Crystal Structure of a Monomeric Thiolase-Like Protein Type 1 (TLP1) from *Mycobacterium smegmatis*

Neelanjana Janardan¹, Rajesh K. Harijan², Rikkert K. Wierenga²*, Mathur R. N. Murthy¹*

¹ Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, India, ² Department of Biochemistry, Biocenter Oulu, University of Oulu, Oulu, Finland

Abstract

An analysis of the *Mycobacterium smegmatis* genome suggests that it codes for several thiolases and thiolase-like proteins. Thiolases are an important family of enzymes that are involved in fatty acid metabolism. They occur as either dimers or tetramers. Thiolases catalyze the Claisen condensation of two acetyl-Coenzyme A molecules in the synthetic direction and the thiolytic cleavage of 3-ketoacyl-Coenzyme A molecules in the degradative direction. Some of the *M. smegmatis* genes have been annotated as thiolases of the poorly characterized SCP2-thiolase subfamily. The mammalian SCP2-thiolase consists of an N-terminal thiolase domain followed by an additional C-terminal domain called sterol carrier protein-2 or SCP2. The *M. smegmatis* protein selected in the present study, referred to here as the thiolase-like protein type 1 (MsTLP1), has been biochemically and structurally characterized. Unlike classical thiolases, MsTLP1 is a monomer in solution. Its structure has been determined at 2.7 Å resolution by the single wavelength anomalous dispersion method. The structure of the protomer confirms that the N-terminal domain has the thiolase fold. An extra C-terminal domain is indeed observed. Interestingly, it consists of six β-strands forming an anti-parallel β-barrel which is completely different from the expected SCP2-fold. Detailed sequence and structural comparisons with thiolases show that the residues known to be essential for catalysis are not conserved in MsTLP1. Consistent with this observation, activity measurements show that MsTLP1 does not catalyze the thiolase reaction. This is the first structural report of a monomeric thiolase-like protein from any organism. These studies show that MsTLP1 belongs to a new group of thiolase related proteins of unknown function.

Citation: Janardan N, Harijan RK, Wierenga RK, Murthy MRN (2012) Crystal Structure of a Monomeric Thiolase-Like Protein Type 1 (TLP1) from *Mycobacterium smegmatis*. PLoS ONE 7(7): e41894. doi:10.1371/journal.pone.0041894

Editor: Anil Kumar Tyagi, University of Delhi, India

Received April 3, 2012; Accepted June 29, 2012; Published July 26, 2012

Copyright: © 2012 Janardan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Department of Biotechnology, Government of India and Academy of Finland Indo-Finnish Grant Grant number: BT/FINNISH/27/MRNM/2009. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mrn@mbu.iisc.ernet.in (MRNM); or rik.wierenga@oulu.fi (RKH)

Introduction

The genus Mycobacterium comprises some of the most devastating pathogens that infect both animals and humans. To date, twenty eight genomes of mycobacterial species have been sequenced completely [http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi]. A striking feature in all these genomes is the abundance of genes coding for enzymes involved in fatty acid metabolism; more than 250 in *M. tuberculosis* compared to only 50 in *Escherichia coli*. Thiolases are a ubiquitous group of enzymes that are involved in fatty acid metabolism. They occur as either dimers or tetramers in the thiolase superfamily. The best characterized member of the thiolase subfamily is the tetrameric biosynthetic bacterial thiolase from *Zoogloea ramigera* [8,9]. The key catalytic residue in all characterized members of the thiolase superfamily is a cysteine residue (Cys89 in *Z. ramigera* thiolase). This nucleophilic cysteine participates in the reaction by forming a covalent intermediate with the substrate. There are two other important catalytic residues, either two histidines or an asparagine and a histidine (Asn316 and His348 in *Z. ramigera* thiolase) that are also conserved in the thiolase superfamily [10]. These two residues are important for the formation of an oxygenation hole, which stabilizes the enolate intermediate of the Claisen condensation reaction. The reaction also requires a base to abstract a proton from the substrate. In the biosynthetic thiolase of *Z. ramigera*, a cysteine (Cys378) has been shown to play this role [8]. In thiolases, these catalytic residues are present in four highly conserved loops with characteristic sequence fingerprints [7,11,12]. An additional sequence fingerprint near the catalytic site is important for substrate discrimination between unbranched and 2-methyl branched fatty acid tails [7]. Each thiolase subunit is about 400 residues long and has three regions, an N-terminal core region, a loop region (referred to as the thiolase finger) and a C-terminal core region [7]. Each of these three regions consists of about 120 residues. The thiolase loop domain shapes the binding pocket of the Coenzyme A (CoA) and only a few tetramers have been reported. The tetramers are dimers of tight dimers. The best characterized member of the thiolase subfamily is the tetrameric biosynthetic bacterial thiolase from *Zoogloea ramigera* [8,9]. The key catalytic residue in all characterized members of the thiolase superfamily is a cysteine residue (Cys89 in *Z. ramigera* thiolase). This nucleophilic cysteine participates in the reaction by forming a covalent intermediate with the substrate. There are two other important catalytic residues, either two histidines or an asparagine and a histidine (Asn316 and His348 in *Z. ramigera* thiolase) that are also conserved in the thiolase superfamily [10]. These two residues are important for the formation of an oxygenation hole, which stabilizes the enolate intermediate of the Claisen condensation reaction. The reaction also requires a base to abstract a proton from the substrate. In the biosynthetic thiolase of *Z. ramigera*, a cysteine (Cys378) has been shown to play this role [8]. In thiolases, these catalytic residues are present in four highly conserved loops with characteristic sequence fingerprints [7,11,12]. An additional sequence fingerprint near the catalytic site is important for substrate discrimination between unbranched and 2-methyl branched fatty acid tails [7]. Each thiolase subunit is about 400 residues long and has three regions, an N-terminal core region, a loop region (referred to as the thiolase loop domain) and a C-terminal core region [7]. Each of these three regions consists of about 120 residues. The thiolase loop domain shapes the binding pocket of the Coenzyme A (CoA)
moiety of the substrate. The N- and C-terminal core regions have similar topologies [12] and appear to be the result of gene duplication. Most of the catalytic residues occur in the C-terminal half but the conserved reactive nucleophile, a cysteine (Cys89 of Z. ramigera), is found in the N-terminal half of the protomer. Catalysis in both the synthetic and the degradative directions starts with the acylation of the catalytic, nucleophilic cysteine (Cys89 in Z. ramigera thiolase) [8].

