Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The 1.8 Å resolution structure of hydroxycinnamoyl-coenzyme A hydratase-lyase (HCHL) from *Pseudomonas fluorescens*, an enzyme that catalyses the transformation of feruloyl-coenzyme A to vanillin

Philip M. Leonard,[a] A. Marek Brzozowski,[a] Andrey Lebedev,[a] Caroline M. Marshall,[a] Derek J. Smith,[b] Chandra S. Verma,[b] Nicholas J. Walton[c] and Gideon Grogan[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, England, [b]Bioinformatics Institute, 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore, and [c]Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, England

Correspondence e-mail: grogan@ysbl.york.ac.uk

Received 25 July 2006
Accepted 25 September 2006

PDB Reference: HCHL, 2j5i, r2j5isf.

The crystal structure of hydroxycinnamoyl-CoA hydratase-lyase (HCHL) from *Pseudomonas fluorescens* AN103 has been solved to 1.8 Å resolution. HCHL is a member of the crotonase superfamily and catalyses the hydration of the acyl-CoA thioester of ferulic acid [3-(4-hydroxy-3-methoxy-phenyl)-prop-2-enoic acid] and the subsequent retro-aldol cleavage of the hydrated intermediate to yield vanillin (4-hydroxy-3-methoxy-benzaldehyde). The structure contains 12 molecules in the asymmetric unit, in which HCHL assumes a hexameric structure of two stacked trimers. The substrate, feruloyl-CoA, was modelled into the active site based on the structure of enoyl-CoA hydratase bound to the feruloyl-CoA-like substrate 4-(*N*,*N*-dimethylamino)-cinnamoyl-CoA (PDB code 1ey3). Feruloyl-CoA was bound in this model between helix 3 of the *A* subunit and helix 9 of the *B* subunit. A highly ordered structural water in the HCHL structure coincided with the thioester carbonyl of feruloyl-CoA in the model, suggesting that the oxyanion hole for stabilization of a thioester-derived enolate, characteristic of coenzyme-A dependent members of the crotonase superfamily, is conserved. The model also suggested that a strong hydrogen bond between the phenolic hydroxyl groups of feruloyl-CoA and *B*Tyr239 may be an important determinant of the enzyme's ability to discriminate between the natural substrate and cinnamoyl-CoA, which is not a substrate.

1. Introduction

The microbial degradation of phenolic compounds from plant residues is of major biological and economic importance since it is an essential process in the decay of wood, in the disposal and recycling of plant wastes and in the industrial production of several commercially significant substances, such as the flavour and aroma compound vanillin (4-hydroxy-3-methoxy-benzaldehyde), by biotransformation. In recent years, considerable progress has been made in the biochemical and molecular-genetic characterization of these microbial catabolic pathways.

One case of special interest is the bacterial degradation of eugenol [2-methoxy-4-(2-propenyl)phenol] and ferulic acid [3-(4-hydroxy-3-methoxy-phenyl)prop-2-enoic acid] to produce vanillin (Walton *et al.*, 2003; Lomascolo *et al.*, 1999; Rabenhorst, 1996). Eugenol is first transformed to ferulic acid by a series of terminal oxidation steps and then to vanillin by a coenzyme A-dependent pathway. Narbad and coworkers isolated a bacterium *Pseudomonas fluorescens* AN103 which was able to grow on ferulic acid as the sole carbon source and

in which a catabolic pathway *via* vanillin was outlined (Narbad & Gasson, 1998). Ferulic acid was first ligated to coenzyme A to form feruloyl-CoA by the action of 4-hydroxycinnamate-CoA ligase-synthetase (HCLS; Fig. 1). The acyl-CoA thioester was then transformed to vanillin (**4**) by the action of a single enzyme, hydroxycinnamoyl-CoA hydratase-lyase (HCHL), which first catalyses the hydration of the double bond to yield 4-hydroxy-3-methoxy-phenyl-β-hydroxy-propionyl-CoA (**3**); the retro-aldol cleavage of (**3**) then gives (**4**) and acetyl-CoA. Equivalent pathways from ferulic acid to vanillin have been described and studied in *Pseudomonas* sp. strain HR199 (Priefert *et al.*, 1997) and also in strains of *Streptomyces setonii* (Muheim & Lerch, 1999) and *Amycolatopsis* (Rabenhorst & Hopp, 1997), the latter yielding 11.5 g vanillin per litre in shake-flask culture.

