J. Mol. Biol. (1997) **273**, 714–728

JMB

The 1.8 Å Crystal Structure of the Dimeric Peroxisomal 3-Ketoacyl-CoA Thiolase of *Saccharomyces cerevisiae*: Implications for Substrate Binding and Reaction Mechanism

Magali Mathieu[1], Yorgo Modis[1], Johan Ph. Zeelen[1], Christian K. Engel[1] Ruben A. Abagyan[2], Anders Ahlberg[3], Bjarne Rasmussen[3] Victor S. Lamzin[4], Wolf H. Kunau[5] and Rik K. Wierenga[1]*

[1]*EMBL, Meyerhofstrasse 1 D69126, Heidelberg, Germany*

[2]*Skirball Institute, New York 10016, NY, USA*

[3]*EMBL Grenoble Outstation F38042, Grenoble, Cedex France*

[4]*EMBL Hamburg Outstation D22603, Hamburg, Germany*

[5]*Ruhr Universität Bochum D44780, Bochum, Germany*

**Corresponding author*

The dimeric, peroxisomal 3-ketoacyl-CoA thiolase catalyses the conversion of 3-ketoacyl-CoA into acyl-CoA, which is shorter by two carbon atoms. This reaction is the last step of the β-oxidation pathway. The crystal structure of unliganded peroxisomal thiolase of the yeast *Saccharomyces cerevisiae* has been refined at 1.8 Å resolution. An unusual feature of this structure is the presence of two helices, completely buried in the dimer and sandwiched between two β-sheets. The analysis of the structure shows that the sequences of these helices are not hydrophobic, but generate two amphipathic helices. The helix in the N-terminal domain exposes the polar side-chains to a cavity at the dimer interface, filled with structured water molecules. The central helix in the C-terminal domain exposes its polar residues to an interior polar pocket. The refined structure has also been used to predict the mode of binding of the substrate molecule acetoacetyl-CoA, as well as the reaction mechanism. From previous studies it is known that Cys125, His375 and Cys403 are important catalytic residues. In the proposed model the acetoacetyl group fits near the two catalytic cysteine residues, such that the oxygen atoms point towards the protein interior. The distance between SG(Cys125) and C3(acetoacetyl-CoA) is 3.7 Å. The O2 atom of the docked acetoacetyl group makes a hydrogen bond to N(Gly405), which would favour the formation of the covalent bond between SG(Cys125) and C3(acetoacetyl-CoA) of the intermediate complex of the two-step reaction. The CoA moiety is proposed to bind in a groove on the surface of the protein molecule. Most of the interactions of the CoA molecule are with atoms of the loop domain. The three phosphate groups of the CoA moiety are predicted to interact with side-chains of lysine and arginine residues, which are conserved in the dimeric thiolases.

Keywords: thiolase; CoA; crystal structure; reaction mechanism; β-oxidation

Introduction

Thiolases form an ubiquitous family of enzymes; they are found in prokaryotes and eukaryotes (Igual *et al.*, 1992). In eukaryotic cells thiolases are found in the cytosol, in microbodies (in particular in peroxisomes) and in mitochondria. Thiolases catalyse the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA. The covalent structure and nomenclature of acetoacetyl-CoA is shown in Figure 1(a). As is shown in

Present address: M.Mathieu, LEBS Bat34, 11, Av. de la Terrasse, F91198 Gif sur Yvette, France.

Abbreviations used: CoA, coenzyme A; DTT, reduced dithiothreitol; Mops, 3-(*N*-morpholino)-propane sulphonic acid; MPD, 2-methyl-2,4-pentanediol; NCS, non-crystallographic symmetry; PDB, Brookhaven Protein Data Bank; rms, root mean square.

0022–2836/97/430714–15 $25.00/0/mb971331

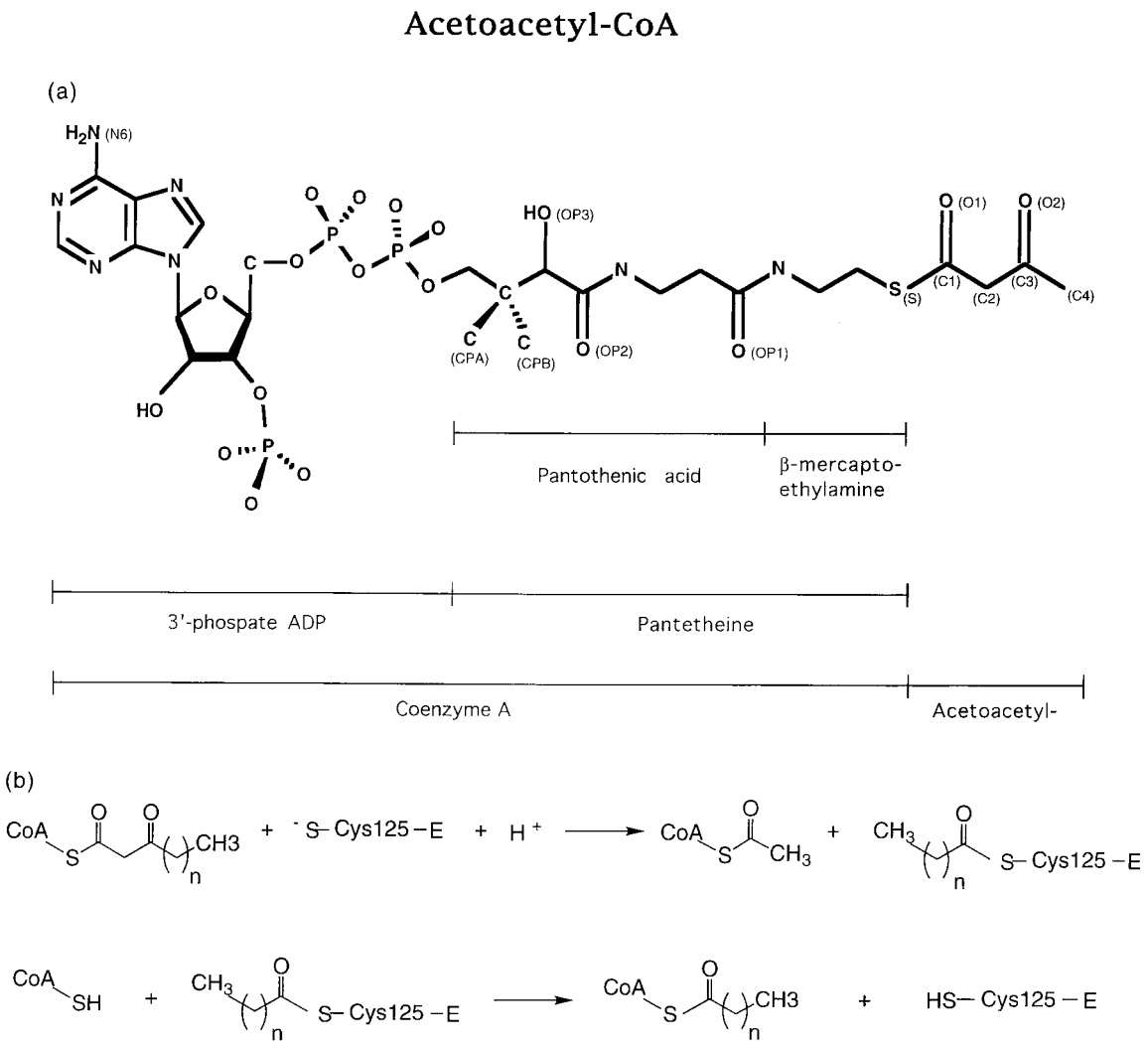

Figure 1. (a) The covalent structure of the acetoacetyl-CoA molecule. This molecule consists of several fragments, as indicated. The acetoacetyl moiety has several different resonance structures (Waterson & Hill, 1972), causing it to be planar, as also seen in crystal structures of this moiety which are present in the Cambridge Structure Database, as discussed (Engel *et al.*, 1996). (b) The two steps of the thiolytic degradation of a 3-ketoacyl-CoA molecule. A covalent intermediate with Cys125 is formed.

Figure 1(b) the cleavage reaction is a two-step reaction. The equilibrium of this reaction is in favour of the cleavage direction (Walsh, 1979), but all thiolases also catalyse the synthesis of acetoacetyl-CoA from two molecules of acetyl-CoA. On the basis of different substrate specificity and different function two classes of thiolases have been defined: thiolase-I (EC 2.3.1.16) and thiolase II (EC 2.3.1.9). Thiolase-I has a broad chain length specificity, being able to cleave acetoacetyl-CoA but also molecules with much longer fatty acid tails. This thiolase, which is important in the β-oxidation pathway for the degradation of fatty acids, will be referred to as the 3-ketoacyl-CoA thiolase. Thiolase II only acts on acetoacetyl-CoA (Kunau *et al.*, 1995). It can cleave acetoacetyl-CoA, but its main function is the synthesis of this compound from two acetyl-CoA molecules. This is a Claisen condensation reaction (Masamune *et al.*, 1989a,b). This thiolase, which is important for several biosynthetic pathways (Thompson *et al.*, 1989; Kunau *et al.*, 1995) will be referred to as the acetoacetyl-CoA thiolase.