In humans, six different thiolases have been identified (CT, T1, T2, TFE, AB and SCP2) [12,13,14] with distinct distribution in cellular compartments, quaternary structure, substrate specificity and enzyme kinetics. The sequences of these six thiolases are similar. Of these thiolases, AB [12], T2 [11] and CT [15] have been well characterized and a crystal structure has been determined for each of these enzymes. In contrast, no crystal structures are available for T1, SCP2 and TFE-thiolases. The mammalian SCP2-thiolase has an additional sterol carrier protein C-terminal domain (SCP2) [16]. The structure of SCP2 is known [17,18].

Examination of the Mycobacterium smegmatis genome revealed the presence of several putative thiolase genes [19]. These genes have been annotated as thiolases on the basis of sequence analysis. However, none of them has been biochemically characterized. The sequence identity between some of these proteins and the other well-characterized thiolases is rather low. The protein encoded by one of the M. smegmatis thiolase-like genes (the thiolase-like protein type-1, M.TLP1) was over expressed in E. coli, purified and crystallized [19]. The crystal structure of this M/TLP1 is described here. Analysis of the structure revealed that the protein has an N-terminal domain with a typical thiolase fold and an additional C-terminal domain. The C-terminal domain is a six stranded anti-parallel β-barrel which is completely different from the structure of SCP2. A detailed analysis of the structure and biochemical properties of this protein revealed that its fold agrees with that of classical thiolase, but the nucleophilic active site cysteine and other characteristic thiolase sequence fingerprints are not conserved. Accordingly, it is found that the protein does not exhibit thiolase activity. Therefore, this protein can be classified as a new subfamily of the thiolase superfamily.

Results and Discussion

Bioinformatics analysis

From a genome database search, two TLP coding genes (M.TLP1 and M.TLP2) were identified in M. smegmatis. The pairwise sequence identity between the two corresponding proteins is 31%. A similar search of the M. tuberculosis genome revealed the presence of only one such gene. The pairwise sequence identities of the protein encoded by this gene with M.TLP1 and M.TLP2 are 33% and 69%, respectively. Subsequent searches revealed the presence of TLP homologs in several other bacterial genomes. However, a protein homologous to TLP was not found in the human genome. A multiple sequence alignment was performed using six human thiolases, several other eukaryotic thiolases and seven TLPs from various bacterial sources including M. smegmatis and M. tuberculosis. The phylogenetic relationships of TLPs with thiolases were analyzed and the results are presented in Figure 2. The TLPs are clustered as a distinct group and the branch point is near the SCP2-thiolase group with high confidence. The mammalian SCP2-thiolase gene codes for an N-terminal thiolase domain and a C-terminal SCP2 domain. Interestingly, the TLP gene also codes for a thiolase like N-terminal domain and an additional C-terminal domain. However, the latter domain is not related to the C-terminal domain of SCP2-thiolase in sequence and its function is yet to be established, as discussed below.

A multiple sequence alignment reveals the presence of a unique YSCF sequence (residues 330–333 of TLP1) fingerprint in TLPs (Figure 3). The pairwise, structure based sequence alignment between the M.TLP1 and the bacterial Z. ramigera thiolase (PDB code: 1DM3) (Figure 4) shows that this sequence fingerprint corresponds to the NEAF sequence (316–319 of Z. ramigera thiolase) motif of thiolases. This motif is known to be important for the geometry of the oxyanion hole [9]. Similarly, four other well characterized, functionally important sequence fingerprints of thiolases (CXS; 89–91; VMG, 287–289; GHP, 347–349 and CXG, 378–380 of Z. ramigera enzyme) were also not conserved in M.TLP1 (Figure 4). The sequence identity between M.TLP1 and Z. ramigera thiolase is 15% (Figure 4).

Enzymological characterization

The molecular mass of an M.TLP1 protomer as calculated from the sequence is 56 kDa. Static Light Scattering (SLS) and size exclusion chromatography experiments show that, unlike other thiolases, M.TLP1 is a monomer of 56 kDa in solution (Figure S1). Similarities between the structures of M.TLP1 and thiolases suggest that M.TLP1 might be able to bind CoA and its derivatives. However, thiolase activity could not be detected in either the forward or the reverse direction. The assays were carried out with two different batches of freshly purified enzyme.

Surface Plasmon Resonance (SPR) binding assays showed that TLP1 binds CoA (Figure S2). The binding curves were fitted to the Langmuir binding equation and the affinity of M.TLP1 for CoA was found to be in the millimolar range (Kd = 0.6 mM ±0.3) (n = 3). The assays were carried out using two different batches of freshly purified enzyme.

