BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Structural characterization of a mitochondrial 3-ketoacyl-CoA (T1)-like thiolase from *Mycobacterium smegmatis*

Neelanjana Janardan,[a] Rajesh K. Harijan,[b,c] Tiila-Riikka Kiema,[b] Rikkert K. Wierenga[b] and M. R. N. Murthy[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560 012, India, [b]Faculty of Biochemistry and Molecular Medicine, Biocenter Oulu, University of Oulu, FIN-90014 Oulu, Finland, and [c]Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA. *Correspondence e-mail: mrn@mbu.iisc.ernet.in

Received 6 August 2015
Accepted 12 October 2015

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: *Mycobacterium smegmatis*; thiolase; crystal structure; coenzyme A.

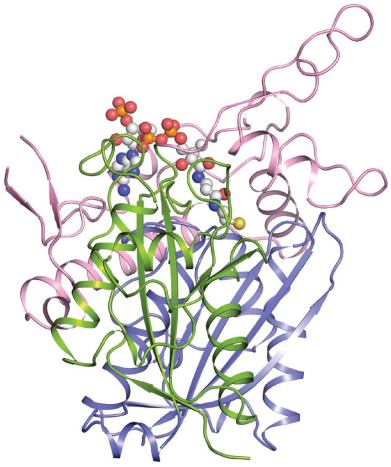
PDB references: MSM-13 thiolase, 4zrc; 5byv; 5cbq; complex with CoA, 5bz4

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

Thiolases catalyze the degradation and synthesis of 3-ketoacyl-CoA molecules. Here, the crystal structures of a T1-like thiolase (MSM-13 thiolase) from *Mycobacterium smegmatis* in apo and liganded forms are described. Systematic comparisons of six crystallographically independent unliganded MSM-13 thiolase tetramers (dimers of tight dimers) from three different crystal forms revealed that the two tight dimers are connected to a rigid tetramerization domain *via* flexible hinge regions, generating an asymmetric tetramer. In the liganded structure, CoA is bound to those subunits that are rotated towards the tip of the tetramerization loop of the opposing dimer, suggesting that this loop is important for substrate binding. The hinge regions responsible for this rotation occur near Val123 and Arg149. The Lα1–covering loop–Lα2 region, together with the Nβ2–Nα2 loop of the adjacent subunit, defines a specificity pocket that is larger and more polar than those of other tetrameric thiolases, suggesting that MSM-13 thiolase has a distinct substrate specificity. Consistent with this finding, only residual activity was detected with acetoacetyl-CoA as the substrate in the degradative direction. No activity was observed with acetyl-CoA in the synthetic direction. Structural comparisons with other well characterized thiolases suggest that MSM-13 thiolase is probably a degradative thiolase that is specific for 3-ketoacyl-CoA molecules with polar, bulky acyl chains.

1. Introduction

Thiolases are enzymes that are involved in fatty-acid metabolism. In the synthetic direction, they catalyze the formation of a carbon–carbon bond *via* a thioester-dependent Claisen condensation reaction (Heath & Rock, 2002). In the degradative direction, 3-ketoacyl-CoA is cleaved into acetyl-CoA and acyl-CoA. Despite low sequence identities, all thiolases share a remarkably similar polypeptide fold and perform similar functions. Thiolases are either dimers or tetramers (constituted by two tight dimers). In the catalytic cycle, a strictly conserved cysteine residue functions as a nucleophile and forms a covalent intermediate with the priming acyl-CoA substrate both in the biosynthetic and degradative reactions. In the biosynthetic direction, with acetyl-CoA as the priming substrate, the acetylated cysteine reacts with another acetyl-CoA molecule, leading to the formation of 3-ketobutyryl-CoA (Modis & Wierenga, 2000). In the degradative direction, with 3-ketoacyl-CoA as the priming substrate, acetyl-CoA is released and the acyl moiety is transferred to CoA.



The six characterized mammalian thiolases have distinct cellular distributions, quaternary structures, substrate specificities and enzyme-kinetic properties (Fukao *et al.*, 2002; Anbazhagan *et al.*, 2014). Cytosolic (CT) thiolase and mitochondrial acetoacetyl-CoA (T2) thiolase are short-chain-specific biosynthetic enzymes. CT thiolase is involved in cholesterol biosynthesis (Middleton, 1974), while T2 thiolase is involved in ketone-body metabolism (Staack *et al.*, 1978) and isoleucine catabolism (Middleton & Bartlett, 1983; Haapalainen *et al.*, 2007). The mitochondrial 3-ketoacyl-CoA (T1) thiolase (Middleton, 1973) and the thiolase dimer of the mitochondrial trifunctional enzyme (TFE; Uchida *et al.*, 1992) are degradative thiolases and catalyze the last step of the β-oxidation pathway. T1 thiolase is specific for medium-chain to long-chain unbranched fatty acids (Staack *et al.*, 1978; Mao *et al.*, 1995; Miyazawa *et al.*, 1981). TFE is specific for long-chain and 2-methyl-branched fatty acids (Mao *et al.*, 1995). Peroxisomal 3-ketoacyl-CoA (AB) thiolase (Antonenkov *et al.*, 1999; Bout *et al.*, 1991) is involved in the peroxisomal β-oxidation of very long chain unbranched fatty acids. SCP2 thiolase is also a peroxisomal enzyme that catalyzes a degradative reaction in bile-acid biosynthesis (Antonenkov *et al.*, 1997). The mammalian SCP2 thiolase has an additional sterol-carrier C-terminal domain (Seedorf *et al.*, 1994) of unknown function. CT, T1 and T2 thiolases are tetrameric, whereas AB, TFE and SCP2 thiolases are dimeric.

The most extensively studied thiolase is the biosynthetic enzyme from *Zoogloea ramigera* (*Z. ramigera* thiolase), which catalyzes the condensation of two acetyl-CoA molecules to give acetoacetyl-CoA (Modis & Wierenga, 2000; Williams *et al.*, 1992). *Z. ramigera* thiolase catalyzes the first step in the biosynthesis of polyhydroxybutyrate (also known as a polyhydroxyalkonaote), an important storage polyester molecule in bacteria (Nishimura *et al.*, 1978). *Z. ramigera* thiolase is a tetramer in which two tightly associated dimers are loosely connected *via* loops that extend from the core of the dimeric protomers. The catalytic site is located in a deep pocket lined by four catalytic loops protruding out of the N-terminal domain (the C*x*S loop) and the C-terminal domain (the NEAF, GHP and C*x*G loops; highlighted in Fig. 1) that contribute residues important for catalysis. Most biosynthetic thiolases characterized to date are tetrameric, while degradative thiolases can be either dimeric or tetrameric. However, the active-site features that differentiate the biosynthetic and degradative thiolases are not well understood.

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a deadly pathogen. Mycobacteria possess a unique and hardy fatty acid-rich cell wall, and fatty-acid metabolism plays an important role in the virulence of mycobacteria (Kotani *et al.*, 1970). Previous bioinformatics analyses have revealed eight and 13 different thiolases, in addition to one and two thiolase-like proteins, in the genomes of *M. tuberculosis* and *M. smegmatis*, respectively (Anbazhagan *et al.*, 2014). Structural and functional characterization of these thiolases and thiolase-like proteins might provide insights into the special features of fatty-acid metabolism in these microorganisms.

Here, we report the crystal structure of a T1 thiolase-like protein from *M. smegmatis* in both the apo and CoA-bound forms. This enzyme was named MSM-13 thiolase on the basis of nomenclature proposed in a previous study (Anbazhagan *et al.*, 2014). Comparison of the structural features and tetrameric organization of MSM-13 thiolase with those of other thiolases revealed that MSM-13 thiolase might function as a degradative thiolase with unique substrate specificity. Moreover, an inherent asymmetry in the tetrameric organization of MSM-13 thiolase was recognized which may have physiological significance.

2. Materials and methods

2.1. Bioinformatics analyses

All thiolase sequences were obtained from the National Center for Biotechnology Information (NCBI) database. The sequences of tetrameric thiolases of known structure were aligned using *ClustalW* (Higgins, 1994). The multiple sequence alignment thus obtained was used to generate a phylogenetic tree by the neighbour-joining method (10 000 bootstrap replicates) in *MEGA*5 (Kumar *et al.*, 1994).

2.2. Cloning, expression and purification of MSM-13 thiolase

The gene coding for MSM-13 thiolase was amplified from *M. smegmatis* genomic DNA using Phusion high-fidelity DNA polymerase (New England Biolabs, Massachusetts, USA) and specific sense (5′-CAT ATG GCT AGC ATG AGC ATG CGC GAC GC-3′) and antisense (5′-GGA TCC TTA CTC GAG TCC CTC TTG CAC CCG-3′) primers. The underlined sequences in the sense and antisense primers represent NheI and XhoI restriction sites, respectively. The amplified fragment was digested with NheI and XhoI (Thermo Scientific, Massachusetts, USA) and cloned into the pET-21b(+) vector (Novagen, Germany). The clone thus obtained was confirmed by DNA sequencing.

Recombinant protein was expressed in an *Escherichia coli* BL21 (DE3) strain transformed with the pGro7 plasmid (Takara Bio Inc., Japan), which encodes the GroEL and GroES chaperones. The transformants were selected on LB agar plates containing 100 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol. Pre-inoculum grown overnight in LB–ampicillin–chloramphenicol medium from a single colony was diluted into 3 l fresh LB medium containing 100 µg ml^{-1} ampicillin, 34 µg ml^{-1} chloramphenicol and 0.3 mg ml^{-1} arabinose. Cells were grown at 37°C until the OD_{600} reached 0.6–0.7. Protein expression was induced with 0.3 m*M* IPTG and the cells were allowed to grow for an additional 12 h at 15°C. The cells were recovered by centrifugation and resuspended in 150 ml cold lysis buffer consisting of 20 m*M* Tris pH 8.0, 400 m*M* NaCl, 10%(*v*/*v*) glycerol, 10 m*M* β-mercaptoethanol. The cell suspension was lysed by sonication on ice. All subsequent purification steps were performed at 4°C. Cell debris was removed from the lysate by centrifugation. An empty column was loaded with Ni^{2+}–NTA beads (bed volume 1 ml) pre-equilibrated with lysis buffer. Clear lysate was

