Crystal structure of a thiolase from *Escherichia coli* at 1.8 Å resolution

M. Ithayaraja,^a N. Janardan,^a Rik K. Wierenga,^b H. S. Savithri^c and
M. R. N. Murthy^{a*}

^aMolecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560 012, India, ^bFaculty of Biochemistry and Molecular Medicine, Biocenter Oulu, University of Oulu, FIN-90014 Oulu, Finland, and ^cBiochemistry Department, Indian Institute of Science, Bangalore, Karnataka 560 012, India. *Correspondence e-mail: mrm@mbu.iisc.ernet.in

Received 15 April 2016

Accepted 25 May 2016

Edited by R. Sankaranarayanan, Centre for Cellular and Molecular Biology, Hyderabad, India

Keywords: *Escherichia coli* thiolase; fatty-acid metabolism; degradative enzyme; active-site geometry; asymmetry.

PDB references: *E. coli* thiolase, 5f0v; complex with CoA, 5f38

Supporting information: this article has supporting information at journals.iucr.org/f

Thiolases catalyze the Claisen condensation of two acetyl-CoA molecules to give acetoacetyl-CoA, as well as the reverse degradative reaction. Four genes coding for thiolases or thiolase-like proteins are found in the *Escherichia coli* genome. In this communication, the successful cloning, purification, crystallization and structure determination at 1.8 Å resolution of a homotetrameric *E. coli* thiolase are reported. The structure of *E. coli* thiolase co-crystallized with acetyl-CoA at 1.9 Å resolution is also reported. As observed in other tetrameric thiolases, the present *E. coli* thiolase is a dimer of two tight dimers and probably functions as a biodegradative enzyme. Comparison of the structure and biochemical properties of the *E. coli* enzyme with those of other well studied thiolases reveals certain novel features of this enzyme, such as the modification of a lysine in the dimeric interface, the possible oxidation of the catalytic Cys88 in the structure of the enzyme obtained in the presence of CoA and active-site hydration. The tetrameric enzyme also displays an interesting departure from exact 222 symmetry, which is probably related to the deformation of the tetramerization domain that stabilizes the oligomeric structure of the protein. The current study allows the identification of substrate-binding amino-acid residues and water networks at the active site and provides the structural framework required for understanding the biochemical properties as well as the physiological function of this *E. coli* thiolase.

1. Introduction

Fatty-acid metabolism is an important biosynthetic pathway that is essential for many biological functions such as membrane assembly, energy production and storage. In each β -oxidation cycle of fatty-acid catabolism, consisting of four discrete steps, the length of the fatty-acid chain is reduced by two carbon units. Thiolases are coenzyme A (CoA)-dependent enzymes and function in both the degradative (EC 2.3.1.16) and biosynthetic (EC 2.3.1.9) directions of fatty-acid metabolism. Degradative thiolases catalyze the breakdown of fatty-acid chains by breaking down 3-ketoacyl-CoA to acetyl-CoA and an acyl-CoA shortened by two C atoms. In the synthetic direction, thiolases catalyze the Claisen condensation of an acetyl-CoA and an acyl-CoA, leading to the formation of a longer acyl-CoA. The thiolase superfamily, of which thiolases are a subfamily, consists of several enzymes involved in fatty-acid metabolism such as ketoacyl-(acyl-carrier-protein)-synthases (KASs), polyketide synthases and chalcone synthases. All of these enzymes share a similar three-dimensional structure and depend on CoA for their biological function.

Extensive structural and functional studies have previously been conducted on thiolases from several organisms (Kursula

et al., 2005; Meriläinen *et al.*, 2009; Kiema *et al.*, 2014). The N- and C-terminal domain core structures of all thiolases possess a similar $\beta\alpha\beta\alpha\beta\beta$ topology, suggesting that these enzymes might have evolved by gene duplication. Although these enzymes may be dimers or tetramers in quaternary organization, they share a similar active-site geometry and probably carry out reactions following similar mechanisms (Kursula *et al.*, 2005). Both in the degradative and synthetic directions, the reaction is a two-step process. In the first step, a nucleophilic cysteine at the active site becomes acylated. In the synthetic direction, this cysteine becomes acetylated by reaction with acetyl-CoA. In the degradative direction, the cysteine becomes acylated by reaction with 3-ketoacyl-CoA, with the concomitant release of acetyl-CoA. In the second reaction of the synthetic pathway, the acyl group covalently linked to the cysteine is transferred to acetyl-CoA, leading to a longer acyl-CoA. In the degradative direction, the acyl group is transferred to CoA (Haapalainen *et al.*, 2006).

Based on cellular localization and functional characterization, the human thiolases have been grouped into six well defined classes: CT, T1, T2, TFE, SCP2 and A/B. Thiolases that belong to the T1 class mainly act in the degradative direction and appear to possess broad chain-length specificity, and most of them are dimeric enzymes. Enzymes of class T2 catalyze the Claisen condensation in the synthetic direction, leading to ketone-body synthesis. These enzymes are tetrameric. In a recent phylogenetic study of 130 thiolase sequences that included all *Escherichia coli* thiolases, 12 distinct classes were identified. This analysis suggested that the *E. coli* thiolase for which the structure is presented in this communication is probably related to the biosynthetic CT class (Anbazhagan *et al.*, 2014).

The fatty-acid content of *E. coli* is about 10% of its dry weight and represents a major energy source for the bacterium (Liu *et al.*, 2012). Although thiolase activity has been demonstrated in *E. coli*, no detailed structural studies of the related enzymes have been reported (Anbazhagan *et al.*, 2014; Kursula *et al.*, 2005). In the present study, a gene coding for thiolase activity was identified in *E. coli*. The gene was cloned and expressed, and the resulting protein was purified and crystallized. The X-ray crystal structure of the enzyme was determined at 1.8 Å resolution. Comparative analysis showed that it is homologous to the well studied *Zoogloea ramigera* thiolase (Kursula *et al.*, 2002). However, enzyme-kinetic studies revealed that unlike *Z. ramigera* thiolase, the *E. coli* enzyme is more efficient with respect to the degradative reaction. Other significant differences from the *Z. ramigera* thiolase include the post-translational modification of a lysine residue in the interface region and the nature of the water network in the vicinity of the active site. The crystal structure of the *E. coli* thiolase obtained in the presence of acetyl-CoA was determined at 1.9 Å resolution. The active-site Cys88 was found to be oxidized in this structure and only CoA was bound at the active site. As in the previously reported structure of a *Mycobacterium smegmatis* thiolase (Janardan *et al.*, 2015), both the unliganded and liganded tetrameric *E. coli* thiolases

were also found to exhibit asymmetry in their tetrameric organization.