Structure determination and model quality

The crystal structure of M.TLP1 was determined using the single wavelength anomalous dispersion (SAD) method. As the trigonal P1 cell could accommodate six protomers of the polyepitide, rotation functions were computed using reflections...
Figure 2. Evolutionary tree analysis of the thiolase sequences. The phylogenetic tree was constructed using the neighbor-joining method, with 10,000 bootstrap replicates in MEGA5 software. Only the region corresponding to the thiolase domain was used for these calculations. The group identifiers correspond to the nomenclature described in the text. The numbers next to each node indicate bootstrap values as percentages. The following sequences (listed with their NCBI accession codes and, for trypanosomatid sequences, their NCBI and GeneDB accession codes) were used for creating the evolutionary tree.

1. Human: HsT1 (NP_006102.2); 2. Mus musculus: MmT1 (NP_803421.1); 3. Salmo salar: SsT1 (ACI33809.1); 4. Monosiga brevicollis: MbT1 (XP_001748310.1); 5. Bdellovibrio bacteriovorus: BbT1 (NP_967398.1); 6. Zoogloea ramigera: ZrCT (1DM3); 7. HsCT (NP_005882.2); 8. MmCT (NP_033364.2); 9. HsT2 (NP_000010.1); 10. Macaca fascicularis: MtT2 (NP_659033.1); 11. Rattus norvegicus: RnT2 (NP_058771.1); 12. MmT2 (NP_659033.1); 13. Saccharomyces cerevisiae: ScAB (1AFW); 14. HsAB (NP_001598.1); 15. Arabidopsis thaliana: AtAB (2WU9); 16. T. cruzi: TcAB (Tc00.1047053511389.150, XP_816035.1); 17. Bodo saltans: BsAB (BSA00133, ACI160321.1); 18. HsTFE (NP_000174.1); 19. TcTFE (Tc00.1047053511389.150, XP_814301.1); 20. L. braziliensis: LmTFE (LmF.31.1640, XP_001685150.1); 21. L. major: LmTFE (LmF.31.1640, XP_001685150.1); 22. L. mexicana: LmTFE (LmF.30.1640, CBZ27218.1); 23. L. infantum: LmTFE (LmF.31.1660, CAM70518.2); 24. HsSCP2 (NP_002970.2); 25. LmSCP2 (LmF.23.0860, XP_001465761.1); 26. LmSCP2 (LmF.23.0860, XP_001683404.1); 27. LmSCP2 (LmF.23.0860, XP_001683404.1); 28. LmSCP2 (LmF.23.0860, XP_001683404.1); 29. LmSCP2 (LmF.23.0860, XP_001683404.1); 30. T. vivax: TcSCP2 (Tb927.8.2020); 31. T. congolense: TcSCP2 (congo1030g10.p1k_8); 32. T. brucei: TbSCP2 (Tb927.8.2540, XP_847087.1); 33. T. gambiae: TgSCP2 (Tbg972.8.2020); 34. M. smegmatis: MsSTLP2 (YP_887911.1); 35. M. tuberculosis: MtSTLP (NP_216383.1); 36. R. jostii: RsSTLP (YP_700393.1); 37. C. crescentus: CsSTLP (YP_002517418.1); 38. H. neptunium: HnSTLP (YP_759708.1); 39. MsSTLP1 (YP_889758.1); 40. Erythrobacter sp.: EbSTLP (YP_01041346.1).

doi:10.1371/journal.pone.0041894.g002
Structure of a TLP from *M. smegmatis*.
Figure 3. Sequence conservation in TLP proteins. The sequence alignment was achieved using ClustalW program. Seven unique TLP proteins were used for the analysis; TLP1 and TLP2 from M. smegmatis (GenBank accession nos. YP_6889758.1, YP_887911.1), TLP from M. tuberculosis (GenBank accession no. XP_004553641.1), A. oryzae, M. acridum (GenBank accession no. XP_001871576.2) and R. erythropolis (GenBank accession no. XP_043886691.1). Conserved residues are highlighted in red. Residues structurally equivalent to the five sequence fingerprint prints of classical thiolases are highlighted in numbered green boxes.

doi:10.1371/journal.pone.0041894.g003

in the 10 Å - 4 Å resolution shell and a radius of integration of 30 Å. Rotation functions corresponding to $\kappa = 180^\circ$ and $\kappa = 120^\circ$ hemispheres had significant peaks consistent with non-crystallographic 32 symmetry [19].

The structure determination was not straightforward and various strategies as described in the methods section had to be used for obtaining a nearly complete structure. About 63 residues of the thiolase domain distributed over four loops could not be built, indicating that these residues belong to truly disordered segments. Electron density is missing for four segments consisting of residues 17–25, 64–77, 135–170 and 253–254 in the thiolase domain (see also Figure 4) as well as for the N-terminal hexa histidine tag. In the rest of the polypeptide, adequate density is not observed for the side chains of only a few surface residues. The data processing and structure refinement statistics are shown in Tables 1 and 2 respectively. The quality of the final electron density map is illustrated in Figure 5. The electron density difference map has an elongated unmodeled feature at an inter-hexamer crystal contact region close the disordered loop 63–76. The uninterpreted density is not connected to the residues at the beginning or end of the disordered loop. The final R-factor and R-free of the refinement at 2.7 Å resolution are 23.3% and 26.1%, respectively (Table 2).

