Protein Engineering, Design & Selection, 2017, vol. 30 no. 3, pp. 227–235
doi: 10.1093/protein/gzw080
Advance Access Publication Date: 6 January 2017
Original Article

Original Article

Crystallographic substrate binding studies of *Leishmania mexicana* SCP2-thiolase (type-2): unique features of oxyanion hole-1

Rajesh K. Harijan[1,5], Tiila-Riikka Kiema[1], Shahan M. Syed[1], Imran Qadir[1], Muriel Mazet[2,3], Frédéric Bringaud[2,3], Paul A.M. Michels[4], and Rik K. Wierenga[1,*]

[1]Faculty of Biochemistry and Molecular Medicine, Biocenter Oulu, University of Oulu, FIN-90014, Finland, [2]Centre de Résonance Magnétique des Systèmes Biologiques (RMSB), UMR5536, Université de Bordeaux, CNRS, 146 rue Léo Saignat, 33076 Bordeaux, France, [3]Present address: Laboratoire de Microbiologie Fondamentale et Pathogénicité (MFP), UMR5234, Université de Bordeaux, CNRS, 146 rue Léo Saignat, 33076 Bordeaux, France, [4]Centre for Immunity, Infection and Evolution and Centre for Translational and Chemical Biology, School of Biological Sciences, The King's Buildings, The University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK, and [5]Present address: Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

[*]To whom correspondence should be addressed. E-mail: rik.wierenga@oulu.fi
Edited by Dagmar Ringe

Received 6 September 2016; Revised 15 November 2016; Editorial Decision 12 December 2016; Accepted 15 December 2016

Abstract

Structures of the C123A variant of the dimeric *Leishmania mexicana* SCP2-thiolase (type-2) (Lm-thiolase), complexed with acetyl-CoA and acetoacetyl-CoA, respectively, are reported. The catalytic site of thiolase contains two oxyanion holes, OAH1 and OAH2, which are important for catalysis. The two structures reveal for the first time the hydrogen bond interactions of the CoA-thioester oxygen atom of the substrate with the hydrogen bond donors of OAH1 of a CHH-thiolase. The amino acid sequence fingerprints (**C**xS, **N**EAF, G**H**P) of three catalytic loops identify the active site geometry of the well-studied CNH-thiolases, whereas SCP2-thiolases (type-1, type-2) are classified as CHH-thiolases, having as corresponding fingerprints **C**xS, **H**DCF and G**H**P. In all thiolases, OAH2 is formed by the main chain NH groups of two catalytic loops. In the well-studied CNH-thiolases, OAH1 is formed by a water (of the Wat-Asn(NEAF) dyad) and NE2 (of the GHP-histidine). In the two described liganded Lm-thiolase structures, it is seen that in this CHH-thiolase, OAH1 is formed by NE2 of His338 (HDCF) and His388 (GHP). Analysis of the OAH1 hydrogen bond networks suggests that the GHP-histidine is doubly protonated and positively charged in these complexes, whereas the HDCF histidine is neutral and singly protonated.

Key words: Coenzyme A, crystal structure, *Leishmania mexicana*, leishmaniasis, ligand binding, lipid metabolism, SCP2-thiolase, sleeping sickness, *Trypanosoma brucei*, β-oxidation

Introduction

Thiolases are enzymes involved in degradative metabolic pathways, such as the β-oxidation cycle for the catabolism of fatty acids, and in biosynthetic pathways, such as sterol biosynthesis (Haapalainen *et al.*, 2006). The function of biosynthetic thiolases (EC 2.3.1.9) is to catalyze the conversion of two molecules of acetyl-CoA into

acetoacetyl-CoA and free CoA, whereas the function of degradative thiolases (EC. 2.3.1.16) concerns the catalytic conversion of 3-ketoacyl-CoA and free Coenzyme A (CoA) into acetyl-CoA and an acyl-CoA, which is two carbon atoms shorter. If in the latter case the substrate is acetoacetyl-CoA, then the products are two molecules of acetyl-CoA. All studied thiolases catalyze the formation of acetoacetyl-CoA from two molecules of acetyl-CoA, and *vice versa*, the reverse reaction. The equilibrium of this reaction strongly favors the degradative direction. As is visualized in Fig. 1, the reaction proceeds via a covalent intermediate and the first step of both reactions is the nucleophilic attack of the catalytic cysteine, which in the latter example, in both directions becomes acetylated.

A recent bioinformatics study (Anbazhagan *et al.*, 2014) showed that there is a large variation in the number of genes in a genome that encode a thiolase. For example there is one gene in the genome of *Trypanosoma* spp., but there are two genes in *Leishmania* spp. and six in human. The leishmanial thiolase studied here is a homolog of the single trypanosomal thiolase, with which it shows 71% sequence identity. The best studied thiolase is the biosynthetic thiolase of the bacterium *Zoogloea ramigera* (Zr-thiolase), which has 23% and 22% sequence identity with this trypanosomal and leishmanial thiolase, respectively (Fig. 2). The eukaryotic organisms *Leishmania* spp. and *Trypanosoma* spp. are parasitic protists belonging to the Trypanosomatidae.

The thiolase bioinformatics study has identified several thiolase subfamilies. For example, in the human genome six thiolases are encoded (CT, T1, T2, AB, SCP2 and TFE), which are located in the cytosol (CT), mitochondria (TFE, T1, T2) and peroxisomes (AB and SCP2). Extensive phylogenetic tree calculations with 130 thiolase sequences have shown that the SCP2-family of thiolases consists of two subfamilies, referred to as type-1 and type-2 (Anbazhagan *et al.*, 2014). The human dimeric peroxisomal SCP2-thiolase (type-1) (SCP2-thiolase (type-1)) is a degradative thiolase, catalyzing the thiolytic cleavage of 24-oxo-trihydroxycoprostanoyl-CoA in the bile-acid synthesis pathway (Antonenkov *et al.*, 1997). The crystal structure of this SCP2-thiolase (type-1) has not been reported. The here described leishmanial SCP2-thiolase is of type-2. The function of this thiolases is not known but structures of this leishmanial as well as its trypanosomal homolog have been reported (Harijan *et al.*, 2013). Thiolases occur as dimers or tetramers (dimers of dimers). The SCP2-thiolases are dimers, whereas the Zr-thiolase (bacterial biosynthetic CT-thiolase of *Zoogloea ramigera*) is a tetramer. Thiolases of each subfamily have different functions and the respective enzyme kinetic properties are also different (Fukao, 2002; Haapalainen *et al.*, 2006). No metal ions are required for the thiolase reaction. Enzymes of the thiolase superfamily share a conserved characteristic structure referred to as the thiolase fold (βαβαβαββ) (Mathieu *et al.*, 1994, 1997). The catalytic residues reside in the βα- and ββ-loops of the two core domains (Haapalainen *et al.*, 2006).