2. Materials and methods

2.1. PCR amplification, cloning, expression and purification

The thiolase gene was amplified from *E. coli* BL21 (DE3) cells by colony PCR with forward primer 5'-CGCTAGCAT-GAAAAATTGTGTCATCGTCAG-3' and reverse primer 3'-GCGCGGATCCTTATTAATTCAACCGTTCAATCAC-5'. The primers were designed to clone the gene coding for the AIF37478 protein of *E. coli* KLY, which is annotated as a thiolase. The 50 µl reaction mixture consisted of 1× Phusion reaction buffer, 1 mM MgCl₂, 0.1 mM DNTs, 0.01 pM of the forward and reverse primers and 1 U of Phusion DNA Polymerase. A gradient PCR reaction was performed using a GeneAmp PCR System 2700 supplied by Applied Biosystems. The reaction was carried out at 98°C for 5 min followed by 30 s denaturation, annealing starting from 54°C with gradual increments of 0.15°C per cycle for 15 s, and extension at 72°C for 45 s. After 30 cycles of amplification, a final round of extension was performed at 72°C for 7 min, after which the reaction products were cooled to 4°C. The amplified fragment was purified using a PCR Clean-Up kit (AMPC002A) procured from Amnion. The gene fragment was ligated between the NheI and BamHI sites of pRSET-C vector, which introduced an N-terminal hexahistidine tag into the ampicillin-resistant construct. The ligated plasmid was transferred by heat shock into *E. coli* TOP-10 cells. The transformed cells were plated on 2% LB-agar containing 100 µg ml⁻¹ ampicillin and grown overnight at 37°C. A single colony was selected from the plate, transferred to 5 ml LB broth and grown for 4 h. The cells were pelleted by centrifugation and the plasmids isolated from the collected cells were screened for the presence of the insert by electrophoresis on 0.8% agarose gel. A positive clone was selected on the basis of an observed band shift with respect to pRSET-C. The clone was further confirmed by insert release upon restriction digestion, PCR amplification and DNA sequencing (data not shown). The clone was transferred into the BL21 pLysS/Rosetta expression system and was grown overnight on 2% LB-agar containing 100 µg ml⁻¹ ampicillin. A few colonies from the plate were inoculated into 5 ml LB-ampicillin broth and allowed to grow at 37°C for 6 h. Protein expression was induced by the addition of 1 mM IPTG. After induction, the culture fractions were collected at intervals of 2 h for 6 h. Protein expression was examined on 12% SDS-PAGE (Laemmli, 1970) using the uninduced fraction as a control. About 2% of the primary inoculum grown overnight was added to 3 l of 4% LB containing 100 µg ml⁻¹ ampicillin and 10%(v/v) glycerol and was allowed to grow at 37°C in an incubator shaker at 160 rev min⁻¹ until the optical density at 600 nm reached about 0.6. Protein expression was then induced by the addition of 0.1 mM IPTG. Subsequently, the temperature was lowered to 15°C and cells were allowed to grow for a further period of 12 h. The cells were recovered by centrifugation at 5500 rev min⁻¹ for 20 min in an SS-34 rotor and were

Table 1
Enzyme-kinetic parameters.

Organism	Acetoacetyl-CoA		Acetyl-CoA		CoA
	K_m (μM)	k_{cat} (s^{-1})	K_m (μM)	k_{cat} (s^{-1})	K_m (μM)
<i>Z. ramigera</i>	15	810	1200	71	9†
<i>E. faecalis</i>	88	1250	600	85	10‡
<i>Bradyrhizobium japonicum</i>	19	NA	100	NA	30‡
Chicken liver acetoacetyl-CoA thiolase	39	NA	270	NA	6.4‡
Mitochondrial pig heart degradative thiolase	3.6	8	300	0.8	13†
Human mitochondrial 3-ketoacyl-CoA thiolase (hT1)	9.2	14.8	250	1.4	NA§
<i>E. coli</i> thiolase	17.6 ± 3.4	220.5 ± 15	138.6 ± 19.4	6.5 ± 0.17	8 ± 2¶

† Modis & Wierenga (2000). ‡ Hedl *et al.* (2002). § Kiema *et al.* (2014). ¶ Present study.

resuspended in 240 ml lysis buffer [100 mM Tris pH 7.5, 400 mM NaCl, 1% Triton X-100, 10%(v/v) glycerol, 5 mM β -mercaptoethanol (β -ME), 5 μM PMSF]. The suspended cells were lysed by sonication using 3:7 on:off pulses at amplitude 32 for 7 min in an ice bath. After removing the cell debris by centrifugation, the supernatant was allowed to bind to Ni-NTA resin for 90 min at 4°C. The protein-bound resin was loaded into a Bio-Rad EconoPac column (catalogue No. 732-1010). The column was washed with 50 mM Tris pH 7.5 containing 200 mM NaCl, 10%(v/v) glycerol and 5 mM β -ME (wash buffer) to elute unbound proteins. Loosely bound proteins were eluted with 50 ml wash buffer containing 50 mM imidazole. The bound recombinant protein was eluted from the column using 10 ml wash buffer containing 250 mM imidazole. The eluted protein was examined by SDS-PAGE and concentrated using a Centricon at 2500 rev min⁻¹ at 4°C using a 30 kDa cutoff filter. The isolated protein was further purified by size-exclusion chromatography using a 120 ml Superdex 200 prep-grade 16/60 preparative FPLC (GE Healthcare) column placed inside a Sanyo LabCool cabinet. The column buffer (4°C) consisted of 25 mM Tris pH 7.5, 100 mM NaCl, 10%(v/v) glycerol, 5 mM β -ME. The protein obtained after size-exclusion chromatography was concentrated to 15 mg ml⁻¹ using a 30 kDa molecular-weight cutoff Centricon (Amicon Ultra-15 Centrifugal Filter Unit, Millipore).