The overall fold

The M/TLP1 polypeptide fold is illustrated in Figure 6. The polypeptide folds into two distinct domains: an N-terminal thiolase like domain (residues 1–407) and an additional C-terminal domain (residues 435–507), with a 27 residue long linker (Figure 6A) connecting the two domains. The N-terminal domain of M/TLP1 exhibits the characteristic thiolase fold. DALI search against the PDB using the N-terminal domain as the template revealed several structural homologs. Functionally, all these proteins are involved in fatty acid metabolism. The N-terminal domain can be further divided into two topologically similar N-terminal and C-terminal halves, both of which exhibit the $\beta\alpha\beta\beta$ topology characteristic of thiolase domains (Figure 6). Between the N- and C-terminal halves is a layer of two $\alpha$ helices contributed by the two halves of the N-terminal domain. The $\beta$ sheets of the N- and C-terminal halves are also surrounded by a helical layer on the side facing the bulk solvent (Figure 6). Thus, the N-terminal domain is organized into a five layered $\alpha/\beta/\alpha/\beta/\alpha$ structure characteristic of the thiolase superfamily. In thiolases, the topologically similar N- and C-terminal halves are connected by the long thiolase loop domain consisting of about 120 residues. In M/TLP1, this region corresponds to residues 133 to 253. The lengths of the loop domains in thiolases and M/TLP1 are comparable. However, there is no electron density for a large segment at the N-terminal region of this loop in M/TLP1 (133–170), indicating that this region of the loop domain is highly flexible. The end of the thiolase loop domain is also disordered in M/TLP1 (Figure 4). Two other disordered loops of the M/TLP1 thiolase domain consisting of 9 and 14 residues, respectively, occur after N81 and N82.

The C-terminal extra domain (residues 435–507) is made up of 6 $\beta$ strands and resembles a barrel-like structure with three anti-parallel $\beta$-strands on either side of the barrel (Figure 6A). There is a short helix located at the top of the barrel resembling a lid. The C-terminal extra domain is connected to the N-terminal thiolase domain by a long loop consisting of a short helix (Figure 6A). The topology of the C-terminal domain is reminiscent of single strand nucleic acid binding proteins. DALI search using this domain against the PDB shows structural similarities to a molybdenum binding protein [PDB id: 1H9K] [20] and a single strand DNA binding protein [PDB id: 3PGZ] (Seattle Structural Genomics Center for Infectious Disease; Unpublished). However, no characteristic DNA binding sequence motif could be identified in the domain.

Organization of protomers in the crystal asymmetric unit

The six protomers in the asymmetric unit are related by a nearly perfect non-crystallographic 32 symmetry (Figure 7). The layer of protomers (labeled A, B, C in Figure 7A) related by the non-crystallographic 3-fold symmetry is staggered with respect to the second layer of protomers (labeled A’, B’, C’) related to the first by the non-crystallographic 2-fold symmetry. In the non-crystallographic 3-fold related layer, the inter-protomer contacts are between the thiolase domain of one protomer and the C-terminal domain of the other protomer (Figure 7). The rotational relationship of neighboring protomers related by the non-crystallographic 3-fold axis are within ±0.5° of 120°. Similarly, the rotations relating protomers of the two layers are within ±0.5° of 180°.

Analysis of inter-protomer contacts

The buried surface areas between protomers related by the non-crystallographic 3-fold and 2-fold axes were computed. The solvent accessible surface area of a protomer is 18,145 Å². Due to the 3-fold related interactions, the area that gets buried in the A/B interface is 561 Å² per protomer. This represents only 3.2% of the total area and is small compared to the buried surface area of tightly associated subunits in protein oligomers. This suggests that the interface between 3-fold related subunits is not strong. On the average, 20 residues from the N-terminal domain of one protomer and 15 residues from the C-terminal domain of the adjacent protomer participate in the interface. The interface is stabilized by 5 hydrogen bonds (Arg 18-Asp 452, Thr 16-Thr 454, Asp 21-Lys 455, 12 H-bonds and 18 salt bridges and is completely different from the interface between A and A’ subunits of M/TLP1. Therefore, the hexameric state observed in the crystal structure of M/TLP1 is likely to be the result of crystal packing. This observation is consistent with the monomeric state of M/TLP1 in solution suggested by size exclusion chromatography and Static Light Scattering (SLS). Modeling studies suggested that the linker region between the N and C-terminal domains of M/TLP1 (Figure 6A) obstructs the formation of an interface similar to that of Z ramigera thiolase due to steric clashes. Also, the tetramerization loop of Z ramigera thiolase, which occurs at the N-terminal end of the thiolase.
Figure 4. Structure based sequence alignment of Z. ramigera thiolase (PDB ID: 1DM3) and MsTLP1 (TLP1, GenBank accession no. YP_889758.1). The secondary structure assignment above and below the sequences refer to Z. ramigera thiolase and MsSTPL1, respectively. Black coils represent $\alpha$-helices and black arrows represent $\beta$-strands. Residues conserved in both proteins are highlighted in red. The sequence identity between the thiolase domains of the Z. ramigera enzyme and MsTLP1 is 15%. The five sequence fingerprints of Z. ramigera thiolase and their corresponding residues in MsTLP1 are colored blue, and labeled in green boxes as in Figure 3. The five segments of the thiolase loop domain with specific functions are highlighted as follows: tetramerization loop (orange), covering loop (purple), cationic loop (grey), adenine loop (light pink), pantetheine loop (yellow). Disordered regions of the MsTLP1 thiolase domain are highlighted in light green.

doi:10.1371/journal.pone.0041894.g004
loop domain (Figure 4), corresponds to a longer, completely disordered region (residues 135–169) in MsTLP1.