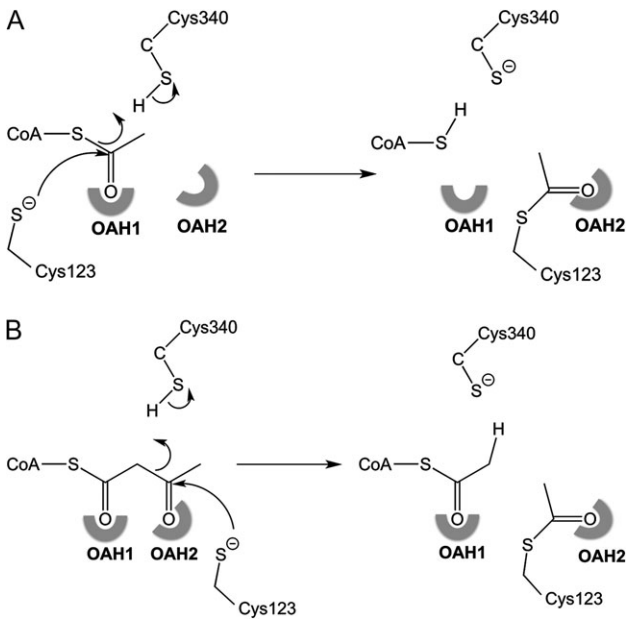

Fig. 1 The mechanism of acetylation of the nucleophilic cysteine, Cys123. Cys340 acts in this step as the catalytic acid, protonating the leaving group. (**A**) The acetylation by acetyl-CoA: the leaving group is CoA. (**B**) The acetylation by acetoacetyl-CoA: the leaving group is acetyl-CoA. The two oxyanion holes (OAH1 and OAH2), which are important for catalysis are highlighted.

Thiolases have four catalytic loops characterized by specific sequence fingerprints, including the CxS, NEAF, GHP and CxG motifs (Fig. 2), whereas a fifth sequence fingerprint (VMG) in the Cβ1-Cα1 loop correlates with the absence of a binding pocket for the methyl group of 2-methyl branched acyl-CoA substrates (Haapalainen *et al.*, 2006, 2007). The catalytic loops are also involved in shaping two oxyanion holes, OAH1 and OAH2. In Zr-thiolase OAH1 is formed by a water, of a Wat-ND2(Asn316) dyad (Asn316 is part of the NEAF loop), and by NE2(His348) (His348 is part of the GHP loop) (Merilainen *et al.*, 2009). OAH2 is formed by the main chain NH groups of the CxS loop and the CxG loop (Kursula *et al*, 2002). The CxS-cysteine is the nucleophilic cysteine of the Nβ3-Nα3 loop (Fig. 2). This nucleophilic cysteine becomes acylated during the reaction (Fig. 1). The NEAF motif belongs to the Cβ2-Cα2 loop. In SCP2-thiolase and dimeric mitochondrial trifunctional enzyme thiolase (TFE-thiolase) the NEAF sequence fingerprint is replaced by HDCF and HEAF, respectively. The histidine of the GHP loop (in the Cβ3-Cα3 loop) is not only a hydrogen bond donor of OAH1 but it is also important for activating the nucleophilic cysteine, by abstracting the proton of this cysteine. The CxG motif belongs to the Cβ4-Cβ5 loop. The cysteine of this loop acts as an acid/base in the classical thiolases (Haapalainen *et al.*, 2006) whereas in the SCP2-thiolases this fingerprint is absent and the acid/base functionality resides in the cysteine of the HDCF loop (Fig. 2).

Here we report on crystallographic substrate binding studies of the dimeric *Leishmania mexicana* SCP2-thiolase (type-2) (Lm-thiolase), which is a mitochondrial enzyme (Mazet *et al.*, 2011). In Lm-thiolase the nucleophilic cysteine of the CxS loop is Cys123 and the acid/base cysteine is Cys340 of the HDCF loop (Harijan *et al.*, 2013). Enzyme kinetic studies of Lm-thiolase have shown that Cys123 and Cys340 are indeed essential for enzymatic activity. It was also shown that, under the assay conditions used (2.0 mM acetyl-CoA and 50 μM acetoacetyl-CoA, 60 μM CoA, respectively), the specific activity in the synthetic direction is higher than in the degradative direction, suggesting that possibly Lm-thiolase could be functionally active also in the synthetic direction. The catalytic properties of biosynthetic thiolases are currently also investigated in the context of metabolic engineering approaches aimed at developing organisms for the synthesis of biofuels (Sheppard *et al.*, 2016) such as butanol (Mann and Lutke-Eversloh, 2013; Reisse *et al.*, 2014; Kim *et al.*, 2015) and bioplastics (Kim and Kim, 2014, 2016), like polyhydroxyalkanoates (Wang *et al.*, 2014). Better understanding of the thiolase reaction mechanism will facilitate the development of biosynthetic thiolases with tailor-made biocatalytic properties.