This paper focuses on yeast peroxisomal 3-ketoacyl-CoA thiolase. This thiolase, first described in 1980 (Miyazawa & Hashimoto, 1980), is a homodimer, but most thiolases are homotetramers (Hartmann & Lynen, 1961; Gehring & Lynen, 1972; Kunau *et al.*, 1995). One thiolase has been reported to be a homohexamer (Kurihara *et al.*, 1989). Some biophysical and kinetic properties of the dimeric rat peroxisomal 3-ketoacyl-CoA thiolase have been described (Miyazawa *et al.*, 1981). The kinetic properties are very similar to the much more thoroughly studied mitochondrial tetrameric acetoacetyl-CoA thiolases (Staack *et al.*, 1978; Izbicka & Gilbert, 1984; Izbicka-Dimitrijevic & Gilbert, 1984; Salam & Bloxham, 1986; Gilbert, 1981; Gilbert *et al.*, 1981). The most detailed reaction mechanism studies concern the bacterial tetrameric acetoacetyl-

CoA thiolase of *Zoogloea ramigera* (Anderson *et al*., 1990; Williams *et al*., 1992). These studies have revealed the importance of two cysteine residues for the catalysis. In the yeast peroxisomal thiolase these are Cys125 and Cys403. From the *Z. ramigera* thiolase studies (but also from studies with tetrameric pig heart mitochondrial 3-ketoacyl-CoA thiolase (Gilbert, 1981; Gilbert *et al*., 1981)) it is known that a covalent intermediate exists, in which the fatty acid moiety is covalently bound to Cys125 (Figure 1). Mutagenesis of this residue into a serine (Thompson *et al*., 1989) reduces the catalytic rate. The importance of Cys403 is also shown by mutagenesis experiments (Palmer *et al*., 1991; Williams *et al*., 1992). Active site labelling studies indicate the importance of His375 for catalysis (Williams *et al*., 1992).

The 2.8 Å crystal structure of the unliganded dimeric peroxisomal 3-ketoacyl-CoA thiolase of the yeast *Saccharomyces cerevisiae* has been reported (Mathieu *et al*., 1994). Here we discuss the structure of this thiolase after refinement at 1.8 Å resolution. From the 1.8 Å structure a plausible model for the mode of binding of acetoacetyl-CoA has been deduced, including a detailed description of the possible reaction mechanism. The mode of binding of acetoacetyl-CoA is predicted by manually docking this molecule onto the surface of thiolase followed by energy minimisation. Based on the structure of the dimeric thiolase, using also the sequence information from the family of thiolases, a model is proposed how two dimeric thiolase molecules can assemble to form the tetrameric thiolase.

Results and Discussion

Quality of the model

In the 1.8 Å structure of thiolase there is continuous density for the subunit-1 main-chain from residues Lys25 to Ile262 and from Gly266 to Glu417. For subunit-2 a model has been built for residues Lys25 to Glu417. Previous studies (Mathieu *et al*., 1994) have shown that the complete protein, residues 1 to 417, has been crystallised; apparently the first 24 residues, which include the peroxisomal targetting sequence (Erdmann, 1994), are disordered. Residues 25 to 154 and 277 to 293 are in the N-domain, residues 155 to 276 define the loop domain and residues 300 to 417 are in the C-domain (Figures 2 and 3). In the 2.8 Å structure 87 residues (in the loop domains of subunit-1 and subunit-2) could not be built, but in the 1.8 Å model the structure of only three residues are not defined by the electron density map (residues 263 to 265 of the loop domain of subunit-1). The 2.8 Å structure (1PXT in the PDB) and the 1.8 Å structure agree very well: leaving out only a few residues at the beginning or end of the gaps in the 2.8 Å structure results in an rms difference for superimposed C^{α}-atoms of 0.35 Å for subunit-1 (leaving out two residues) and 0.44 Å for subunit-2 (leaving out three residues).

There are a few peaks above 4 σ in the final difference Fourier map. The highest peaks (6 σ) are correlated with structural heterogeneity of side-chains, in particular near Thr307 (subunit-1) and Arg430 (subunit-2). High difference Fourier density near the sulphur atom of Cys353 (subunit-2) suggests that this sulphur is partially oxidised. There are no difference Fourier peaks in the active site pockets, therefore there is no evidence for oxidation of any of the active site cysteine or methionine residues.

The model has good geometry (Table 4). The Ramachandran plot shows six outliers. It concerns, in both subunits, Phe50, Gln124, and Thr406. The (ϕ,ψ) values are (73, − 42), (49, − 135) and (92,7), respectively, for the subunit-1 residues and (71, − 40), (51, − 133) and (86,11), respectively, for the subunit-2 residues. The structures of these residues are well defined by the electron density map. They are at or near the active site. Gln124 is just before the catalytic residue Cys125 and Thr406 is just after the catalytic residue Cys403. Gln124 and Thr406 are also at the dimer interface.

The two subunits are related to each other by a non-crystallographic 2-fold axis (K = 180.0, as calculated from the superposition of all C^{α} atoms). The two subunits superimpose very well with an rms difference for all C^{α} atoms of subunit-1 and subunit-2 of 0.34 Å. Figure 4 shows that the largest differences between the C^{α}-traces of the two subunits are near residue 75, which is in the loop after $N^{\alpha 1}$, as well as in the loop domain, residues 155 to 276. Figure 5 shows the *B*-factor plots for subunit-1 and subunit-2. The average *B*-factors of the loop domain residues are higher than for the N-domain and the C-domain (Figure 5), except for residues near Met186. Met186 is in the active site pocket. The lowest *B*-factors are observed in the N-domain, consistent with the fact that the N-domain is at the dimer interface at the centre of the molecule.

The fully refined structure of this 3-ketoacyl-CoA thiolase provides a good model for a detailed analysis of the active site pocket and for modelling the acetoacetyl-CoA into its assumed binding pocket.

Sequence alignment

Sequence alignments have been made using the sequences of seven dimeric 3-ketoacyl-CoA peroxisomal thiolases and 14 tetrameric thiolases, as found in sequence databanks. Two tetrameric thiolases are mitochondrial 3-ketoacyl-CoA β-oxidation enzymes and the other 12 are acetoacetyl-CoA thiolases. These sequences show a high degree of sequence similarity. The residues conserved in all 21 thiolases are indicated in Figure 3 (11% sequence identity; Asp/Glu, Arg/Lys, and Ser/Thr are grouped together). Within the family of dimeric thiolases there is 38% sequence identity (Figure 3). The positions of the residues conserved

in the 14 tetrameric thiolases (14% sequence identity) cluster around the active site and the dimer interface. The sequence similarity between the dimeric and tetrameric thiolases indicate that the fold of the tetrameric thiolases and the dimeric thiolases may be very similar, which also implies that the tetrameric thiolases will be an assembly of two yeast thiolase dimers.