2.2. Size-exclusion chromatography

The purified protein (125 μ l) at a concentration of 15 mg ml⁻¹ was applied onto a 24 ml Superdex 200 10/300 GL analytical column to determine the oligomeric state of the protein. The protein and standard protein molecular-weight markers (Bio-Rad, catalogue No. 151-1901) were eluted using 25 mM Tris pH 7.5, 100 mM NaCl, 10%(v/v) glycerol, 5 mM β -ME with a flow rate of 0.3 ml min⁻¹ at 4°C. From the elution profile, a standard plot of log(molecular weight) *versus* elution volume (V_e) was obtained. The molecular weight of the oligomeric species was estimated using the standard plot.

2.3. Mass spectrometry and peptide mass fingerprinting

The accurate molecular weight of a protomer of the *E. coli* thiolase was determined by MALDI mass spectrometry. As a

further attempt to confirm the identity of the protein, the band corresponding to the protein after separation by SDS-PAGE was subjected to in-gel trypsin digestion (Shevchenko *et al.*, 2007). The digest was subjected to peptide mass fingerprinting (PMF) in MALDI mass spectrometry. The data pertaining to 52 peaks were submitted for analysis with the Mascot web server by setting carbamidomethylation (C) and oxidation (M) as fixed and variable parameters, respectively.

2.4. Enzymatic studies

The degradative activity of the enzyme with acetoacetyl-CoA as the substrate was estimated at 30°C by monitoring the depletion of the Mg²⁺-acetoacetyl-CoA complex, which has an absorption peak at 303 nm, for 1 min using a Jasco V-630 spectrophotometer. The reaction was initiated by adding 10 ng of the enzyme to 200 μ l of a reaction cocktail consisting of 50 mM Tris pH 7.8, 10 mM MgCl₂, 0.1 mM CoA and variable concentrations of acetoacetyl-CoA (5–70 μM). K_m , V_{max} and k_{cat} were obtained by fitting the data to the Michaelis–Menten equation. In addition, to estimate the kinetic constants for CoA, the acetoacetyl-CoA concentration was kept at a constant value of 40 μM and the concentration of CoA was varied (0.5–45 μM). The enzyme activity was estimated assuming the extinction coefficient (ϵ) of Mg²⁺-acetoacetyl-CoA to be 21 400 M⁻¹ cm⁻¹ and the activity was expressed in units corresponding to 1 μM substrate catalyzed by the enzyme per minute under the given condition (Kiema *et al.*, 2014).

Synthesis of acetoacetyl-CoA was estimated using a coupled enzyme assay by monitoring the conversion of acetoacetyl-CoA to butyryl-CoA by short-chain L-3-hydroxyacyl-CoA dehydrogenase (scHAD) in the presence of NADH (Barycki *et al.*, 1999). The biosynthetic reaction was initiated by the addition of 200 ng enzyme to a reaction mixture consisting of 50 mM Tris pH 7.8, 0.1 mM dithiothreitol, 0.1 mM NADH, 40 mM KCl, 1 U scHAD (one unit is the amount of enzyme that catalyzes the synthesis of 1 μM of acetoacetyl-CoA per minute) and variable concentrations of acetyl-CoA (0.1–3 mM). The reaction was monitored at 340 nm at 30°C for 1 min. The enzyme activity was calculated assuming the extinction coefficient of NADH (ϵ) to be 6220 M⁻¹ cm⁻¹. The results of the enzymatic assays are provided in Table 1.

Table 2

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Thiolase, apo (PDB entry 5f0v)	Thiolase, CoA complex (PDB entry 5f38)
Data collection		
Space group	C2	$P2_12_12_1$
Wavelength (Å)	0.95	1.54
Resolution (Å)	35.65–1.89 (1.9–1.8)	36.86–1.90 (1.94–1.90)
Unit-cell parameters (Å, °)	$a = 190.93$, $b = 75.31$, $c = 147.41$, $\alpha = \gamma = 90.0$, $\beta = 131.4$	$a = 73.72$, $b = 76.03$, $c = 266.42$, $\alpha = \beta = \gamma = 90.0$
Observed reflections	548625 (61588)	1904480 (72292)
Unique reflections	135774 (17322)	150897 (7148)
Data completeness (%)	99.4 (97.2)	99.5 (96.2)
$\langle I/\sigma(I) \rangle$	7.1 (2.5)	6.2 (2.1)
Multiplicity	4.6 (4.3)	12.6 (10)
$R_{\text{merge}}^\dagger$ (%)	14.8 (42.0)	11.7 (79.4)
Protomers in the asymmetric unit	4	4
Refinement		
Wilson B factor (Å ²)	18.9	17.1
Average B factor (Å ²)		
Protein atoms	17	11.86
Ligands (CoA)	—	35
Water	37	46
No. of molecules		
Amino acids	1579	1577
Ligands	17	49
Waters	1134	1125
$R_{\text{work}}^\ddagger$ (%)	16.1	19.6
$R_{\text{free}}^\S$ (%)	19.6	23.5
R.m.s.d., bonds (Å)	0.018	0.016
R.m.s.d., angles (°)	1.78	1.71
Ramachandran plot		
Favoured (%)	96.7	95.9
Allowed (%)	3.0	3.64
Outliers (%)	0.3	0.46

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i th measurement of reflection hkl and $\langle I(hkl) \rangle$ is its mean intensity. $^\ddagger R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. $^\S R_{\text{free}}$ is calculated using a randomly selected subset of the R_{work} data (5%) which were excluded from refinement.

2.5. Crystallization and data collection

Initial crystallization trials were carried out using kits supplied by Hampton Research and Jena Bioscience. The *E. coli* thiolase was found to crystallize under the following conditions: (i) PEG/Ion condition No. 3 [0.2 M ammonium fluoride, 20% (w/v) PEG 3350] and (ii) PEGRx 1 condition No. 39 [0.1 M bis-tris propane pH 9.0, 30% PEG 6000]. Subsequently, condition (ii) was optimized to 0.1 M bis-tris propane pH 8.0, 35% (w/v) PEG 6000 with additives such as LiCl, 1,5-diaminopentane or guanidinium chloride. The optimized condition (ii) gave crystals that diffracted to reasonably high resolution. The crystals remained stable when soaked in buffers containing substrates such as CoA, acetyl-CoA and acetoacetyl-CoA (data not shown). The crystals obtained under condition (i) also diffracted to high resolution. A crystal of approximately 0.1 mm in each dimension obtained under condition (i) was mounted on a cryo-loop and flash-cooled in a stream of nitrogen gas at 100 K. The cooled crystal was exposed to X-rays from a Rigaku RU-200 rotating-anode X-ray generator. X-ray diffraction data were collected as 300