The enzymatic transformation of (**2**) into (**4**) represents an interesting mode of enzymatic activity that is reminiscent of the hydration of double bonds in enoyl-CoA and related substrates in fatty-acid oxidation pathways by the enzyme enoyl-CoA hydratase (ECH), sometimes known as crotonase (Engel *et al.*, 1996). ECH lends its name to a low-sequence-homology superfamily of enzymes known as the crotonase or low-similarity hydratase/isomerase (LSI/H) superfamily, whose varied catalytic chemistry, allied to pronounced structural similarity, has been the subject of much recent interest (Holden *et al.*, 2001). The gene encoding HCHL from AN103 was cloned and expressed in *Escherichia coli* by Gasson *et al.* (1998). The amino-acid sequence revealed HCHL to be a member of the crotonase superfamily, bearing 24% sequence identity to the ECH from rat liver. Crotonase superfamily members have been observed to catalyse a number of interesting chemical reactions, including the stereospecific hydration of double bonds performed by ECH and also dehalogenation (Benning *et al.*, 1996), double-bond isomerization in fatty acids (Modis *et al.*, 1998; Mursula *et al.*, 2001) and cyclization/aromatization reactions in the synthesis of vitamin K intermediates (Truglio *et al.*, 2003). The spectrum of reactions catalysed by members of the superfamily provides persuasive evidence that the function of enzymes encoded by certain open reading frames may not be predicted from amino-acid sequence alone, as the divergent catalytic chemistry performed is often enabled by different amino-acid residues throughout the sequences which are often not conserved between superfamily members (Gerlt & Babbitt, 2001). A number of recent reports also describe the ability of certain crotonase homologues to cleave carbon–carbon bonds. 2-Ketocyclohexanecarbonyl-CoA hydrolase (BadI) from *Rhodopseudomonas palustris* catalyses the retro-Dieckmann condensation of its natural substrate by an as yet unknown mechanism (Pelletier & Harwood, 1998; Eberhard & Gerlt, 2004). A formally similar reaction is catalysed by 6-oxo camphor hydrolase (OCH), which cleaves a C—C bond in 6-oxo camphor with almost absolute prochiral selectivity for one face of the symmetrical substrate (Grogan *et al.*, 2001; Whittingham *et al.*, 2003; Leonard & Grogan, 2004). The reaction catalysed by HCHL is again difficult to rationalize mechanistically as, in addition to the hydration step, a carbon–carbon bond is cleaved. In order to illuminate the mechanism of this interesting and useful member of the crotonase superfamily, a study of the X-ray crystal structure was undertaken and a preliminary report of crystallization conditions and the analysis of a data set has been published (Leonard *et al.*, 2004). In this paper, we present a solution of the structure of HCHL, solved by molecular-replacement techniques using the ECH from *Thermus thermophilus* (PDB code 1uiy) as a model. The structure of HCHL reveals that it has the intra-trimer association fold displayed by a subset of known crotonase structures, including enoyl-CoA hydratase and 4-chlorobenzoyl-CoA dehalogenase, and that the putative oxyanion hole thought to stabilize intermediate enolates in coenzyme A-dependent members of this superfamily is conserved. A model of the natural substrate, feruloyl-CoA, in the active site, is strongly indicative of a catalytic role for Glu143 and a substrate-specificity determining role for Tyr239 from the neighbouring monomer.

