Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the putative thioesterase protein TTHA1846 from *Thermus thermophilus* HB8 complexed with coenzyme A and a zinc ion

Toshiaki Hosaka,[a] Kazutaka Murayama,[a,b] Miyuki Kato-Murayama,[a] Akiko Urushibata,[a] Ryogo Akasaka,[a] Takaho Terada,[a] Mikako Shirouzu,[a] Seiki Kuramitsu[c,d] and Shigeyuki Yokoyama[a,c,e]*

[a]Protein Research Group, RIKEN Systems and Structural Biology Center, Tsurumi-ku, Yokohama 230-0045, Japan, [b]Biomedical Engineering Research Organization, Biofunctional Science, Tohoku University, Aoba-ku, Sendai 980-8574, Japan, [c]RIKEN SPring-8 Center, Harima Institute, Sayo, Hyogo 679-5148, Japan, [d]Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan, and [e]Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Received 4 February 2009
Accepted 27 April 2009

PDB Reference: TTHA1846, 2cye, r2cyesf.

TTHA1846 is a conserved hypothetical protein from *Thermus thermophilus* HB8 with a molecular mass of 15.1 kDa that belongs to the thioesterase superfamily (Pfam 03061). Here, the 1.9 Å resolution crystal structure of TTHA1846 from *T. thermophilus* is reported. The crystal structure is a dimer of dimers. Each subunit adopts the so-called hot-dog fold composed of five antiparallel β-strands flanked on one side by a rather long α-helix and shares structural similarity to a number of thioesterases. Unexpectedly, TTHA1846 binds one metal ion and one ligand per subunit. The ligand density was modelled as coenzyme A (CoA). Its structure was confirmed by MALDI–TOF mass spectrometry and electron-density mapping. X-ray absorption fine-structure (XAFS) measurement of the crystal unambiguously characterized the metal ion as zinc. The zinc ion is tetrahedrally coordinated by the side chains of Asp18, His22 and Glu50 and the CoA thiol group. This is the first structural report of the interaction of CoA with a zinc ion. From structural and database analyses, it was speculated that the zinc ion may play an inhibitory role in the enzymatic activity.

1. Introduction

TTHA1846 consists of 133 amino-acid residues and a *PSI-BLAST* search (Altschul *et al.*, 1997) using the TTHA1846 sequence identified more than 100 similar proteins from bacteria and archaea, with sequence identities ranging from 26% to 41% and *E*-values below 8×10^{-8} (Fig. 1*a*). *CD-Search* (Marchler-Bauer & Bryant, 2004) revealed that TTHA1846 and its homologous proteins contain a hot-dog fold. In the Pfam database (Bateman *et al.*, 2000), TTHA1846 belongs to the acyl-carrier protein thioesterase and thioesterase superfamily, which is equivalent to COG0824 (predicted thioesterase) in the National Center for Biotechnology Information database of Clusters of Orthologous Groups.

Thioesterases are widespread in eukaryotes, bacteria and archaea and are involved in reactions such as thioester hydrolysis in fatty-acid metabolism and degradation of the environmental pollutant 4-chlorobenzoate (Benning *et al.*, 1998; Hunt & Alexson, 2002; Thoden *et al.*, 2003; Zhuang *et al.*, 2004). These binding enzymes adopt a core hot-dog fold domain, mainly comprising a long central α-helix and an enclosing antiparallel β-strand, in various oligomeric states (Dillon & Bateman, 2004; Forwood *et al.*, 2007; Leesong *et al.*, 1996; Li *et al.*, 2000). The hot-dog fold protein family includes various CoA-binding oligomeric enzymes (Kunishima *et al.*, 2005). Coenzyme A (CoA) exists widely among prokaryotes and eukaryotes and its main function is to carry acyl groups (such as an acetyl group) or thioesters. A number of crystal