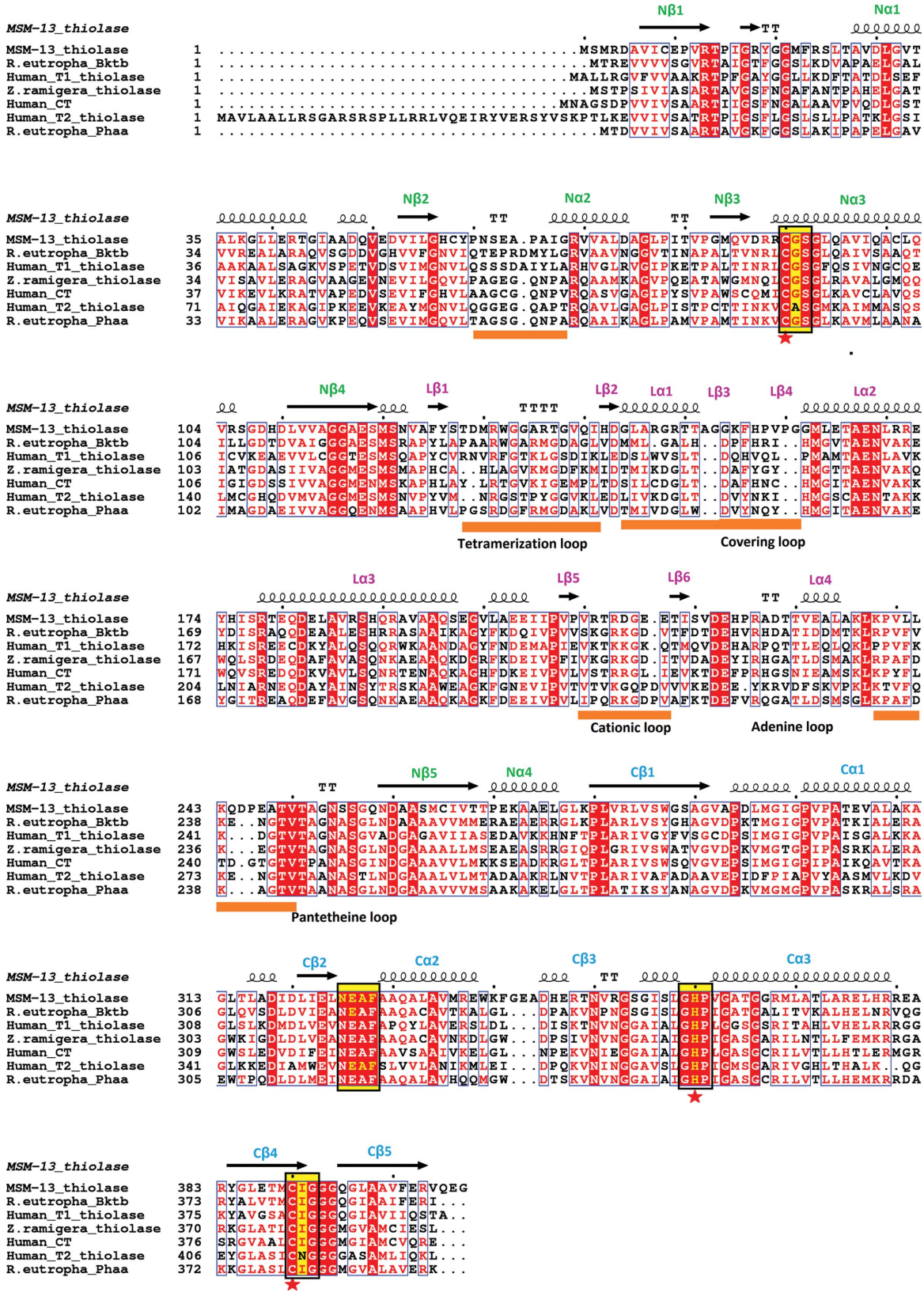


Figure 1
Sequence alignment of MSM-13 thiolase and other well characterized tetrameric thiolases of known structure. The following sequences were downloaded from the NCBI sequence database and aligned using *ClustalW*: MSM-13 thiolase (YP_886561.1), *R. eutropha* BktB (WP_010810126.1), human T1 thiolase (NP_006102.2), *Z. ramigera* thiolase (AAA27706.1), human CT thiolase (AAH00408.1), human T2 thiolase (P24752.1) and *R. eutropha* PhaA (WP_010810132.1). The secondary structure of MSM-13 thiolase is shown above the alignment (N-terminal domain, green; loop domain, pink; C-terminal domain, blue). The four catalytic loops are highlighted in yellow and the catalytic residues are marked with red stars. The active-site regions of MSM-13 that are structurally different from those of *Z. ramigera* thiolase are highlighted by orange bars.

Table 1
Crystallographic statistics of data collection and structure refinement.

Values in parentheses are for the highest resolution shell.

	Form I	Form II	Form III	Form II, with CoA
Beamline	BM14, ESRF	BM14, ESRF	Home source	Home source
Wavelength (Å)	0.9537	0.9537	1.5418	1.5418
Resolution range (Å)	54.8–2.4 (2.5–2.4)	45.1–2.1 (2.2–2.1)	69.0–2.7 (2.8–2.7)	47.5–2.4 (2.5–2.4)
Space group	$C222_1$	$P6_1$	$P12_11$	$P6_1$
Unit-cell parameters				
a (Å)	113.5	190.9	75.4	190.0
b (Å)	181.4	190.9	104.1	190.0
c (Å)	271.7	264.6	104.5	265.0
α (°)	90	90	90	90
β (°)	90	90	107.5	90
γ (°)	90	120	90	120
Total observations	173300 (14451)	564928 (55668)	83401 (7213)	205752 (19521)
Unique reflections	94222 (8598)	290242 (28891)	42124 (3859)	201832 (19346)
Multiplicity	1.8 (1.7)	1.9 (1.9)	2.0 (1.9)	1.0 (1.0)
Completeness (%)	91.8 (84.5)	99.9 (100.0)	100.0 (90.9)	98.9 (93.8)
Mean $I/\sigma(I)$	7.7 (1.7)	12.2 (3.4)	11.0 (4.0)	8.2 (1.8)
Wilson B factor (Å^2)	37.2	28.5	42.3	39.0
$R_{\rm merge}$†	0.07 (0.52)	0.04 (0.22)	0.06 (0.20)	0.10 (0.76)
$CC_{1/2}$	0.9 (0.5)	0.9 (0.7)	0.8 (0.6)	—‡
Reflections used for $R_{\rm free}$	4603	14680	2129	10163
$R_{\rm work}$§	0.21 (0.29)	0.16 (0.23)	0.23 (0.28)	0.16 (0.35)
$R_{\rm free}$¶	0.26 (0.35)	0.20 (0.28)	0.27 (0.34)	0.19 (0.38)
No. of non-H atoms				
Total	17678	36621	11628	36512
Protein	17300	35097	11517	35014
CoA bound to subunits B, D, F, H and K	—	—	—	240
Water	378	1524	111	1258
No. of protein residues	2375	4764	1572	4773
R.m.s.d., bonds (Å)	0.01	0.01	0.01	0.01
R.m.s.d., angles (°)	1.4	1.2	1.7	1.3
Ramachandran angles, favoured (%)	99.0	97.0	94.7	96.0
Ramachandran angles, allowed (%)	0.9	3.5	4.5	4.9
Ramachandran angles, outliers (%)	0.1	0.3	0.8	0.3
Clashscore	6.9	6.8	9.4	7.8
Average B factor (Å^2)				
Macromolecules	39.4	32.1	45.5	41.1
Ligands				56.7
Solvent	37.2	39.8	19.5	36.9
PDB code	5cbq	5byv	4zrc	5bz4

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is its mean intensity. ‡ $CC_{1/2}$ was not evaluated owing to the lack of redundancy in measured observations. § $R_{\rm work} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure factors. ¶ $R_{\rm free}$ is calculated as for $R_{\rm work}$ but for a randomly selected subset of the data (5%) which were excluded from refinement.

allowed to pass through the loaded column at a flow rate of 1 ml min^{-1} to allow binding of the recombinant His-tagged protein to the beads. The protein-bound matrix was washed with at least three column volumes of wash buffer 1 [100 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0, 10%(*v*/*v*) glycerol, 10 m*M* β-mercaptoethanol, 10 m*M* imidazole]. Protein was eluted using 200 m*M* imidazole along with 100 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0, 10%(*v*/*v*) glycerol, 10 m*M* β-mercaptoethanol. The eluted protein was concentrated to 1 ml by low-speed centrifugation using a 30 kDa molecular-weight cutoff Amicon Ultra Centrifugal filter (Merck Millipore, Germany). Further purification was achieved by size-exclusion chromatography using an ÄKTA Basic 10 liquid-chromatography system (GE Healthcare, Uppsala, Sweden) with a Superose 6 HR 10/30 column in 20 m*M* Tris–HCl buffer pH 8.0 containing 100 m*M* NaCl, 10%(*v*/*v*) glycerol and 5 m*M* β-mercaptoethanol. Purified protein was concentrated to 4 mg ml^{-1}. Protein concentration was estimated by the Bradford method (Bradford, 1976). The molecular weight and purity of the enzyme were checked by 12% SDS–PAGE (Laemmli, 1970) and mass spectrometry.

2.3. Crystallization of the apoenzyme

Conditions suitable for crystallization were screened by the microbatch method at 22°C using crystallization kits from Hampton Research (Li *et al.*, 2007). Crystallization drops contained 2 µl each of protein (concentration 5 mg ml^{-1}) and precipitant solution. Crystallization plates were covered with a 1:1 mixture of silicone and paraffin oils. Crystals were obtained under many conditions, which were subsequently optimized individually to improve the size and diffraction quality of the crystals. Three different crystal forms (Table 1) could be optimized under different conditions to yield crystals of quality suitable for X-ray diffraction data collection: form I ($C222_1$; 20% PEG 3350, 0.15 *M* calcium chloride dehydrate),

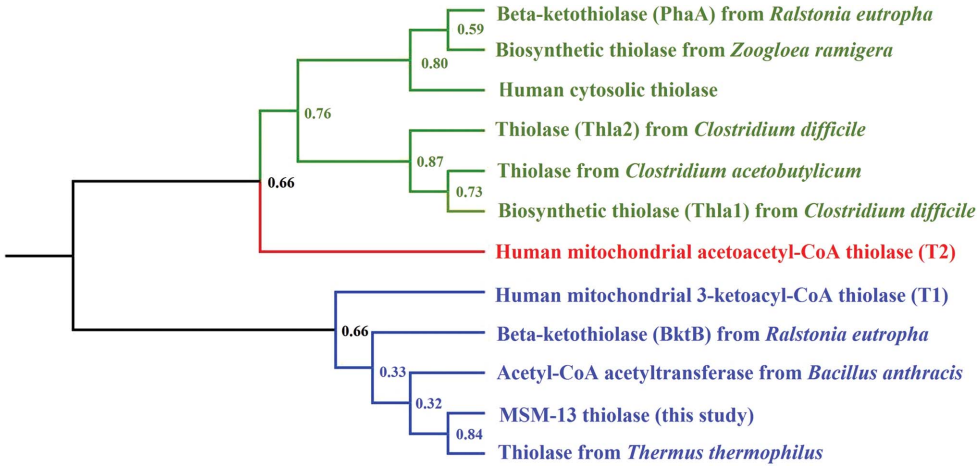


Figure 2
Phylogenetic tree depicting the relationship between various tetrameric thiolases for which the crystal structures are known. The tree depicts three different classes of tetrameric thiolases: the CT biosynthetic thiolases (green), the T2 thiolases (red) and the T1 thiolases (blue). Each class is also represented by a well characterized human tetrameric thiolase. MSM-13 thiolase falls within the T1 thiolase category.

form II ($P6_1$; 45% PEG 200, 0.1 *M* MES monohydrate pH 6.0, 0.07 *M* calcium chloride dehydrate) and form III ($P2_1$; 0.2 *M* sodium acetate trihydrate, 0.1 *M* sodium cacodylate trihydrate pH 6.5, 30% PEG 8000, 5% *n*-octyl-β-D-glucoside). In the case of the form III crystal, where an additive (5% *n*-octyl-β-D-glucoside) was used to improve the crystal quality, the volume of the precipitant solution was decreased to 1.5 µl to accommodate 0.5 µl of the additive. Crystals appeared in 5 d and grew to full size in 10 d. Crystals were mounted on a cryoloop and flash-cooled in a stream of nitrogen gas at 100 K. Ethylene glycol was used as the cryoprotectant [mother liquor supplemented with 20%(*v*/*v*) ethylene glycol] for all three crystal forms.