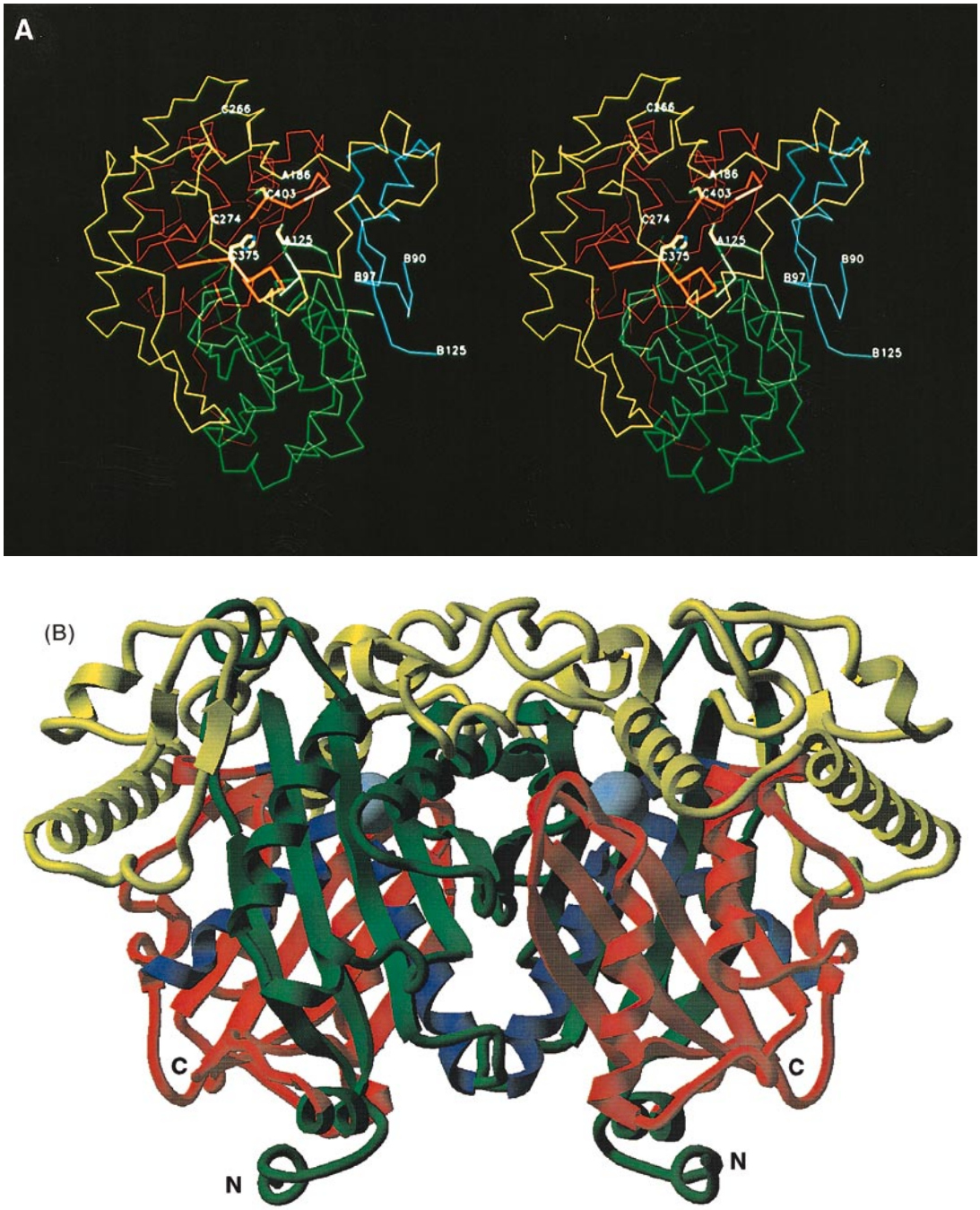

Figure 2. (A) Stereo view into the active site pocket, along the dimer 2-fold axis (top view). A complete C^{α}-trace of subunit-1 is shown. The N-domain is in green, the C-domain is in red and the loop domain is in yellow. Subunit-2 is in blue. The dimer 2-fold axis is near residue B97; the predominant contacts across the dimer interface are between the two N-domains. The loop domain circles around the active site; the side-chains of the catalytic residues Cys125, His375 and Cys403 are explicitly drawn. The three catalytic loops (123–128, 373–381 and 402–408) are highlighted. The catalytic loops 123–128 and 375–381 continue in the two central helices, $N^{\alpha 3}$ and $C^{\alpha 3}$, respectively. (B) Side view of the dimer (rotated 90° around the horizontal with respect to (A); both subunits are color-coded as for subunit-1 in (A). The N terminus (25) and the C terminus (417) are solvent exposed. The two central helices $N^{\alpha 3}$ (125 to 140, near the dimer interface) and $C^{\alpha 3}$ (375 to 390) are in blue. The N termini of these two helices are near the active site, which is marked by a blue sphere, centred at SG(Cys125). The green loop (of the N-domain) contributing to the flat appearance of the surface (top side) consists of residues 48 to 51. This Figure was made with ICM (Abagyan, 1997).

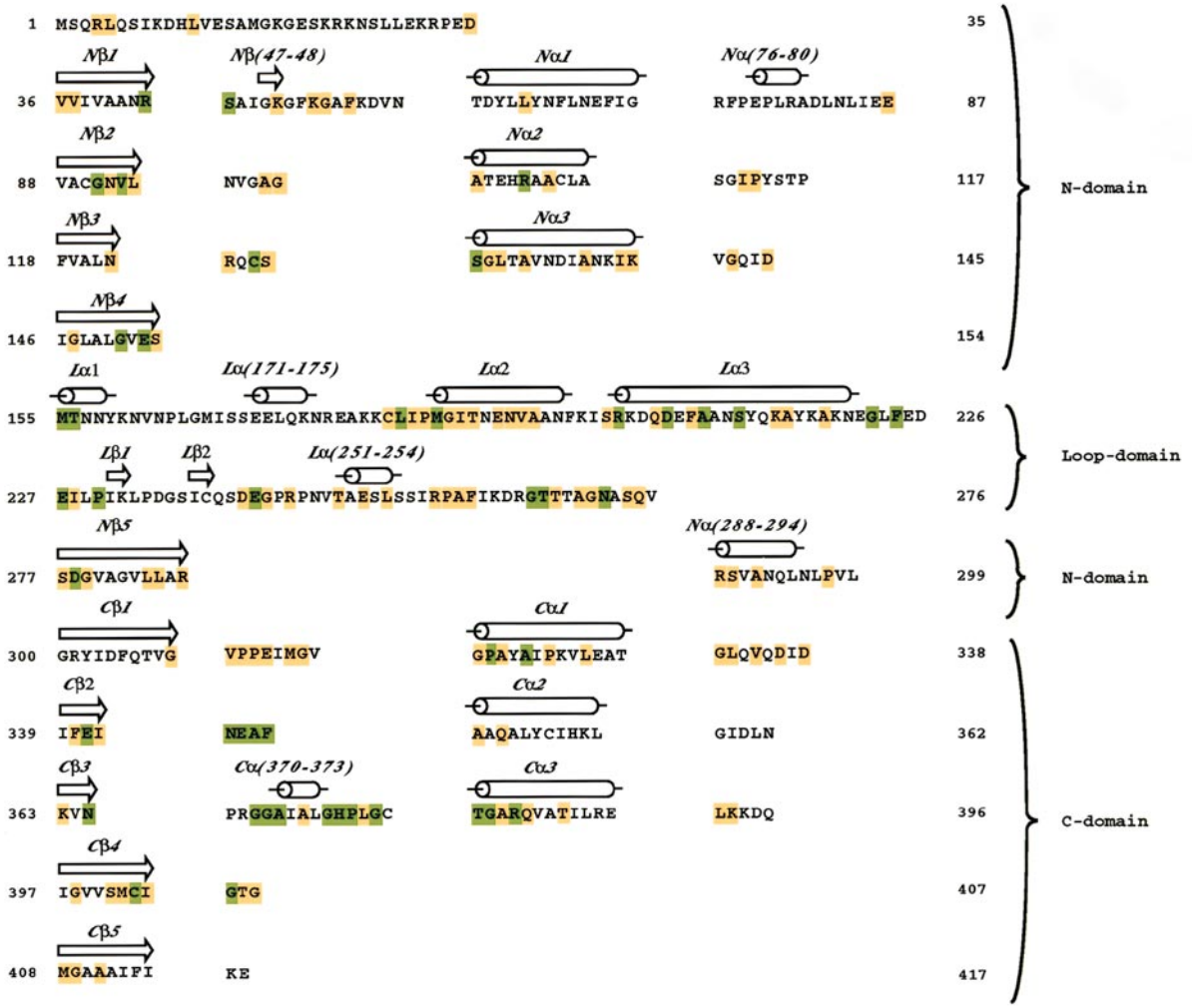

Figure 3. The sequence and secondary structure of yeast thiolase. The secondary structure assignment is according to Kabsch & Sander (1983). The residues conserved (Ser/Thr, Asp/Glu and Lys/Arg have been grouped together) in all sequences are coloured green; the residues conserved only in the dimeric thiolases are coloured yellow. The layout of the Figure emphasises the regular βα-repeats of the N-domain and C-domain, which have the same topology. The nomenclature of the secondary structure elements is the same as reported previously (Mathieu *et al.*, 1994).

Crystal contacts

Subunit-1 has relatively few crystals contacts (91 contacts within 4 Å), whereas subunit-2 has more (206). A very extensive crystal contact is the contact of loops 74–83 and 289–295 of subunit-2, which penetrate into the entrance of the active site pocket of subunit-2 of a crystallographically related molecule, formed by the loop regions 50–52 and 259–262. The latter loop region is mobile in subunit-1 but structured in subunit-2. The structures of the loop region 74–83 of subunit-1 and subunit-2 are different (Figure 4).

Cavities, buried charged residues and the dimer interface

The cavity calculations (Connolly, 1985) were done in the presence and absence of solvent molecules. Calculations in the presence of the solvent molecules show only small cavities, with volumes less than 25 $Å^3$. In the absence of solvent molecules the largest cavity (170 $Å^3$) is at the dimer interface. There are several other cavities at the dimer interface with volumes of approximately 40 $Å^3$. All these cavities are filled with water molecules. There is one water-filled cavity (39 $Å^3$) near each active site. This cavity is filled with two water molecules. These water molecules are part of a hydrogen bonding network, which includes the side-chains of Asn343, Glu341 and Glu391. The buried carboxylate groups of Glu341 and Glu391 are hydrogen-bonded with each other, indicating that one of them has to be protonated, as observed in other protein structures as well (Flocco & Mowbray, 1995). These side-chains in fact form a cluster of buried polar groups with hydrogen bonding interactions towards the active site as well as to the other side of the molecule. There are several other buried side-chains of glutamate (five per monomer), aspartate (two per monomer), arginine (four per monomer) and lysine residues (one per

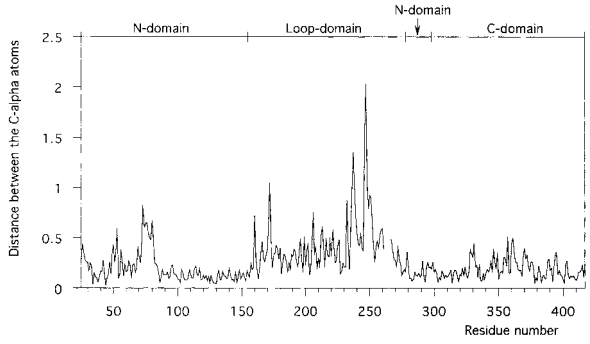

Figure 4. The C^{α}-distances between superimposed subunit-1 and subunit-2 (shown on the *y*-axis) as a function of residue number (shown on the *x*-axis).

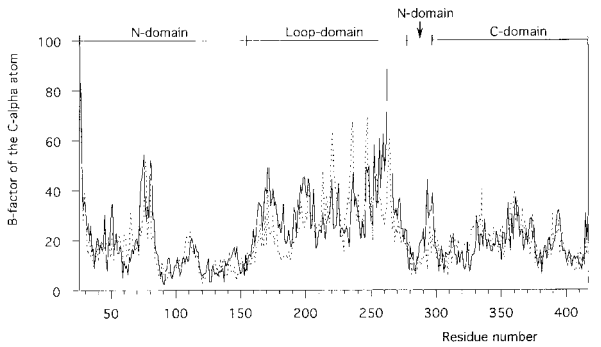

Figure 5. The *B*-factors of the C^{α}-atoms of subunit-1 (continuous line) and subunit-2 (dotted line).

monomer). Most of these side-chains are involved in salt bridges (Table 1). All these side-chains participate in good hydrogen bonding networks and all but one are hydrogen-bonded to at least one buried water molecule (Table 1).

The two central helices ($N^{\alpha 3}$ and $C^{\alpha 3}$) are completely buried in the dimer (Figure 2). Nevertheless, both helices have a polar and an apolar side. In the $N^{\alpha 3}$ helix the polar residues are facing the large dimer interface cavity (filled with water molecules). In the $C^{\alpha 3}$ helix the polar residues (Arg383, Gln384 and Thr387) participate in the buried cluster of polar residues mentioned above. This cluster of polar residues is buried between the β-sheet of the C-domain and the $C^{\alpha 3}$ helix. This is a peculiar but conserved feature of the thiolase fold, because the residues involved in this cluster are highly conserved in the sequences (Figure 3).