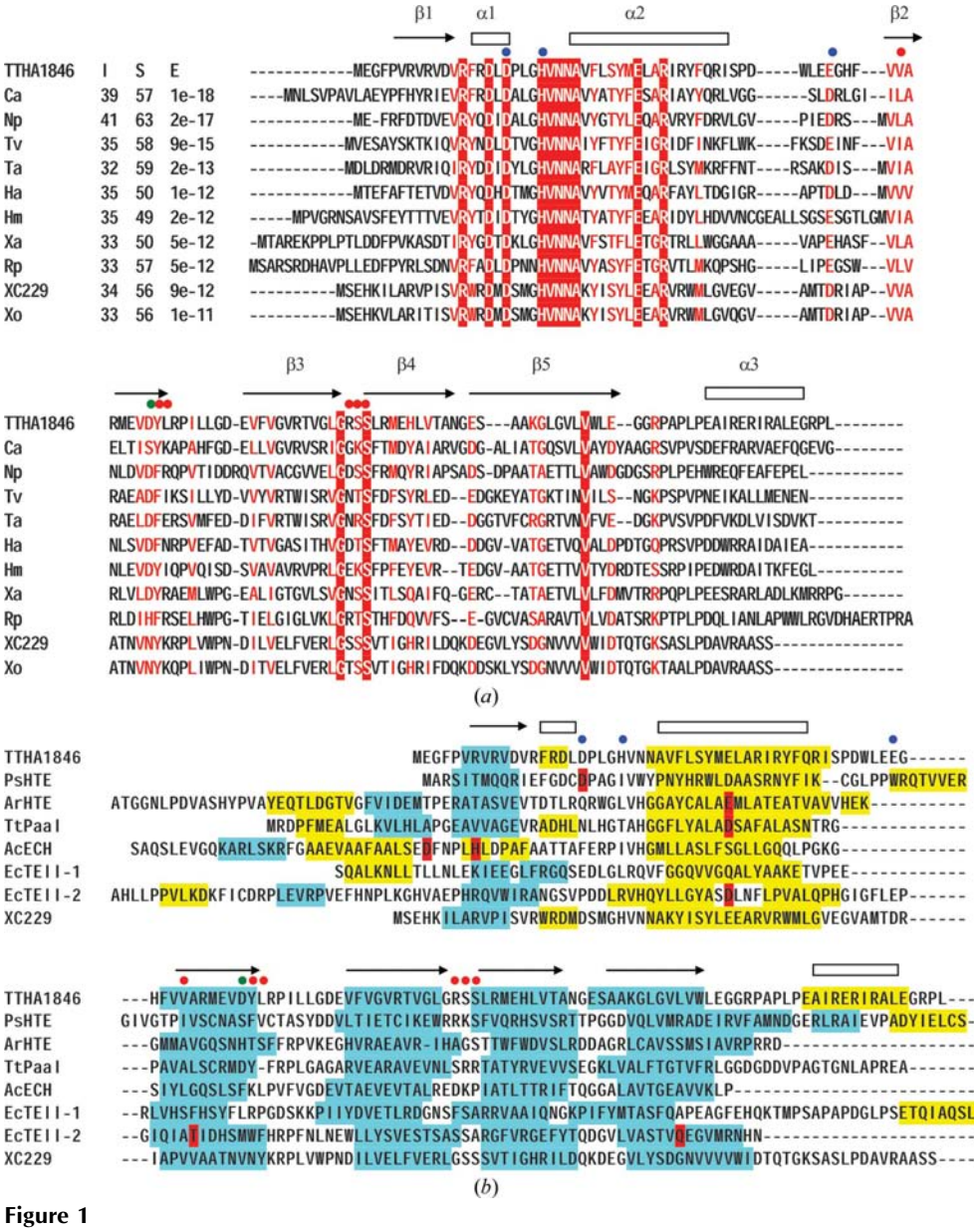

Figure 1
Amino-acid sequence alignments of TTHA1846. (*a*) Sequence comparison between TTHA1846 and its ten closest homologues. The figure was constructed using *PSI-BLAST* (Altschul *et al.*, 1997) followed by *ClustalW* (Thompson *et al.*, 1994). The first column shows the protein identifier. These proteins are annotated as thioesterase-family members or conserved hypothetical proteins: Ca, thioesterase superfamily from *Chloroflexus aurantiacus*; Np, conserved hypothetical protein from *Natronomonas pharaonis*; Tv, hypothetical protein from *Thermoplasma volcanium*; Ta, predicted thioesterase from *Thermoplasma acidophilum*; Ha, Vng1407c from *Halobacterium* sp.; Hm, thioesterase of unknown specificity from *Haloarcula marismortui*; Xa, conserved hypothetical protein from *Xanthobacter autotrophicus*; Rp, thioesterase superfamily from *Rhodopseudomonas palustris*; XC229, putative acyl-CoA thioesterase from *Xanthomonas campestris* (PDB code 2fuj); Xo, hypothetical protein XOO1618 from *Xanthomonas oryzae* pv. *oryzae*. 'I', 'S' and 'E' represent percentage identity, percentage similarity and *E*-value from the *PSI-BLAST* results, respectively. Strictly conserved and similar residues are represented within a red box and by red letters, respectively. Residues that recognize the CoA of the ligand with their main chains and side chains are indicated by filled green circles and red circles, respectively. Residues that contact the zinc ion with their side chains are indicated by filled blue circles. (*b*) A secondary-structure-based sequence alignment between TTHA1846 and six similar structures. These proteins are annotated as thioesterase-family members: PsHTE, 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3; ArHTE, 4-hydroxybenzoyl-CoA thioesterase from *Arthrobacter* sp. strain SU; TtPaaI, phenylacetate-degradation protein PaaI from *Thermus thermophilus* HB8; AcECH, prokaryotic enoyl-CoA hydratase 2 from *Aeromonas caviae*; EcTEII, medium-chain acyl-CoA thioesterase II from *E. coli*; XC229, putative acyl-CoA thioesterase from *X. campestris*. Residues that are involved in helices and strands are represented within yellow boxes and cyan boxes, respectively. The active-site residues are highlighted by red boxes.

structures of hot-dog fold proteins have been reported, including those of β-hydroxydecanoyl thiol ester dehydrase from *Escherichia coli* (EcHTD; Leesong *et al.*, 1996), the 4-hydroxybenzoyl-CoA thioesterases from *Pseudomonas* sp. strain CBS-3 (PsHTE; Benning *et al.*, 1998; Thoden *et al.*, 2002) and from *Arthrobacter* sp. (ArHTE; Thoden *et al.*, 2003), medium-chain acyl-CoA thioesterase II from *E. coli* (EcTEII; Li *et al.*, 2000), TtPaaI from *T. thermophilus* (Kunishima *et al.*, 2005), the putative acyl-CoA thioesterase from *Xanthomonas campestris* (XC229; Chin *et al.*, 2006), human thioesterase superfamily member 2 (Cheng *et al.*, 2006), acyl-CoA thioesterase from *Haemophilus influenzae* (HI0827; Willis *et al.*, 2008) and the bacterial acyl-CoA thioesterase HpYbgC from *Helicobacter pylori* (Angelini *et al.*, 2008). Here, we report the crystal structure of the thioesterase-family member TTHA1846 from *T. thermophilus* HB8 in complex with CoA. We discuss the putative function of TTHA1846 based on its structure and compare its structure with those of other family members.

2. Materials and methods

2.1. Protein expression and purification

The TTHA1846 gene was amplified by polymerase chain reaction (PCR) from *T. thermophilus* HB8 genomic DNA and cloned into the plasmid vector pET11a (Novagen). The selenomethionine-labelled TTHA1846 protein was expressed in the methionine auxotroph *E. coli* B834(DE3)pRARE strain. The cells were cultured at 310 K in LeMaster medium containing 50 µg ml^{-1} selenomethionine (SeMet; Hendrickson *et al.*, 1990). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to

0.5 m*M* final concentration once the culture reached an optical density (at 600 nm) of 0.7. After 3 h further growth, the cells were harvested and stored at 193 K until use.

Wet cells (4.25 g) were suspended in 100 ml lysis buffer (20 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl and 5 m*M* 2-mercaptoethanol) and then sonicated on ice. The crude extract was then incubated at 343 K for 30 min and centrifuged. The protein was then applied onto a HiTrap Q Sepharose HP column (GE Healthcare Biosciences; 5 ml bed volume) and eluted with a linear gradient of sodium chloride (0–1 *M* NaCl, 20 m*M* Tris–HCl pH 8.5, 5 m*M* 2-mercaptoethanol). A saturated ammonium sulfate solution was added to the fractions containing TTHA1846 to a final concentration of 0.8 *M*. The protein was applied onto a HiTrap Phenyl column (GE Healthcare Biosciences; 5 ml bed volume) and was eluted with a linear gradient of elution buffer (20 m*M* Tris–HCl pH 8.0, 5 m*M* 2-mercaptoethanol). The pooled fractions containing TTHA1846 were collected and dialyzed against 20 m*M* Tris–HCl pH 8.5, 5 m*M* 2-mercaptoethanol. The solution was then loaded onto a Resource Q column (6 ml bed volume) and eluted with a linear gradient of sodium chloride (0–1 *M* NaCl, 20 m*M* Tris–HCl pH 8.5, 5 m*M* 2-mercaptoethanol). The TTHA1846 protein was applied onto a HiPrep 26/10 Desalting column (GE Healthcare Biosciences) equilibrated with 20 m*M* Tris–HCl pH 7.0 containing 5 m*M* 2-mercaptoethanol. TTHA1846 was further purified on a Resource Q column again and was eluted with a linear gradient of 0–1 *M* NaCl. Finally, TTHA1846 was purified by gel filtration on a Superdex-75 column equilibrated with 20 m*M* Tris–HCl pH 8.0 containing 2 m*M* DTT and 150 m*M* NaCl. The purified protein was concentrated to approximately 22 mg ml^{-1} using a Centricon YM-10 filter (Millipore). The estimated yield was 3.84 mg purified protein per litre of culture.

2.2. Crystallization and data collection

For crystallization, the sitting-drop vapour-diffusion method was used by mixing 2 µl protein solution with 2 µl reservoir solution and equilibrating against 500 µl reservoir solution at 293 K. Crystals were produced in 0.1 *M* sodium acetate, 0.8 *M* ammonium sulfate and 3% PEG 400 at pH 5.0 (Hampton Research). Crystals with a rod-like morphology were obtained within 3 d and were used for data collection.