2.4. Data collection and structure determination

Data for crystal forms I and II were collected at a wavelength of 0.9537 Å on beamline BM14 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data for crystal form III were collected at the home source using a MAR345 image-plate detector system mounted on a Rigaku RU200 X-ray generator equipped with a 300 µm focal cup and a copper anode (wavelength 1.5418 Å). Data for crystal forms I and II were processed using *DENZO* and *SCALEPACK* from the *HKL*-2000 suite (Otwinowski & Minor, 1997). Data for crystal form III were processed using *iMosflm* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) from the *CCP*4 suite (Winn *et al.*, 2011). The number of protomers in the asymmetric unit was estimated by calculating the Matthews coefficient individually for all of the crystal forms. A monomeric polyalanine model of *Z. ramigera* thiolase (PDB entry 1dm3; Modis & Wierenga, 2000), from which water molecules and ligands were deleted, was used as the phasing model to determine the structure of the form III crystal by molecular replacement using *Phaser* (McCoy *et al.*, 2007). The resulting translation function had an LLG and a *Z*-score of 959 and 25, respectively. The *AB* dimer of the refined crystal form III was used as the phasing model to determine the structures of the other two crystal forms. For the hexagonal crystal data, the choice of space group between $P6_1$ and $P6_5$ was made by *Phaser* on the basis of the LLG value (2380 for $P6_1$ and 191 for $P6_5$). All three structures were built to completion by alternative rounds of automated model building with *AutoBuild* (from the *PHENIX* suite; Terwilliger *et al.*, 2008), manual model building with *Coot* (Emsley *et al.*, 2010) and refinement with *phenix.refine* (from the *PHENIX* suite; Afonine *et al.*, 2012). NCS restraints were used in the refinement of all of the structures.

2.5. Crystallization, data collection and structure determination of the MSM-13 thiolase–CoA complex

Prior to crystallization, the protein (5 mg ml^{-1}) was incubated with 5 m*M* CoA at 4°C for 30 min. Crystallization was set up in microbatch plates using the microbatch-under-oil method at 22°C. Crystallization drops contained 2 µl each of protein–CoA complex and precipitant solution. Crystals appeared in the same condition as the apo form II crystals [40%(*w*/*v*) PEG 200, 0.1 *M* MES monohydrate pH 6.0, 0.12 *M* calcium chloride dihydrate]. The crystals grew to their full size in 10 d. Prior to cooling, the crystals were soaked in a solution consisting of mother liquor, 25%(*v*/*v*) ethylene glycol and 5 m*M* CoA. Data were collected using a MAR345 image-plate detector system mounted on a Rigaku RU200 X-ray generator equipped with a 300 µm focal cup. Data were processed using *DENZO* and *SCALEPACK* from the *HKL*-2000 suite (Table 1). The crystals belonged to space group $P6_1$ and were isomorphous to the apo hexagonal crystals. The final reflection data set was directly used to refine the coordinates of the apo hexagonal MSM-13 thiolase structure to acceptable R_{work} and R_{free} values using *phenix.refine*. CoA molecules were built into the corresponding densities in the difference Fourier maps using *LigandFit* (Terwilliger *et al.*, 2006) from the *PHENIX* suite.

2.6. Structure analysis

Among the three apo structures of MSM-13 thiolase, form II was of the highest resolution and was therefore used in all analyses. The *B* subunit of the MSM-13–CoA complex structure was used to analyze the ligand-binding properties because this subunit exhibited the most significant density for CoA. The MSM-13 thiolase structure was compared with the following tetrameric thiolases: *Z. ramigera* thiolase (PDB entry 1dm3, chain *B*; Modis & Wierenga, 2000), human T1 thiolase (PDB entry 4c2j, chain *A*; Kiema *et al.*, 2014), human T2 thiolase (PDB entry 2ibw, chain *B*; Haapalainen *et al.*, 2007), human CT thiolase (PDB entry 1wl4, chain *A*; Kursula *et al.*, 2005), *Ralstonia eutropha* PhaA (PDB entry 4o9c, chain *C*; Kim & Kim, 2014) and *R. eutropha* BktB (PDB entry 4nzs, chain *B*; Kim *et al.*, 2014). The structure was also compared with those of the dimeric yeast peroxisomal AB thiolase (PDB entry 1afw; chain *A*; Mathieu *et al.*, 1997) and *M. tuberculosis* TFE thiolase (PDB entry 4b3j; chain *C*; Venkatesan & Wierenga, 2013). All structural analyses were performed using *Coot* and *PyMOL* (http://www.pymol.org/). The *PISA* server (Krissinel & Henrick, 2007) was used to calculate buried surface areas. All structural illustrations were prepared using *PyMOL*.

2.7. Thiolase assay

MSM-13 thiolase activity was examined at 25°C using a Jasco V-660 spectrophotometer. In the degradative direction, the activity measurement was performed using the Mg^{2+}–acetoacetyl-CoA complex method, as described previously (Haapalainen *et al.*, 2007). The exact concentration of CoA was determined by Ellman's test (Ellman, 1959) as follows: acetyl-CoA and acetoacetyl-CoA were hydrolyzed with hydroxylamine and the concentration of free CoA released was estimated by monitoring the absorbance at 412 nm. The 500 µl reaction cocktail consisted of 50 m*M* Tris–HCl pH 7.8, 25 m*M* $MgCl_2$, 60 µ*M* CoA and 50 µ*M* acetoacetyl-CoA. The disappearance of the Mg^{2+}–acetoacetyl-CoA complex was monitored at 303 nm for 3 min after the addition of 500 ng MSM-13 thiolase. An assay without added enzyme served as a control. The initial rate was calculated by estimating the change in absorption between 10 and 70 s. In the synthetic direction, the formation of acetoacetyl-CoA was estimated as described previously (Haapalainen *et al.*, 2007). The reaction mixture (500 µl) consisted of 50 m*M* Tris–HCl pH 7.8, 40 m*M* KCl, 0.2 m*M* NADH, one unit of 3-*R*-hydroxyacyl-CoA dehydrogenase (one unit was defined as the amount of enzyme required to convert 1 µmol of the substrate into product in 1 min), 0.5 m*M* DTT and 2 m*M* acetyl-CoA. The reaction was initiated by the addition of 500 ng MSM-13 thiolase. The rate of NADH oxidation was monitored at 340 nm. For both the biosynthetic and degradative reactions, *Z. ramigera* thiolase was used as a positive control.

3. Results and discussion

3.1. Bioinformatics analysis

A *BLAST* search against the PDB with the amino-acid sequence of MSM-13 thiolase as the query revealed several

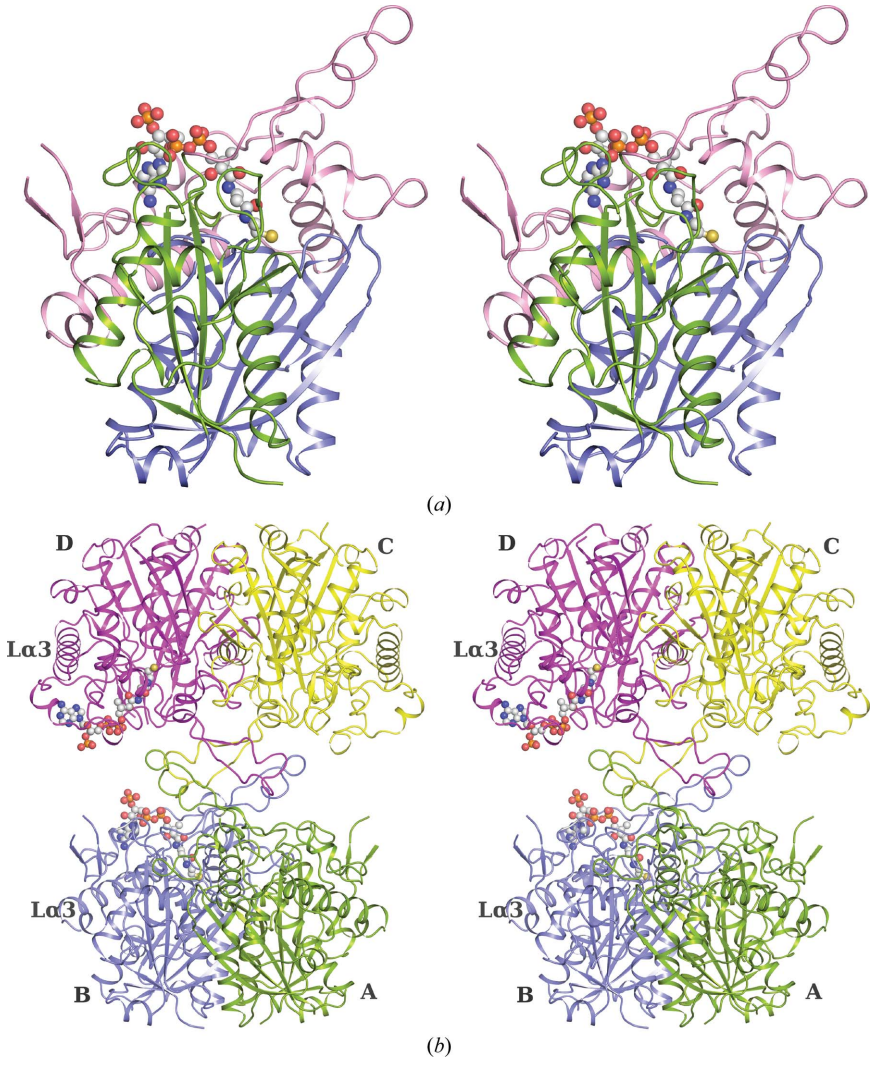


Figure 3
Structure of a single protomer and the quaternary organization of MSM-13 thiolase. (*a*) A single MSM-13 thiolase protomer. The structure can be divided into three subdomains: the N-terminal domain (blue), the loop domain (pink) and the C-terminal domain (green). The CoA molecule, represented as a ball-and-stick model, is bound to the loop domain. Its reactive sulfur moiety (yellow sphere) is bound near the catalytic site. (*b*) Tetrameric organization of MSM-13. Chains *A*, *B*, *C* and *D* are coloured green, blue, yellow and magenta, respectively. The *B* and *D* subunits have a bound CoA molecule. The *CD* dimer (top) is rotated along a vertical axis such that the cationic loops of the *CD* dimer point towards the active sites of the *AB* dimer (bottom) and *vice versa*. The cationic loop of the *B* subunit is best ordered.

Table 2
Distances (in Å) between the corresponding C^{α} positions of residue Ala185 (in Lα3) of the opposing subunits of independent MSM-13 thiolase tetramers.