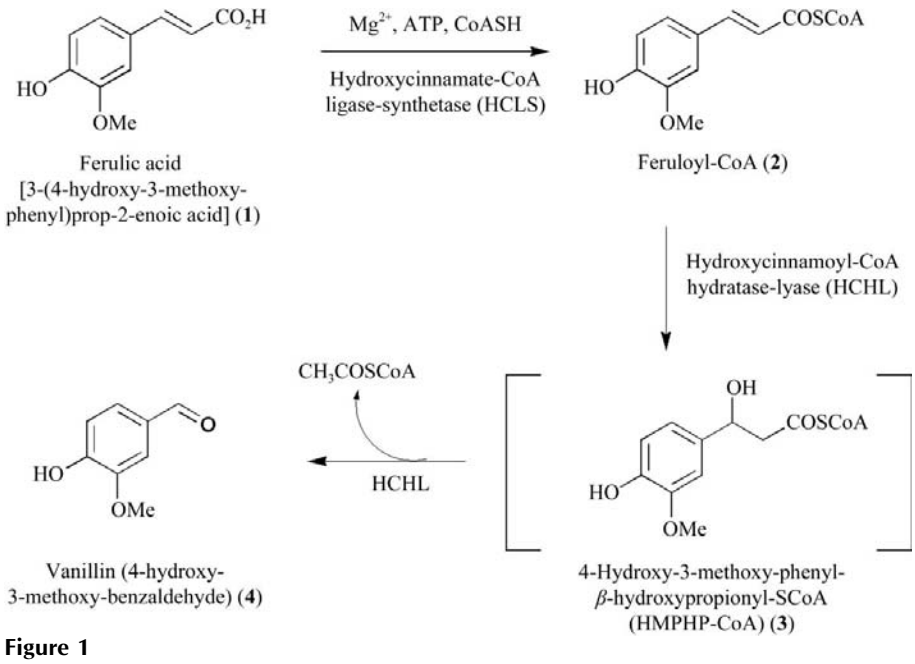


Figure 1
Metabolism of ferulic acid [3-(4-hydroxy-3-methoxy-phenyl)prop-2-enoic acid] (**1**) to vanillin (4-hydroxy-3-methoxybenzaldehyde) (**4**) by *P. fluorescens* AN103.

2. Materials and methods

2.1. Chemicals

All chemicals were obtained from Sigma–Aldrich (Poole, England) unless otherwise specified. Butane-1,4-diol and vanillin (4-hydroxy-3-methoxy-benzaldehyde) were of 99% purity. Feruloyl-

Table 1
Data-collection and refinement statistics for HCHL.

Values in parentheses refer to the highest resolution shell.

Beamline	ID14-EH1
Wavelength (Å)	0.934
Resolution (Å)	30–1.8 (1.85–1.8)
Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 154.24, b = 167.49, c = 130.81
Unique reflections	308927 (25526)
Completeness (%)	100 (100)
R_{sym}† (%)	4.5 (33.2)
Multiplicity	4.2 (4.2)
$\langle I/\sigma(I)\rangle$‡	21.9 (3.3)
Protein atoms	23434
Solvent waters	1884
R_{cryst}	0.179
R_{free}	0.215
R.m.s.d. 1–2 bonds (Å)	0.018
R.m.s.d. 1–3 angles (Å)	1.639
Average main-chain B (Å^2)	20
Average side-chain B (Å^2)	25
Average solvent B (Å^2)	32

† $R_{sym} = \sum_{\mathbf{h}}\sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_l \langle I_{\mathbf{h}}\rangle$, where I_l is the lth observation of reflection $\mathbf{h}$ and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations l of reflection $\mathbf{h}$. ‡ $\langle I/\sigma(I)\rangle$ indicates the average of the intensity divided by its average standard deviation.

CoA was prepared using the method previously described (Gasson *et al.*, 1998). Reagents for molecular biology, including competent *E. coli* BL21(DE3) cells, were purchased from Novagen.

2.2. Plasmids

The plasmid construct containing the gene encoding HCHL, pFI3009, was obtained from Dr Arjan Narbad at the Institute of Food Research, Norwich, England.

2.3. Overexpression and purification

HCHL protein was purified from an overexpressing *E. coli* strain BL21(DE3) which had been transformed with the plasmid pFI1039 encoding the gene for *P. fluorescens* biovar V, strain AN103 HCHL (Gasson *et al.*, 1998), which had been constructed from the pSP72 plasmid (Novagen). The expression of the gene and isolation of the protein are described in Leonard *et al.* (2004).