Data collection was performed at 100 K using Paratone-N (Hampton Research) as a cryoprotectant. MAD data were collected at three different wavelengths on the RIKEN Structural Genomics Beamline 1 (BL26B1) at SPring-8 (Hyogo, Japan; Yamamoto *et al.*, 2002) and were recorded on a Jupiter 210 CCD detector (Rigaku). All diffraction data were processed with the *HKL*-2000 program (Otwinowski & Minor, 1997).

2.3. Structure determination and refinement

The program *SOLVE* (Terwilliger & Berendzen, 1999) was used to locate the selenium sites and to calculate the phases and *RESOLVE* (Terwilliger, 2002) was used for density modification and partial model building. The rest of the model was built with the program *XtalView* (Hosaka *et al.*, 2003; McRee, 1999) and was refined with the program *Crystallography & NMR System* (*CNS*; Brünger *et al.*, 1998). Refinement statistics are presented in Table 1. The quality of the model was inspected using the program *PROCHECK* (Laskowski *et al.*, 1993). The graphical figures were created using the program *PyMOL* (DeLano, 2002). The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession code 2cye.

Table 1
X-ray data-collection, phasing and refinement statistics.

Values in parentheses are for the outer shell (1.97–2.02 Å).

	Peak	Edge	Remote
Data collection			
Wavelength (Å)	0.97919	0.97949	0.96200
Resolution (Å)	50.0–1.9	50.0–1.9	50.0–1.9
Unique reflections	41880	41892	41903
Redundancy	4.0	4.0	3.9
Completeness (%)	99.3	99.4	99.4
$I/\sigma(I)$	10.9 (3.47)	13.0 (3.45)	13.4 (4.15)
$R_{\rm merge}$†	10.6 (33.7)	9.1 (34.1)	8.9 (28.6)
MAD analysis			
Resolution (Å)		50.0–1.9	
No. of Se sites‡		16	
$\rm FOM_{MAD}$§		0.46	
Refinement			
Resolution (Å)		47.0–1.9	
No. of reflections		41903	
No. of protein atoms		1042	
No. of water molecules		296	
$R_{\rm work}$¶		20.3	
$R_{\rm free}$††		24.2	
R.m.s.d. bond lengths (Å)		0.008	
R.m.s.d. bond angles (°)		1.5	
Average *B* factor (Å^2)		17.3	
Ramachandran plot			
Most favoured regions (%)		94.0	
Additional allowed regions (%)		5.4	
Generously allowed regions (%)		0.0	
Disallowed regions (%)		0.0	

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity. ‡ Number of selenium sites located with *SOLVE*. § Figure of merit after *SOLVE* phasing. ¶ $R_{\rm work} = \sum_{hkl} |F_{\rm o}| - |F_{\rm c}| / \sum_{hkl} F_{\rm o}$, where $F_{\rm o}$ and $F_{\rm c}$ are observed and calculated structure factors, respectively. †† $R_{\rm free}$ is calculated for a randomly selected 5% of reflections excluded from refinement.

2.4. Characterization of the ligand

MALDI–TOF mass-spectrometry experiments were performed on a Voyager-DE STR reflection MALDI–TOF mass spectrometer (GE Healthcare Biosciences). For analyses, saturated cyano-4-hydroxycinnamic acid (Nacalai Tesque Inc.) in 100% aqueous acetonitrile plus 0.3% trifluoroacetic acid was used as the matrix. Samples were prepared by mixing 1 µl acidified sample solution with 1 µl matrix. Ligand mass spectra were obtained in the positive-ion mode at an acceleration voltage of 20 kV by accumulating 400 laser shots.

The X-ray absorption fine structure (XAFS) of the crystal was measured around each absorption edge of zinc and ferric ions.

2.5. Sedimentation-equilibrium experiments

The sedimentation-equilibrium experiments were performed at 293 K with six-channel centerpieces, with loading concentrations of 1.68, 0.84 and 0.42 mg ml^{-1}, using an Optima XL-I analytical ultracentrifuge (Beckman Coulter). The sample buffer was 20 m*M* Tris–HCl pH 8.0, 300 m*M* sodium chloride and 5 m*M* 2-mercaptoethanol. Data were obtained at 6000, 8000 and 11 000 rev min^{-1} and the absorbance wavelength was 280 nm. A total equilibration time of 14 h was used for each speed. The estimated partial specific volume of the protein was 0.7442 ml g^{-1} and the estimated solvent density was 1.0113 g ml^{-1}. The data were fitted using the manufacturer's software.

3. Results and discussion

3.1. Overall protein structure

The TTHA1846 crystals have the symmetry of the trigonal space group *C*2, with unit-cell parameters $a = 89.53$, $b = 55.82$, $c = 110.15$ Å. The structure was refined to 1.9 Å resolution by the MAD method. The crystallographic data are summarized in Table 1. The final model includes 518 amino-acid residues of the four TTHA1846 monomers (chain *A*, 130 amino acids; chain *B*, 130 amino acids; chain *C*, 133 amino acids; chain *D*, 125 amino acids), 296 water molecules, four CoA molecules and four zinc ions per asymmetric unit. Owing to poor electron density, the atomic models of residues 1–2 of chain *D*, 44–45 of chains *A* and *B*, 128–133 of chain *D* and 133 of chains *A* and *B* were not built. A ribbon representation of the tetramer with bound CoA and zinc ions is depicted in Fig. 2(*a*). There are four protein chains per asymmetric unit and the structures of the four chains are essentially identical, with a root-mean-square deviation (r.m.s.d.) of only 0.23–0.42 Å for the C$^{\alpha}$ atoms. The subunit definition is provided in Fig. 2(*a*). The TTHA1846 monomer structure consists of three α-helices (α1–α3) and five β-strands (β1–β5) (Fig. 2*b*). The secondary-structure elements are ordered as follows: β1-α1-α2-β2-β3-β4-β5-α3. The five β-strands form an antiparallel β-strand that wraps around the long central α-helix (α2) constructed by 18 residues (Figs. 1*a* and 2*b*).

As shown in Fig. 2(*a*), the α1 and α2 helices of one dimer abut the α-helix motif in the second dimer to form the protomer (*A*/*C* and *B*/*D*). Chains *A* and *D* interact with each other *via* their α1 helices and the loop between α1 and α2. Calculation of the solvent-accessible area using the *PROTORP* program (http://www.bioinformatics.sussex.ac.uk/protorp/; Reynolds *et al.*, 2009) revealed that the buried surface area of a monomer in the *AC* dimer is 1041.2 Å^2, which corresponds to 13.9% of the surface area of a monomer. The buried surface area in the *AB* dimer is 1026.6 Å^2. In contrast, for the *AD* dimer the value is 380.7 Å^2, corresponding to 5.1% of the monomer surface.