	ABCD tetramer		*EFGH* tetramer		*JKLM* tetramer	
	Narrow cleft	Wide cleft	Narrow cleft	Wide cleft	Narrow cleft	Wide cleft
Form I	54.6 (*BD*)	56.2 (*AC*)	55.4 (*FF′*)	59.6 (*EE′*)		
Form II	53.8 (*BD*)	57.4 (*AC*)	54.0 (*FH*)	57.5 (*EG*)	54.6 (*KM*)	57.3 (*JL*)
Form II (+ CoA)	53.8 (*BD*)	57.0 (*AC*)	53.4 (*FH*)	57.5 (*EG*)	54.4 (*KM*)	56.9 (*JL*)
Form III	57.0 (*BD*)	57.7 (*AC*)				

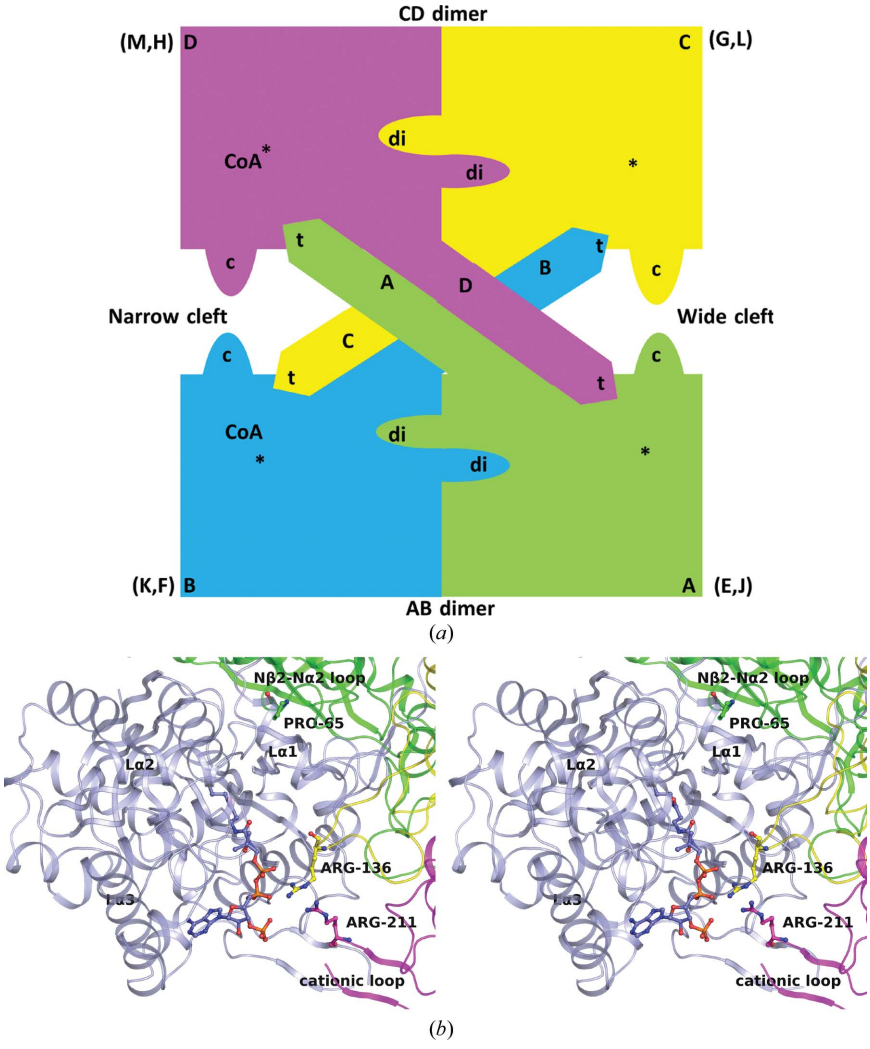


Figure 4
Tetrameric organization of MSM-13 thiolase. (*a*) Schematic diagram of the MSM-13 thiolase tetramer (similar view as in Fig. 3). The arrows indicate the tetramerization loops. The interactions of the *A* and *C* tetramerization loops with the opposing subunits are different from those of the *B* and *D* tetramerization loops. A small rotation along an axis perpendicular to the page applied to the *AB* dimer (clockwise) and the *CD* dimer (anticlockwise) generates an asymmetric tetramer with a narrow cleft (lined by the active sites of the *B* and *D* subunits) and a wide cleft (lined by the active sites of the *A* and *C* subunits). The active sites of the *B*, *F* and *K* subunits are complexed with CoA with full occupancy. The *D* and *H* subunits are also complexed with CoA but with lower occupancy. The active site of each subunit (*) is completed by loops extending from the other three subunits. For example, the subunit *B* active site is completed by the dimer interface ('di') Nβ2–Nα2 loop of the *A* subunit, the tetramerization loop ('t') of the *C* subunit and the cationic loop ('c') of the *D* subunit. (*b*) The active site of the *B* subunit (blue). Arg136 (*C* chain, yellow), at the tip of the tetramerization loop, points towards the pyrophosphate and the 3′-phosphate moiety of the bound CoA in the opposing *B* subunit. The cationic loop of the *D* subunit (magenta, labelled at residue Arg211) also comes close to the substrate-binding pocket of the *B* subunit. The substrate-specificity pocket is lined by the Nβ2–Nα2 loop (green, labelled at residue Pro65) of the *A* subunit.

homologous proteins. Fig. 1 depicts a sequence alignment of MSM-13 thiolase and other well studied thiolases. MSM-13 thiolase shares amino-acid sequence identities of 41% with *Z. ramigera* thiolase, 40% with human T1 thiolase and 37% with human CT thiolase (Fig. 1). The sequences with highest identity were found in *Bacillus anthracis* (43% identity; PDB entry 3ss6; Center for Structural Genomics of Infectious Diseases, unpublished work), *Thermus thermophilus* (45% identity; PDB entry 1ulq; RIKEN Structural Genomics/Proteomics Initiative, unpublished work), *R. eutropha* PhaA (44% identity; PDB entry 4o9c; Kim & Kim, 2014) and *R. eutropha* BktB (45% identity; PDB entry 4nzs; Kim *et al.*, 2014). Both PhaA and BktB are reported to function as biosynthetic thiolases. PhaA catalyzes the synthesis of acetoacetyl-CoA from two molecules of acetyl-CoA (Kim & Kim, 2014), while BktB catalyzes the synthesis of 3-ketovaleryl-CoA from propionyl-CoA and acetyl-CoA (Kim *et al.*, 2014). Structural comparisons have revealed that BtkB has a larger substrate-specificity pocket than PhaA (Kim *et al.*, 2014).

To identify the relationship between MSM-13 and other tetrameric thiolases, a phylogenetic tree was constructed using the aligned sequences of all tetrameric thiolases deposited in the PDB (Fig. 2). Three major branches could be recognized in the phylogenetic tree. Each branch included one of the three distinct human tetrameric thiolases (CT, T1 and T2). Therefore, the branches were labelled CT, T1 and T2. MSM-13 and its homologues from *B. anthracis*, *T. thermophilus* and *R. eutropha* (BktB) clustered in the T1 branch.

3.2. Biochemical properties of MSM-13 thiolase and crystal structures of the apo and liganded forms

MSM-13 thiolase was expressed in a soluble form in *E. coli* in the presence of the GroEL–GroES chaperone complex and was purified to apparent homogeneity by immobilized metal-ion affinity and size-exclusion chromatography. Apo MSM-13 thiolase could be

crystallized in three forms (I–III). Crystal form I belonged to space group $C222_1$, with one full tetramer (protomers *A*/*B*/*C*/*D*) and one-half of another tetramer (protomers *E*/*F*) in the asymmetric unit. The other half (protomers *E*′/*F*′) is generated by crystallographic twofold symmetry. Form II belongs to space group $P6_1$, with three tetramers (*A*/*B*/*C*/*D*, *E*/*F*/*G*/*H* and *J*/*K*/*L*/*M*) in the asymmetric unit. Form III belongs to space group $P2_1$, with one tetramer (*A*/*B*/*C*/*D*) in the asymmetric unit. Together, the three crystal forms provide structures of 22 independent copies of the protomer and six independent tetramers of unliganded MSM-13 thiolase (Table 2). The resolution of the structures range between 2.7 and 2.16 Å. Each of these structures could be refined to reasonable R_{work} and R_{free} values. The quality of the final electron-density maps allowed unambiguous modelling of most of the residues (4–404) in all three crystal forms. A few N-terminal residues and the side chains of some residues on the exterior surface of the protein did not have adequate density. Residues 212–216, which form the tip of the cationic loop, are disordered in most protomers. Except for Arg89, all residues are in the allowed or generously allowed region of the Ramachandran map. Arg89 precedes the catalytic cysteine Cys90 (as deduced on the basis of comparisons with *Z. ramigera* thiolase) and adopts a strained conformation, as in other thiolases.