2.4. Crystallization and data collection

HCHL was crystallized as described previously (Leonard *et al.*, 2004). In brief, crystallization conditions were established using the hanging-drop vapour-diffusion technique with the Clear Strategy Screen (CSS) from Molecular Dimensions Ltd (Brzozowski & Walton, 2001). The optimum conditions were obtained with a protein concentration of 10 mg ml^{-1} in 11%(*w*/*v*) PEG 20 000 with 8%(*v*/*v*) PEG 550 monomethyl ether, 0.8 *M* sodium formate, 0.2%(*v*/*v*) butane 1,4-diol and 10 m*M* vanillin in 0.05 *M* 2-(*N*-morpholino)ethanesulfonic acid buffer pH 5.6. A 1.8 Å resolution data set was collected at ESRF Grenoble station ID14-EH1 (Leonard *et al.*, 2004; see Table 1 for details).

2.5. Structure solution

Three structures of HCHL sequence homologues, rat liver (24% sequence identity; PDB code 1dub) and *T. thermophilus* (PDB code 1uiy; 28% sequence identity) enoyl-CoA hydratases and 4-chlorobenzoyl-CoA dehalogenase from *Pseudomonas* sp. (PDB code 1nzy; 28% sequence identity), were used in initial molecular-replacement trials with *MOLREP* (Vagin & Teplyakov, 1997). Neither full trimer/hexamer models nor their truncated versions yielded any contrast in rotation and translation functions over various resolution ranges. In addition to a likely different orientation of the two trimers within the hexamer of HCHL in relation to model molecules, molecular replacement was complicated further by an NCS translation vector (0.34, 0.3, 0.5 and its equivalents; the height of the corresponding Patterson peaks are ∼23% of the origin peak at 3 Å resolution cutoff) that was sufficiently strong to not be treated properly in the currently available MR programs. Therefore, a one-dimensional search was undertaken where the target function was the maximum value of the cross-rotation function (CRF) and the variable parameter was the relative rotation of trimers within the model of the hexamer. A set of models was generated from the hexamers of the above three proteins with the relative rotations of trimers in the range 0–120° (sufficient range for point group 32) and with an increment of 2°. From all models tried, a model based on the structure 1uiy with relative rotation of trimers by ∼9° gave the best contrast in the CRF calculated using *MOLREP* at 5 Å resolution cutoff. The first six peaks in the CRF for this particular model were in the range 6.2–7.2 SRF/σ(SRF) compared with the seventh peak of 3.6. The hexamer rotated according to one of these six equivalent peaks was used in a translation search. The first two translation solutions were correct and subsequent rigid-body refinement and an initial round of refinement by *REFMAC* (Murshudov *et al.*, 1997) gave an R value of 0.43 and an R_{free} of 0.52.

2.6. Model building and refinement

5% of the total reflections were flagged for cross-validation before refinement. These data were used to monitor the modelling process at various stages of refinement for the weighting of geometrical and temperature-factor restraints. All computing was undertaken using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Model building and refinement were carried out with *REFMAC* (Murshudov *et al.*, 1997) in conjunction with *ARP*/*wARP* (Perrakis *et al.*, 1999) in the whole (30–1.8 Å) resolution range. *Coot* (Emsley & Cowtan, 2004) was used for manual corrections to the model. The N-termini of HCHL were visible in electron-density maps from Thr3/Tyr4. The C-termini were only modelled to Leu248/Asp249/Thr250 in particular monomers and, as they pointed towards large cavities within the crystal space, their lack of structural definition suggested disorder or truncation of the last ∼27 residues (see §3). A total of 1884 water molecules were identified. Although preliminary maps indicated some residual electron density that could be allocated to a vanillin molecule, the final maps did not allow

modelling or refinement of this ligand. The final R_{cryst} and R_{free} are 0.179 and 0.215, respectively, with 94.1% residues in the most favoured regions, 5.6% in additional allowed regions and 0.3% in generously allowed regions as indicated by *PROCHECK* (Laskowski *et al.*, 1993).