The analytical ultracentrifugation results generated an estimated molecular mass of TTHA1846 in solution of ~65.0 kDa, indicating that TTHA1846 exists as a tetramer in solution since the calculated molecular mass of the TTHA1846 tetramer is 60.4 kDa. Therefore, TTHA1846 forms a homotetrameric structure composed of two protomers. The interactions in the tetramer are mainly between loops α1–α2 and α2 itself. The hot-dog fold proteins are all defined by the orientation of the two hot-dog folds in the γ protomers. However, these dimeric subunits form tetramers by two different modes. The TTHA1846 protomers associate on the helical side of the protomers, with the β-sheet sides exposed at the top and bottom of the tetramer (Fig. 2*a*). This orientation is called the $\varepsilon\gamma$ mode (Kunishima *et al.*, 2005; Chin *et al.*, 2006). This arrangement mode was also observed in 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* (PsHTE; Benning *et al.*, 1998), the putative acyl-CoA thioesterase from *X. campestris* (XC229; Benning *et al.*, 1998; Chin *et al.*, 2006) and the bacterial acyl-CoA thioesterase HpYbgC from *H. pylori* (HP0496; Angelini *et al.*, 2008). However, the tetramer orientations in TTHA1846 and HpYbgC are not the same. The angle of the two protomers differs by about 30°, but the buried surface area of each monomer is similar.

Table 2
Structural similarity search data by *DALI*.

PDB code	*Z* score	R.m.s.d. (Å)	Identity (%)	Protein name
1z54	17.7	1.8	26	Hypothetical protein TT1821
2oiw	17.6	1.4	31	Putative 4-hydroxybenzoyl CoA thioesterase
1s5u	17.5	1.6	21	Protein YBGC
2oaf	17.2	1.9	17	Thioesterase superfamily
2av9	16.9	1.9	28	Thioesterase
2nuj	16.4	2.0	25	Thioesterase superfamily
2o6t	16.7	2.1	28	Thioesterase
2o5u	16.5	2.3	28	Thioesterase
2o6b	16.3	2.1	28	Thioesterase
2egi	16.2	1.6	20	Hypothetical protein AQ_1494
3ck1	16.1	2.0	21	Putative thioesterase
2egj	16.1	1.7	20	Hypothetical protein AQ_1494
2gf6	16.1	2.2	20	Conserved hypothetical protein
2egr	16.0	1.7	20	Hypothetical protein AQ_1494
1lo8	16.0	1.5	15	4-Hydroxybenzoyl CoA thioesterase
2ali	16.0	1.9	32	Hypothetical protein PA2801
1lo7	15.9	1.5	15	4-Hydroxybenzoyl CoA thioesterase
2hlj	15.9	2.1	22	Hypothetical protein
1lo9	15.8	1.5	14	4-Hydroxybenzoyl CoA thioesterase
1bvq	15.7	1.5	15	4-Hydroxybenzoyl CoA thioesterase
2pzh	15.5	1.8	22	Hypothetical protein HP_0496

The results of a *DALI* (Holm & Sander, 1998) structural similarity search are shown in Table 2. In contrast to the lack of significant sequence identity among the hot-dog fold proteins (below 30%), the similarities in their main-chain foldings are obvious. The r.m.s.d.s for the corresponding C$^{\alpha}$ atoms between TTHA1846 and the known hot-dog fold proteins are within 2.5 Å (Benning *et al.*, 1998; Kunishima *et al.*, 2005; Leesong *et al.*, 1996; Li *et al.*, 2000).

3.2. The CoA binding site

Between the protomer subunits, strong electron density corresponding to a ligand molecule was clearly present (Fig. 3). This ligand density was thought to be derived from the ligand-bound form of the *T. thermophilus* enzyme, because we did not add such large molecules to the protein solution during the purification and crystallization. Therefore, the ligand origi-

nated from the *E. coli* host and bound tightly to TTHA1846 during purification.

The molecule bound to TTHA1846 was determined to have a molecular mass of 769.52 Da by mass spectrometry. This molecular mass is almost identical to the calculated theoretical molecular mass of CoA (767.55 Da). The hot-dog fold protein family is known to include various CoA-binding oligomeric enzymes. According to this viewpoint, we modelled this ligand density as CoA. The CoA fitted well to the composite OMIT σ_A-weighted $|F_o| - |F_c|$ electron-density maps (Fig. 3), where F_o and F_c are the observed and calculated structure factors, respectively.

In the TTHA1846 tetramer, the CoA occupies all the active sites. As shown in Fig. 2(*a*), the CoA is primarily wedged between the two subunits of the protomer pair (chains *A*/*C* and *B*/*D*) in the tetramer and specifically between the two major α-helices. The pyrophosphate moieties of the CoA ligands project outwards and the zinc binding site (the putative active site) in one protomer is ∼25 Å away from the symmetry-related active site in the second protomer. A schematic view of the CoA binding site is presented in Fig. 4. Only the backbone carbonyl or peptidic NH groups lie within hydrogen-bonding distance of the O and N atoms of the CoA β-mercaptoethylamine and pantothenate units. The pyro-