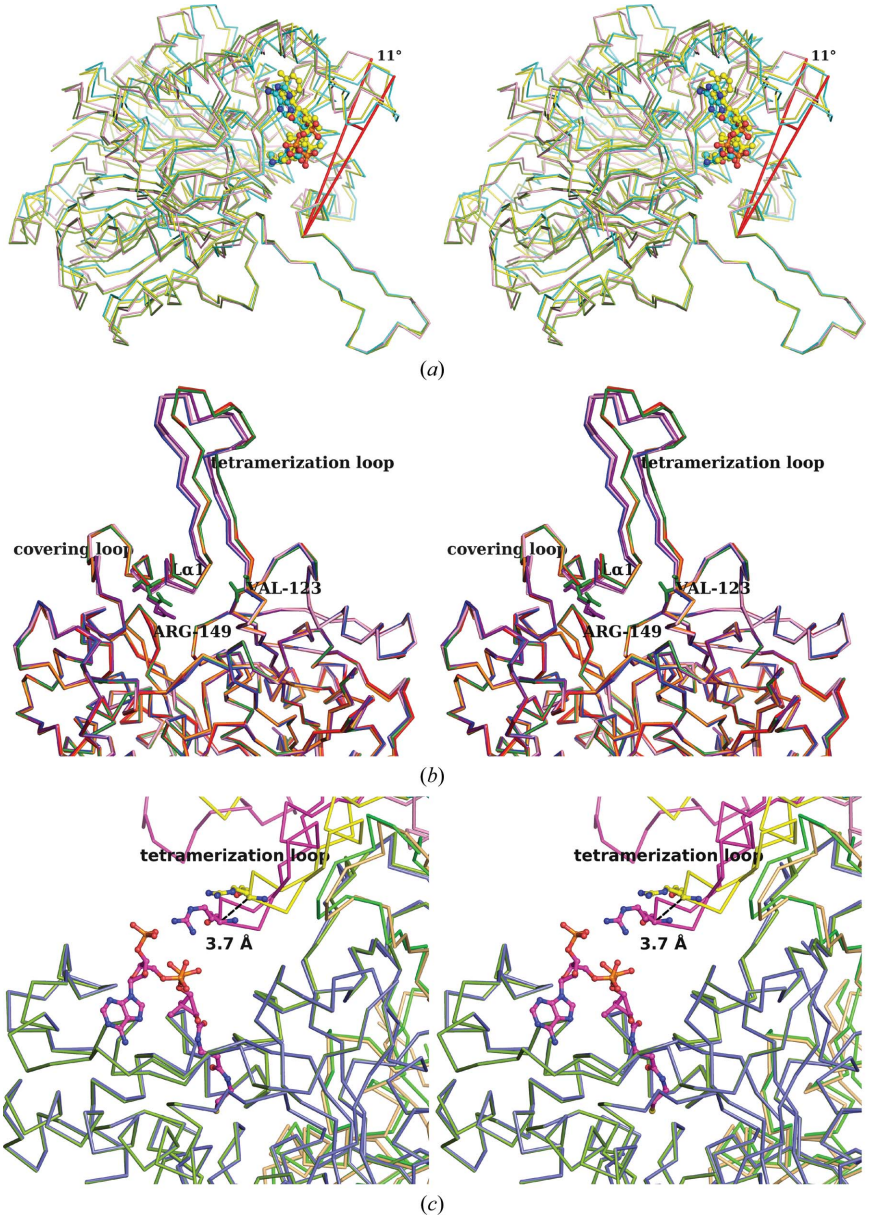


Figure 5
Asymmetry of the MSM-13 thiolase tetramer. (*a*) Superposition of the tetramerization loops of the *A* (green), *B* (cyan), *C* (pink) and *D* (yellow) subunits. The *B* and *D* subunits (with bound CoA) rotate around a hinge by approximately 11° with respect to the *A* and *C* (unliganded) subunits. (*b*) Superposition of the *A*, *E* and *J* subunits and of the *B*, *F* and *K* subunits. The hinge regions are near Val123 just before the tetramerization domain and near Arg149 (in Lα1) just after the tetramerization loop. (*c*) Superposition of the *A* subunit on the *B* subunit. The tip of the subunit *C* tetramerization loop (red) is closer to the subunit *B* active site (green) than the tip of the subunit *D* tetramerization loop (yellow) is to subunit *A* (green).

The crystal structure of CoA-bound MSM-13 thiolase was determined at 2.4 Å resolution. This structure is isomorphous to the apo form II structure, with three tetramers in the asymmetric unit. In the final difference electron-density map, strong density for CoA could be seen in the active sites of the *B* (Fig. 7*a*), *F* and *K* subunits, while weaker density was observed in the *D* and *H* subunits. There was no density for CoA in the other subunits.

3.3. Structure of the MSM-13 thiolase protomer

As in other thiolases, each protomer of MSM-13 thiolase consists of three domains: an N-terminal domain, a C-terminal domain and a loop domain, which is an insertion in the N-terminal domain immediately after the β4 strand (Figs. 1 and 3*a*). Both the N- and C-terminal domains exhibit the βαβαβαββ topology characteristic of thiolases. Between the N- and C-terminal halves is a layer of two α-helices contributed by the N-terminal domain (Nα3, just after the C*x*S motif) and the C-terminal domain (Cα3, just after the GHP motif). The β-sheets of the N- and C-terminal halves are surrounded on the outside by helical layers exposed to the bulk solvent, giving the protomer a five-layered appearance. The CoA-binding site is mostly constituted by the loop domain, which hosts several important CoA-binding segments such as the tetramerization, covering, cationic, adenine and pantetheine loops (highlighted in Fig. 1; Modis & Wierenga, 1999).

Gel-filtration analysis suggested that MSM-13 thiolase is a tetramer in solution (data not shown). Consistent with this

finding, in all three crystal forms the protomers associate into tetramers (Fig. 3*b*), as observed in *Z. ramigera* thiolase. Superposition of protomers within a crystal form and between different crystal forms resulted in r.m.s.d.s ranging between 0.3 and 0.5 Å for equivalent C^{α} atoms, suggesting that the structures of independent protomers are nearly identical. The only difference between the structures of the subunits is in the relative positioning of the protomers with respect to the tetramerization loop (see subsequent sections).

3.4. Quaternary organization of MSM-13 thiolase

The overall tetrameric organization of MSM-13 thiolase closely resembles that of *Z. ramigera* thiolase. The average interface area of the 11 independent dimers in the three apo structures is 2939 $Å^2$ per subunit, as calculated by the *PISA* server, suggesting tight association of the protomers in the dimer. This interface is stabilized by approximately 62 hydrogen bonds and 23 salt bridges. Structural comparisons revealed that the mode of assembly of two protomers into a tight dimer is nearly identical in all tight dimers. In all cases, two such tight dimers assemble to form a tetramer which is stabilized by a characteristic cage-like structure (referred to as the tetramerization motif in *Z. ramigera* thiolase; Modis & Wierenga, 1999). The surface area buried owing to the formation of the tetramerization motif is ∼1900 $Å^2$ per subunit in MSM-13 thiolase. The corresponding area in *Z. ramigera* thiolase is 2200 $Å^2$. In the tetramer, the cationic loop of each subunit points towards the active site of the opposing subunit (Fig. 3*b*).

Residues 126–142 at the N-terminus of the loop domain of the four protomers interact to form the tetramerization motif of MSM-13 thiolase. In this assembly, the tetramerization loops of the *A* and *C* subunits interact with each other, and the tips of these loops point toward the active sites of the *D* and *B* subunits, respectively. Similarly, the tetramerization loops of the *B* and *D* subunits interact with each other and point towards the active sites of the *C* and *A* subunits, respectively (Fig. 4*a*). The tetramerization motif of MSM-13 thiolase is stabilized by a number of hydrogen bonds. The core of the motif is stabilized by hydrophobic interactions mediated by four methionine residues (Met130) extending from the four subunits. The methionines stack one above the other in two pairs. Other side chains that contribute to the hydrophobic core of the tetramerization motif are Phe125, Val139 and Ile141. The tip of the loop is formed by Gly134-Ala135-Arg136-Thr137. Although these residues have relatively high *B* factors, their main-chain conformation is well defined. The long side chain of Arg136 of the *C* and *A* subunits points towards the phosphate moieties of the CoA bound in the opposing binding pockets of the *B* and *D* subunits, respectively.

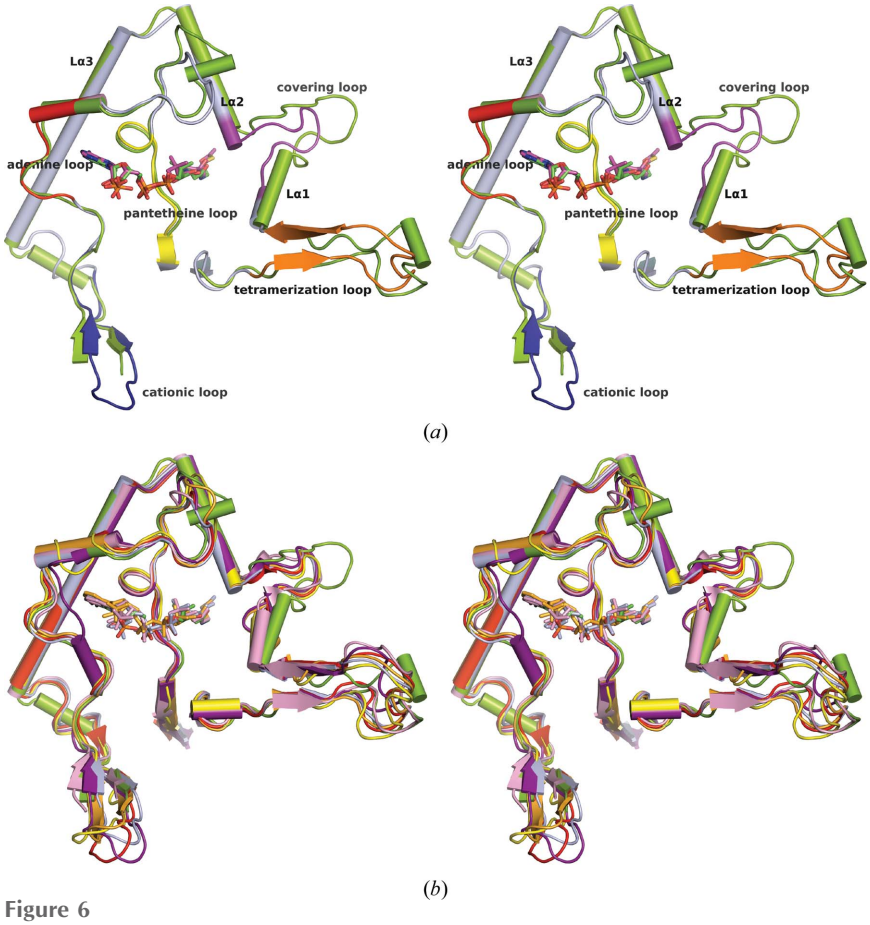


Figure 6
Comparison of the loop domains of tetrameric thiolases. (*a*) Variations in the loop domains of MSM-13 (green) and *Z. ramigera* thiolases. The various regions of the *Z. ramigera* loop domain are coloured as follows: tetramerization loop, orange; covering loop, purple; cationic loop, dark blue; adenine loop, red; pantetheine loop, yellow. The conformations of the tetramerization and covering loops are significantly different in MSM-13 thiolase. (*b*) Superposition of the loop domains of MSM-13 thiolase (green), *Z. ramigera* thiolase (light blue), human T1 thiolase (pink), human T2 thiolase (purple), CT thiolase (orange), BktB (yellow) and PhaA (red).

3.5. Asymmetry in the MSM-13 thiolase tetramer

Within a tetramer, the 'bottom' tight dimers (*A*/*B*, *E*/*F* and *J*/*K*; Fig. 4*a*) superimpose well on the 'top' tight dimers (*C*/*D*, *G*/*H* and *L*/*M*, respectively), suggesting structural invariance of the tight dimers. However, an asymmetry was recognized in the MSM-13 thiolase tetramer when the positioning of the two tight dimers of a tetramer was examined with respect to the tetramerization motif. This analysis was initially performed with the three tetramers of the apo form II crystals. Careful examination of the three tetramers of crystal form II revealed that the *AB* dimers and *CD* dimers are assembled at slightly different orientations with respect to the tetramerization domain. Fig. 5(*a*) clearly demonstrates that the orientation of the *B* and *D* subunits with respect to the