Structural comparison of the thiolase loop domain of MsTLP1 and Z. ramigera thiolase

The most extensively studied thiolase structure is that from Z. ramigera. Of the 392 Cα atoms of Z. ramigera thiolase, 306 could be superposed on corresponding Cα atoms of the N-terminal domain of MsTLP1 with a root mean square deviation (r.m.s.d.) of 2.4 Å (Figure 8A). It is noteworthy that most of the structural differences observed between MsTLP1 and Z. ramigera thiolase are in the thiolase loop domain (119–249 of Z. ramigera thiolase). In Z. ramigera thiolase, five segments of this domain appear to be important for catalysis and substrate specificity [11,12]. Figure 8B shows a structural superposition of the thiolase loop domains of Z. ramigera thiolase and MsTLP1 highlighting the five functionally important segments. The tetramerization loop occurring at the N-terminus of this domain is essential for the tetrameric organization of Z. ramigera thiolase. Residues from this loop also interact with the substrate. The covering loop occurs immediately after the tetramerization loop and covers the active site pocket. The pantetheine loop interacts with the pantetheine part of bound CoA. The covering loop and the pantetheine loop together shape...

Table 1. Data collection statistics for Mycobacterium smegmatis TLP1.

| Space group | Se-Peak-1 | Se-Peak-2 | Se-Peak-3 |
|-------------|----------|----------|----------|
| Wavelength(Å) | 0.9789   | 0.9789   | 0.9789   |
| Temperature (K) | 100      | 100      | 100      |
| Resolution range (Å) | 50.00-2.60(2.64-2.60) | 50.00-2.60(2.64-2.60) | 50.00-2.70(2.75-2.70) |
| Unit cell parameters (Å, °) | α = 96.30, β = 101.95, γ = 102.40, χ = 116.32, θ = 101.20, γ = 97.71 | α = 96.30, β = 101.95, γ = 102.40, χ = 116.32, θ = 101.20, γ = 97.71 | α = 96.30, β = 101.95, γ = 102.40, χ = 116.32, θ = 101.20, γ = 97.71 |
| Observed reflections | 197,397 | 196,381 | 172,105 |
| Unique reflections | 53,404 | 38,502 | 41,976 |
| Redundancy | 3.7(2.8) | 5.1(3.7) | 4.1(2.8) |
| Mosaicity (deg) | 0.7 | 0.7 | 0.7 |
| Data completeness (%) | 97.1(84.9) | 97.3(86.5) | 95.3(77.2) |
| I/σ(I) | 16.6(2.3) | 18.8(2.5) | 12.8(4.6) |
| Rmerge (%) | 0.08(0.3) | 0.09(0.4) | 0.09(0.4) |
| No. of monomers per asymmetric unit | 6 | 6 | 6 |

Values in parenthesis refer to the highest resolution shell.

I is the integrated intensity and σ(I) is the estimated standard deviation of that intensity.

Rmerge = ∑hkl |I(hkl)−<I(hkl)>|/∑hkl I(hkl), where I(hkl) is the intensity of the hkl measurement of reflection hkl and <I(hkl)> is its mean.

doi:10.1371/journal.pone.0041894.t001

Table 2. Refinement statistics and model quality.

| Resolution (Å) | 2.7 (2.77-2.70) |
| R (%) | 23.3 (39.6) |
| Rfree (%) | 26.1 (42.8) |
| No. of atoms | 20423 |
| Protein atoms | 20211 |
| Solvent atoms | 212 |
| Root mean square deviations from ideal values |
| Bond length (Å) | 0.01 |
| Bond angle (degrees) | 1.9 |
| Average B factors (Å²) | |
| Main chain atoms | 46.3 |
| Side chain atoms and waters | 47.1 |
| Residues in Ramachandran plot (%) |
| Most favored region | 95.6 |
| Allowed region | 4.4 |
| Generously allowed region | 0.0 |
| Disallowed region | 0.0 |
| PDB code entry | 4EGV |

doi:10.1371/journal.pone.0041894.t002
the entrance to the catalytic pocket of *Z. ramigera* thiolase. The cationic loop is solvent-exposed and is thought to capture the negatively charged substrate. The adenine binding loop promotes binding of the adenosine moiety of CoA. The tetramerization loop and the pantetheine loop that occur at the end of the thiolase loop domain are disordered in *M. smegmatis* TLP1. From sequence comparison (Figure 4), it is evident that the tetramerization loop is longer in *M. smegmatis* TLP1 by about eight residues when compared to the corresponding loop of *Z. ramigera* thiolase. The cationic loop is substantially shortened in *M. smegmatis* TLP1 and hence does not resemble...
the corresponding loop of \( \zeta \) ramigera thiolase. These differences in the conformation of the loops surrounding the active site pocket are likely to be functionally significant. The adenine loop, however, is in a similar position and is of the same length in both the proteins.