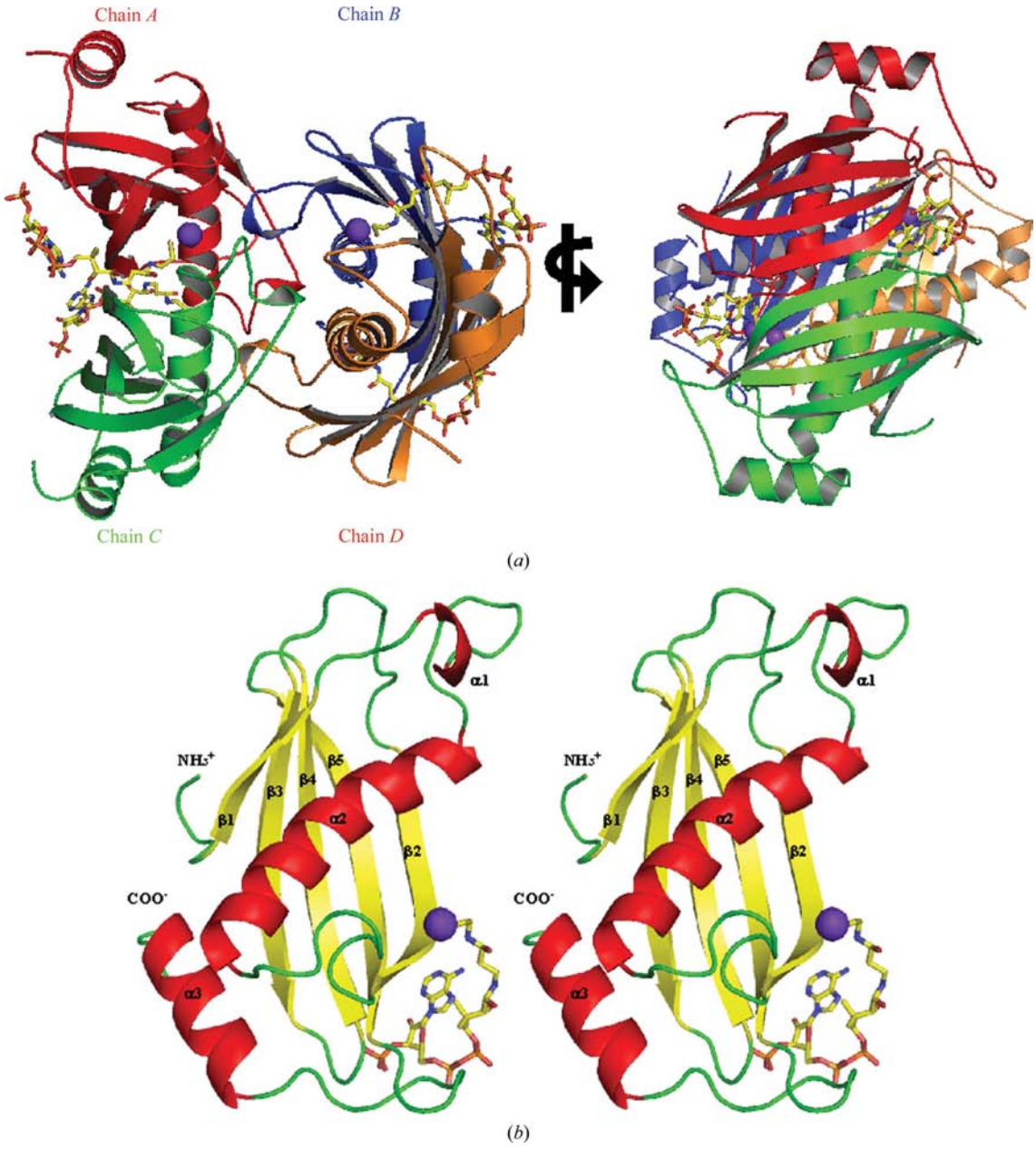

Figure 2
The crystal structure of the tetrameric TTHA1846 from *T. thermophilus*. (*a*) Ribbon diagram of the TTHA1846 tetramer. Chains *A*, *B*, *C* and *D* are coloured red, blue, green and orange, respectively. The CoA is represented by a stick model. The zinc ion is depicted by a purple sphere. (*b*) Stereo ribbon diagram of the TTHA1846 monomer. The α-helices and the β-strands are coloured red and yellow, respectively. The CoA is represented by a stick model. The zinc ion is depicted by a purple sphere.

phosphate moiety of CoA is positioned in a depression located on the solvated surface of one of the paired subunits. The side chains of three consecutive residues, Arg83, Ser84 and Ser85, from subunit *C* lie within hydrogen-bonding distance of the 3′-phosphate group of the CoA ribose from subunit *A*; therefore, this group was tightly fixed. The amino groups at position 6 of the adenine ring form hydrogen bonds to the O^{δ} atoms of Asp61 from subunit *A*. Six water molecules are located within hydrogen-bonding distance of the CoA. There are four well ordered water molecules located within 3.5 Å of the acyl pantetheine moiety and one of them is linked to the thiol group of the CoA.

3.3. The zinc ion binding site

In each subunit, strong electron density was located in the centre of the protomer. XAFS measured near each absorption edge of the possible metal ions only yielded a notable signal at the absorption edge of a zinc ion, although a ferric ion was also considered for coordination of the side chain and thiol group of CoA (Harding, 2001, 2004; Murata *et al.*, 2005). This location was considered to support the existence of a zinc ion according to the statistical data (Harding, 2001, 2004). No zinc ions were added to any of the solutions used for the purification and crystallization of the SeMet form of TTHA1846.

Each zinc ion is surrounded by a tetrahedral coordination geometry with distances of 2.11–2.41 Å. The coordination is accomplished by the O^{δ} atom of Asp18 and the N^{ε} atom of His22 from one monomer, the O^{ε} atom of Glu50 from the other monomer of the protomer and the thiol group of CoA (Fig. 4). The zinc ion is predominantly held in place by electrostatic interactions with these three side chains and the thiol group of the CoA. The bound ion is not able to access the bulk solvent because it is sandwiched between the protein and the CoA (Fig. 5). This is the first structural report of the thiol group of CoA interacting with a zinc ion and therefore we have revealed a novel zinc-binding motif. The amino-acid residues that contact the ligand and the zinc ion are well conserved among the closest homologues (Fig. 1*a*).

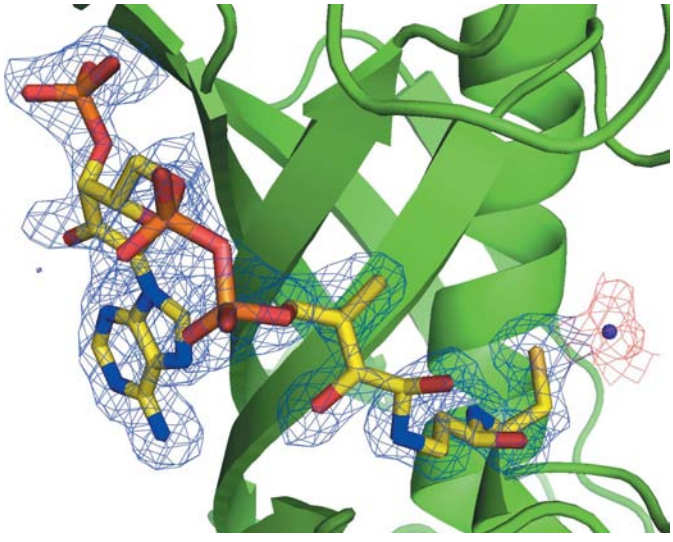

Figure 3
Intersubunit CoA and zinc binding pocket. The CoA is represented by a stick model. The zinc ion is depicted by a purple dot. Composite OMIT σ_A-weighted $|F_o| - |F_c|$ electron-density maps are contoured at 1σ and are rendered as blue and red meshes within 1.5 Å of the CoA and the zinc ion, respectively. F_o is the native structure-factor amplitude and F_c is the calculated structure-factor amplitude from the model lacking the coordinates for the ligand. The map was calculated to 1.9 Å resolution. The figure was generated with *PyMOL* (http://pymol.sourceforge.net).

3.4. The putative active site

It is possible to predict the location of the active site of TTHA1846 on the basis of its structural similarity to other thioesterases for which the active sites have been identified. Structural comparisons of TTHA1846 with three thioesterases, 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* (PsHTE) with its inhibitor 4-hydroxyphenacyl-CoA (PDB code 1lo7; Thoden *et al.*, 2002), that from *Arthrobacter* with its inhibitor 4-hydroxyphenacyl-CoA (ArHTE; PDB code 1q4t; Thoden *et al.*, 2003) and the phenylacetate-degradation protein PaaI, chains *A* and *E*, from *T. thermophilus* HB8 (TtPaaI) with its ligand hexanoyl-CoA (PDB code 1wn3; Kunishima *et al.*, 2005), revealed that the CoA ligand binds to these four thioesterases in distinctively different manners (Fig. 6). In particular, the adenine rings in these enzymes are pointing in different directions (Fig. 6). These