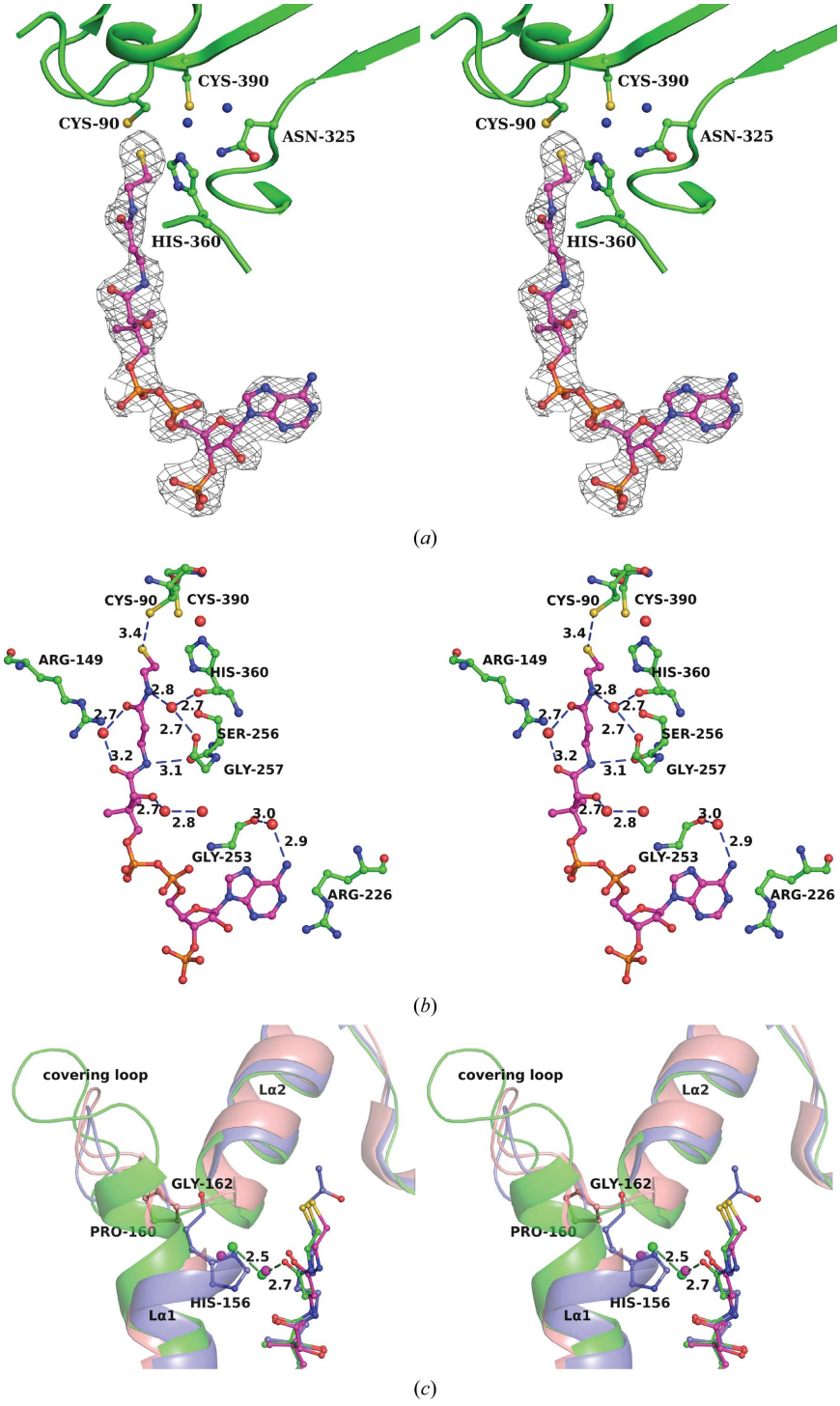


Figure 7
Interactions between CoA and MSM-13 thiolase. (*a*) The active site. The four catalytic residues (Cys90, Asn325, His360 and Cys390) and two catalytic waters (blue spheres) are conserved in MSM-13 thiolase. The $2F_o - F_c$ OMIT electron-density map (calculated after OMIT refinement, contoured at 1.0σ) for CoA, along with a stick representation of the ligand, is also shown. (*b*) Protein–CoA interactions. The ligand is shown in pink. The residues that interact with the ligand are shown as green sticks. The water molecules present around the ligand are shown as red spheres. Hydrogen bonds (within 3.5 Å) are indicated by dashed lines. (*c*) Hydrogen-bonding pattern around the O5P moiety of CoA. The O5P moiety is hydrogen-bonded to a water molecule in degradative thiolases (MSM-13 thiolase, green; T1 thiolase, pink). The water molecules are shown as spheres. In the biosynthetic *Z. ramigera* thiolase (blue) the O5P moiety points to His156, but does not hydrogen-bond to any waters or residues. The extra space for the two waters is provided by the substitution of His156 (*Z. ramigera* thiolase) by smaller residues such as Pro160 (T1 thiolase) and Gly162 (MSM-13).

tetramerization loop is different from that of the *A* and *C* subunits. However, systematic superposition of residues 129–139 from pairs of protomers revealed that there are no structural rearrangements within the tetramerization loops of the *A*, *B*, *C* and *D* subunits (r.m.s.d. of 0.3–0.4 Å; Fig. 5*a*), suggesting that the tetramerization domain is invariant and functions as a rigid motif. The conformational flexibility observed in the tetramer can therefore be attributed to small changes in the main-chain dihedrals near Val123 (just before the tetramerization loop) and Arg149 in Lα1 (just after the tetramerization loop) (Fig. 5*b*).

This asymmetry was quantified by measuring the distances between Ala185 of opposing subunits (Table 2). Ala185 occurs in the middle of the long Lα3 helix, which is a well defined helix of the loop domain located at the edge of the subunit (Fig. 3*b*). The relative orientation of the *AB* and *CD* tight dimers with respect to the tetramerization domain is such that the *B* and *D* subunits are rotated slightly towards the cleft between the two dimers. Consequently, the *A* and *C* subunits are rotated away from the cleft. This unique assembly generates a 'narrow' and 'wide' cleft between the dimers (Fig. 4*a*). The distances listed in Table 2 confirm that this asymmetry is preserved in all three tetramers of the form II crystal. Interestingly, this asymmetry is also preserved in the tetramers of the form I and form III crystals (Table 2).

The same asymmetry is also observed in the CoA-complexed form II (Table 2). It may be noted here that CoA is bound only to the active sites of those subunits that point towards the 'narrow' cleft, suggesting that active sites lining the 'narrow' cleft have a higher affinity for CoA. Furthermore, the structure of Lα1, as well as the side-chain conformation of Arg149, is different in the liganded 'narrow-cleft' active sites and the unliganded 'wide-cleft' active sites. In the CoA-complexed active sites the side chain of Arg149 points inwards (towards the bound CoA), whereas in the uncomplexed active sites this residue points towards the solvent (Fig. 5*b*). The higher affinity of the 'narrow-cleft' active sites for CoA may be attributable to the fact that the 'narrow-cleft' active sites are closer to the tip of the opposing tetramerization loop (near Arg136), facilitating interaction of the side chain of Arg136 of the opposing loop with

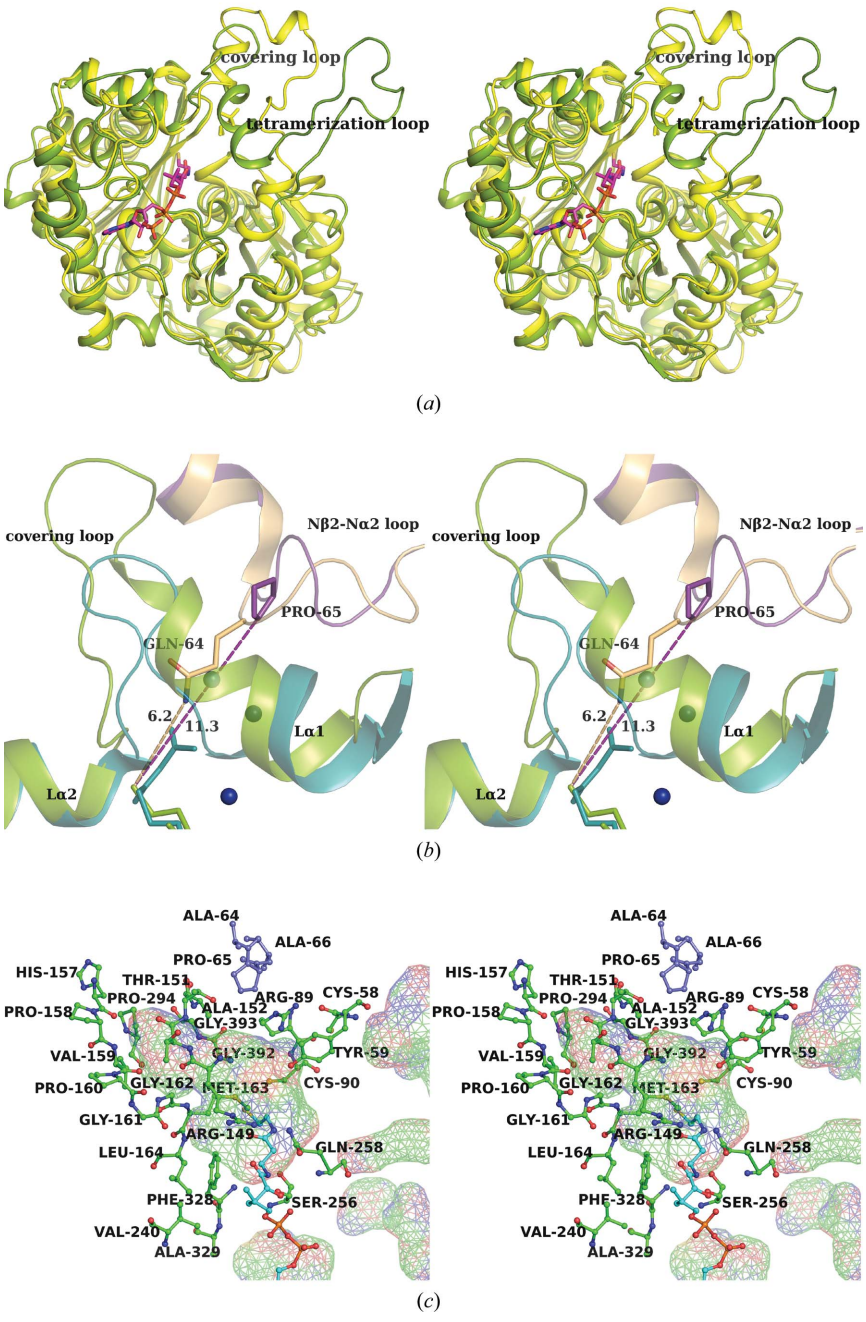


Figure 8
The larger substrate-specificity cavity in MSM-13 thiolase. (*a*) Comparison of MSM-13 thiolase (green) and yeast peroxisomal AB thiolase (yellow). In the peroxisomal AB thiolase, space for the very long chain fatty-acyl moiety is created by the extended conformation of the region corresponding to Lα1 and the covering loop of MSM-13 thiolase. In this region, an MPD molecule is bound in the yeast AB thiolase. (*b*) Larger substrate-specificity cavity in MSM-13 thiolase (green and purple for adjacent subunits) when compared with that in *Z. ramigera* thiolase (blue and brown for adjacent subunits). The extra space in the active site of MSM-13 thiolase, filled by three waters (blue spheres), is created by the displacement of the Lα1–covering loop and the Nβ2–Nα2 loop (of the other subunit). A conserved glutamine (Gln64) of *Z. ramigera* thiolase that protrudes into this cavity is replaced by a smaller proline residue, Pro65, in MSM-13 thiolase. The distance between the terminal S atom of CoA and the conserved Gln64 side chain of *Z. ramigera* thiolase is 6.2 Å (brown dashes). The corresponding distance between the –SH moiety of CoA and Pro65 in MSM-13 thiolase is 11.3 Å (purple dashes). (*c*) The substrate-specificity cavity for binding the acyl tail in MSM-13 thiolase. Chain *B* residues are shown in green and chain *A* residues are shown in blue. The protein Connolly surface of the *AB* apo dimer was calculated with *PyMOL* using the surface-cavity mode with a probe radius of 1.4 Å and is represented as a wire frame. The cavity is lined by the polar residues Tyr59, Arg89 and Arg149. The bound CoA molecule is shown in blue.

the phosphates of the bound CoA (Fig. 5*c*), thereby stabilizing the binding of the ligand to the active site.