The dimer interface is extensive. The molecular surface area buried at the dimer interface is 1400 $Å^2$ per subunit. There are two salt bridges at the dimer interface. Most of the monomer-monomer contacts concern residues 161 to 179 of the loop domain and the residues of helix $N^{\alpha 2}$, strand $N^{\beta 3}$ and helix $N^{\alpha 3}$. As can be seen in Figure 2, the dimer interface has a rather flat surface; the only exceptions are the loop residues 161 to 179, which protrude out of the monomer and lay on top of helix $N^{\alpha 2}$ of the other subunit. Although these dimer interface regions are not directly involved in the catalysis, they are part of the active site architecture. For example, the catalytic Cys125 is in the loop between $N^{\beta 3}$ and $N^{\alpha 3}$. The extensive interactions seen at the dimer interface contribute to the observed rigid architecture of the two active sites.

The active site pocket

The catalytic residues Cys125, His375 and Cys403 are at the bottom of a deep cylindrical pocket with a radius of approximately 10 Å. Cys125 and His375 are solvent-exposed, whereas Cys403 is almost completely buried (Figure 6). Also at the bottom of this pocket is Gln124 (Figure 6). Several different regions can be identified in the walls of this pocket. There are two hydrophobic areas, which are located on opposite sides of the pocket. One region (the ''upper'' side) is near Cys403, consisting of Phe346, Met186,

Table 1. Structural data of basic and acid residues whose polar side-chain atoms are buried

	Secondary structure element	Contacting waters	Salt bridge[a] interactions	Conserved in dimers and tetramers	At dimer interface
Glu153	$N^{\beta 4}$	1	–	Yes	–
Glu227	Loop ($L^{\alpha 3}$-$L^{\beta 1}$)	2	Arg 43/W..R367	Yes	–
Glu341	$C^{\beta 2}$	3	W…R383	Yes	–
Glu344	Loop ($C^{\beta 2}$-$C^{\alpha 2}$)	1	W…K214	Yes	–
Glu391	$C^{\alpha 3}$	1	–	No[b]	–
Asp134	$N^{\alpha 3}$	2	Lys138	No[b]	Yes
Asp 278	$N^{\beta 5}$	1	–	Yes	–
Arg 43	$N^{\beta 1}$	0	Glu227, Asp242	Yes	–
Arg104	$N^{\alpha 2}$	1	–	Yes	Yes
Arg123	Loop ($N^{\beta 3}$-$N^{\alpha 3}$)	1	Glu87[c]	Yes	Yes
Arg383	$C^{\alpha 3}$	2	W…Glu341	Yes	–
Lys138	$N^{\alpha 3}$	1	Asp 134, Glu 87	No[b]	Yes

[a] W indicates a water mediated salt bridge.
[b] These residues are also not conserved in the dimer-only family of sequences.
[c] Of the other subunit.

Met315, Ala345, Phe261, and Ala273. The other region (the ''lower'' side) is near Leu94 and includes residues Val162, Val93, Leu377, Tyr159, Val276, Met155, and Phe50 (Figure 6). These rather extensive hydrophobic walls are separated from each other by the side-chain of Ser274, which is a conserved residue (Figure 3). The two hydrophobic walls, together with Ser274, make up half of the cylindrical pocket. The remaining semi-circle of the cylindrical pocket is shaped by residues 163 to 185 of the loop domain of the same subunit and residues 99 to 108 of helix $N^{\alpha 2}$ of the other subunit, in particular Thr101 (Figure 6), which side-chain is hydrogen-bonded to the side-chain of Gln124.

There are no side-chains of charged residues (Arg, Lys, Asp, Glu) within 9 Å of SG(Cys125) of the catalytic site, except for the conserved residue Glu153 (see also Table 1). The carboxylate oxygen atoms of this residue are just 9 Å away from SG(Cys125). It is interesting to point out that within this 9 Å sphere there is a high concentration

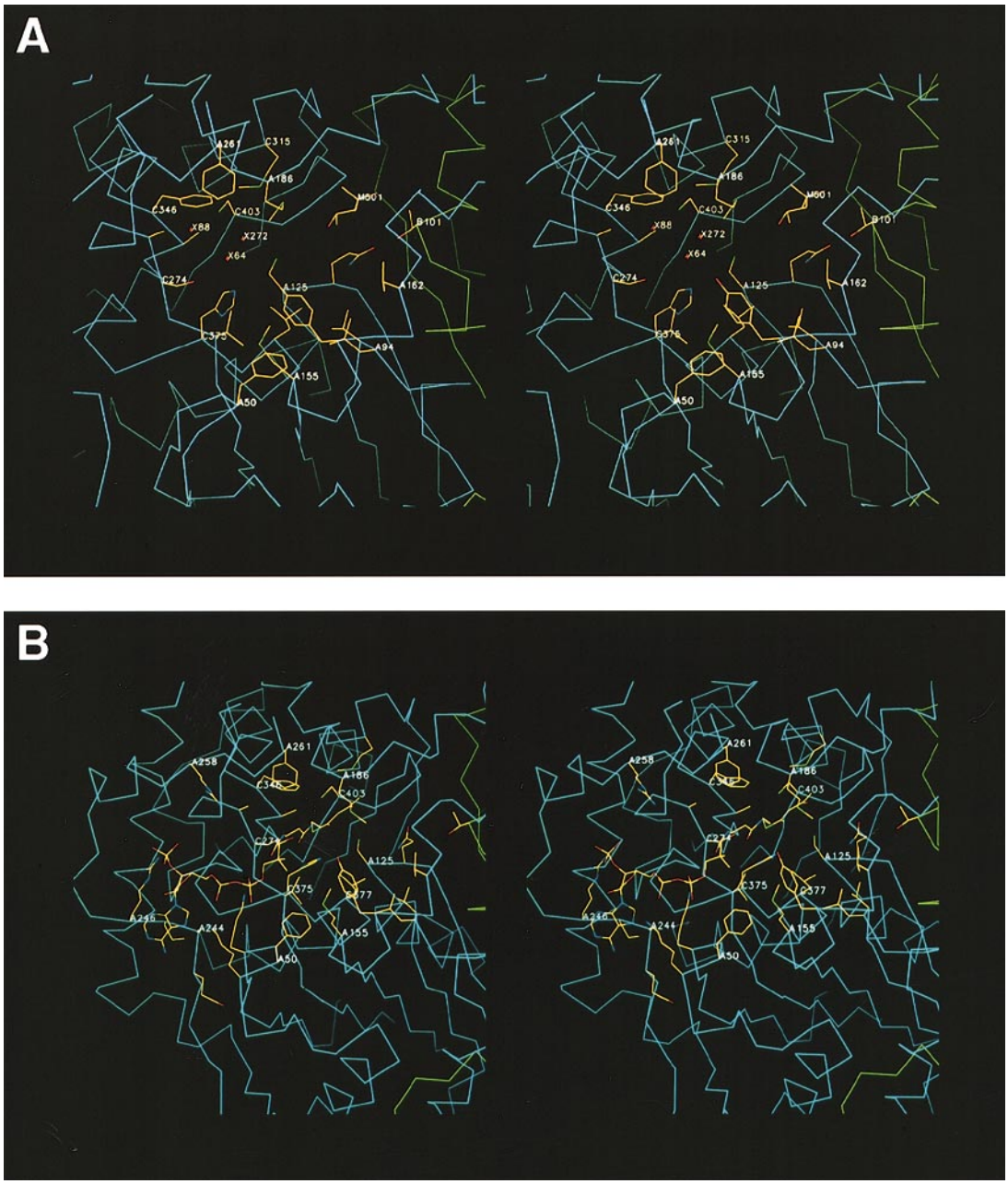

Figure 6. (A) Stereo view into the active site pocket (top view). The C^{α}-traces of subunit-1 and subunit-2 are in blue and green, respectively. The side-chains of the residues forming the hydrophobic walls (see the text) are shown. M501 labels the MPD molecule. X64 and X272 are the two water molecules between SG(Cys125) and SG(Cys403). Water X88 is behind SG(Cys403). Thr101, of the other subunit, is in hydrogen bonding contact with Gln124. (B) Stereo view into the active site of the modelled complex (top view). The interactions of the labelled side-chains with the modelled acetoacetyl-CoA group are discussed in the text.

Table 2. The polar interactions of the catalytic residues within 4.2 Å

Atom of the catalytic residue	Contacting atom	Distance (Å)
SG (Cys125)	N (Cys125)	2.9
	N (Cys405)	3.5
	O (Cys403)	4.1
	NE2 (His375)	4.2
	WAT272	3.0
	WAT423	3.2
	WAT64	3.9
SG (Cys403)	N (Cys403)	3.2
	NE2 (Gln349)	3.6
	OE1 (Gln349)	4.0
	WAT88	2.8
	WAT272	3.5
	WAT64	3.6
NE2 (His375)	ND2 (Asn343)	3.6
	SG (Cys125)	4.2
	WAT64	3.7
	WAT272	3.9
	WAT85	4.0
ND1 (His375)	OG1 (Thr380)	3.0
	ND2 (Asn343)	3.5

of methionine side-chain atoms (five residues out of a total of six).