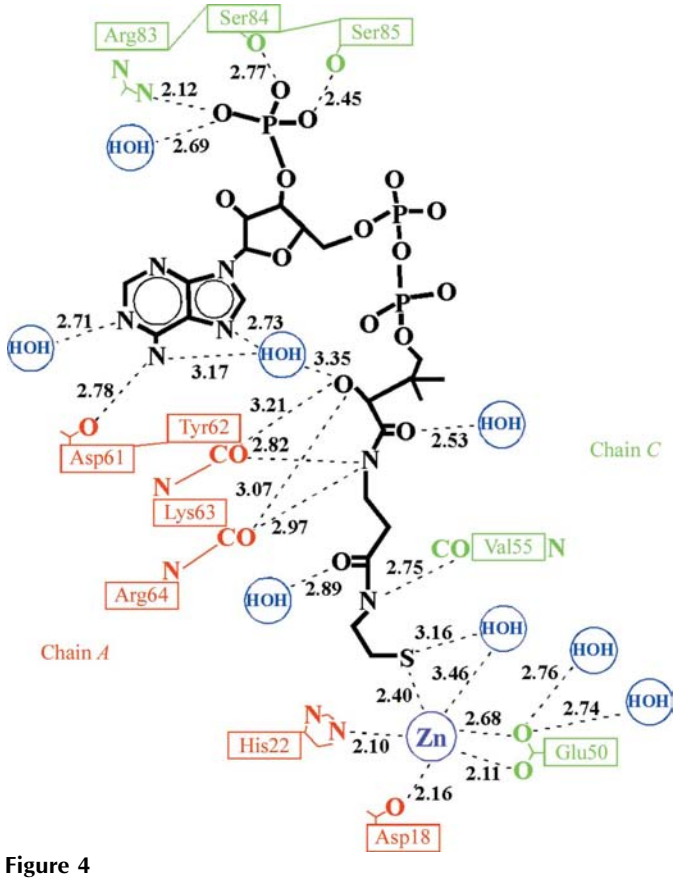

Figure 4
Ligand-recognition scheme. The ligand recognition by the TTHA1846 dimer is shown. Polar interactions are denoted by dotted lines, with distances in Å. Chains *A* and *C* and waters are coloured red, green and blue, respectively.

observations indicate that the nucleotide moiety of the CoA ligand can adapt to various quaternary structures, in contrast to the more restrained conformations of the phosphopantetheine moiety. Of these three thioesterase proteins, the global structure of our structure is very similar to that of PsHTE (Figs. 1*b* and 6).

The ligand binding sites of TTHA1846 and PsHTE are composed of only two subunits (Figs. 6 and 7*a*). In TTHA1846 and PsHTE the 4′-phosphopantetheine portion of CoA contacts the β-strand as it reaches the solvent and the 3′-phosphate group hydrogen bonds to the backbone peptidic NH groups positioned in the reverse turn, connecting the third and fourth β-strands. In contrast, in TtPaaI and ArHTE three subunits contribute to ligand binding. In TtPaaI and ArHTE the nucleotide is directed toward the centre of the protomer–protomer interface, where it interacts with residues contributed by a third monomer in the tetramer (Fig. 6). Therefore, the tetrameric state does not seem to be crucial for ligand binding in TTHA1846 and PsHTE, whereas it is important in TtPaaI and ArHTE. Despite the common hot-dog scaffold in these thioesterases, the substrate binding sites are not similar and the differences are apparently a function of the differing N- and C-terminal regions (Fig. 1*b*). In addition, as described above, we speculate that the difference is caused by the tetrameric state, as both TTHA1846 and PsHTE adopt the $\varepsilon\gamma$ mode.

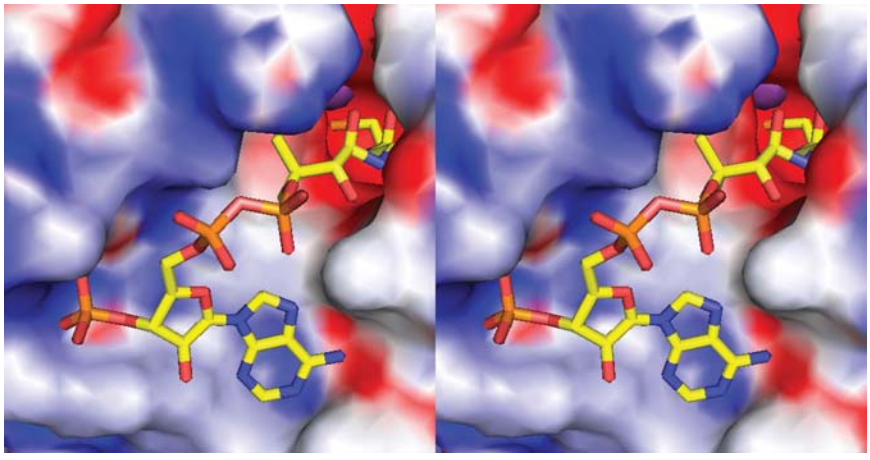

Figure 5
The CoA and zinc ion binding site. The solvent surface of the binding site is coloured according to the electrostatic potential (red, negative; blue, positive; grey, neutral). CoA is depicted as a stick model. The purple sphere is a zinc ion.

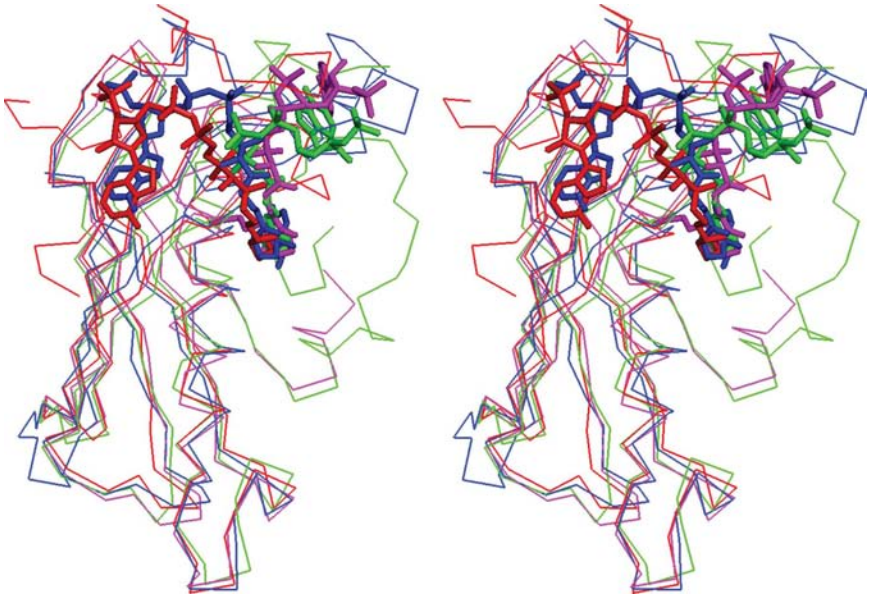

Figure 6
Stereo superimposition diagram of four thioesterase monomers with their ligands. TTHA1846, PsHTE (PDB code 1lo7), ArHTE (PDB code 1q4s) and TtPaaI (PDB code 1wn3) are displayed in red, blue, green and magenta, respectively. Their ligands are depicted as stick models.

In contrast, the conformational differences in the pantetheine unit are limited (Fig. 6). Each thiol group is positioned in the same area of the protomer crevice. Therefore, the active sites of these four enzymes are formed by their dimerization. The active site is a tunnel-shaped pocket formed at a monomer–monomer interface and two active sites exist per protomer. Although the environment of the tunnel is predominantly hydrophobic, three charged or polar residues, Asp18, His22 and Glu50, are localized at the bottom of the tunnel, forming a complex hydrogen-bond network with three waters (Fig. 4). An inspection of the active-site structure suggests that Asp18 and His22 are the most promising candidates for the catalytic residues. Asp18 and His22 are conserved in other putative bacterial thioesterases (Fig. 1*a*), indicating the importance of these residues for the enzymatic function.