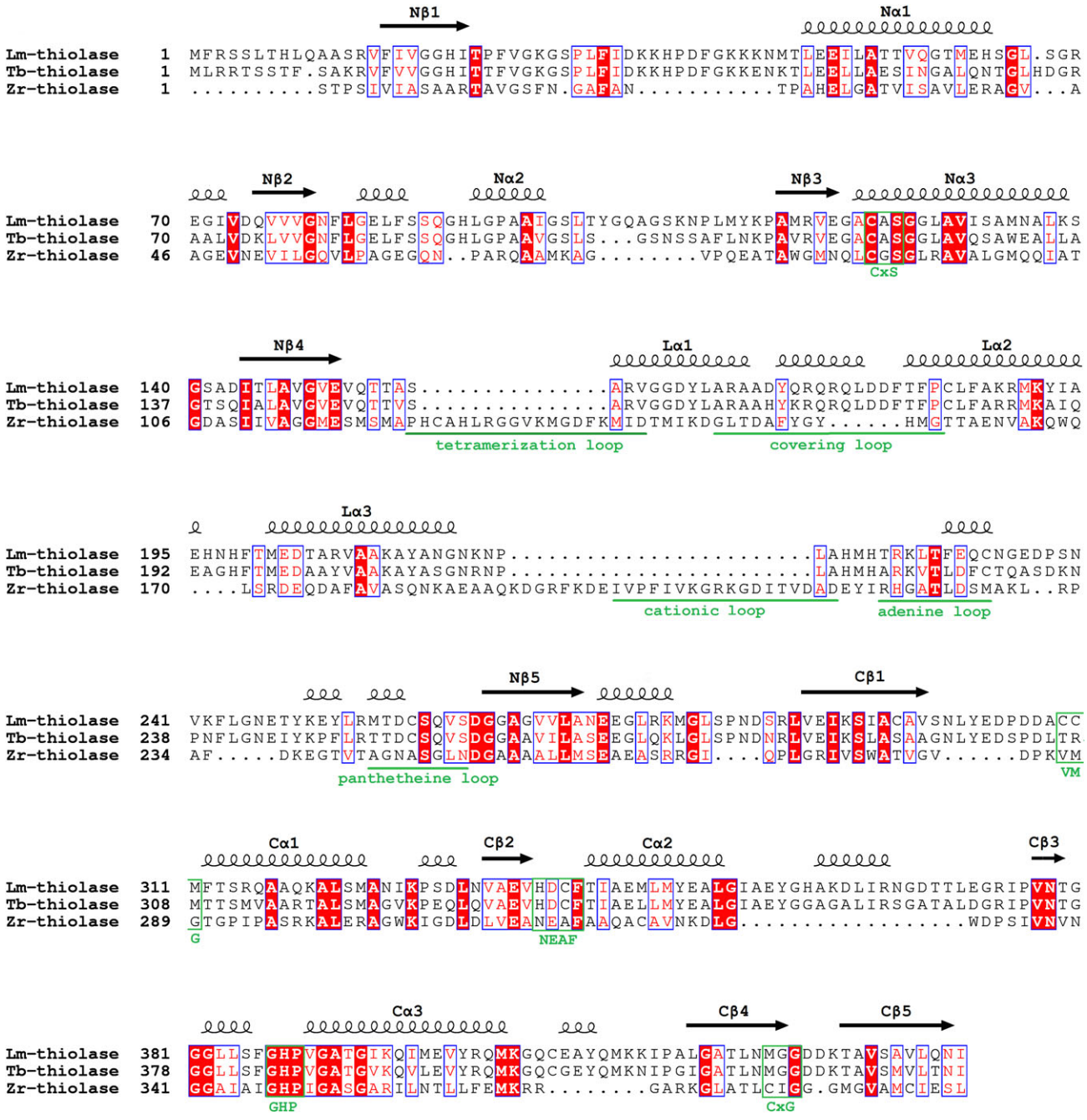

Fig. 2 The structure based sequence alignment of Lm-thiolase (PDB code 3ZBG), trypanosomal SCP2-thiolase (type-2) (Tb-thiolase) (PDB code 4BI9) and Zr-thiolase (PDB code 1DM3). The secondary structure assignment, as indicated above the sequence, refers to the Lm-thiolase structure. The sequence fingerprints of the active site loops, as discussed in the text, are given below the sequences.

Several other enzymes have been found to be structurally related to thiolases, and consequently a thiolase superfamily has been identified, which includes not only the degradative and biosynthetic thiolases but also other enzymes such as chalcone synthases (CHS), 3-ketoacyl-acyl carrier protein (ACP) dependent synthases (KAS-I, KAS-II, KAS-III), polyketide synthases (PKS) as well as 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase (Jiang *et al.*, 2008). The thiolases of this superfamily can be subdivided in a CNH subfamily having the CxS, the NEAF and the GHP sequence motifs, and a CHH subfamily having the CxS, HDCF (or HEAF) and GHP sequence fingerprints (Jiang *et al.*, 2008). The TFE- and SCP2-thiolases belong to the CHH subfamily and the tetrameric cytosolic thiolase (CT-thiolase), tetrameric mitochondrial thiolase (T1-thiolase), tetrameric mitochondrial T2-thiolase (T2-thiolase) and dimeric peroxisomal thiolase (AB-thiolase) belong to the CNH subfamily. The structural properties of OAH1 of the CNH-thiolases have been well described (Merilainen *et al.*, 2009), but the structural properties of OAH1 of the CHH-thiolases have not yet been reported. We describe here the OAH1 properties of Lm-thiolase, which is a CHH-thiolase, as characterized in the crystal structures

of C123A Lm-thiolase complexed with acetyl-CoA and acetoacetyl-CoA, respectively.

Results and discussion

The crystal structures of C123A Lm-thiolase complexed with acetyl-CoA and acetoacetyl-CoA, respectively, have been determined at 2.25 Å and 1.98 Å resolution, respectively, using crystals obtained by cocrystallization experiments (Tables I and II). The crystal structures of the unliganded wild-type and unliganded C123A Lm-thiolase and the C123A Lm-thiolase (complexed with CoA) are known (Harijan *et al.*, 2013). Cocrystallization experiments with wild-type Lm-thiolase in the presence of CoA or its derivatives did not yield structures of a complexed active site, presumably because the active site cysteine (Cys123) oxidizes during crystallization, forming a bulky sulfonate moiety, which would clash with the CoA molecule as bound to C123A Lm-thiolase (Harijan *et al.*, 2013). The crystals of C123A Lm-thiolase grown in the presence of acetyl-CoA and acetoacetyl-CoA have the same space group and similar cell dimensions as previously described for the corresponding CoA complexed form, being $P2_1$, with four protomers (subunits A, B, C and D) generating the AB-dimer and the CD-dimer in the asymmetric unit. The density for the acetyl moiety of acetyl-CoA and the acetoacetyl moiety of acetoacetyl-CoA is clearly defined in the respective omit maps for subunit B (Fig. 3), whereas subunit C is unligated, like in the CoA complexed crystal form. In the subunits A and D the density of the acyl moieties is fragmented. In the acetyl-CoA complex, one conformation has been built in subunit A and D (being similar to subunit B), whereas in the acetoacetyl-CoA complex structure two conformations have been built in these active sites, different from the mode of binding seen in subunit B. For the structure description the mode of binding in subunit B will be used, unless mentioned otherwise. Except for a few residues at the N-terminus, the subunits of both structures, starting at residue Ala12, have been modeled completely in the electron density maps. A few loop regions have high B-factors, in particular near residues 40, 100, 236, 277 and 372. The 236 region is near the adenine binding pocket. None of these flexible regions are near the catalytic site.