**Sequence fingerprints of thiolases**

All enzymes in the thiolase family have five highly conserved sequence fingerprints as shown in Figure 4 [7,11,14]. The corresponding residues in \( \text{MstLP1} \) were identified by structure based sequence alignment and careful manual examination (Figure 4). The most significant difference is that the active site nucleophilic cysteine, Cys89 of the CXS motif of \( \zeta \) ramigera thiolase, has been replaced by Gly102 in \( \text{MstLP1} \). This fingerprint is labeled 1 in Figures 3 and 4. The VMG motif (residues 287–289 of thiolase; fingerprint 2) determines the size of the substrate that can be accommodated in the active site cavity. In T2-thiolase [10], the active site cavity can accommodate a branched fatty acid with a 2-methyl group. The inability of \( \zeta \) ramigera thiolase to bind long chain acyl CoA moieties as well as 2-methyl branched fatty acid CoAs has been attributed to the presence of two residues, Met157 and Met288, which protrude
into the binding pocket [21,22]. The *Z. ramigera* thiolase-Met157 has been replaced by a smaller residue Ala178 in *M*.TLP1. The VMG-motif that includes Met288 is significantly displaced in M.TLP1 such that a much larger volume is available for ligand binding (Figure 9). Therefore, this region of the putative binding pocket of M.TLP1 may bind a more extended ligand. Substantial differences are also observed between *Z. ramigera* thiolase and M.TLP1 in other motifs that define the nature of the active site. Asn316 in the NEAF motif (fingerprint 3; residues 316–319) of *Z. ramigera* thiolase makes a hydrogen bond with the catalytic water [9]. This motif has been replaced by YSCF (residues 330–333) in M.TLP1. The GHP motif (residues 347–349 of *Z. ramigera* thiolase, fingerprint 4), which is involved in the formation of the oxyanion hole, has been replaced by GGL (residues 364–366 of M.TLP1). Finally, Cys379 of the CXG motif (residues 378–380 of *Z. ramigera* thiolase, fingerprint 5), which acts as the base for proton abstraction from the substrate in thiolase, has been replaced by Asn395 of the corresponding NGG motif in M.TLP1 (395–397). The altered motifs YSCF, GGL and NGG are, however, conserved in homologs of M.TLP of other prokaryotes (Figure 3) suggesting that these residues may be important for the function. The conservation of the fingerprint sequences within TLP homologs and classical thiolases but not between the two families may be a key feature that determines the functional differences between proteins belonging to these families.

The putative Coenzyme A binding groove of MsTLP1

The surface characteristics of M.TLP1 are shown in Figure 10A. A large groove is observed extending across the full length of the M.TLP1 molecule. Comparison with the *Z. ramigera* thiolase structure shows that this groove corresponds to the binding site for a CoA molecule or a fatty acyl CoA molecule. As in *Z. ramigera* thiolase, several positively charged residues, notably Arg227, Arg240, Arg248, line this putative CoA-binding pocket, presumably to stabilize binding of the negatively charged phosphate groups of the ligand (Figure 10A). An analysis of the loops of TLP1 (Figure 4) shows that the shape and binding properties of the groove extending beyond the CoA-binding site will be affected by the disordered loops occurring at the beginning (135–169) and end (residues 153, 154) of the loop domain, as visualized in Figure 10B. In particular, the disordered loop at the beginning of the loop domain (135–169) is rather long and could cover part of the binding pocket for CoA or a fatty acyl CoA molecule after complex formation. Both loops may become ordered on binding the physiological ligand, thereby shielding the bound ligand from bulk solvent.

The C-terminal domain

The C-terminal domain of M.TLP1 does not have the same fold as the SCP2 domain of human SCP2-thiolase. The residues of the linker region between the thiolase domain and the C-terminal domain prevent formation of dimers resembling those of classical thiolases. This concerns for example, clashes between the M.TLP1 linker helix (resides 419–427) and the loop region just before the Nβ3 dimer interface β-strand of the other subunit of the hypothetical dimer (residues 73–83), as shown in Figure 10B.

The interface area between the N-terminal thiolase domain and the C-terminal domain is 1035 Å². This constitutes 22% of the C-terminal domain and 4% of the N-terminal domain surface areas, respectively. There are 19 hydrogen bonds and one salt bridge (between Asp487 of the C-terminal domain and Arg190 of the thiolase domain) between the two domains. Interestingly, Arg190 is largely conserved in the homologs of TLP. Asp487 is also conserved in the homologs of TLP. In three of the seven homologs shown in Figure 3, it has been replaced by glutamate and hence the charge is conserved. Thus it appears that most of the TLP homologs have a C-terminal domain anchored to the thiolase domain by similar interactions. This domain is not close to the putative active site, except for residues DWTP (452–455) of the long loop occurring after EB1 (Figure 4, Figure 8A). This loop is hydrogen bonded to the thiolase domain via an anti-parallel β-sheet interaction between residues TVR (448–450) and IVD (242–244) of the region just before Nβ5 of the thiolase domain. It is interesting to note that residues of this domain have a slightly higher B-factor (54.3 Å²) as compared to the N-terminal thiolase domain (41.6 Å²). The role of the C-terminal domain is yet to be elucidated. Deletion studies are underway to shed more light on the functional importance of this domain.

Concluding remarks

All enzymes of the thiolase superfamily that have been structurally characterized so far share four features: 1) conservation of the core α/β/α/β/α-layered structure of the thiolase domain, 2) conservation of the extensive dimerization interface, 3)
Interestingly, none of the catalytic residues of the thiolase subfamily is conserved in M\textsubscript{st}TLP1. Nevertheless, the fold of the thiolase part of M\textsubscript{st}TLP1 closely resembles the \textit{Z. ramigera} thiolase structure. The location of the ligand binding pocket for CoA in the two proteins also appears to be conserved. M\textsubscript{st}TLP1 has a unique extra C-terminal domain of unknown function. The structure of this domain is completely different from that of the SCP2 domain of SCP2-thiolase. Examination of TLP1 homology suggests that the interactions between the thiolase-like domain and the C-terminal additional domains are largely conserved. The physiological function of the full length M\textsubscript{st}TLP1 protein has not yet been established. Further studies have been initiated to establish the intriguing role of M\textsubscript{st}TLP1 in the lipid metabolism of mycobacterium and other microorganisms.