The essential component is believed to be an active-site carboxylate residue that mediates the hydrolysis reaction *via* either nucleophilic or general base catalysis. Most of the catalytic residues of proteins with the hot-dog fold are found on either the upstream loop or the long central α-helix. The catalytic residues are Asp17 in PsHTE (corresponding to Asp18), Glu73 in ArHTE (corresponding to Glu33; Benning *et al.*, 1998), Glu73 in ArHTE (corresponding to Glu33; Thoden *et al.*, 2003) and Asp48 in TtPaaI (corresponding to Glu33; Kunishima *et al.*, 2005) (Fig. 1*b*). In the case of TTHA1846, Asp18 is in the putative active loop and Glu33 is in the long α-helix. The zinc ion was bound to Asp18 and Glu50. However, Glu50 is located in the loop connecting the major α-helix to the second β-sheet. In contrast, the side chain of Glu33 in the second protomer is more than 6.3 Å away from the zinc ion. Therefore, the

Glu33 residue is not involved in the TTHA1846 thioesterase activity and thus the Asp18 residue and/or the zinc ion could act as catalytic components. Asp17 in PsHTE, which corresponds to Asp18 in TTHA1846, was shown to function as a direct nucleophile that attacks the thioester carbonyl C atom of 4-hydroxyphenacyl-CoA. Therefore, these two thioesterases may share a similar catalytic mechanism.

3.5. The role of the zinc ion

As a divalent cation, zinc plays a structural role in a wide range of important proteins. It is generally thought that zinc has a completely filled *d* shell with ten electrons and can interact strongly with a variety of ligand types, including sulfur from a cysteine, nitrogen from a histidine and oxygen from a glutamate, an aspartate and water (Berg & Shi, 1996; Wang *et al.*, 2003). In TTHA1846, the zinc ion is predicted to play an important role in the activity. We propose two possible roles for the zinc ion: (i) it functions in the active centre, as in a metalloprotease, or (ii) it inhibits substrate binding.

To probe the active site of TTHA1846, we constructed a superimposed structure of the TTHA1846–4-hydroxyphenacyl-CoA complex using the 4-hydroxyphenacyl-CoA coordinates from PsHTE (Fig. 7*a*). Between the protomer subunits, the CoA β-mercaptoethylamine and pantothenate units can indeed be fitted well into the crevices. The zinc ion, which binds to the inner side of the tunnel-shaped pocket, was located in the vicinity of the 4-hydroxyphenacyl moiety. This substrate binding pocket begins at the subunit interface between strands β3 and β4 from the other subunit of the protomer and heads toward the long central α-helix. At about 11 Å from the entrance, each pocket comes to a dead end at the zinc ion. Thus, there was no space for other probable substrates between the thiol group of CoA and the zinc ion in TTHA1846. Therefore, we proposed that the zinc ion does not function as an active centre but as an inhibitor of specific substrate binding or translocation of the new substrate.

3.6. Putative substrate

To date, the function of TTHA1846 has been uncertain owing to the lack of significant sequence similarity with proteins of known function in the databases (Fig. 1*a*). The TTHA1846 structure presented here will provide the basis for the design of further studies to clarify the functions of this family of proteins.

XC229 (PDB code 2fuj) is the substrate-unbound crystal structure of a hot-dog fold thioesterase protein. The amino-acid sequence of XC229 is homologous to that of TTHA1846 (34% identity; Figs. 1*a* and 1*b*; Chin *et al.*, 2006), but the largest differences between the TTHA1846 and XC229 thioesterases exist within the loop between α2 and β2 (Fig. 7*b*). In TTHA1846, Glu50, which is located on the loop, interacts with the zinc ion. On the other hand, in XC229 Asp52 (corresponding to Glu50 in TTHA1846) is far from the putative catalytic site (Asp20 in XC229). The *B* factors of the loop and of the overall structure were 19.9 and 16.1 Å^2, respectively, in TTHA1846. Meanwhile, in XC229, the *B* factors of the loop and of the overall structure were 45.8 and 23.5 Å^2, respectively. As indicated by their *B* factors, the loop in XC229 is more mobile than that in TTHA1846.

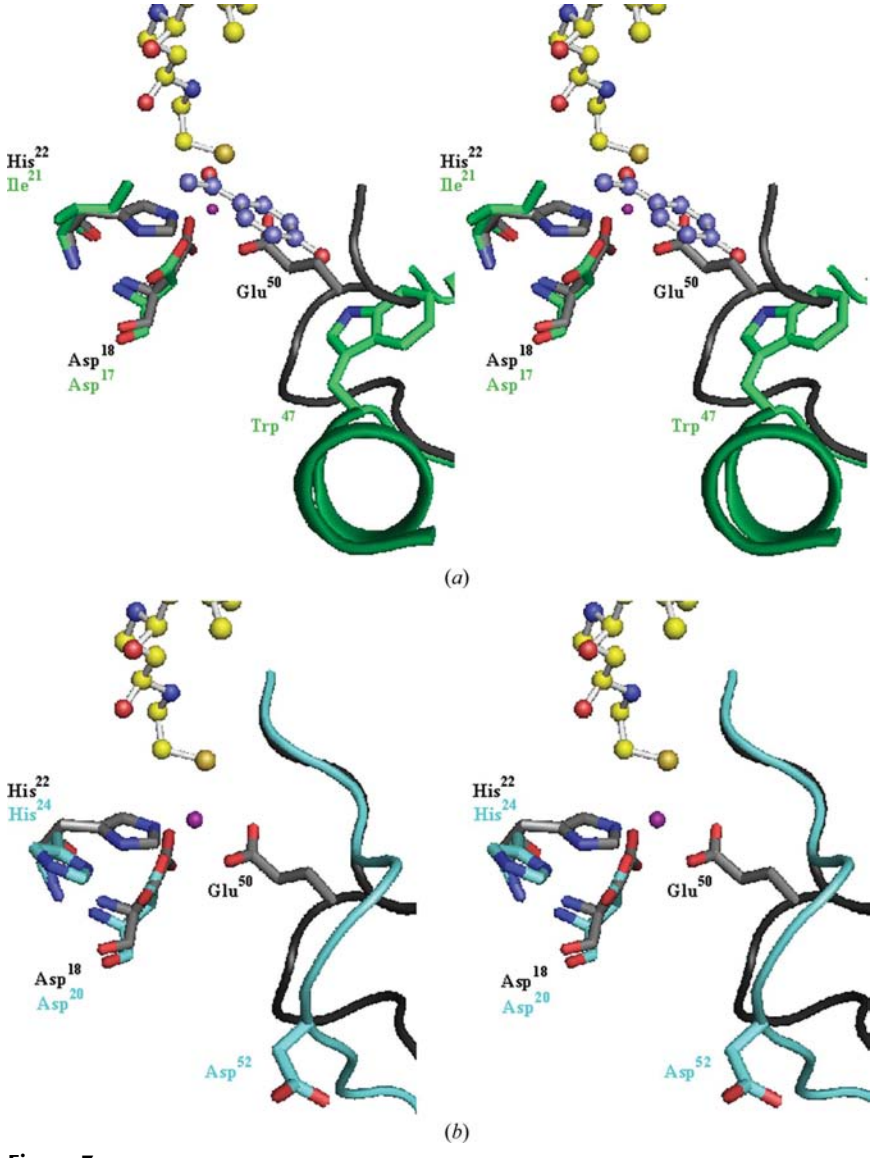

Figure 7
Superimposition of the active site. (*a*) TTHA1846 with CoA and PsHTE (PDB code 1lo7) with its inhibitor 4-hydroxyphenacyl-CoA are coloured black and green, respectively. (*b*) TTHA1846 with CoA and XC229 (PDB code 2fuj) are coloured black and cyan, respectively. The CoA is depicted as a ball-and-stick model. The purple dot is the zinc ion. The 4-hydroxyphenacyl moiety is depicted as a blue ball-and-stick model.