The three catalytic residues are in three loops with highly conserved sequences (Figure 3). The Cys125 loop (N-domain) and the His375 loop (C-domain) continue into the buried central helices, $N^{\alpha 3}$ and $C^{\alpha 3}$, respectively (Figure 2(A)). The Cys403 loop is a tight turn between the two antiparallel strands ($C^{\beta 4}$ and $C^{\beta 5}$) of the C-terminal domain. The distance between SG(Cys125) and SG(Cys403) is 5.8 Å. Other contact distances in the active site are given in Table 2. The side-chain atom ND1 of His375 is hydrogen-bonded to Thr380 (Table 2). NE2(His375) interacts with SG(Cys125). The His375 hydrogen bonding suggests that the NE2 atom is not protonated, indicating that the His375-Cys125 interaction will activate the SG(Cys125) for nucleophilic attack (see also Figures 7 and 9).

Unexpectedly, one MPD molecule (used as a cryoprotectant) is bound near the active site pocket of each subunit (Figure 6(A)). The MPD binding site is between residues 164 to 166 and 182 to 186 of the loop domain. The mode of binding is well defined by the electron density map. The MPD molecule interacts with the protein only *via* van der Waals interactions. The two MPD oxygen atoms make hydrogen bonds to protein-bound water molecules. The binding affinity of MPD for thiolase is unknown. MPD is a very poor inhibitor of the thiolase reaction (U. Dorpmund & W.-H. Kunau, unpublished observations); this could be because of low affinity or because MPD binding does not effect substrate binding.

In the unliganded active site pocket there is an extensive network of water molecules. This network includes WAT272 and WAT64, which bind between SG(Cys125) and SG(Cys403) and which are also near NE2 (His375) (Table 2 and Figure 6(A)). WAT88, which is buried behind

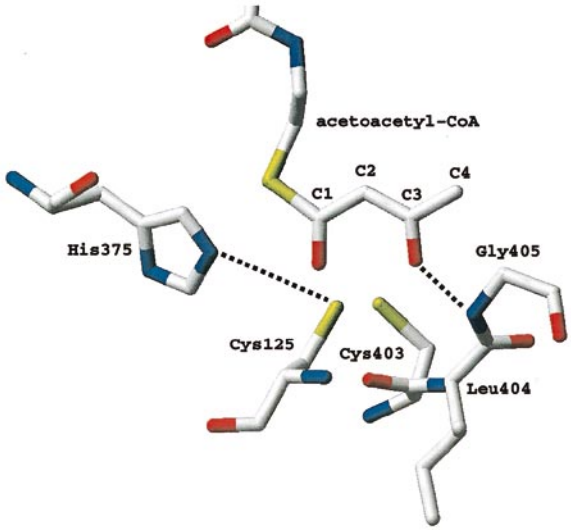

Figure 7. Schematic picture indicating the relative position of the modelled acetoacetyl group and the catalytic residues. Important hydrogen bonds are highlighted. The (SG(Cys125)–C1) and the (SG(Cys125)–C3) distances are 3.7 Å and 3.7 Å, respectively. The (SG(Cys403)–C2) distance is 4.5 Å.

SG(Cys403) could be important for the reaction mechanism, as it interacts tightly with SG(Cys403) (Table 2 and Figure 6).

The predicted mode of binding of acetoacetyl-CoA

Attempts to bind CoA to the crystalline thiolase have failed, because crystals transferred to a mother liquor with CoA deteriorate to the extent that data collection is impossible (Mathieu, 1995). Therefore, it has been attempted to predict the mode of binding of acetoacetyl-CoA into the active site pocket. The modelling of the CoA moiety is difficult, because there are no known CoA binding motifs, as was pointed out in a recent review on structural features of the known structures of protein-CoA complexes (Engel & Wierenga, 1996). The modelling was done in two steps. First the acetoacetyl-CoA molecule was manually docked onto the surface of thiolase; subsequently the docked complex was regularised and energy minimised with the ICM package (Abagyan *et al.*, 1994). Only side-chain atoms and CoA atoms were allowed to move in the minimisation calculations. The final structure of acetoacetyl-CoA has good geometry and there are no clashes between the docked acetoacetyl-CoA molecule and the protein atoms.

The docking experiments of the acetoacetyl moiety are facilitated by the notion that this group is planar (Figure 1). The knowledge about the reaction mechanism suggests that: (i) the C3 carbon should be close to Cys125, because the Cys125-S-acyl complex is known to be an intermediate (Figure 1); and (ii) the C2 carbon atom has to be

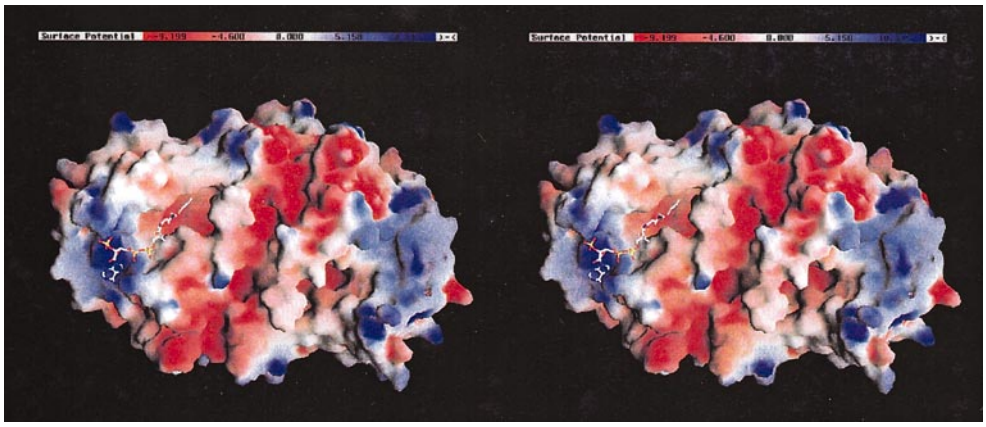

Figure 8. The surface of thiolase, colour-coded according to the electrostatic potential by GRASP (Nicholls *et al.*, 1991) (top view). Also shown is the predicted mode of binding of acetoacetyl-CoA (the side-chain of Met186 has been removed from the calculation, in order to see the atoms of the acetoacetyl moiety).

near to Cys403 (see also Figure 9). These geometric requirements suggest that the acetoacetyl group has to be oriented as shown in Figure 6(B), with the two oxygen atoms pointing towards the bottom of the active site pocket and the S-atom of the CoA moiety pointing upward. The acetoacetyl group is locked into a position near Cys125 and Cys403. The two water molecules, WAT272 and WAT64, seen in the unliganded structure (Figure 6) will be displaced on binding of the substrate. In this mode of binding O1 is near SG(Cys125) and NE2(His375), whereas O2 is hydrogen bonded to N(Gly405) (Figure 7). The O2-N(Gly405) distance is 3 Å. This mode of binding is only possible when the Met186 side-chain adopts a conformation that is different from what is observed in the crystal structure.

The CoA group points out of the binding pocket as shown in Figures 6(B) and 8; the fatty acid tail (extending from the C4 atom) is pointing upwards as well, and can therefore fill the remaining volume of the binding pocket. The CoA moiety has three phosphate groups, which presumably interact with conserved positively charged side-chains on the surface of thiolase. In the proposed mode of binding the phosphate groups can interact with Lys48, Lys51, Arg246 and Arg258 (Figures 6(B) and 8). These residues are conserved in the family of dimeric thiolases (Figure 3). In the predicted mode of binding the CoA-molecule adopts an extended conformation, as seen in other CoA-protein complexes (Engel & Wierenga, 1996). One of the few common features in the structures of CoA-protein complexes are the hydrogen bonds of the adenine amino group (N6) with the protein. This is also observed in the modelled complex, in which the N6 atom points to the main-chain oxygen atoms of Glu243, Gly244 and Pro245.

The β-mercaptoethylamine moiety is modelled between the side-chains of Phe346 on the upper side and Met155 and Val276 on the lower side of the binding pocket (Figure 6(B)). The OP1 and OP2 atoms of the pantothenic acid point inwards to the side-chain of Ser274 and N(Val276), respectively (Figure 6(B)). The two methylgroups (CPA and CPB) are in van der Waals contact with Phe50, Val276 and Tyr159. The docked CoA molecule fits into the groove formed by loops 48–51 (loop Nβ1–Nα1 in the N-domain) and 259–262 of the loop domain (Figure 6(B)). An analysis of the van der Waals surfaces of the docked complex shows that this groove is somewhat too wide for a perfect fit, suggesting that small loop movements might be induced by the CoA binding; possibly this concerns loop 259–262, which has very high *B*-factors in this structure (Figure 5). It is interesting to point out that the crystal packing is such that in subunit-2 the pocket between these two loops is involved in crystal contact interactions. The presence of these crystal contacts would explain why the thiolase crystals dissolve in solutions with CoA, making this crystal form unsuitable for crystallographic binding studies.

In the proposed model CoA interacts almost exclusively with residues of the loop domain. The exceptions are residues Lys48, Lys51 and Phe50 which are in the ($N^{\beta 1}$–$N^{\alpha 1}$)-loop of the N-domain.

The mode of binding of the acyl chain of the 3-ketoacyl-CoA moiety has not been modelled. Kinetic measurements indicate that the acyl chain does contribute to the binding energy, because the K_m value of 3-ketohexadecanoyl-CoA (1.9 μM) is seven times smaller than for acetoacetyl-CoA (14.0 μM; Miyazawa *et al.*, 1981). It seems likely that the acyl moiety will interact with hydrophobic side-chains of the binding pocket. As described above there are two hydrophobic ''walls'' in the binding pocket. The acetoacetyl moiety binds close to the hydrophobic wall near Cys403. In this binding mode there is still space left for binding of the acyl moiety to the hydrophobic wall (near Leu377) of the cyclindrically shaped active site pocket (Figure 6). In this way the acyl chain would curl around the acetoacetyl moiety and bind at the bottom of the active site pocket. This mode of binding would also explain why acyl chains of several different chain lengths can bind.