3.6. Comparison of the loop domains of MSM-13 and other tetrameric thiolases

The loop domain is critical for substrate recognition, and its N-terminal region is important for tetramerization. Fig. 6(*a*) presents a superposition of the loop domains of MSM-13 and *Z. ramigera* thiolases. The loop domains of MSM-13 and *Z. ramigera* thiolases are largely similar, with the exception of the tetramerization loop, the conformation of which is markedly different between the two structures. Moreover, the segment of the loop domain between Lα1 and Lα2 (the covering loop) is longer by four residues in MSM-13 thiolase (Fig. 1). In contrast, the cationic, adenine and pantetheine loops of MSM-13 thiolase are remarkably similar to those of *Z. ramigera* thiolase. These observations are consistent with the fact that the CoA molecule is bound to the ligand-binding pocket in the same conformation in both structures.

Similar comparisons with other well characterized tetrameric thiolases revealed that the loop domains of *Z. ramigera* thiolase, CT, BktB and PhaA (all of which are biosynthetic thiolases) share a similar structure. The loop domains of T1 and T2 thiolases (both of which are degradative thiolases) deviate slightly from that of *Z. ramigera* thiolase; the loop domain of MSM-13 thiolase exhibits the maximum deviation (Fig. 6*b*).

3.7. The CoA-binding pocket and catalytic centre of MSM-13 thiolase

The density for CoA is most well defined in the *B*, *F* and *K* subunits (Fig. 7*a*). As in *Z. ramigera* thiolase, the pantetheine moiety of CoA binds to MSM-13 thiolase in a long tunnel starting at the active-site pocket and leading towards the solvent. The terminal sulfur of CoA is at a distance of 3.4 Å from the catalytic cysteine. The pantetheine moiety is lined by residues of the pantetheine loop. No large conformational changes were observed in the structure of MSM-13 thiolase upon CoA binding, except for the altered conformation of the side chain of Met163 (in Lα2) near the catalytic centre. Additionally, the SG moiety of Cys90

rotates slightly towards the NE2 moiety of His360 upon CoA binding, as in *Z. ramigera* thiolase. The residues that interact with CoA are mainly contributed by the loop domain. The adenosine moiety binds in a shallow groove, and the N6 amino group of the adenine ring points inwards and is weakly hydrogen-bonded to the side-chain hydroxyl moiety of Thr229.

Several unique arginines interact with the bound CoA in MSM-13 thiolase (Fig. 7*b*). Arg226 is stacked with the adenine moiety. Arg18 lines the CoA-binding groove. In addition, there is an arginine at the tip of the cationic loop (Arg216 of the opposing dimer), as well as at the tip of the tetramerization loop (Arg136 of the opposing dimer). Both of these arginines point towards the phosphate moieties of CoA. The tip of the cationic loop is disordered in the unliganded and liganded protomers. Although the tip of the tetramerization loop (near Arg136) is ordered in all protomers, the side chain of Arg136 is not always ordered. Interactions with these arginines might help stabilize the negatively charged substrate in the ligand-binding cavity.

Systematic comparisons of well characterized tetrameric thiolases revealed that in biosynthetic thiolases the O5P atom of CoA is not hydrogen-bonded to either water or a protein atom (Kursula *et al.*, 2005). However, in degradative thiolases such as TFE thiolase and T1 thiolase this O atom is hydrogen-bonded to either a protein side chain (TFE) or a water molecule (T1). In MSM-13 thiolase the O5P moiety of CoA is hydrogen-bonded to a water molecule (Fig. 7*c*). Moreover, a plausible hydrogen-bonding interaction (at 3.6 Å) was observed between Arg149 and the O5P of the pantetheine moiety of CoA. Therefore, MSM-13 thiolase is similar to degradative thiolases in this respect. How the presence of this additional hydrogen bond impacts the catalytic cycle remains to be elucidated.

As in other thiolases, the active site of MSM-13 thiolase is located in a deep pocket in the interdimer space close to the tetramerization motif (Figs. 3 and 5). As illustrated in Fig. 4(*b*), the active site of each subunit is completed by loops from the other three subunits. Specifically, the subunit *B* active site is completed by the Nβ2–Nα2 loop of the *A* subunit, the tetramerization loop of the *C* subunit and the cationic loop of the *D* subunit.

The catalytic cysteine that becomes covalently attached to the priming substrate is conserved in MSM-13 thiolase (Cys90; equivalent to Cys89 in *Z. ramigera* thiolase). The S atom of Cys90 is hydrogen-bonded to the main-chain N atoms of Cys90 and Gly392. In *Z. ramigera* thiolase, Cys89 is activated by His348 by the abstraction of a proton. Another residue important for catalysis in *Z. ramigera* thiolase is Cys378, which acts as a base in the condensation reaction (Modis & Wierenga, 2000). His348 and Cys378 of *Z. ramigera* thiolase are structurally equivalent to His360 and Cys390, respectively, of MSM-13 thiolase. The conservation of the geometry of these active-site residues suggests that they play similar roles in the two enzymes. Moreover, the two catalytic waters known to be essential for catalysis are preserved (Fig. 7*a*). These waters function in oxyanion hole 1 (OAH1) as hydrogen-bond donors by interacting with the thioester O atom of acyl-CoA. They are also important in modulating the properties of the catalytic base (Meriläinen *et al.*, 2009; Cys390 in MSM-13 thiolase).

3.8. The distinct substrate-specificity pocket of MSM-13 thiolase

The conformations of the following three loop regions have been shown to determine the substrate specificity of thiolases: (i) the Lα1–covering loop–Lα2 region, (ii) the Nβ2–Nα2 region of the adjacent subunit and (iii) the Cβ1–Cα1 loop (Haapalainen *et al.*, 2007; Kiema *et al.*, 2014).

The Lα1–covering loop–Lα2 motif shields the active site from bulk solvent. In biosynthetic thiolases, the conformation of this region is such that there is no space to accommodate a bulky acyl group. In MSM-13 thiolase, Lα1 and the covering loop are displaced away from the active site by about 4 Å when compared with biosynthetic thiolases. The extra space created by this displacement hosts three water molecules. In the dimeric AB peroxisomal thiolase from yeast, this motif adopts a completely different, more extended conformation (Fig. 8*a*). This change in conformation is proposed to provide the extra space that is required for the binding of a long-chain fatty-acyl tail. In the unliganded crystal structure of yeast AB thiolase this extra space is occupied by an MPD molecule.

The Nβ2–Nα2 loop is highly conserved in biosynthetic thiolases. A conserved glutamine (Gln64 in *Z. ramigera* thiolase; Fig. 8*b*) protrudes from this loop into the active-site pocket of the neighbouring subunit, thereby allowing only small acyl chains to bind in the pocket. This glutamine residue is at a distance of 6.2 Å from the terminal S atom of acetoacetyl-CoA bound to *Z. ramigera* thiolase, leaving no space for longer acyl chains. In MSM-13 thiolase the Nβ2–Nα2 loop from the adjacent protomer is displaced away from the active site and adopts a different conformation when compared with that in *Z. ramigera* and other biosynthetic thiolases (Fig. 8*b*). The conserved glutamine that protrudes from this loop into the active-site cavity in synthetic thiolases is replaced by a smaller proline residue in MSM-13 thiolase (Fig. 1). Superposition of the MSM-13 and *Z. ramigera* thiolase active-site pockets (Fig. 8*b*) revealed that the distance between this proline and bound CoA is of the order of ∼11 Å in MSM-13 thiolase, leaving enough space for the binding of a more bulky substrate.

The Cβ1–Cα1 loop has a conserved VMG motif that allows the binding of straight-chain acyl-CoA substrates. In thiolases that act on substrates with a branched 2-methyl-acyl chain, this motif is different. For example, in human T2 thiolase, which accommodates a 2-methyl group in its specificity pocket, the corresponding sequence is DFP. MSM-13 thiolase has the sequence LMG at this region, suggesting that it is specific for straight-chain acyl-CoA substrates.

The putative fatty acyl-binding cavity, as calculated by *PyMOL*, is visualized in Fig. 8(*c*). A similar cavity extending from the nucleophilic cysteine was also identified in human T1 thiolase (Kiema *et al.*, 2014). However, the shape and relative location of this cavity is different in MSM-13 thiolase. Phe328,

Arg149, Glu258, Ser256, Val240, Met163, Glu258, Tyr59, Arg89, Ala327 and Ala152 line this cavity; residues of the Nβ2–Nα2 loop of the adjacent subunit, including Pro65, seal this cavity. In T1 thiolases, this fatty acyl-binding cavity is lined by hydrophobic residues. In contrast, the cavity in MSM-13 is lined by several polar residues, including Arg89, Arg149 and Tyr59. Arg89 precedes the nucleophilic cysteine and its main-chain dihedrals adopt a strained conformation. Other tetrameric thiolases have a shorter hydrophobic residue (Leu, Ile or Val) at the corresponding position (Fig. 1). The other unique arginine, Arg149, protrudes from Lα1 (immediately after the tetramerization loop). A short residue such as Gly, Ala or Ser is found in other tetrameric thiolases at the corresponding position (Fig. 1). These two arginines, along with Tyr59 (of the Nβ2–Nα2 loop of the same subunit; an isoleucine or leucine residue in all other tetrameric thiolases), generate a unique water-filled specificity pocket near the active site that is not seen in any of the other tetrameric thiolases (Fig. 9). The polar features of this pocket suggest that the priming substrate of MSM-13 thiolase might have a bulky polar acyl moiety.

Biochemical analysis revealed that MSM-13 thiolase exhibits residual thiolytic cleavage activity with acetoacetyl-CoA as the substrate. No catalytic activity was detected in the synthetic direction with acetyl-CoA (data not shown). It is possible that acetoacetyl-CoA and acetyl-CoA, both of which harbour only small acyl chains, are poor substrates for MSM-13 thiolase; the enzyme might specifically catalyze the degradation of substrates with bulky polar acyl moieties. Further studies with these substrates will provide insights into the specificity and catalytic activity of MSM-13 thiolase.

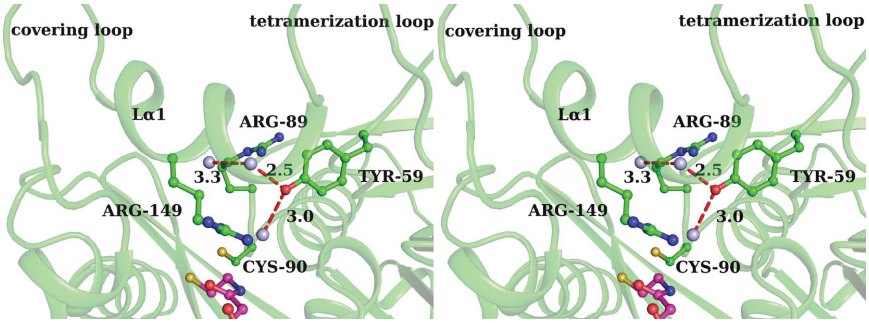


Figure 9
The unique substrate-specificity cavity of MSM-13 thiolase. Three residues near the MSM-13 thiolase substrate-specificity cavity (Tyr59, Arg89 and Arg149) generate a distinct cavity occupied by three water molecules (blue spheres) in MSM-13 thiolase. The red dotted lines highlight the hydrogen-bond interactions of these waters.