The mode of binding of the acyl moieties of acetyl-CoA and acetoacetyl-CoA in the active site of C123A Lm-thiolase

The pantetheine part of acetyl-CoA and acetoacetyl-CoA is bound in the pantetheine binding tunnel and the acyl moiety is deeply buried in the catalytic cavity. The N4 atom of the pantetheine moiety is hydrogen bonded to OG(Ser259), like in the Zr-thiolase complexes (Fig. 4). The thioester oxygen of acetyl-CoA and acetoacetyl-CoA is bound in OAH1, formed by NE2(His338) and NE2(His388). In the acetyl-CoA complex a water molecule is interacting with N(Cys123) and N(Gly428), being bound in OAH2. His338 of OAH1 is part of the Cβ2-Cα2 HDCF loop and His388 belongs to the Cβ3-Cα3 GHP loop. Figures 4 and 5 visualize the anchoring hydrogen bond interactions between ND1(His388, of the GHP-histidine) and OG1(Thr393) and between ND1(His338, of the HDCF histidine) and ND2(Asn425). The latter hydrogen bond network is extended further as OD1(Asn425) is hydrogen bonded to N(Gly127) (Fig. 5). In Zr-thiolase OAH1 is formed by NE2(His348) of the GHP loop (anchored to Ser353, like in Lm-thiolase) and by the water of the Wat-Asn(NEAF) dyad, in which this water is hydrogen bonded to ND2(Asn316).

In the acetoacetyl-CoA complex the 3-keto oxygen atom is hydrogen bonded in OAH2, like also observed in the structure of acetoacetyl-CoA complexed to the C89A variant of Zr-thiolase (PDB code 1M1O) (Kursula *et al.*, 2002). In the latter structure the acetoacetyl moiety is bound in a planar conformation, whereas in the Lm-thiolase complex, it is bound in a twisted conformation. In this well defined B active site mode of binding the acetoacetyl moiety clashes with the putative position of the Cys123 sulfur atom in the wild-type active site, as visualized in Fig. 4. For example, in the superposition visualized in Fig. 4B, the distance between SG(Cys89) (1DM3) and C2 (of the acetoacetyl moiety of acetoacetyl-CoA) (5LNQ) is 2.2 Å. Therefore, the mode of binding of the acetoacetyl moiety in the competent active site is not precisely known.

Comparison of the structures of the acetyl-CoA and acetoacetyl-CoA complexes with the CoA complexed structure shows that the mode of binding of CoA in the latter structure does not exactly match with the mode of binding of the CoA part of the acetyl- and acetoacetyl-CoA in the complexed structures. In the CoA complex the sulfur of the CoA part is bound deeper in the active site, by 2 Å. The structure of the CoA complex concerns its complex with the C123A variant and in this complex the S(CoA) atom is at van der Waals distance from Cβ(Ala123) (at 3.3 Å), which is not compatible with the mode of binding to the wild-type active site, causing clashes with the Cys123 side chain. In the acetyl- and acetoacetyl-CoA complex the S(CoA)-atom has moved further outwards, allowing for the thioester oxygen to bind in OAH1. There are no other structural differences in the active site when comparing the acetyl-CoA and acetoacetyl-CoA complexed active site with the unliganded active site, captured in subunit C of these complexes or in the B active sites of the unliganded crystal forms of the wild-type (PDB code 3ZBG) and the C123A variant (PDB code 3ZBK). Some conformational heterogeneity of the Phe182 side

Table I. Crystallization and crystal handling

C123A Lm-thiolase[a]	Well solution buffer[b]	Cryobuffer[c]	Space group	Comments
acetyl-CoA complex	100 mM Tris, pH 8.0, 0.2 M lithium chloride and 20% w/v PEG 6000	100 mM Tris, pH 8.0, 0.2 M lithium chloride, 20% w/v PEG 6000 and 1 mM acetyl-CoA	$P2_1$	PDB 5LOT
acetoacetyl-CoA complex	100 mM Bis-Tris propane, pH 7.5, 0.2 M sodium fluoride and 20% w/v PEG 3350	100 mM Bis-Tris propane, pH 7.5, 0.2 M sodium fluoride, 20% w/v PEG 3350 and 1 mM acetoacetyl-CoA	$P2_1$	PDB 5LNQ

[a]In all experiments the protein solution has been 5.0 mg/ml enzyme in gel filtration buffer (25 mM Tris, pH 7.8, 50 mM NaCl, 1 mM EDTA and 1 mM DTT).
[b]The crystallization was done at room temperature using the sitting drop vapor diffusion method, by mixing 1.5 µl protein solution (containing also 1 mM ligand) and 1.5 µl well solution.
[c]The crystals did grow within 3–4 days and, following a short soak (maximally 30 s) in cryobuffer, were immediately cryofrozen.

Table II. Data collection and refinement statistics

Dataset[a]	*L. mexicana* C123A SCP2-thiolase	
	acetyl-CoA complex	acetoacetyl-CoA complex
Unit cell data		
Space group	$P2_1$	$P2_1$
Cell parameters (Å, °)	a = 79.70, b = 89.89, c = 126.33 α, γ = 90, β = 105.8	a = 74.14, b = 89.80, c = 123.70 α, γ = 90, β = 104.8
V_m (Å^3/Dalton)	2.3	2.1
Number of subunits in the asymmetric unit	4	4
Data collection		
Beam line	Home source	ID29 (ESRF)
Wavelength (Å)	1.5418	0.9763
Temperature (K)	100	100
Resolution range (Å)	47.26–2.25 (2.35–2.25)	56.03–1.98 (2.01–1.98)
Total number of observed reflections	565 093 (80 155)	412 038 (20 926)
Number of unique reflections	21 309 (8805)	108 328 (5369)
R_{merge} (%)[b]	11.6 (48.7)	15.4 (72.8)
R_{pim} (%)[c]	4.5 (33.9)	9.1 (43.0)
CC1/2 (%)	99.4 (64.4)	98.8 (57.0)
<I/σ(I)>[d]	10.6 (1.4)	6.9 (2.0)
Completeness (%)	98.2 (89.0)	99.2 (99.1)
Multiplicity	7.1 (2.4)	3.8 (3.9)
Wilson B-factor (Å^2)	14.2	16.2
Refinement		
R_{work} (%)[e]	21.3	18.9
R_{free} (%)[f]	25.5	22.6
No. of atoms	13 684	14 016
Protein atoms	12 892	12 916
Ligand atoms	153	270
Solvent atoms	639	830
Model quality		
RMS deviation from ideal value		
Bond length (Å)	0.012	0.015
Bond angle (°)	1.5	1.8
Average B-factor		
Protein atoms (A/B/C/D) (Å^2)	34.7 / 24.4 / 31.8 / 29.7	22.6 / 21 / 27 / 26.3
Ligand atoms (A/B/C/D) (Å^2)[g]	97 / 21 / – / 64	43, 147 / 25 / – /41, 122
Waters (Å^2)	26.6	29.4
Ramachandran plot[h]		
Most favored regions (%)	98.1	98.2
Allowed regions (%)	1.9	1.7
Outlier regions (%)[i]	0.0	0.1
PDB code entry	5LOT	5LNQ

[a]Values in parentheses refer to the highest resolution shell.