frames to ~ 2.0 Å resolution using a MAR DTB235 imaging-plate detector with an exposure time of 300 s per frame and an oscillation angle of 1° . The crystal-to-detector distance was set to 150 mm. Examination of the X-ray diffraction data processed and scaled using *iMosflm* 7.1.1 revealed that the crystal belonged to the monoclinic space group $C2$. The same crystal was sent to ESRF, Grenoble, France for data collection on beamline BM14. The data obtained at the synchrotron were processed to 1.8 Å resolution using *DENZO* and scaled with *SCALA* from the *CCP4i* suite (Winn *et al.*, 2011). Ligand-bound *E. coli* thiolase crystals could only be obtained by co-crystallization of the protein with the ligand. X-ray diffraction data from a crystal obtained under optimized condition (ii) containing 2 mM acetyl-CoA was collected using home-source X-rays. The data obtained were processed to 1.9 Å resolution using *XDS* (Kabsch, 2010). The crystal belonged to space group $P2_12_12_1$. The X-ray diffraction data statistics are summarized in Table 2.

2.6. Structure determination and refinement

The crystal structure of thiolase in space group $C2$ corresponding to the data collected at the home source was determined first. Initial phases of reflections were obtained by molecular replacement using the 2.0 Å resolution structure of the *Clostridium difficile* thiolase (PDB entry 4e1l; Center for Structural Genomics of Infectious Diseases, unpublished work), with which the *E. coli* thiolase shares 60% amino-acid sequence identity, as the phasing model. Molecular-replacement computations were performed using the molecular-replacement program *Phaser* (McCoy *et al.*, 2007). The electron-density map obtained was used for model building with the *AutoBuild* option of the *PHENIX* suite (Adams *et al.*, 2010). The model was further manually adjusted into the density using *Coot* (Emsley *et al.*, 2010). The structure was subjected to rigid-body refinement followed by restrained refinement with *REFMAC5* (Winn *et al.*, 2001). The geometry of the final structure was examined using the various options available in *Coot*. The final structure was used for refinement against the data collected at the synchrotron. It was also used as the phasing model in molecular replacement to determine the structure of the crystal obtained in the presence of acetyl-CoA, which belonged to a different space group ($P2_12_12_1$). The refinement statistics are presented in Table 2.

2.7. Sequence and structure analysis

Pairwise superposition of the structures was achieved using the *SSM* protocol (Krissinel & Henrick, 2004) available in *Coot* and the r.m.s.d.s of corresponding C^α atoms were recorded. The structures were also examined using the *DALI* server (Holm & Rosenström, 2010), *Chimera* (Pettersen *et al.*, 2004) and *PyMOL* (Schrödinger).

3. Results and discussion

3.1. Cloning, expression and purification

A thiolase-coding gene from *E. coli* was successfully cloned into the pRSET-C vector between *NheI* and *BamHI* sites. The

expressed protein was purified by Ni-NTA and gel-filtration chromatography to apparent homogeneity. Analytical size-exclusion chromatography revealed that the single peak eluting at about 13 ml corresponds to a protein of molecular weight 158 kDa, suggesting that the *E. coli* thiolase is a tetramer in solution (data not shown).

3.2. Enzymatic properties

The results of the kinetic experiments obtained for the *E. coli* thiolase are compared with the kinetic properties of well studied thiolases from other sources in Table 1. From the kinetic data, it is clear that the *E. coli* enzyme catalyzes both the synthetic and degradative reactions, although the efficiency is much higher in the biodegradative direction. The kinetic constants are closer to those of the human mitochondrial 3-ketoacyl-CoA thiolase, which is a degradative enzyme, than to those of the *Z. ramigera* thiolase, which functions as a biosynthetic enzyme. Therefore, the *E. coli* enzyme could be considered to be a biodegradative thiolase. The structural features presented later are also consistent with this enzyme functioning in the biodegradative direction.

3.3. Mass spectrometry and peptide mass fingerprinting (PMF)

The present *E. coli* thiolase consists of 394 amino acids with a calculated molecular weight (MW) of 40 352 Da. However, the expressed protein with four additional residues from the vector and a hexahistidine tag has a calculated molecular weight of 41 951 Da. In mass-spectrometric analysis, the peak observed corresponds to a MW of 42 004 Da (Supplementary Fig. S1, inset) and approximately matches that of the *E. coli* thiolase with these additional residues. The expressed protein was subjected to peptide mass fingerprinting (PMF; Supplementary Fig. S1). Peaks corresponding to about 52 cleaved peptides were observed, of which 21 peptide masses were in agreement with the *E. coli* thiolase sequence (Supplementary Table S1). The agreement of the masses of the PMF peaks with those of tryptic peptides of the *E. coli* thiolase confirmed that the expressed protein is indeed the protein annotated as an *E. coli* thiolase.

3.4. Structure determination and refinement

The crystals of the apo form of *E. coli* thiolase belonged to the monoclinic space group *C2* (Table 2; Supplementary Fig.

S2). Calculation of the Matthews coefficient suggested that the crystal asymmetric unit is likely to contain four protomers, corresponding to a calculated Matthews coefficient of $2.43 \text{ \AA}^3 \text{ Da}^{-1}$. Molecular-replacement calculations with the data collected using home-source X-rays successfully placed four copies (chains A–D) of the *C. difficile* thiolase protomer in the asymmetric unit of the *E. coli* thiolase cell. The structure determined was refined at a resolution of $2.0 \text{ \AA}$. The same crystal diffracted to a higher resolution of $1.8 \text{ \AA}$ when beamline BM14 of ESRF, Grenoble was used to collect X-ray diffraction data. However, the synchrotron data processed with *iMosflm* were indexed using related but different basis vectors that corresponded to space group *I2*. Refinement with synchrotron data was carried out after reindexing so as to be consistent with the home-source data. The quality of the final map is illustrated in Fig. 1. In the final map, significant electron density was not observed for the two C-terminal residues and residues Arg208, Lys209 and Lys210. In chain D, density that could correspond to residues Asp237, Lys238 and Ala239 was weak. In other chains, these residues had significant density.