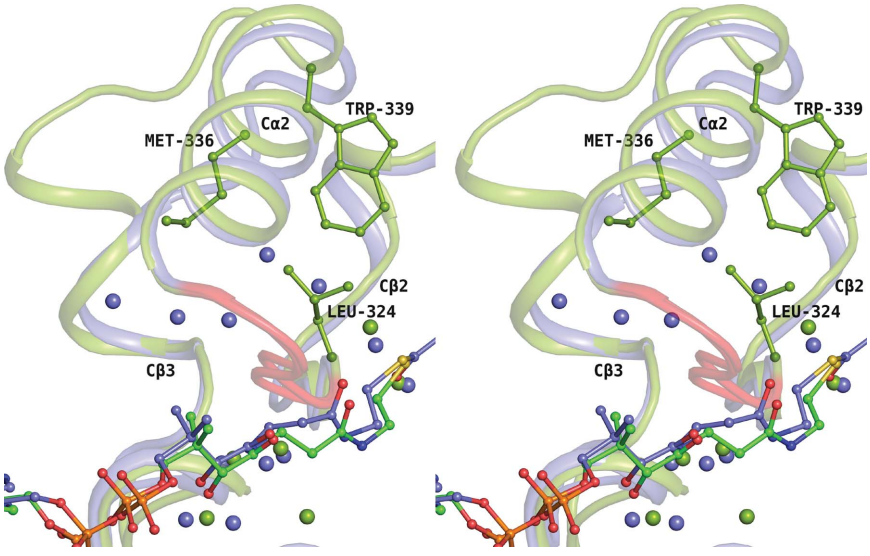


Figure 10
Differences in the water networks between biosynthetic (*Z. ramigera*) and degradative (MSM-13) thiolases. Biosynthetic thiolases (*Z. ramigera* thiolase, blue) have two networks of buried water molecules extending from the active-site pocket: (i) a trail of water molecules extending from the OAH1 water molecule ('above' the NEAF–Cβ2–Cα2 loop) and (ii) a cluster of water molecules ('below' the NEAF–Cβ2–Cα2 loop). The cluster of water molecules is present in MSM-13 but the trail of waters is absent because of the presence of the bulky hydrophobic residues Leu324 (Cβ2), Met336 (Cα2) and Trp339 (Cα2). The NEAF–Cβ2–Cα2 loops of *Z. ramigera* and MSM-13 thiolases are highlighted in red.

3.9. Water-mediated hydrogen-bond networks in the active sites of thiolases

Water molecules play an important role in the catalytic function of thiolases. Apart from the two conserved catalytic waters (Fig. 7*a*), two water networks have been identified near the active sites of thiolases: (i) a trail of waters extending from the catalytic site to the bulk solvent and (ii) a cluster of waters that are not connected to the bulk solvent. These networks of buried hydrogen-bonded water molecules are located near the NEAF–Cβ2–Cα2 loop (Fig. 10). The cluster of waters is present in all thiolases, whereas the trail of waters is present only in the two biosynthetic enzymes CT thiolase and *Z. ramigera* thiolase. In *Z. ramigera* thiolase, the trail of ordered water molecules extending from the water at OAH1 (Wat82 in *Z. ramigera* thiolase) to the bulk solvent is lined by the side chains of residues Ala315 and Asn327 near the Cβ2–Cα2 loop. The loop between Cβ2 and Cα2 contains the highly conserved NEAF motif. The asparagine side chain of the NEAF motif fixes the water at OAH1. This oxyanion hole has been proposed to stabilize the negative charge of the CoA-thioester intermediate (Meriläinen *et al.*, 2009). The connected network of hydrogen-bonded water molecules has been proposed to be responsible for the high catalytic turnover of synthetic thiolases (Meriläinen *et al.*, 2009). In T2 and yeast peroxisomal thiolases, both of which are degradative thiolases, this water trail is not present.

Instead, the side chains of hydrophobic residues fill up the remaining space.

Inspection of the ordered waters in MSM-13 thiolase revealed that the first water network, the trail of waters, is absent, whereas the water cluster is present (Fig. 10). Leu324 of Cβ2, just before the NEAF motif, protrudes into this channel, filling the space occupied by the water trail in *Z. ramigera* thiolase. Two other bulky residues, Met336 and Trp339, are also present in the vicinity of this channel. Leu324, Met336 and Trp339 are replaced by Ala315, Asn327 and Leu330, respectively, in the biosynthetic *Z. ramigera* thiolase. Thus, in terms of hydrogen-bonded water networks, MSM-13 thiolase is similar to biodegradative enzymes such as human T2 and T1 thiolases, suggesting that MSM-13 is possibly a degradative thiolase.

4. Concluding remarks

Phylogenetic analysis revealed that tetrameric thiolases may be classified into three distinct classes (CT, T1 and T2). MSM-13 thiolase clustered into the T1 branch, indicating that this thiolase is closely related to human T1 thiolase. The overall assembly of MSM-13 thiolase is very similar to that of the well characterized biosynthetic thiolase from *Z. ramigera*. The mode of assembly of tetrameric thiolases generates an unusual quaternary structure, which provides a unique access path for substrates diffusing from the bulk solvent to the active site. This is achieved by connecting the two tight dimers *via* flexible hinge regions to a small, rigid tetramerization domain formed by the four tetramerization loops. Analysis of the structures of nine tetramers of MSM-13 present in three different crystal forms revealed that the tetramers are asymmetric. The structural variations result from differences in the positioning of the *AB* and *CD* dimers with respect to the rigid tetramerization domain. This asymmetry generates differences in the geometry of the active sites with respect to two arginines: Arg136 (at the tip of the tetramerization loop of the opposing subunit) and Arg149 (of Lα1).

The four catalytic residues and two waters known to be essential for catalysis in thiolases are conserved in MSM-13 thiolase. Structural comparison of the active site of MSM-13 thiolase with that of *Z. ramigera* thiolase revealed that the Lα1 helix and covering loop, as well as the loop occurring after Nβ2 (of the adjacent subunit), are displaced away from the active site, generating a larger substrate-specificity pocket in MSM-13 thiolase that differs from the specificity pocket found in human T1 thiolase. In MSM-13 thiolase, two arginines (Arg89 and Arg149) and a tyrosine (Tyr59) point into this pocket.

Biosynthetic tetrameric thiolases have two distinct networks of buried waters near the catalytic site, whereas degradative thiolases have only one such network. The water network in the active site of MSM-13 thiolase resembles that of degradative thiolases. The larger and more polar substrate-specificity pocket and the low catalytic efficiency for the standard thiolase substrates (acetoacetyl-CoA) suggest that MSM-13 thiolase has a unique substrate specificity. Further studies are needed to identify the preferred substrate and physiological function of MSM-13 thiolase.

Acknowledgements

MRN thanks DBT, DST and the Bose fellowship for financial support. MRN and RKW thank DBT for sanctioning an Indo–Finnish grant. NJ thanks CSIR for supporting her work through 2010–2015. RKW gratefully acknowledges the support of the Academy of Finland (131795). We thank the staff of BM14 at ESRF for expert support. RKW thanks the University of Oulu for a travel grant to complete this manuscript.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
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.
Antonenkov, V. D., Van Veldhoven, P. P. & Mannaerts, G. P. (1999). *Lipids*, **34**, S157.
Antonenkov, V. D., Van Veldhoven, P. P., Waelkens, E. & Mannaerts, G. P. (1997). *J. Biol. Chem.* **272**, 26023–26031.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bout, A., Franse, M. M., Collins, J., Blonden, L., Tager, J. M. & Benne, R. (1991). *Biochim. Biophys. Acta*, **1090**, 43–51.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Ellman, G. D. (1959). *Arch. Biochem. Biophys.* **82**, 70–77.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fukao, T., Nakamura, H., Nakamura, K., Perez-Cerda, C., Baldellou, A., Barrionuevo, C. R., Castello, F. G., Kohno, Y., Ugarte, M. & Kondo, N. (2002). *Mol. Genet. Metab.* **75**, 235–243.
Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007). *Biochemistry*, **46**, 4305–4321.
Heath, R. J. & Rock, C. O. (2002). *Nat. Prod. Rep.* **19**, 581–596.
Higgins, D. G. (1994). *Methods Mol. Biol.* **25**, 307–318.
Kiema, T.-R., Harijan, R. K., Strozyk, M., Fukao, T., Alexson, S. E. H. & Wierenga, R. K. (2014). *Acta Cryst.* D**70**, 3212–3225.
Kim, E.-J. & Kim, K.-J. (2014). *Biochem. Biophys. Res. Commun.* **452**, 124–129.
Kim, E.-J., Son, H. F., Kim, S., Ahn, J.-W. & Kim, K.-J. (2014). *Biochem. Biophys. Res. Commun.* **444**, 365–369.
Kotani, S., Yanagida, I., Kato, K. & Matsuda, T. (1970). *Biken J.* **13**, 249–275.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Kumar, S., Tamura, K. & Nei, M. (1994). *Comput. Appl. Biosci.* **10**, 189–191.
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.
Li, J., Shen, W., Liao, M. & Bartlam, M. (2007). *Acta Cryst.* F**63**, 24–26.
Mao, L. F., Chu, C., Luo, M. J., Simon, A., Abbas, A. S. & Schulz, H. (1995). *Arch. Biochem. Biophys.* **321**, 221–228.
Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A., Rasmussen, B., Lamzin, V. S., Kunau, W. H. & Wierenga, R. K. (1997). *J. Mol. Biol.* **273**, 714–728.
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.
Middleton, B. (1973). *Biochem. J.* **132**, 731–737.
Middleton, B. (1974). *Biochem. J.* **139**, 109–121.
Middleton, B. & Bartlett, K. (1983). *Clin. Chim. Acta*, **128**, 291–305.
Miyazawa, S., Furuta, S., Osumi, T., Hashimoto, T. & Ui, N. (1981). *J. Biochem.* **90**, 511–519.
Modis, Y. & Wierenga, R. K. (1999). *Structure*, **7**, 1279–1290.
Modis, Y. & Wierenga, R. K. (2000). *J. Mol. Biol.* **297**, 1171–1182.
Nishimura, T., Saito, T. & Tomita, K. (1978). *Arch. Microbiol.* **116**, 21–27.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Seedorf, U., Scheek, S., Engel, T., Steif, C., Hinz, H. J. & Assmann, G. (1994). *J. Biol. Chem.* **269**, 2613–2618.
Staack, H., Binstock, J. F. & Schulz, H. (1978). *J. Biol. Chem.* **253**, 1827–1831.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W. & Cohn, J. D. (2006). *Acta Cryst.* D**62**, 915–922.
Uchida, Y., Izai, K., Orii, T. & Hashimoto, T. (1992). *J. Biol. Chem.* **267**, 1034–1041.
Venkatesan, R. & Wierenga, R. K. (2013). *ACS Chem. Biol.* **8**, 1063–1075.
Williams, S. F., Palmer, M. A., Peoples, O. P., Walsh, C. T., Sinskey, A. J. & Masamune, S. (1992). *J. Biol. Chem.* **267**, 16041–16043.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.