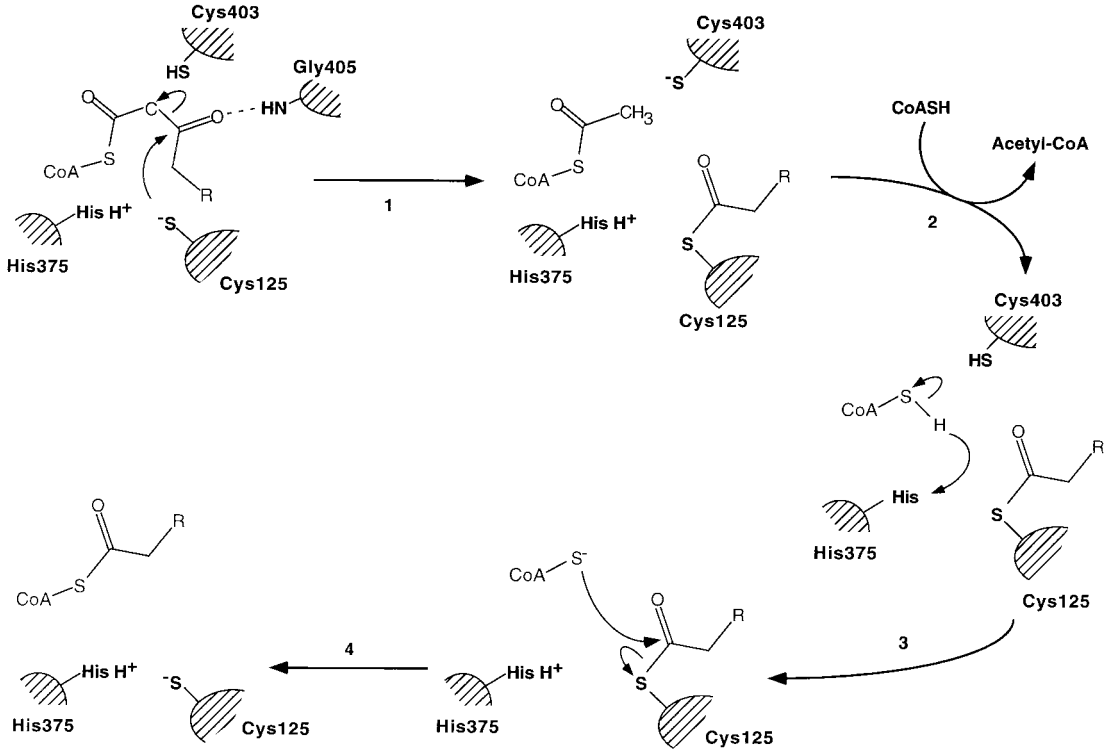

Figure 9. The proposed reaction mechanism for the degradative reaction of thiolase. Four steps are emphasised: 1, the formation of the covalent intermediate; 2, the replacement of acetyl-CoA by CoA; 3, the activation of CoA; and 4, the formation of the product.

The reaction mechanism

The different steps of the reaction mechanism are shown in Figure 9. The first step is the nucleophilic attack of SG(Cys125) on the C3 carbon atom, which is facilitated by the predicted hydrogen bond between O2 and N(Gly405) (Figure 7). This results in breaking the C2–C3 bond, producing a C2 carbanion which is subsequently protonated, possibly by a proton of the SG atom of Cys403, as deduced from mutagenic studies with *Z. ramigera* acetoacetyl-CoA thiolase (Palmer *et al.*, 1991; Williams *et al.*, 1992). This generates acetyl-CoA, which dissociates out of the active site pocket and the fatty acid tail remains bound to Cys125. After release of the acetyl-CoA, a free CoA molecule binds to the enzyme (step 2). It is assumed that CoA binds in the same way as the CoA moiety of the substrate, with the S atom close to His375. In this way it is predicted that the CoA-S atom can be activated by either the ND1(His375) or the NE2(His375) atom (step 3). Activation by ND1(His375) would require a 180° rotation of the side-chain with respect to the position observed in this crystal form. The crystal structure of the thiolase-CoA complex could clarify this point. The activated CoA molecule reacts with the acyl-moiety of the covalent intermediate which results in the formation of a new molecule of acyl-CoA (step 4).

It is interesting to point out that in the first step the CoA-S atom is separated by two carbon atoms from the C atom which becomes bound to SG(Cys125), while in the fourth step the CoA-S atom becomes covalently linked to a SG(Cys125)-bound C atom (Figure 9). This suggests structural rearrangements of the S-acyl chain between step 1 and step 4. Water molecules are not directly involved in the reaction mechanism, but WAT88 (near SG(Cys403)) could be important for the reactivity of SG(Cys403) (Table 2 and Figure 6(A)).

The active site of thiolase can also catalyse the synthesis of acetoacetyl-CoA from two molecules of acetyl-CoA. The first step of this reaction is the binding of acetyl-CoA, which is converted into a CoA molecule and a thiolase-Cys125-S-acetyl complex. For this reaction to occur SG(Cys125) has to react with the C1 carbon (not the C3 carbon atom as in the degradative reaction). In our model the distance of SG(Cys125) to the C1 atom is 3.7 Å (Figure 7). Therefore, it seems that acetyl-CoA and acetoacetyl-CoA can bind in the same way, but react differently with the SG(Cys125) atom. The hydrogen bond between O2 and N(Gly405) (Figure 7), which exists only in the presence of the acetoacetyl moiety, will favour the reaction with the C3 atom, as is required for the catalysis of the cleavage reaction.

The predicted structure of the thiolase tetramer

Tetrameric thiolases can be either 3-ketoacyl-CoA thiolases or acetoacetyl-CoA thiolases. The tetrameric thiolases are more widespread than the

dimeric thiolases, which occur only in the peroxisomes (Igual *et al*., 1992). The high sequence similarity between the tetrameric and dimeric thiolases suggests that the tetrameric thiolases are assembled as a dimer of two thiolase dimers. It is therefore very likely that the tetrameric thiolase has tetrahedral symmetry, with three perpendicular, intersecting 2-fold axes. One of these axes is the dimer 2-fold axis. This allows for only two possible modes of assembly: either the dimer/dimer interface is on the same side as the active sites (Figure 2), or it is on the opposite side, near the N and C termini. As can be deduced from Figure 2(A) (top view) and Figure 2(B) (side view) the first possibility would require interactions of the loop domains with each other at the dimer/dimer interface, whereas the latter possibility would require the N and C termini to interact with each other. For two reasons the latter possibility is not very likely. (i) One would expect to see some conserved residues in the tetrameric sequences at the dimer/dimer interface, but none of this is seen on this side of the dimer (data not shown). (ii) The convex shape of the surface provides only a small contact area for the potential dimer/dimer interface. However the flat shape of the active site face provides a much better surface for the dimer-dimer interactions. The flat appearance of this surface (Figure 2(B)) is due to four loops, which are residues 48 to 70 (N-domain), residues 155 to 177 (loop domain), residues 234 to 237 (loop domain) and residues 256 to 266 (loop domain). There are several conserved sequence features in the three loops, which are near the 2-fold axis (48–70, 155–177, 234–237). (i) In the loop 48–70 there is a strictly conserved pattern in the dimeric sequences, being K/R(48)-G/A-G/F-K/R(51)-G-X-F-K/R(55). The equivalent tetrameric sequences are all one residue shorter and have a different pattern of conservation. (ii) At the beginning of the loop domain (residues 155 to 177) all tetrameric thiolase sequences have an insertion with a unique sequence pattern. (iii) The length of the loop region near Gly236 is very well conserved in the tetrameric thiolases, but in the dimeric thiolases it is either six residues shorter, as in yeast, or four residues longer (as in the other dimeric thiolases). Moreover in the tetrameric thiolases a completely conserved (K/R, G) pattern is found at this site.

The loop region near residue 262, which is further away from the molecular centre, also contributes to the flat appearance of the surface on this side of the molecule and therefore might also contribute to the postulated dimer/dimer interface.

The proposed dimer/dimer interface is on the same side of the molecule as the active site and therefore also near the binding site for the CoA substrate molecule. However, it is not incompatible with the proposed mode of binding of the CoA substrate molecule, because the CoA substrate molecule is predicted to bind in a large groove below the outer surface (Figure 8).

The proposed mode of assembly exposes the N termini and the C termini in the tetramer to the solvent (Figure 2(B)). The N terminus contains the targetting sequence for entry into either the peroxisome or the mitochondria (Erdmann & Kunau, 1994), therefore the import apparatus would be able to recognise this targetting signal also in the folded tetrameric protein.

Concluding Remarks

The analysis of the structure has revealed some peculiar structural features. For example, the channel of buried polar residues near the active site pocket and the 12 buried charged residues per monomer. Most of these buried residues are completely conserved (Table 1). It will be interesting to analyse, by site-directed mutagenesis, the importance of these residues for the structure and function of thiolase.

The active site of thiolase catalyses reactions with 3-ketoacyl-CoA, acetyl-CoA and CoA. The binding of 3-ketoacyl-CoA is the first step in the degradation reaction and a covalent bond between SG(Cys125) and the C3 atom of the substrate is formed. The binding of acetyl-CoA is the first step in the synthetic reaction, in this case a covalent bond is formed between SG(Cys125) and the C1 atom of the substrate. Binding of CoA to the thiolase–Cys125-S–acyl complex is the second step of the degradative reaction and it leads to the formation of the acyl-CoA product molecule.

Here we have proposed a very plausible mode of binding of acetoacetyl-CoA as well as a reaction mechanism for thiolase. The proposed mode of binding of acetoacetyl-CoA and the proposed reaction mechanism do not infer conformational changes of loops at the catalytic site. However, it is required that the side-chain of Met186 adopts another conformation on binding of the acetoacetyl group. From the modelling it is unclear if this would induce main-chain conformational changes in this region. Possibly loop conformation changes occur in the 259 to 262 region, allowing for an optimal fit of the CoA part; this region has high *B*-factors in the unliganded structure.