[b]$R_{merge} = (\Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|)/\Sigma_{hkl}\Sigma_i \langle I_i(hkl)\rangle$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection (hkl) and $\langle I(hkl)\rangle$ is its mean intensity.

[c]$R_{pim} = (\Sigma_{hkl}[1/(N_{hkl}-1)]^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|) / \Sigma_{hkl}\Sigma_i \langle I_i(hkl)\rangle$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection (hkl), $\langle I(hkl)\rangle$ is its mean intensity and N is the number of measurements.

[d]I is the integrated intensity and σ(I) is its estimated standard deviation.

[e]$R_{work} = (\Sigma_{hkl}|Fo-Fc|)/\Sigma_{hkl}Fo$ where Fo and Fc are the observed and calculated structure factors.

[f]R_{free} is calculated as for R_{work} but from a randomly selected subset of the data (5%), which were excluded from the refinement calculation.

[g]A double conformation has been built in subunits A and D of the acetoacetyl-CoA complex.

[h]Calculated by MOLPROBITY (Chen *et al.*, 2010).

[i]The residues in the outlier regions are part of high B-factor loops.

chain has been observed in the structure of the unliganded C123A Lm-thiolase complex (subunit A, PDB code 3ZBK) (Harijan *et al.*, 2013). This residue is located at the beginning of helix Lα2, immediately after the covering loop, and it covers the binding pocket of the acyl moiety. However, in its alternative conformation the side chain is rotated into another position, such that a narrow tunnel, filled with two bound water molecules, is formed that connects this acyl binding pocket to the bulk solvent. This space could possibly allow the binding of acyl chains longer than the C4 moiety of acetoacetyl-CoA. This putative tunnel is not present in Zr-thiolase, as the more extended covering loop of Lm-thiolase (Fig. 2) and its different Cβ4-Cβ5 loop conformation provide this additional space in Lm-thiolase.

In the unliganded active sites of the wild-type enzyme Cys123 is oxidized to a sulfonate moiety and two different oxygen atoms of this sulfonate group bind in OAH1 and OAH2 (Harijan *et al.*, 2013), respectively, whereas in the unliganded C123A variant both OAHs are occupied by a water (Fig. 4).

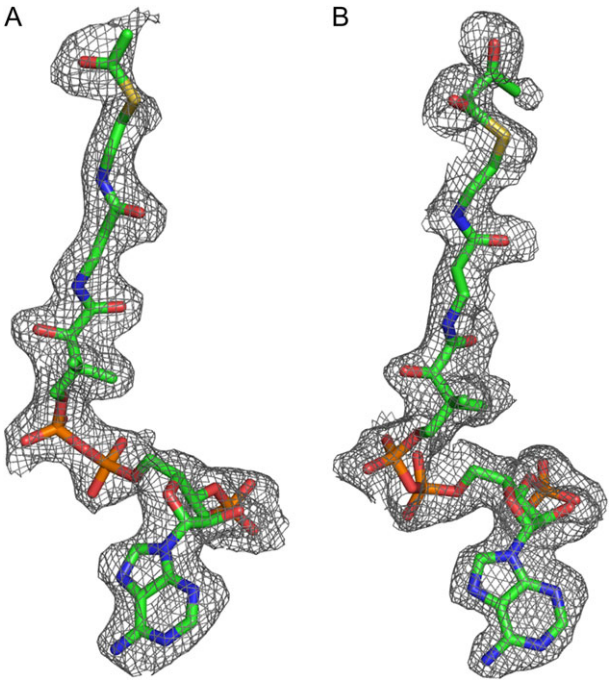

Fig. 3 Omit maps of the subunit B active site of C123A Lm-thiolase. (**A**) The omit (Fo–Fc) difference electron density map of the acetyl-CoA structure. Included is the acetyl-CoA as a stick model. (**B**) The omit (Fo–Fc) difference electron density map of the acetoacetyl-CoA structure. Included is the acetoacetyl-CoA as a stick model. The (Fo–Fc) difference maps have been calculated after 10 cycles of omit refinement by REFMAC5, leaving out the subunit B active site ligand. The contour levels are at 2.5σ and 2.0σ for the structure of the acetyl-CoA complex (PDB code 5LOT) and the structure of the acetoacetyl-CoA complex (PDB code 5LNQ), respectively.

The substrate interactions in OAH1 and OAH2 of the SCP2-thiolase (type-2)

The two structures capture the mode of binding of the substrates acetyl-CoA and acetoacetyl-CoA as the transfer of their acetyl moieties to the nucleophilic Cys123 is not possible because of the C123A mutation. As highlighted in Fig. 1, the acetyl moiety that is transferred to the nucleophilic cysteine is the C1-C2 fragment in case the substrate is acetyl-CoA and it is the C3-C4 fragment in case acetoacetyl-CoA is the substrate. It can be expected that the transferred acetyl moiety will adopt the same conformation as captured in the structure of the acetylated Zr-thiolase (Cys89), having its thioester oxygen bound in OAH2 (PDB code 1DM3). Therefore, some conformational changes of the S-acetoacetyl and S-acetyl moieties during catalysis in the Lm-thiolase active site are required such that the acetyl moiety that is transferred to SG(Cys123) adopts the mode of binding as seen in the Zr-thiolase structure (PDB code 1DM3) and such that SG(Cys340) can protonate the leaving group, being the terminal C2 methyl group of acetyl-CoA or the sulfur atom of CoA, respectively (Fig. 1).