X-ray diffraction data extending to $1.9 \text{ \AA}$ resolution were recorded from a crystal obtained in the presence of acetyl-CoA using home-source X-rays. The crystals belonged to space group *P2₁2₁2₁*. Calculation of the Matthews coefficient suggested that the asymmetric unit of the crystal could contain four protomers, with a Matthews coefficient of $2.3 \text{ \AA}^3 \text{ Da}^{-1}$. The structure of the complex was determined using the apo structure as the phasing model in the molecular-replacement program *Phaser* from the *CCP4* suite (Winn *et al.*, 2011). The structure was refined using *REFMAC5* (Table 2). Including TLS parameters with local NCS constraints during the final cycles of refinement improved *R* and *R_{free}*. As in the native structure, electron density was not significant for the two C-terminal residues and residues Val206–Phe212 and Lys238–Gly240 in chain A, Lys209 in chain B and Arg208–Lys210 in chain C. In addition, there was no significant density for residues Gly170–Met175 in chain A. The statistics of data collection and refinement pertaining to the two data sets are listed in Table 1.

3.5. Protomer structure of *E. coli* thiolase

The protomeric structure and quaternary organization of *E. coli* thiolase are illustrated in Figs. 2(a) and 2(b). The structures of the four protomers (A–D) in the *C2* asymmetric

Figure 1
Electron-density ($2F_o - F_c$) map corresponding to residues 80–86. The map is contoured at 1.6σ .

unit of unliganded *E. coli* thiolase are very similar, with an r.m.s. deviation of less than 0.5 Å in the positions of corresponding C α atoms upon pairwise structural superposition. The four protomers of the asymmetric unit constitute the functional *E. coli* thiolase enzyme. Each protomer of *E. coli* thiolase consists of three domains: an N-terminal domain (residues 1–117 and 253–261), a loop domain (residues 118–252) and a C-terminal domain (residues 273–393). The nomenclature for the different segments of the protomer used previously to describe the structure of the *Z. ramigera* thiolase (Kursula *et al.*, 2002) are also used here. The N-terminal and C-terminal domains have a similar $\beta\alpha\beta\alpha\beta\beta$ secondary-structure topology and are interconnected by the loop domain.

The N β 5 strand appears between N β 1 and N β 4, forming a mixed five-stranded β -pleated sheet with strand order β 1– β 5– β 4– β 2– β 3. A similar topology is also found in the C-terminal domain. However, in the C-terminal domain C β 3 is only weakly hydrogen-bonded to C β 2, as in the human (hT1) and *Z. ramigera* (PDB entry 1dm3) thiolases (Kiema *et al.*, 2014). The loop domain is also involved in the formation of the tetramerization loop or domain (residues 122–142) that stabilizes the oligomeric structure of the enzyme. Helices N α 3 and C α 3 are between the β -sheets of the N-terminal and C-terminal domains. The β -sheets of the two domains are also covered by α -helices on the side facing the bulk solvent, giving the protomer a five-layered appearance (Fig. 2*a*). The

Figure 2

(*a*) Cartoon representation of *E. coli* thiolase. (*a*) Unliganded protomer. The N-terminal, loop and C-terminal domains are shown in pink, grey and cyan, respectively. N α 5 penetrating between N α 1 and N α 4 is shown in yellow. (*b*) Unliganded tetramer. Chains A–D are coloured cyan, green, yellow and pink, respectively. (*c*) The B subunit in surface representation with the bound ligand shown in blue stick representation. Blue and red represent positively and negatively charged atoms, while yellow represents uncharged atoms. (*d*) Structural superposition of the present structure and the *E. coli* thiolase (PDB entry 4wys) reported by Kim *et al.* (2015). An enlarged view of the ligand bound to the B subunit is also shown. Arg134 and the ligand in the liganded forms are illustrated in stick representation. In the liganded structure, Arg134 interacts with the phosphates of the ligand.

loop domain has been shown to define the active-site geometry and substrate specificity and to play a crucial role in CoA binding and catalysis (Haapalainen *et al.*, 2007).

Compared with the N-terminal and C-terminal domains, the *B* factors associated with the loop domain show larger variations (Supplementary Fig. S3). Apart from the loop domain, the *B* factors associated with residues 39–47, 260–272, 300–309, 327–335 and 368–372 are also higher when compared with the residues adjacent to these segments. Strong electron-density evidence was found for post-translational modification of Lys86 to *N*^ε-dimethyllysine in all subunits of both the native and CoA-bound structures. Such a modification has not been observed in any of the published thiolase structures. Lys86 and Lys93 of one protomer are close to Glu50 of a neighbouring protomer that occurs at the interface of tight dimers (*AB* and *CD*). Since there is no other negatively charged residue close to these residues, the modification of Lys86 may be necessary for the stability of the dimeric interface.

3.6. Liganded structure of *E. coli* thiolase

The liganded structure of the *E. coli* thiolase, which crystallized in space group $P2_12_12_1$, also contains a tetramer in the asymmetric unit. The protomer structure is not altered significantly by ligand binding. The r.m.s. deviation on pairwise structural superposition of protomers selected from liganded and unliganded structures varies between 0.2 and 0.5 Å. Arg134 is observed to be in the disallowed region of the Ramachandran plot (Ramachandran & Sasisekharan, 1968) in both the unliganded and liganded structures. The residue is exposed to the solvent and interacts with the pantetheine moiety of CoA.

It was observed that the active sites of the *B*, *C* and *D* subunits of the liganded structure had noncovalently bound CoA and the active-site Cys88 was transformed to an oxidized form in all of the subunits. CoA could be fully built only in protomer *B* (Figs. 2*c* and 2*d*), while it had to be truncated (the adenosine moiety was not modelled) owing to a lack of adequate density in the *C* and *D* protomers. The ligand could not be modelled in the *A* chain. Residues 222–240 were also partially disordered in the *A* chain and the associated *B* factors were high for this segment (Supplementary Fig. S3). The disorder of the 222–240 loop might be the reason why CoA is not bound to the *A* subunit. These residues are not close to other molecules related by crystallographic symmetry only in the *A* subunit, while they appear to be stabilized by intermolecular interactions in the *B*, *C* and *D* subunits.