It is proposed that CoA and acetyl-CoA bind in the same way as acetoacetyl-CoA. The modelling experiments with acetoacetyl-CoA suggest that the acetoacetyl group fits near the side-chains of the two catalytic cysteines. The proposed reaction mechanism, as inferred from the mode of binding of the acetoacetyl group, is in good agreement with previously published experimental data and it also explains the importance of the His375 side-chain. This side-chain can activate either the S(CoA) atom in the second step of the degradation reaction, or it can activate the SG(Cys125) atom for nucleophilic attack on either the C3 atom or the C1 atom of the acetoacetyl group or the acetyl group, respectively. The reaction between SG(Cys125) and

the C3 atom of 3-ketoacyl-CoA, as required in the degradative reaction, is favoured by the predicted hydrogen bond between O3 and N(Gly405).

Further crystallographic studies to verify the proposed mode of binding of acetoacetyl-CoA and the reaction mechanism have been initiated.

Materials and Methods

Crystallisation and data collection

The 3-ketoacetyl CoA thiolase from yeast (*S. cerevisiae*) peroxisomes was expressed, purified and crystallised as described (Mathieu *et al.*, 1994). The crystals are grown by a dialysis method (Zeelen & Wierenga, 1992) in a 25 mM 3-(*N*-morpholino)-propanesulphonic acid (Mops) buffer (pH 6.5), 1 mM DTT, 1 mM NaN_3, 1 mM EDTA. The space group is $P2_12_12_1$ and the cell dimensions (at room tempertature) are $a = 71.78$ Å, $b = 93.72$ Å and $c = 120.45$ Å. Fresh crystals diffract to 1.8 Å. The data were collected with a big MAR image plate at the ERSF (Grenoble) at station BM4 from a thiolase crystal flash-cooled at 100 K. MPD was used as a cryoprotectant. A freshly grown thiolase crystal was transferred from its mother liquor into a cryosolution which was made by mixing equal volumes of MPD and the original mother liquor. The resulting MPD concentration was approximately 4 M. After 20 minutes equilibration in this solution the crystal was mounted using the loop technique (Teng, 1990) and flash-frozen in a cold nitrogen gas stream at 100 K. The dataset was collected in two passes: a high resolution pass (20 seconds exposure time per 0.8° frame) and a low resolution pass (four second exposure time per 1.2° frame) to achieve good completeness at low resolution. The wavelength of the X-ray beam was 0.995 Å. The cell dimensions of the cryo-cooled thiolase crystal ($a = 71.17$ Å, $b = 92.65$ Å, $c = 116.72$ Å) are smaller compared to a room temperature crystal. The data were processed with DENZO (Otwinowski, 1993) and SCALEPACK (Gewirth *et al.*, 1995). The refined value of the mosaicity for this dataset is 0.35°. The data were initially merged into a complete dataset with an *R*-factor of 5.0% and a completeness of 98.1% between 17 Å and 1.77 Å.

Structure refinement

The refinement of the thiolase structure against this 1.8 Å dataset was done first with X-PLOR (Brünger, 1992). In later stages also ARP (Lamzin & Wilson, 1993) and TNT (Tronrud, 1992) were used. A test set of reflections (5% of the data, 3790 reflections) was used for *R*-free calculations. Each aspect of the refinement strategy was verified by monitoring the variation of *R*-free. The starting model for the refinement was the previously determined structure at 2.8 Å resolution. Its position in the new cell was found by molecular replacement calculations and rigid body refinement.

Many cycles of refinement and model building were required, initially at 2.6 Å resolution but eventually at 1.8 Å resolution. The models were built in SIGMAA-weighted maps (Read, 1986), calculated with CCP4 programs (Collaborative Computational Project, 1994). The program O (Jones *et al.*, 1991; Kleywegt & Jones, 1996) was used for the model building. NCS restraints were used initially for the positional and *B*-factor refinement. For the refinement at 1.8 Å resolution no NCS-restraints were used.

In the starting model several regions of the loop domain were missing. Missing residues were included in subsequent models whenever the new maps clearly indicated their structure. Also water and two MPD molecules were added to the structure. Eventually, using TNT, a model was obtained with an *R*-factor of 21.1% (*R*-free = 28.7%). This model has 5842 protein atoms, 430 water molecules and 16 MPD atoms. At this stage the dataset was reprocessed, taking particular care to have a final dataset as complete as possible and also using better overload criteria. Overloaded observations were removed during the reprocessing, which resulted in better merging statistics, apparently because of the non-linearity properties of the detector. The data between 10 Å and 17 Å were removed because these data were unreliable due to a slightly misaligned beam stop. In addition,

Table 3. Statistics of the final processing of the 1.8 Å dataset

Resolution limits: inner (Å)	outer (Å)	*R*-merge (%)	I/σ	Completeness (%)	Total number of reflections
10.00	4.72	3.4	43	99.9	3822
4.72	3.81	3.4	54	100.0	3726
3.81	3.35	4.1	68	100.0	3658
3.35	3.05	4.7	65	100.0	3664
3.05	2.84	5.3	57	100.0	3625
2.84	2.68	5.6	51	99.8	3617
2.68	2.55	5.4	45	98.9	3556
2.55	2.44	5.5	45	99.2	3592
2.44	2.34	5.4	45	99.5	3608
2.34	2.26	5.7	41	99.6	3570
2.26	2.19	6.0	39	99.4	3593
2.19	2.13	6.2	37	99.5	3543
2.13	2.08	6.7	33	99.3	3591
2.08	2.03	7.7	30	99.6	3532
2.03	1.98	9.0	27	99.6	3573
1.98	1.94	9.9	25	99.6	3571
1.94	1.90	11.9	23	99.0	3563
1.90	1.86	12.0	21	99.2	3521
1.86	1.83	12.2	19	99.6	3530
1.83	1.80	12.2	17	86.8	3115
Total		4.7		98.9	71,570

Table 4. Crystallographic data

A. *Refinement statistics*				
Total number of reflections	71,365			
R-factor (%)	18.7			
Total number of protein atoms	5874			
Total number of MPD atoms	16			
Total number of water atoms	448			
B. *Geometry statistics*				
rms bond length (Å)	0.017			
rms bond angle (°)	2.4			
rms ΔB (bonded atoms), all atoms (Å^2)	6.0			
rms ΔB (bonded atoms), main-chain atoms (Å^2)	5.4			
rms ΔB (bonded atoms), side-chain atoms (Å^2)	6.9			
Average *B* (all protein atoms) (Å^2)	26.3			
Average *B* (main-chain atoms) (Å^2)	22.8			
Average *B* (side-chain atoms) (Å^2)	30.2			
average *B* (water) (Å^2)	35.9			
Average *B* (MPD) (Å^2)	31.3			
$\chi_1\chi_2$ imperfection (°)[a]	33.3			
C. *Ramachandran plot*[b]				
Most favoured regions	597 (89.4%)			
Additionally allowed regions	65 (9.7%)			
Generally allowed (4) and disallowed regions (2)	6 (0.9%)			
D. *Non-crystallographic symmetry*[c]				
	N	L	C	All
rms ΔB (all atoms) (Å^2)	9.0	15.3	8.5	11.6
rms ΔB (main-chain atoms) (Å^2)	7.9	12.3	7.4	9.5
rms ΔB (side-chain atoms) (Å^2)	10.0	18.0	9.6	13.7
rms (ΔC^α) (Å)	0.22	0.50	0.19	0.34
rms ($\Delta\phi$) (°)	5.5	10.3	6.5	9.4
rms ($\Delta\psi$) (°)	5.4	12.4	6.4	10.7

[a] The $\chi_1\chi_2$ imperfection value (Noble *et al.*, 1993) is the rms difference between observed $\chi_1\chi_2$-values and the nearest preferred cluster value, as seen in a database of well-refined structures.
[b] As defined by PROCHECK (Laskowski *et al.*, 1993).
[c] N, N-domain (residues 25 to 154, 277 to 299); C, C-domain (residues 300 to 417); L, loop-domain (residues 155 to 276).

some bad frames were removed from the processing. This reprocessing and rescaling resulted in a considerably better dataset. The *R*-merge (on *I*) is 4.7% and the completeness between 10 Å and 1.8 Å is 98.9% (Table 3). The overall *R*-merge (4.7%) is entirely dominated by the low resolution data, due to the very high redundancy of the data in the low resolution shells. Nevertheless, the data at high resolution (*R*-merge 12.2%) are of good quality. The reprocessing of the data improved the crystallographic *R*-factor significantly.

Further refinement, against the new dataset, with ARP and TNT resulted eventually in a model with an *R*-factor of 20.2% for data between 10 Å and 1.8 Å, as calculated by TNT (now using all the data). This model includes 5865 protein atoms, 472 water and two MPD molecules. The *R*-free value of this model was assessed by carrying out a simulated annealing X-PLOR run starting at 2000 K with data between 8 Å and 1.8 Å (without *B*-factor refinement). The *R*-factor and *R*-free are 21.6% and 26.1% for the work set and the test set, respectively. Subsequently the anisotropy in the native dataset was corrected which reduced the crystallographic *R*-factor by 1%. Further checking of the model against the map suggested additional minor changes. In addition the TNT weights were optimised. The final model has an *R*-factor of 18.7%. In this model there are 5874 protein atoms, 448 water and two MPD molecules. Residues 263 to 265 of subunit-1 are missing. The *R*-free value of this model was again verified by a simulated annealing X-PLOR run starting at 2000 K for data between 8 Å and 1.8 Å (without *B*-factor refinement). The *R*-factor and *R*-free values are now 20.3% and 24.2% for the workset and the test set, respectively. The refinement statistics are summarised in Table 4.