The role of OAH1 and OAH2 for the chemistry of the CNH-thiolases has been described (Haapalainen *et al.*, 2006; Merilainen *et al.*, 2009), stabilizing the negatively charged oxygen atom of the enolate intermediate (OAH1) and the tetrahedral intermediate (OAH2). The GHP-histidine activates the nucleophilic cysteine by abstracting its proton and therefore it becomes positively charged (Kursula *et al.*, 2002). This geometry of the GHP-histidine is preserved in Lm-thiolase and therefore the OAH1 hydrogen bonding network in the active site of Lm-thiolase also suggests that His388 (GHP) is doubly protonated and positively charged like in the Zr-thiolase complexes. His338 (HDCF) appears to be neutral (like the Wat-Asn(NEAF) dyad in Zr-thiolase), as it is hydrogen bonded to ND2(Asn425), such that Asn425 is the hydrogen bond donor of the ND1(His338)-ND2(Asn425) hydrogen bond (Fig. 5). This implies that NE2(His338) must be protonated and that this atom functions as a hydrogen bond donor of OAH1. Asn425 is located in Cβ4 (Fig. 2) and it is a conserved residue for this subfamily of thiolases (Harijan *et al.*, 2013). The distance between NE2(His388) and NE2(His338) is 4 Å, disfavoring that both these side chains are positively charged. The hydrogen bond network of His338 shows how it mimics the hydrogen bond of the Wat-Asn(NEAF) dyad, both being neutral. Key features of the extended hydrogen bond networks of OAH1 of Zr-thiolase and Lm-thiolase are schematically shown in Fig. 5. A buried water molecule is part of the His(GHP) hydrogen bond network in both thiolases. Also conserved are a buried arginine (Arg356) in Zr-thiolase, corresponding to a buried lysine (Lys396) in Lm-thiolase and a conserved buried glutamate (Glu314) in Zr-thiolase, corresponding to Glu336 in Lm-thiolase. As has been described before, the waters that interact with Glu314/Glu336 and Arg356/Lys396 fill a polar cavity and are not hydrogen bonded to bulk water (Kursula *et al.*, 2002). A trail of water molecules, extending from the water of the Wat-Asn (NEAF) dyad (towards the back side of the molecule) in Zr-thiolase (Merilainen *et al.*, 2009), is absent in Lm-thiolase (Fig. 5).

Concluding remarks

A hallmark of the thiolase active site is the presence of two conserved oxyanion holes, OAH1 and OAH2. Spatially, these two oxyanion holes are close together (Fig. 4). In all studied thiolases OAH2 is formed by two main chain NH hydrogen bond donors, being part of the CxS and CxG loops, respectively. The OAH1 hydrogen bond donors of the best characterized thiolases, the CNH-thiolases, are formed by the Wat-Asn(NEAF) dyad and by the NE2 moiety of the GHP-histidine. The structures reported here show that for this leishmanial SCP2-thiolase (type-2), which is a CHH-thiolase, the Wat-Asn(NEAF) hydrogen bond donor of the CNH-thiolase OAH1 is replaced by the NE2-atom of the HDCF histidine. For OAH1 of this SCP2-thiolase (type-2) the hydrogen bond donors are therefore two protonated NE2(His)-atoms, of which His338(HDCF) is a singly protonated, uncharged histidine and His388(GHP) is a doubly protonated, positively charged histidine. Structural studies of TFE-thiolases and SCP2-thiolases (type-1) have been initiated to establish if for these CHH-thiolases the OAH1 geometry is the same as described here for the SCP2-thiolase (type-2).

Materials and methods

Cloning, expression and purification of C123A Lm-thiolase

The gene coding for the C123A variant of *L. mexicana* SCP2-thiolase (type-2) (Lm-thiolase) was cloned and over-expressed in *Escherichia coli* and purified as described earlier (Harijan *et al.*, 2013).

Cocrystallization experiments

Previous studies have shown that the catalytic cysteine of the unliganded wild-type Lm-thiolase was oxidized to cysteine sulfonate during the crystallization experiments (Harijan *et al.*, 2013). In this study, the C123A Lm-thiolase was used for the crystallographic