3.7. Comparison of the present structure with the previously determined *E. coli* thiolase structure with PDB code 4wys

During the preparation of this manuscript for publication, the structure of *Clostridium acetobutylicum* thiolase was published (Kim *et al.*, 2015). This publication also included a short reference to an *E. coli* thiolase structure determined by the same authors. The structure has been deposited in the PDB as entry 4wys. Interestingly, in the present study the unliganded *E. coli* thiolase crystallized in the monoclinic space

group *C2*, while PDB entry 4wys, which is also unliganded, belongs to the orthorhombic space group $P2_12_12_1$. The unit-cell parameters of 4wys are significantly different from the unit-cell parameters of our *E. coli* thiolase crystals obtained in the presence of CoA, which also belonged to space group $P2_12_12_1$. Structural superposition of the present thiolase and PDB entry 4wys shows that the polypeptide folds are very similar (Fig. 2*d*). The r.m.s. deviation between equivalent C^α atoms of selected pairs of subunits upon structural superposition is in the range 0.25–0.5 Å. A couple of residues at the N-terminus of PDB entry 4wys are disordered, while in the present structure not only is the N-terminus fully ordered but the densities for an additional two residues resulting from the cloning strategy are also visible. The largest shifts are seen for the tetramerization domain, the adenine-binding loop and the cationic loop. With these changes, the active-site groove appears to be narrower in the present structure. The side chain of Arg134 is oriented in very different directions in the two structures. This residue has a conformation that is appropriate for interaction with the phosphate group of the bound ligand of an opposing subunit only in the present structure. The post-translational modification of Lys86 observed in the present structure is also not observed in 4wys.

3.8. Quaternary structure of *E. coli* thiolase

Gel-filtration chromatography as well as the crystal structure suggest that the functional unit of the *E. coli* thiolase is a tetramer. In the crystal structure, the tetramer appears as a loose dimer of tight dimers, as observed for the *Z. ramigera* and other tetrameric thiolase structures (Kursula *et al.*, 2005). Analysis using PISA revealed that the accessible surface area (ASA) of the *E. coli* thiolase protomer is 15 410 Å², while that of the tetramer is 47 355 Å². The area that becomes buried owing to the association of the *AB* and *CD* protomers that constitute the tight dimers is 9958 Å². This interface is very stable owing to the presence of about 40 hydrogen bonds and eight salt bridges. Nβ3 (residues 81–85; GFTVN) in the tight dimer interface is antiparallel to the Nβ3 strand of the adjacent protomer such that the two protomers are held together strongly by hydrogen bonding and salt bridges. The modified Lys86 is situated close to the interface as well as to the active site and makes contact with the neighbouring subunit. This modification is likely to increase the stability of the tight dimeric interface. The area that becomes buried owing to the formation of the weak interfaces of *AC* and *BD* is 1112 Å², while the interface area that becomes buried between the *AD* and *BC* subunits is 2586 Å². These minor interface areas suggest that the two tight dimers are only loosely associated. The total surface area that becomes buried on tetramerization is 13 656 Å². There are significant differences between the tight interfaces of the *E. coli* and *Z. ramigera* thiolases. Residues 81–85 (GFTVN) constituting Nβ3 at the tight dimeric interface of the *E. coli* thiolase are not conserved in the *Z. ramigera* thiolase (TAWGM). The only conservative substitution found in these residues is Phe82 of the *E. coli* thiolase, which is equivalent to Trp83 of the *Z. ramigera* thiolase.

The tetrameric structure of the *E. coli* thiolase has approximate 222 symmetry. As evaluated by the *SSM superpose* program in *Coot*, protomers *A* and *B* forming a tight dimer are related by a rotation of 179.7° (axis *P*). In contrast, protomers *A* and *C* (axis *Q*) are related by a rotation of 173.4° and protomers *A* and *D* (axis *R*) are related by a rotation of 177.4°. Thus, large departures from exact twofold rotation (180°) are observed for the weak dimers. In the liganded structure, the same angles (*A/B*, *A/C* and *A/D*) are 179.5, 176.0 and 176.5°, respectively (Table 3). These observations suggest a departure from exact 222 symmetry in both the native and liganded structures. Superposition of *A* on *B* leads to superposition of *B* on *A*, as expected for a dimer with nearly exact twofold symmetry. This rotation does not lead to superposition of *C* on *D* or of *D* on *C*. However, superposition of *C* on *D* leads to superposition of *D* on *C*, suggesting that the *C* and *D* subunits constituting the second tight dimer are also related by an exact twofold symmetry, similar to subunits *A* and *B*. This implies that the axis relating *A* and *B* (*P*1) and the axis relating *C* and *D* (*P*2) are not coincident. In contrast, the twofolds relating *A* and *C* (*Q*1) and *B* and *D* (*Q*2) are nearly coincident. Similarly, the twofolds relating *A* and *D* (*R*1) and *B* and *C* (*R*2) are nearly coincident. Thus, the departure from exact 222 symmetry arises mainly because the twofolds relating the two tight dimers, although exact, are not coincident. Since the tight dimers share a large interface and are stabilized by a large number of interactions, it is not surprising that the twofold symmetry relating *A* and *B* as well as *C* and *D* is not disturbed owing to crystal-packing forces. However, since the other interfaces are not as strong, it is likely that they are perturbed by crystal-packing forces. Interestingly, the tetrameric organization in the *E. coli* thiolase with PDB code 4wys determined previously is much closer to 222 symmetry.

3.9. Asymmetry in the quaternary organization of thiolases from other sources

The observations presented in the previous section prompted a systematic analysis of the departure from exact 222 symmetry in thiolase structures from other organisms. Table 3 lists the rotational angles required for the structural superposition of pairs of protomers related by the *P* (*A/B* and *C/D*), *Q* (*A/C* and *B/D*) and *R* (*A/D* and *B/C*) axes and the angle between the *P*1/*P*2, *Q*1/*Q*2 and *R*1/*R*2 directions. To obtain Table 3, the nomenclature used to designate the protomers of human T2 thiolase (PDB entry 2ib7; Haapalainen *et al.*, 2007) had to be changed appropriately so as to be consistent with the nomenclature used to describe the protomers of the present structure. As expected, it was found that the subunits sharing the large interface corresponding to the *P* axis are invariably related by a rotation of close to 180°, suggesting that the crystal-packing forces do not distort the tight dimers. The largest departure from 180° is observed for the *Q* axis that relates subunits *A* and *B* to *C* and *D*, respectively. The departure from collinearity of the *P*1 and *P*2 axes could be as large as 10°. This distortion leads to a large difference between the active sites of *A* and *B* on one hand and *C* and *D* on the other. As observed in other liganded

structures, the residue Arg134 is involved in anchoring the phosphate moiety of the substrate. This is in a suitable position and orientation to interact with the phosphates in the *C* and *D* subunits but not in the *A* and *B* subunits. Therefore, the asymmetry reported in the structure of *M. smegmatis* (Janardan *et al.*, 2015) thiolase might also be an artifact of crystal packing. It is plausible that all of the subunits of *M. smegmatis* thiolase are liganded in solution when incubated with acetyl-CoA. However, after crystallization, the asymmetry induced in the quaternary structure may decrease the binding affinity in the *A* and *C* subunits, inducing the bound ligand to become detached and diffuse into the solvent. However, it should be noted that similar asymmetry was observed in three different crystal forms of *M. smegmatis* thiolase with very different packing arrangement of molecules. Therefore, the observed asymmetry could be intrinsic and could have functional significance. These observations clearly warrant further investigations.