Structure analysis and sequence alignment

The structure has been analysed with O and WHAT IF (Vriend, 1990). The program ICM (Abagyan *et al.*, 1994) was used for the modelling experiments. Unless otherwise stated the structure description refers to subunit-1. Sequence alignments have been made with Pile-Up of the GCG package (Devereux *et al.*, 1984). The cavities and molecular surfaces were calculated with the MSP package (Connolly, 1985); atomic radii and probe radius (1.2 Å) were the same as used previously (Wierenga *et al.*, 1992). The polar residues Arg, Lys, Asp, and Glu were considered to be buried when the polar side-chain atoms did not contribute to the outer surface of the protein.

Acknowledgements

We thank Ursula Dorpmund for the purification of thiolase, Dr Horst Schulz for stimulating discussions and Dr Cusack for his interest in this project. The high resolution dataset was collected at station BM4 at the ESRF Grenoble. The coordinates (1AFW) and the structure factors (R1AFWSF) of the refined structure have been deposited at the PDB. The coordinates of the modelled acetoacetyl-CoA complex have also been deposited in the PDB (1AFY).

References

Abagyan, R. A. (1997). *The ICM 2.6 Manual* (http://molsoft.com). Molsoft LLC, New York.

Abagyan, R. A., Totrov, M. & Kuznetsov, D. (1994). ICM-A new method for protein modelling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506.

Anderson, V. E., Bahnson, B. J., Wlassics, I. D. & Walsh, C. T. (1990). The reaction of acetyldithio-CoA, a readily enolized analog of acetyl-CoA with thiolase from *Zoogloea ramigera*. *J. Biol. Chem.* **265**, 6255–6261.

Brünger, A. T. (1992). *X-PLOR: version 3.1, a System for X-ray Crystallography and NMR.* Yale University Press, New Haven and London.

Computational Project N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

Connolly, M. L. (1985). Computation of molecular volume. *J. Am. Chem. Soc.* **107**, 1118–1124.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

Engel, C. K. & Wierenga, R. (1996). The diverse world of coenzyme A binding proteins. *Curr. Opin. Struct. Biol.* **6**, 790–797.

Engel, C. K., Mathieu, M., Zeelen, J. Ph. , Hiltunen, J. K. & Wierenga, R. K. (1996). Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 Å resolution: a spiral fold defines the CoA-binding pocket. *EMBO J.* **15**, 5135–5145.

Erdmann, R. (1994). The peroxisomal targeting signal of 3-oxoacy 1-CoA thiolase from *Saccharomyces cerevisiae*. *Yeast,* **10**, 935–944.

Erdmann, R. & Kunau, W.-H. (1994). Purification and immunolocalization of the peroxisomal 3-oxoacyl-CoA thiolase of *Saccharomyces cerevisiae*. *Yeast,* **10**, 1173–1182.

Flocco, M. M. & Mowbray, S. L. (1995). Strange bedfellows: interactions between acidic side-chains in proteins. *J. Mol. Biol.* **254**, 96–105.

Gehring, U. & Lynen, F. (1972). Thiolase. In *The Enzymes* (Boyer, P. D., ed.), vol. 7, pp. 391–405, Academic Press, New York and London.

Gewirth, D., Otwinowski, Z. & Minor, W. (1995). *The HKL Manual* (4th Edit.) (Gewirth, D., ed.), Yale University, New Haven.

Gilbert, H. F. (1981). Proton transfer from acetyl-coenzyme A catalyzed by thiolase I from porcine heart. *Biochemistry,* **20**, 5643–5649.

Gilbert, H. F., Lennox, B. J., Mossman, C. D. & Carle, W. C. (1981). The relation of acyl transfer to the overall reaction of thiolase I from porcine heart. *J. Biol. Chem.* **256**, 7371–7377.

Hartmann, G. & Lynen, F. (1961). Thiolase. In *The Enzymes* (Boyer, P. D., ed.), vol. 5, pp. 381–386, Academic Press, New York and London.

Igual, J. C., Gonzalez-Bosch, C., Dopazo, J. & Pérez-Ortin, J. E. (1992). Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes. *J. Mol. Evol.* **35**, 147–155.

Izbicka, E. & Gilbert, H. F. (1984). Fluorescence energy transfer measurements of spatial relationships between sulfhydryl groups of Thiolase I from porcine heart. *Biochemistry,* **23**, 6383–6388.

Izbicka-Dimitrijevic, E. & Gilbert, H. F. (1984). Multiple oxidation products of sulfhydryl groups near the active site of thiolase I from porcine heart. *Biochemisty,* **23**, 4318–4324.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110–119.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577–2637.

Kleywegt, G. J. & Jones, T. A. (1996). Efficient rebuilding of protein structures. *Acta Crystallog. sect. D,* **52**, 829–832.

Kunau, W. -H., Dommes, Y. & Schulz, H. (1995). β-Oxidation of fatty acids in mitochondria, peroxisomes and bacteria: a centry of continued progress. *Progr. Lipid Res.* **34**, 267–342.

Kurihara, T., Ueda, M. & Tanaka, A. (1989). Peroxisomal acetoacetyl-CoA thiolase and 3-ketoacyl-CoA thiolase from an *n*-alkane-utilizing yeast, *Candida tropicalis*: purification and characterization. *J. Biochem.* **106**, 474–478.

Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D,* **49**, 129–147.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a Program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Masamune, S., Palmer, M. A. J., Gamboni, R., Thompson, S., Davis, J. T., Williams, S. F., Peoples, O. P., Sinskey, A. J. & Walsh, C. T. (1989a). Bio-Claisen condensation catalyzed by thiolase from *Zoogloea ramigera*. Active site cysteine residues. *J. Am. Chem. Soc.* **111**, 1879–1881.

Masamune, S., Walsh, C. T., Sinskey, A. J. & Peoples, O. P. (1989b). Poly-(R)-3-hydroxybutyrate (PHB) biosynthesis: mechanistic studies on the biological Claisen condensation catalyzed by β-ketoacyl thiolase. *Pure Appl. Chem.* **61**, 303–312.

Mathieu, M. (1995). Structure determination of thiolase, Ecole Centrale, Paris.

Mathieu, M., Zeelen, J. P., Pauptit, R. A., Erdmann, R., Kunau, W.-H. & Wierenga, R. K. (1994). The 2.8 Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: a five-layered αβαβα structure constructed from two core domains of identical topology. *Structure,* **2**, 797–808.

Miyazawa, S. & Hashimoto, T. (1980). The presence of a new 3-oxoacyl-CoA thiolase in rat liver peroxisomes. *Eur. J. Biochem.* **103**, 589–596.

Miyazawa, S., Furuta, S., Osumi, T., Hashimoto, T. & Ui, N. (1981). Properties of peroxisomal 3-ketoacyl-CoA thiolase from rat liver. *J. Biochem.* **90**, 511–519.

Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. genet.* **11**, 281–296.

Noble, M. E. M., Zeelen, J. P., Wierenga, R. K., Mainfroid, V., Goraj, K., Gohimont, A.-C. & Martial, J. A. (1993). Structure of triosephosphate isomerase from *Escherichia coli* determined at 2.6 Å resolution. *Acta Crystallog. sect. D,* **49**, 403–417.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: ''Data Collection and Processing''* (Sawyer, L., Isaacs, N. & Baily, S., eds), pp. 56–62, SERC, Daresbury Laboratory, England.

Palmer, M. A. J., Differding, E., Gamboni, R., Williams, S. F., Peoples, O. P., Walsh, C. T., Sinskey, A. J. & Masamune, S. (1991). Biosynthetic thiolase from *Zoogloea ramigera*. Evidence for a mechanism involving Cys378 as the active site base. *J. Biol. Chem.* **266**, 8369–8375.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A,* **42**, 140–149.

Salam, W. H. & Bloxham, D. P. (1986). Identification of subsidiary catalytic groups at the active site of β-ketoacyl-CoA thiolase by covalent modification of the protein. *Biochim. Biophys. Acta,* **873**, 321–330.

Staack, H., Binstock, J. F. & Schulz, H. (1978). Purification and properties of a pig heart thiolase with broad chain length specificity and comparison of thiolases from pig heart and *Escherichia coli*. *J. Biol. Chem.* **253**, 1827–1831.

Teng, T.-Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallog.* **23**, 387–391.

Thompson, S., Mayerl, F., Peoples, O. P., Masamune, S., Sinskey, A. J. & Walsh, C. T. (1989). Mechanistic Studies on β-ketoacyl thiolase from *Zoogloea ramigera*: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry,* **28**, 5735–5742.

Tronrud, D. E. (1992). Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crystallog. sect. A,* **48**, 912–916.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.

Walsh, C. T. (1979). *Enzymatic Reaction Mechanisms* (Bartlett, A. C. & McCombs, L. W., eds). W. H. Freeman and Company, San Francisco.

Waterson, R. M. & Hill, R. L. (1972). Enoyl coenzyme A hydratase (Crotonase). *J. Biol. Chem.* **247**, 5258–5265.

Wierenga, R. K., Noble, M. E. M. & Davenport, R. C. (1992). Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. *J. Mol. Biol.* **224**, 1115–1126.

Williams, S. F., Palmer, M. A. J., Peoples, O. P., Walsh, C. T., Sinskey, A. J. & Masamune, S. (1992). Biosynthetic thiolase from *Zoogloea ramigera*. Mutagenesis of the putative active-site base Cys378 changes the partitioning of the acetyl *S*-enzyme intermediate. *J. Biol. Chem.* **267**, 16041–16043.

Zeelen, J. Ph. & Wierenga, R. K. (1992). The growth of yeast thiolase crystals using a polyacrylamide gel as dialysis membrane. *J. Cryst. Growth,* **122**, 194–198.

Edited by R. Huber

(Received 9 April 1997; received in revised form 30 July 1997; accepted 4 August 1997)