Substrate-bound crystal structures are available for several hot-dog fold thioesterases, including PsHTE, ArHTE and TtPaaI. PsHTE has also been crystallized with the inhibitor 4-hydroxyphenacyl-CoA (PDB code 1lo7) and the inactive mutant (Asp17Asn) was crystallized with the substrate 4-hydroxybenzoyl-CoA (PDB code 1lo9), allowing a detailed description of its active site (Thoden *et al.*, 2002). The secondary structure and the direction of the adenine ring of TTHA1846 are similar to those of PsHTE (Figs. 1*b* and 6). Although the sequence identity and similarity between TTHA1846 and PsHTE are only 17.2% and 36.1%, respectively, as determined by *ClustalW* (Thompson *et al.*, 1994), the tertiary structures are actually quite similar (r.m.s.d. 1.7 Å; Figs. 6 and 7*a*). Several amino-acid residues close to the zinc ion binding site, including Asp18 and His22, are also conserved. The Asp17 residue (corresponding to Asp18 in TTHA1846) is located 3.2 Å away from the carbonyl group of the substrate, where it can act as a nucleophile, as determined from mutagenesis (Thoden *et al.*, 2002). The major difference between the active sites is represented by a loop connecting α2 to strand β2, which is closer in TTHA1846 compared with PsHTE. The side chain of Trp47 (corresponding to Glu50 in TTHA1846) is located on the loop between α2 and β2 and interacts with the aromatic moiety of the substrate (4-hydroxyphenacyl; Fig. 7*a*). Therefore, this loop may interact with the substrate moiety and may be involved in the binding specificity and stability.

As shown in Fig. 7(*a*), the inhibitor binding region of PsHTE is broader than that of TTHA1846, indicating that TTHA1846 may recognize a smaller substrate than that of PsHTE. Thus, Glu50 defines the size and charge of the substrate binding pocket, which is sufficient to accommodate a hydrophilic substrate but unable to accommodate an acyl chain. To accommodate acyl substrates, the tunnel may need to protrude further into the hydrophobic core of the protein. The substrate for HpYbgC (PDB code 2pzh) is a long acyl chain such as palmitoyl-CoA, stearoyl-CoA or myristoyl-CoA or an aromatic chain such as benzoyl-CoA, while PsHTE is only active towards aromatic chain CoAs (Angelini *et al.*, 2008). In fact, HpYbgC presents a long tunnel associated with the region binding the acyl moiety of the substrate, whereas this tunnel is absent in TTHA1846 (Fig. 5). The electrostatic potential distribution around Glu50 showed that negatively charged regions exist at the bottom of the CoA binding site (Fig. 5), suggesting that TTHA1846 interacts with some positively charged molecules.

TTHA1846 seems to bind both the CoA ligand and a zinc ion strongly (as they survived the extensive purification procedures). How can such an inactive conformation be reactivated? We think that the affinity for the actual substrate is very high and that this substrate does not exist in *E. coli*. Further studies will be needed to identify the substrate and to clarify the catalytic mechanism of TTHA1846.

This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, the Ministry of Education, Culture, Sports, Science and Technology of Japan. We thank Drs N. Handa, M. Kukimoto-Niino and S. Kishishita for helpful advice. We thank Ms Yukiko Kinoshita for preparation of the protein. We also thank the beamline staff for their assistance during data collection at beamline BL26B1 of SPring-8.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Angelini, A., Cendron, L., Goncalves, S., Zanotti, G. & Terradot, L. (2008). *Proteins*, **72**, 1212–1221.
Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). *Nucleic Acids Res.* **28**, 263–266.
Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D. & Holden, H. M. (1998). *J. Biol. Chem.* **273**, 33572–33579.
Berg, J. M. & Shi, Y. (1996). *Science*, **271**, 1081–1085.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cheng, Z., Song, F., Shan, X., Wei, Z., Wang, Y., Dunaway-Mariano, D. & Gong, W. (2006). *Biochem. Biophys. Res. Commun.* **349**, 172–177.
Chin, K. H., Chou, C. C., Wang, A. H. & Chou, S. H. (2006). *Proteins*, **64**, 823–826.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, Palo Alto, California, USA.
Dillon, S. C. & Bateman, A. (2004). *BMC Bioinformatics*, **5**, 109.
Forwood, J. K., Thakur, A. S., Guncar, G., Marfori, M., Mouradov, D., Meng, W., Robinson, J., Huber, T., Kellie, S., Martin, J. L., Hume, D. A. & Kobe, B. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 10382–10387.
Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.
Harding, M. M. (2004). *Acta Cryst.* D**60**, 849–859.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Holm, L. & Sander, C. (1998). *Nucleic Acids Res.* **26**, 316–319.
Hosaka, T., Meguro, T., Yamato, I. & Shirakihara, Y. (2003). *J. Biochem. (Tokyo)*, **133**, 817–823.
Hunt, M. C. & Alexson, S. E. (2002). *Prog. Lipid Res.* **41**, 99–130.
Kunishima, N., Asada, Y., Sugahara, M., Ishijima, J., Nodake, Y., Sugahara, M., Miyano, M., Kuramitsu, S., Yokoyama, S. & Sugahara, M. (2005). *J. Mol. Biol.* **352**, 212–228.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. (1996). *Structure*, **4**, 253–264.
Li, J., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). *Nature Struct. Biol.* **7**, 555–559.
Marchler-Bauer, A. & Bryant, S. H. (2004). *Nucleic Acids Res.* **32**, W327–W331.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Murata, T., Yamato, I., Kakinuma, Y., Leslie, A. G. & Walker, J. E. (2005). *Science*, **308**, 654–659.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Reynolds, C., Damerell, D. & Jones, S. (2009). *Bioinformatics*, **25**, 413–414.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Thoden, J. B., Holden, H. M., Zhuang, Z. & Dunaway-Mariano, D. (2002). *J. Biol. Chem.* **277**, 27468–27476.

Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. & Holden, H. M. (2003). *J. Biol. Chem.* **278**, 43709–43716.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Wang, F., Wang, X., Liu, Y., Tian, W. X. & Zhou, H. M. (2003). *Int J. Biochem. Cell Biol.* **35**, 391–400.
Willis, M. A., Zhuang, Z., Song, F., Howard, A., Dunaway-Mariano, D. & Herzberg, O. (2008). *Biochemistry*, **47**, 2797–2805.
Yamamoto, M., Kumasaka, T., Ueno, G., Ida, K., Kanda, H., Miyano, M. & Ishikawa, T. (2002). *Acta Cryst.* A**58**, C302.
Zhuang, Z., Song, F., Takami, H. & Dunaway-Mariano, D. (2004). *J. Bacteriol.* **186**, 393–399.