Materials and Methods

Cloning, expression and purification

The gene coding for \textit{M. smegmatis} TLP1 (YP_889758) was cloned, over-expressed in \textit{E. coli} and purified as described earlier [19].

Selenomethionine incorporation

The M\textsubscript{st}TLP1 clone was transformed into BL21 (DE3) Rosetta strain of \textit{E. coli}. A single colony was picked and grown in 1 ml LB-ampicillin overnight at 37° C. This was then inoculated into 500 ml minimal media with 100 μg/ml ampicillin and grown for 12 hrs at 37° C. Methionine synthesis was inhibited by the addition of an amino acid mixture containing 50 mg/l of leucine, isoleucine, valine, lysine, threonine and phenylalanine. After incubation for half an hour, protein expression was induced with 1.0 mM isopropyl-beta-thio galactopyranoside (IPTG). Selenomethionine was added at the time of induction. Purification of the protein was carried out as described for the native enzyme. Selenomethionine incorporation was confirmed by accurate mass determination using Electrospray ionisation mass spectrometry (ESI-MS).

Crystallography, data collection and data processing

The crystals were obtained as described previously [19]. Briefly, the crystals were grown at room temperature using the microbatch method. A crystallization droplet consisted of 2 μl of protein solution (5 mg/ml in 50 mM Tris-HCl pH 8.0 containing 100 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol) and 2 μl of crystallization solution (100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 20% Polyethylene glycol (PEG) 4000, 10% isopropanol). Selenium incorporated M\textsubscript{st}TLP1 crystals were obtained under the same conditions. Three data sets of selenomethionine labeled crystals (Se-M\textsubscript{st}TLP1) were collected to 2.7 Å resolution at a wavelength of 0.978 Å (peak) at beam line BM14 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, using three different crystals. The data were processed using DENZO and SCALEPACK from the HKL-2000 suite [23]. Data processing revealed that the crystal belonged to the triclinic space group P1 and the size of the unit cell is compatible with 4–8 protomers of MW 56 kDa corresponding to Matthews coefficients of 4.0 Å\textsuperscript{3}Da\textsuperscript{–1}, 2.0 Å\textsuperscript{3}Da\textsuperscript{–1}, respectively. The final statistics for data collection and processing are summarized in Table 1.

Structure solution and refinement

The structure of M\textsubscript{st}TLP1 was determined as follows. Using anomalous differences, 63 selenium positions were identified and refined using the program SHELXD [24]. Initial selenium
position based phases obtained by the PHENIX program were improved by solvent flattening using the program DM of CCP4 [25] suite. The resultant map was used for automated model building with the program Autobuild of the Phenix suite [26]. The first round of automated model building could generate only 40% of the structure. The model obtained was manually adjusted using the interactive graphics program COOT [27] and refined using REFMAC5 [28] of the CCP4 suite. Subsequent rounds of manual model building and refinement extended the structure by an additional 20%. At this stage, a phase combination of the initial phases obtained by anomalous data and the model based phases was carried out using the program Phase-Combine of CCP4 suite [25]. The electron density map obtained using the combined phases was used for a new round of automated model building. This extended the structure by an additional 5%. At this stage, an anomalous difference Fourier map was calculated and selenium positions were re-determined. The 62 highest peaks in the difference Fourier map were accepted as representing selenium atoms. Between this set and the earlier set of selenium positions, 50 were common. The new selenium positions were subjected to extensive refinement using PHASER [29]. Phases obtained using these positions were used for map calculation and automated model building. At this stage, 80% of the structure could be built.

These calculations were based on only one of the three data sets collected at the peak wavelength. At this stage, all the three peak data sets collected at ESRF were merged and scaled and the resulting data were used to perform Molecular Replacement-Single wavelength Anomalous Dispersion (MR-SAD) using PHASER [29] along with the 80% model already built as the partial structure for initial phase calculation. This extended the model by an additional 10% and brought the structure to its current state of completion, i.e. 90%.

Structure analysis

The geometry of the final model was examined using PROCHECK [30]. All structural superpositions were achieved using the SSM superpose feature of COOT [27]. Average B-factors for protein atoms, water molecules and ligands were calculated using BVERAGE of the CCP4 suite. The PISA [31] server was used for interface area calculations. Surface area calculations and analysis of contacts were performed using programs available in the CCP4 suite [25]. All figures were prepared using PYMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). Bacterial Z. ramigera thiolase (PDB ID: 1DM3) has been used for all sequence and structural alignments. 1DM3 represents the acetylated Z. ramigera thiolase structure complexed with acetyl CoA.

Bioinformatics analysis

Only the thiolase domain of MTLPL1 was used in all sequence analyses. The amino acid sequence of MTLPL1 was used as a starting point in an extensive bioinformatics analysis of the M. smegmatis genome: (i) Proteins homologous to MTLPL1 were identified in the non-redundant protein sequence database of SWISSPROT [32] using BLAST [33]. (ii) Additional sequences that might be related to MTLPL1 were identified using the six characterized human thiolases as query sequences in the BLAST program. Two TLP homologs, TLP1 and TLP2 (YP_887911.1 and YP_889738.1), were identified in the M. smegmatis genome. The catalytic cysteine is not conserved in either of these proteins although they have been annotated as thiolases. These proteins appear to have an N-terminal thiolase-like domain and an extra domain at the C-terminus. (iii) A BLAST search against the human genome using the MTLPL1 amino acid sequence did not reveal any hits. However, homologs of MTLPL1 could be identified in many prokaryotic genomes. TLP homologs were also observed to be present in all other Mycobacterium spp. Interestingly, only one TLP sequence was found in the M. tuberculosis genome. (iv) A multiple sequence alignment was performed by ClustalW [34] using amino-acid sequences of the six human thiolases, several other eukaryotic thiolases and seven TLPs from various sources including two mycobacterial spp. This alignment was used to generate a phylogenetic tree using the neighbor-joining method with 10,000 bootstrap replicates in MEGA5 [35].