3.10. Active-site geometry

The mode of CoA binding to the *B* subunit of *E. coli* thiolase is illustrated in Fig. 3. By comparison with other well characterized thiolases, residues 122–124, 146–156, 201–218, 219–229 and 241–260 of the loop domain could be identified as constituting the tetramerization domain, covering loop, cationic loop, adenine-binding loop and pantetheine loop, respectively. Most of the residues in these segments are known to interact with the substrate. The important catalytic residues Cys88, Asn317, His349 and Cys379 are found in the four active-site motifs Nβ3–Nα3, Cβ2–Cα2, Cβ3–Cα3 and Cβ4–Cβ5 and correspond to the fingerprint sequences CXS, NEAF, GHP and CXG, respectively. These fingerprint residues are highly conserved among all members of the biosynthetic/degradative thiolase superfamily (Anbazhagan *et al.*, 2014). All of these residues are also conserved in *E. coli* thiolase.

Although the *E. coli* thiolase was crystallized in the presence of acetyl-CoA, the final electron-density map revealed density only for CoA. Also, the electron density suggested that the active-site Cys88 may have been oxidized to its sulfenic acid form. It is likely that the high pH of crystallization resulted in the hydrolysis of acetyl-CoA and acetylation of the active-site cysteine. The modified cysteine might have subsequently become oxidized to sulfenic acid. In contrast, such oxidation was not observed in the *Z. ramigera* thiolase (PDB entry 1dm3) as the protein was crystallized at the acidic pH of 5.0 (Modis & Wierenga, 2000).

The active-site pocket is hydrated in the absence of CoA. On CoA binding, 13 water molecules are expelled from the active site. The normalized *B* factors [$(B - \langle B \rangle)/\sigma(B)$] of the cationic loop (residues 206–213) are lower in the ligand-bound structure (Supplementary Fig. S3) when compared with the unliganded structure, suggesting that the cationic loop is involved in promoting and stabilizing the binding of the substrate (Modis & Wierenga, 1999). As observed in other thiolases, the binding of CoA is stabilized by a hydrogen-bonding network involving water molecules.

Table 3

Rotational relationship of pairs of subunits related by the molecular P (relating the A/B and C/D subunits), Q (relating the A/C and B/D subunits) and R (relating the A/D and B/C subunits) symmetry axes.

The values represent the rotation required for structural superposition of the subunits listed and the departure from linearity of the *P*, *Q* and *R* axes.

Protein	Rotation <i>A/B</i> (°)	Rotation <i>C/D</i> (°)	Angle between axes (°)	Rotation <i>A/C</i> (°)	Rotation <i>B/D</i> (°)	Angle between axes (°)	Rotation <i>A/D</i> (°)	Rotation <i>B/C</i> (°)	Angle between axes (°)
<i>E. coli</i> thiolase (present study)	179.7	179.6	7.13	173.4	173.3	0.84	177.4	177.4	0.32
<i>E. coli</i> thiolase (PDB entry 4wys)	179.5	179.9	1.20	178.8	178.8	0.23	179.6	179.8	0.34
<i>Mycobacterium tuberculosis</i> thiolase (<i>P6₁</i> , <i>ABCD</i>)	179.1	179.8	4.42	175.7	175.4	0.29	179.4	179.3	0.57
<i>M. tuberculosis</i> thiolase (<i>P6₁</i> , <i>EFGH</i>)	178.8	179.9	9.10	174.9	174.4	0.59	178.6	178.6	0.75
<i>M. tuberculosis</i> thiolase (<i>P6₁</i> , <i>JKLM</i>)	179.1	179.9	5.84	174.4	174.0	0.43	179.6	179.7	0.56
<i>M. tuberculosis</i> thiolase (<i>C222₁</i>)	179.4	179.7	0.41	178.9	178.4	0.30	179.9	179.3	0.53
<i>M. tuberculosis</i> thiolase (<i>P2₁</i>)	179.4	179.9	1.89	177.8	178.5	0.28	179.8	179.9	0.55
PDB entry 1dm3	179.9	179.0	1.15	178.3	179.7	0.13	178.4	178.8	0.14
PDB entry 1ulq	179.3	179.7	2.56	179.1	178.3	0.85	178.1	178.2	0.54
PDB entry 4c2k	179.1	179.9	0.65	179.7	179.4	0.10	179.3	179.9	0.41
PDB entry 4n44	179.2	179.4	0.83	179.8	178.8	0.33	179.4	179.6	0.73
PDB entry 4o99	179.9	179.7	3.40	178.9	178.6	0.08	176.9	176.8	0.27
PDB entry 1ibz	179.7	179.8	2.77	178.2	177.7	0.39	178.1	177.5	0.10

Arg220, Gln183, Thr224, Gly244 and Ser247 are some of the important residues near the active site that interact with the substrate in the *Z. ramigera* thiolase (Modis & Wierenga,

1999). These residues are structurally equivalent to the *E. coli* thiolase residues Lys221, Gln184, Ser224, Gly245 and Ser248, respectively. The equivalent residues are either identical or

Figure 3

(a) Binding of CoA at the active site and the surrounding residues. Dashed lines represent hydrogen bonds. Oxyanion hole OAH1 is stabilized by His349 and Wat18.

conservative substitutions. Thus, the interactions that mediate substrate binding are likely to be similar in the *E. coli* and *Z. ramigera* thiolases.