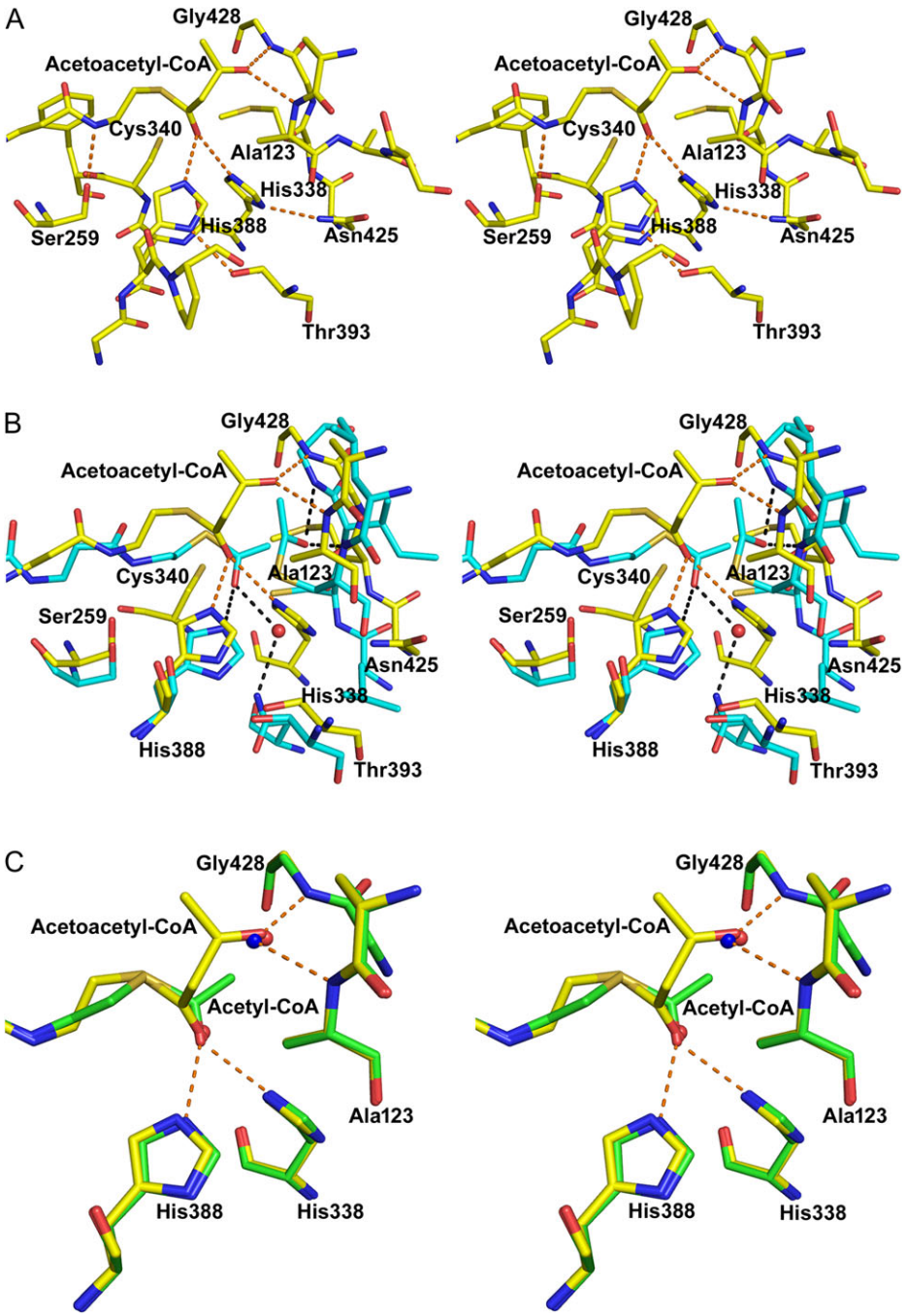

Fig. 4 The C123A Lm-thiolase active site complexed with acetoacetyl-CoA (subunit B) (carbons in yellow color). (**A**) The OAH1 and OAH2 interactions with acetoacetyl-CoA are highlighted. Also the hydrogen bond interactions of Ser259 (with N4P of acetoacetyl-CoA), Thr393 (with ND1 of His388) and Asn425 (with ND1 of His338) are shown by dotted lines. (**B**) Superimposed with the acetylated active site of Zr-thiolase, complexed with acetyl-CoA (subunit B, PDB code 1DM3) (carbons in blue color). Highlighted are also the hydrogen bond interactions of the hydrogen bond donors of OAH1 and OAH2 with orange dashed lines (Lm-thiolase) and black dashed lines (Zr-thiolase). (**C**) Superimposed with (i) acetyl-CoA, as bound in the C123A Lm-thiolase structure (subunit B) (carbons in green color), including also the water molecule bound in OAH2 of this active site (colored blue) and (ii) unligated active site of the acetoacetyl-CoA structure (subunit C), visualizing the presence of the two waters bound in OAH1 and OAH2 (colored red), respectively.

ligand binding studies with acetyl-CoA and acetoacetyl-CoA. 5.0 mg/ml C123A Lm-thiolase was incubated with 1.0 mM acetyl-CoA and 1.0 mM acetoacetyl-CoA in the gel filtration buffer (25 mM Tris (2-amino-2-hydroxymethyl-propane-1,3-diol), pH 7.8, 50 mM NaCl, 1 mM EDTA and 1 mM DTT) in Eppendorf tubes for 2 h on ice. Then, the crystallization properties of the incubated protein-ligand mixtures were screened by the sitting drop vapor diffusion crystallization method at 22°C using the Hampton (Crystal screen I & II), Molecular Dimension (Proplex, JCSG and PACT) and home-made factorial crystallization conditions (Zeelen *et al.*, 1994). The drops were set up in 96-well Corning plates using the TECAN Freedom EVO-100 crystallization robot. Each crystallization drop

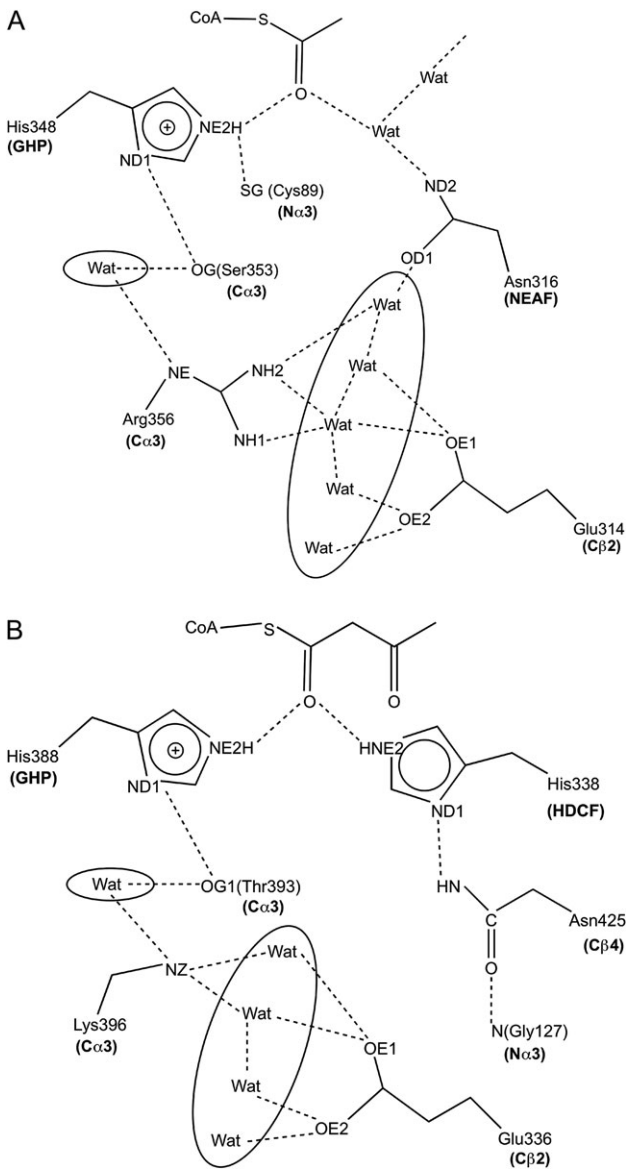

Fig. 5 The hydrogen bond network extending from OAH1. (**A**) In a CNH-thiolase, as observed in the structure of acetylated Zr-thiolase complexed with acetyl-CoA (PDB code 1DM3). (**B**) In a CHH-thiolase, as observed in the structure of the C123A Lm-thiolase, complexed with acetoacetyl-CoA (PDB code 5LNQ). The waters included are buried inside the matrix of the protein, either as a single water or as a cluster of waters.

contained 1.5 µl of protein-ligand buffer and 1.5 µl of well solution. The volume of the well solution was 80 µl. Good quality crystals of the acetyl-CoA complex were obtained in 0.1 M Tris, pH 8.0, 0.2 M lithium chloride and 20% w/v polyethylene glycol (PEG) 6000 and the crystals of the acetoacetyl-CoA complex were grown in 0.1 M 1,3-bis(tris(hydroxymethyl)methylamino)propane (Bis-tris propane), pH 7.5, 0.2 M sodium fluoride and 20% w/v PEG 3350. Further details of the crystallization experiments and the cryofreezing protocols are summarized in Table I.

Data collection and data processing

The crystals grew within 3–4 days and were then immediately stored in liquid nitrogen. The data from the acetyl-CoA complex crystal were collected to 2.25 Å resolution at a wavelength of 1.5418 Å using a Bruker Microstar rotating anode X-ray generator with a Cu anode and Platinum 135 CCD detector system. The crystal was aligned using the 4-circle KAPPA goniometer. These data were processed using the PROTEUM2 software suite (Bruker AXS Inc.) in space group $P2_1$. The diffraction data of the acetoacetyl-CoA complex crystal were collected to 1.98 Å resolution at beam line ID29, ESRF, Grenoble, France. The data were processed using the XDS suite (Kabsch, 2010) and scaled with the AIMLESS program of the CCP4 suite (Winn *et al.*, 2011) in the space group $P2_1$.

