (2015) Membranes serve as allosteric activators of phospholipase A2, enabling it to extract, bind, and hydrolyze phospholipid substrates. Proc. Natl. Acad. Sci. U.S.A. 112, E516–E525 CrossRef Medline
21. Marsh D. (2003) Lipid-binding proteins: structure of the phospholipid ligands. Protein Sci. 12, 2109–2117 CrossRef Medline
22. Bansal-Mutalik, R., and Nikaio, H. (2011) Quantitative lipid composition of cell envelopes of Corynebacterium glutamicum elucidated through reverse micelle extraction. Proc. Natl. Acad. Sci. U.S.A. 108, 15360–15365 CrossRef Medline
23. Hoffmann, C., Leis, A., Niederweis, M., Plištko, J. M., and Engelhardt, H. (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. Proc. Natl. Acad. Sci. U.S.A. 105, 3963–3967 CrossRef Medline
M. tuberculosis Rv3802 structure and lipid binding

24. Parish, T., Liu, J., Nikaido, H., and Stoker, N. G. (1997) A Mycobacterium smegmatis mutant with a defective inositol monophosphate phosphatase gene homolog has altered cell envelope permeability. *J. Bacteriol.* 179, 7827–7833 CrossRef Medline

25. Korduláková, J., Gilleron, M., Mikusova K., Puzo, G., Brennan, P. J., Gicquel, B., and Jackson, M. (2002) Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis PimA is essential for growth of mycobacteria. *J. Biol. Chem.* 277, 31335–31344 CrossRef Medline

26. Morita, Y. S., Velasquez, R., Taig, E., Waller, R. F., Patterson, J. H., Tull, D., Williams, S. J., Billman-Jacobe, H., and McConville, M. J. (2005) Compartmentalization of lipid biosynthesis in mycobacteria. *J. Biol. Chem.* 280, 21645–21652 CrossRef Medline

27. Fischer, K., Scotet, E., Niemeyer, M., Koebenick, H., Hurwitz, R., Kursar, M., Bonneville, M., and Kaufmann, S. H., and Schaible, U. E. (2004) Mycobacterial phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10685–10690 CrossRef Medline

28. Álvesa-Jové, D., Svetlíková, Z., Tersa, M., Sancho-Vaello, E., Carreras-González, A., Bonnet, P., Arrasate, P., Eguskiza, A., Angala, S. K., Cifuentes, J. O., Korduláková, J., Jackson, M., Mikusová, K., and Guerin, M. E. (2016) Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nat. Commun.* 7, 10906 CrossRef Medline

29. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326 CrossRef

30. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kaprál, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oefner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221 CrossRef Medline

31. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258 CrossRef Medline

32. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501 CrossRef Medline

33. Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W., and Cohn, J. D. (2006) Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallogr. D Biol. Crystallogr.* 62, 915–922 CrossRef Medline

34. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21 CrossRef Medline

35. Delano, W. L. (2013) *The PyMOL Molecular Graphics System*, version 1.6, Schrödinger, LLC, New York

36. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217 CrossRef Medline

37. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539 Medline

38. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* 43, W389–W394 CrossRef Medline

39. Krogh, A., Larsson, B., von Heine, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580 CrossRef Medline