3.11. Comparative analysis of water networks

A conspicuous structural feature of thiolases is the arrangement of water networks, which differs in biosynthetic and degradative thiolases (Fig. 4), although how these water networks influence the catalytic properties is not well understood. Two hydrogen-bonded water networks have been identified in synthetic thiolases, one on either side of C β 2. One of these ordered water networks penetrates through the space between C α 2 and C β 2, leading to the active site at one end and bulk solvent at the other end, as observed in *Z. ramigera* thiolase and other biosynthetic thiolases. However, in degradative thiolases this network is limited to two water molecules. Careful examination of the *E. coli* thiolase structure shows that the water network is not extensive owing to the occurrence of a cluster of hydrophobic residues. Phe324 and Phe333 occupy the space between C α 2 and C β 2 and limit the number of water molecules to two, as in other biodegradative thiolases. The other water network is present in both biosynthetic and biodegradative thiolases. Thus, the hydration at the active site

of the *E. coli* thiolase is consistent with kinetic studies, which show that it is more efficient as a degradative enzyme.

3.12. Conservation of oxyanion holes and implications for the catalytic mechanism

It has been shown in the *Z. ramigera* thiolase that the enolate and tetrahedral intermediate transition states formed during catalysis are stabilized by two oxyanion holes, OAH1 and OAH2 (Kursula *et al.*, 2002; Haapalainen *et al.*, 2006; Janardan *et al.*, 2015). The residues constituting these oxyanion holes are also conserved in the *E. coli* thiolase. OAH1 is formed by His349 (N ϵ^2) and Wat18. OAH2 is formed by the backbone NH group of Cys88 and Gly381 (distance of 3.2 Å) along with Wat17 (Fig. 5). In the synthetic direction, two water molecules, equivalent to Wat17 and Wat18 in the *E. coli* thiolase, have been proposed to be important for initiation of the condensation reaction. In the liganded structure of the *E. coli* thiolase it is observed that these water molecules move towards Asn317 of the NEAF motif. These similarities suggest that the mechanism of the synthetic reaction catalyzed by the *E. coli* enzyme might be similar to that of the *Z. ramigera* enzyme.

Figure 4

(a) The water network observed in the active site of the biosynthetic *Z. ramigera* thiolase (PDB entry 1dm3). (b) The water network in the *E. coli* thiolase. Phe324 and Phe333 interfere with the continuity of the hydrogen-bonded network of water molecules.

Figure 5
The oxyanion hole OAH2 is formed by the NH group of the oxidized Cys88, Gly381 and Wat17.

4. Conclusions

The X-ray crystal structure of the *E. coli* thiolase was determined both in the native form and in complex with acetyl-CoA. The liganded structure was obtained by co-crystallization with acetyl-CoA. In the liganded structure, it was observed that the active-site cysteine was oxidized to sulfenic acid and that free CoA was noncovalently held at the active site. However, the overall structure was not significantly altered by ligand binding. The post-translational modification of an interface lysine observed in the *E. coli* thiolase might contribute to the stability of the tight dimers. Enzyme-kinetic studies suggest that the *E. coli* thiolase can catalyze both the synthetic and degradative reactions, although the efficiency of the degradative reaction is much higher. The water network and interface region are two notable features that distinguish the *E. coli* thiolase from the well studied *Z. ramigera* thiolase. The water structure at the active site of the *E. coli* thiolase is consistent with its higher efficiency as a degradative enzyme. Compared with an independently determined structure of the same enzyme (PDB entry 4wys), the two present structures show a significant departure from exact 222 molecular symmetry. The significance, if any, of this departure from exact symmetry needs to be established by further work.

Acknowledgements

MRNM thanks DBT and DST and a Bose fellowship for financial support. MRNM and RK thank DBT for the sanction

of an Indo-Finnish grant. NJ thanks CSIR for supporting her work and IR thanks UGC for a Dr D. S. Kothari Postdoctoral Fellowship. The linker enzyme was a kind gift from Dr Joseph Barycki.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Anbazhagan, P., Harijan, R. K., Kiema, T. R., Janardan, N., Murthy, M. R. N., Michels, P. A., Juffer, A. H. & Wierenga, R. K. (2014). *Tuberculosis*, **94**, 405–412.
- Barycki, J. J., O'Brien, L. K., Bratt, J. M., Zhang, R., Sanishvili, R., Strauss, A. W. & Banaszak, L. J. (1999). *Biochemistry*, **38**, 5786–5798.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007). *Biochemistry*, **46**, 4305–4321.
- Haapalainen, A. M., Meriläinen, G. & Wierenga, R. K. (2006). *Trends Biochem. Sci.* **31**, 64–71.
- Hedl, M., Sutherland, A., Wilding, E. I., Mazzulla, M., McDevitt, D., Lane, P., Burgner, J. W. II, Lehnbeuter, K. R., Stauffacher, C. V., Gwynn, M. N. & Rodwell, V. W. (2002). *J. Bacteriol.* **184**, 2116–2122.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Janardan, N., Harijan, R. K., Kiema, T.-R., Wierenga, R. K. & Murthy, M. R. N. (2015). *Acta Cryst.* **D71**, 2479–2493.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kiema, T.-R., Harijan, R. K., Strozyk, M., Fukao, T., Alexson, S. E. H. & Wierenga, R. K. (2014). *Acta Cryst.* **D70**, 3212–3225.
- Kim, S., Jang, Y.-S., Ha, S.-C., Ahn, J.-W., Kim, E.-J., Lim, H. J., Cho, C., Ryu, S. Y., Lee, K. S., Lee, S. Y. & Kim, K.-J. (2015). *Nature Commun.* **6**, 8410.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Kursula, P., Ojala, J., Lambeir, A. M. & Wierenga, R. K. (2002). *Biochemistry*, **41**, 15543–15556.
- Kursula, P., Sikkilä, H., Fukao, T., Kondo, N. & Wierenga, R. K. (2005). *J. Mol. Biol.* **347**, 189–201.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Liu, H., Yu, C., Feng, D., Cheng, T., Meng, X., Liu, W., Zou, H. & Xian, M. (2012). *Microb. Cell Fact.* **11**, 41.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Meriläinen, G., Poikela, V., Kursula, P. & Wierenga, R. K. (2009). *Biochemistry*, **48**, 11011–11025.
- Modis, Y. & Wierenga, R. K. (1999). *Structure*, **7**, 1279–1290.
- Modis, Y. & Wierenga, R. K. (2000). *J. Mol. Biol.* **297**, 1171–1182.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
- Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V. & Mann, M. (2007). *Nature Protoc.* **1**, 2856–2860.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.