2.7. *CHARMM* modelling

The crystal structure of HCHL was used as the starting model for the calculations. A dimer (chains *A* and *B*) was used for this study since part of the ligand lies at the interface. In addition, all crystallographic water molecules within 5 Å of either chain were used. The dimer was superposed onto the crystal structure of enoyl-CoA hydratase (PDB code 1ey3; Bahnson *et al.*, 2002) with an r.m.s.d. of 1.7 Å for C^{α} atoms. The associated substrate, 4-(*N*,*N*-dimethylamino)cinnamoyl-coenzyme A, was taken and modifed structurally to generate feruloyl-coenzyme A and then superposed into the corresponding pocket in HCHL. Seven water molecules within the pocket of HCHL that were involved in steric clashes with the feruloyl-CoA were then removed.

Energy minimization was performed on the complex using the program *CHARMM* (Brooks *et al.*, 1983; MacKerell *et al.*, 1998) to relax the structure and remove close steric overlaps between the protein atoms and the modelled feruloyl-CoA. This involved adding all H atoms to the system using *HBUILD* (Brünger & Karplus, 1988), followed by a protocol of 100 steps of steepest-descent minimization and 300 steps of conjugate-gradient minimization with the backbone, side chains and ligand restrained with harmonic potentials of 42, 21 and 4.2 kJ mol^{-1} Å^{-3}, respectively. Following this, a second round of minimization was carried out as above, but with an additional 800 steps of ABNR minimization and backbone, side-chain and ligand restraints lowered to 21, 8.4 and 0.42 kJ mol^{-1} Å^{-3}, respectively.