Oligomeric state of MTLPL1

The oligomeric state of MTLPL1 in solution was determined using a static light scattering (SLS) system (Wyatt Minidawn) with its flow cell connected to an AKTA purifier operated at 17°C. MTLPL1 (0.5 ml of 4.0 mg/ml) protein solution in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 10% glycerol, 5 mM 2-mercapto-ethanol was loaded on a Superdex 200 (10/300) GL column (GE Healthcare) attached to the AKTA purifier. The elution profiles (Supplementary Figure S1) were also monitored using RI and UV detectors and analysed using the ASTRA program. The molecular mass expected from the sequence was calculated using the online tool PROTPARAM [36].

Thiolase activity assays

The thiolase activity in the degradative (Figure 1) and synthetic directions were estimated as described previously [8] at 30°C using a Shimadzu UV-1800 299 spectrophotometer. The concentrations of the substrates CoA, acetyl CoA and acetoacetyl CoA, in the respective stock solutions, were verified using the Ellman’s test [37]. In all the assays, a cocktail without MTLPL1 served as the negative control and measurements with active trypanosomal thiolase were used as the positive control.

In the thiolytic direction, the 500 μl reaction cocktail contained 50 mM Tris-HCl pH 7.8, 25 mM MgCl₂, 60 μM CoA and 50 μM acetoacetyl CoA. The disappearance of the Mg²⁺/acetoacetyl CoA complex was measured at 303 nm for 3 min after the addition of 10 μg of MTLPL1. In the synthetic direction, the activity assay was carried out using a short chain 3-hydroxyacyl CoA dehydrogenase (SC-HAD) as the linker enzyme [9]. In this assay, the formation of acetoacetyl CoA was measured by incubating the enzyme with acetyl CoA and measuring the product formation by reducing it with nicotinamide adenine dinucleotide (NADH), as catalyzed by the linker enzyme (as described previously [11]). The reaction was initiated by the addition of 10 μg of MTLPL1 to a mixture containing 50 mM Tris-HCl, pH 7.8, 40 mM KCl, 0.2 mM NADH, 1 U of SCHAD (1 U is 1 μmol/[min. mg] of protein), 0.5 mM dithiothreitol (DTT) and 2 mM acetyl CoA. The total reaction volume was 500 μl and the rate of NADH oxidation was monitored at 340 nm for 5 minutes.

CoA affinity studies

The binding studies between MTLPL1 with CoA were performed with a Biacore 3000 (Biacore, Uppsala, Sweden) optical biosensor at 25°C. 2 μM of MTLPL1 protein was immobilized by amine coupling on the surface of a CMS sensor chip (GE Healthcare) to a level of ~3,000 response units using the GE Healthcare standard immobilization protocol. Binding and dissociation were measured at a flow rate of 20 μl per minute. The ligand solution was run over the sensor surface at five different ligand concentrations (0.3, 0.6, 1.25, 2.50 and 5 μM) in a running buffer of 10 mM HEPES pH 7.4, 130 mM NaCl, 3 mM ethylenediaminetetraacetic acid (EDTA) and 0.005% P-20 The
sensor surface was regenerated with borate buffer pH 8.5. The binding curves were corrected for non-specific binding by subtracting the signal obtained for the negative control flow cell. Kinetic constants for association and dissociation were derived from linear transformations of the binding data. The data were fit to the simple 1:1 Langmuir interaction model using the BIA EVALUATION software.

Supporting Information

Figure S1 The oligomeric state analysis of the recombinantly expressed, purified MsTLP1 has been determined by using static light scattering (SLS) in combination with size-exclusion chromatography. A Superdex 200 10/300 GL column (GE Healthcare) was used for the size-exclusion chromatography. The elution profile is provided by the UV signal (thin red line). The SLS signal is combined with the RI-signal for the calculation of the molar mass (thick red line) by the ASTRA program.

(TIF)

Figure S2 Kinetics of binding of CoA to MsTLP1 determined using SPR. The SPR sensogram was obtained by flowing different CoA solutions (see inset) over the TLP1 immobilized sensor chip. The data show that CoA binds to TLP1 with a Kd in the millimolar range.

(Acknowledgments)

The intensity data were collected using the X-ray Facility for Structural Biology at the Molecular Biophysics Unit, Indian Institute of Science, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT), Government of India and at the BM 14 beam line, ESRF, Grenoble, France. We acknowledge the award of Shyama Prasad Mukerji Fellowship by the Council for Scientific and Industrial Research (CSIR), Govt. of India to NJ. We thank Prof. D. Chatterji for the gift of M. smegmatis genomic DNA. We thank Anja Paul for technical assistance and Dr. Piotr Prus for helping with the Biacore measurements.

Author Contributions

Conceived and designed the experiments: MRNM RKW. Performed the experiments: NJ RH. Analyzed the data: NJ RH. Wrote the paper: NJ. Obtained the grants for this research: MRNM RKW.