The data analysis calculations with the SFCHECK (Winn *et al.*, 2011) and XTRIAGE (Adams *et al.*, 2010) programs suggest that these crystals are not twinned. The Matthews coefficient (V_m) calculations reveal that the unit cell of the $P2_1$ crystal forms is compatible with four protein subunits in the asymmetric unit with 47% solvent content for the acetyl-CoA complex crystal and 42% for the acetoacetyl-CoA complex crystal. For both data sets the CC1/2 value for the data in the highest resolution shell is approximately 60% (Table II).

Structure solution and refinement

The structures of both complexes were solved by molecular replacement using PHASER (McCoy *et al.*, 2007). The dimeric unliganded Lm-thiolase structure was used as the initial phasing model. The model obtained from PHASER (McCoy *et al.*, 2007) was manually adjusted and completed using the graphics program COOT (Emsley *et al.*, 2010). Refinement was performed with the REFMAC5 program (Murshudov *et al.*, 1997), using standard protocols for the NCS refinement. The CoA molecule was left out from the models in the beginning of the refinement. After completion of the water structure, first the CoA moiety of the ligand was built in the corresponding density and further refinement was done. Finally, the acetyl and acetoacetyl moieties of the ligands were fitted in the respective active site densities. The final refinement statistics of both structures are summarized in Table II.

Structure analysis

The crystal structures of unliganded wild-type Lm-thiolase (PDB code 3ZBG), unliganded C123A Lm-thiolase (PDB code 3ZBK) and C123A Lm-thiolase (complexed with CoA) (PDB code 3ZBN) have been used for comparisons, using in each case subunit B. The structure of Zr-thiolase, in particular the complex of the acetylated enzyme with bound acetyl-CoA (PDB code 1DM3, subunit B), has been used as the reference for structural comparisons. The structure of C89A Zr-thiolase complexed with acetoacetyl-CoA (PDB code 1M1O, subunit B) has also been used for structural comparisons. All structural superpositions were achieved using the SSM protocol of COOT (Emsley *et al.*, 2010). The geometries of the final models were examined using the program PROCHECK of the CCP4 suite (Winn *et al.*, 2011). Further structure analyses, including the calculation of the B-factor profiles by BAVERAGE, were done using the CCP4 suite (Winn *et al.*, 2011). All figures were made with the molecular graphics program PyMOL. For both Lm-thiolase complex structures subunit B was used for the structural analyses and comparisons.

Acknowledgements

The intensity data were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, beam line ID29 and we thank the ID29

beam line scientists for their expert support. The use of the facilities and expertize of the Biocenter Oulu core facility, a member of Biocenter Finland and Instruct-FI, is also gratefully acknowledged. We also would like to acknowledge Dr Kristian Koski for collecting the data at ESRF and the expert support by Ville Ratas.

Funding

This research was supported through Academy of Finland grant 131795 and the European Commission's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570).

References

Adams,P.D., Afonine,P.V., Bunkoczi,G., *et al*. (2010) *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

Anbazhagan,P., Harijan,R.K., Kiema,T.R., Janardan,N., Murthy,M.R., Michels,P.A., Juffer,A.H. and Wierenga,R.K. (2014) *Tuberculosis (Edinb)*, **94**, 405–412.

Antonenkov,V.D., Van Veldhoven,P.P., Waelkens,E. and Mannaerts,G.P. (1997) *J. Biol. Chem.*, **272**, 26023–26031.

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

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

Fukao,T. (2002) *Wiley Encyclopedia of Molecular Medicine*, **Vol. 5**. John Wiley & Sons, Inc, New York, pp. 6–9.

Haapalainen,A.M., Merilainen,G., Pirila,P.L., Kondo,N., Fukao,T. and Wierenga,R.K. (2007) *Biochemistry*, **46**, 4305–4321.

Haapalainen,A.M., Merilainen,G. and Wierenga,R.K. (2006) *Trends Biochem. Sci.*, **31**, 64–71.

Harijan,R.K., Kiema,T.R., Karjalainen,M.P., Janardan,N., Murthy,M.R., Weiss, M.S., Michels,P.A. and Wierenga,R.K. (2013) *Biochem. J.*, **455**, 119–130.

Jiang,C., Kim,S.Y. and Suh,D.Y. (2008) *Mol. Phylogenet. Evol.*, **49**, 691–701.

Kabsch,W (2010) *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.

Kim,E.J. and Kim,K.J. (2014) *Biochem. Biophys. Res. Commun.*, **452**, 124–129.

Kim,J. and Kim,K.J. (2016) *Int. J. Biol. Macromol.*, **82**, 425–431.

Kim,S., Jang,Y.S., Ha,S.C., *et al*. (2015) *Nat. Commun.*, **6**, 8410.

Kursula,P., Ojala,J., Lambeir,A.M. and Wierenga,R.K. (2002) *Biochemistry*, **41**, 15543–15556.

Mann,M.S. and Lutke-Eversloh,T. (2013) *Biotechnol. Bioeng.*, **110**, 887–897.

Mathieu,M., Modis,Y., Zeelen,J.P., Engel,C.K., Abagyan,R.A., Ahlberg,A., Rasmussen,B., Lamzin,V.S., Kunau,W.H. and Wierenga,R.K. (1997) *J. Mol. Biol.*, **273**, 714–728.

Mathieu,M., Zeelen,J.P., Pauptit,R.A., Erdmann,R., Kunau,W.H. and Wierenga,R.K. (1994) *Structure*, **2**, 797–808.

Mazet,M., Harijan,R.K., Kiema,T.R., Haapalainen,A.M., Morand,P., Morales,J., Bringaud,F., Wierenga,R.K. and Michels,P.A. (2011) *Int. J. Parasitol.*, **41**, 1273–1283.

McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) *J. Appl. Crystallogr.*, **40**, 658–674.

Merilainen,G., Poikela,V., Kursula,P. and Wierenga,R.K. (2009) *Biochemistry*, **48**, 11011–11025.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

Reisse,S., Garbe,D. and Bruck,T. (2014) *Biochimie*, **103**, 16–22.

Sheppard,M.J., Kunjapur,A.M. and Prather,K.L. (2016) *Metab. Eng.*, **33**, 28–40.

Wang,Y., Yin,J. and Chen,G.Q. (2014) *Curr. Opin. Biotechnol.*, **30**, 59–65.

Winn,M.D., Ballard,C.C., Cowtan,K.D., *et al*. (2011) *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

Zeelen,J.P., Hiltunen,J.K., Ceska,T.A. and Wierenga,R.K. (1994) *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 443–447.