3. Results and discussion

3.1. Fold of the monomer

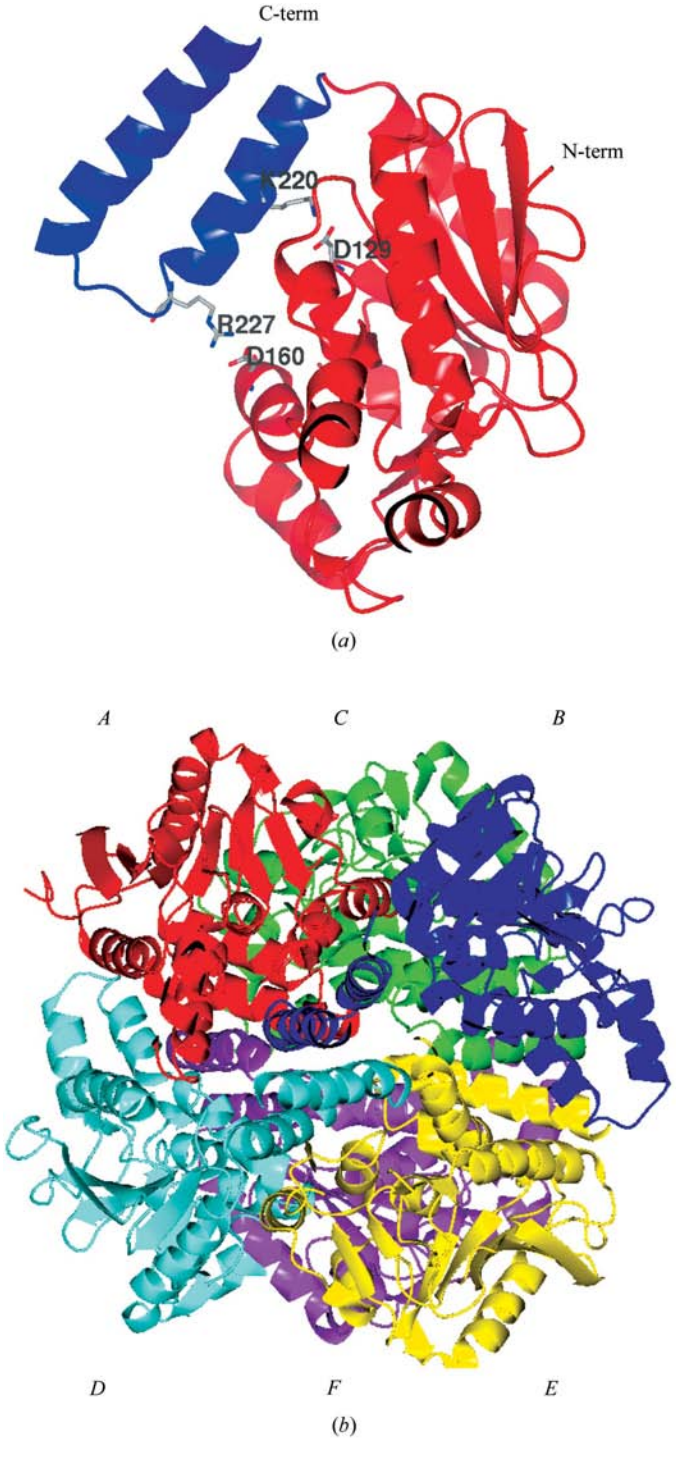
The HCHL monomer encoded by Orfa as cloned by Gasson *et al.* (1998) was a polypeptide consisting of 276 amino acids. The most C-terminal residues that were considered suitable for building in the maps were Leu248–Thr250, suggesting that approximately 27 C-terminal amino acids of HCHL were either highly disordered or absent. MALDI mass-spectrometric analysis of the isolated protein gave a weight of 28 658 Da, compared with the predicted weight of 31 008 Da for the full mature gene product, suggesting a deletion, perhaps through proteolysis, of the C-terminus of HCHL during protein cultivation and isolation. The most C-terminal helix of HCHL that would be predicted on the basis of structural homology to ECH and other members of the LSI/H superfamily is therefore absent in the final monomer. The monomer reveals that HCHL displays a distinct C-terminal domain fold, as observed for ECH (Engel *et al.*, 1996), dienoyl-CoA isomerase (Modis *et al.*, 1998), 4-chlorobenzoyl dehalogenase (Benning *et al.*, 1996) and the human AUH protein (Kurimoto *et al.*, 2001). The N-terminal section is comprised of a four-spiral turn of mixed α and β character that results in an inverted prism structure, with a five-strand β-sheet at the top in which the most N-terminal strand is antiparallel to the remaining four, which run parallel to each other (Fig. 2*a*). There are eight α-helical regions in the N-terminal domain, the most C-terminal of which loops back over the β-sheet at the top of the monomer as shown. The C-terminal domain of the truncated form of HCHL presented is comprised of two α-helices which form a hairpin joined by a turn between residues Cys226 and Thr230. The monomer superimposes with that of ECH, with an r.m.s. $C^{\alpha}-C^{\alpha}$ distance of 1.64 Å for 215 matched residues, indicative of the very close structural similarity between the monomer cores of these proteins, despite the comparatively low amino-acid sequence identity of 24%. Amongst the interactions stabilizing the monomer structure are salt bridges that fix the C-terminal and N-terminal domains between Asp129 and Lys220, which is conserved in ECH, and also between Arg227 and Asp160 and between Arg103 and Glu228, neither of which is conserved in the ECH monomer (Fig. 2*a*).

When the truncated enzyme was incubated with the substrate feruloyl-CoA and the reaction monitored by TLC, the enzymatic production of vanillin was clearly observed when compared with the appropriate buffer-only controls, indicating that the HCHL used in crystallization trials was active in the absence of the C-terminal amino acids from Thr250. The structure of enoyl-CoA hydratase bound to acetoacetyl-coenzyme A (PDB code 1dub; Engel *et al.*, 1996) features the equivalent missing C-terminus from residues Asp270 to His290. In the ECH structure, this C-terminal region is peripheral to the hexamer, yet includes amino-acid side chains that contribute to relevant monomer–monomer interactions and and also to coenzyme A binding. The Arg272 hydrogen binds to the main-chain carbonyl of Gln104 in the neighbouring monomer; Arg273 forms a salt bridge with the side chain of Asn105 in the neighbouring monomer. The terminal amino group of Lys282 is bound to the ribose phosphate of coenzyme A and the side chain of Phe279 protrudes into the substrate-binding channel and appears to constrain the adenine ring. Arg272, Lys282 and Phe79 are all conserved in HCHL, yet HCHL is at least partially active in the absence of these residues. There are to our knowledge no data which describe the activity of an ECH deletion mutant which lacks the relevant portion of the C-terminus.

3.2. Factors affecting trimer assembly

The structure of the HCHL monomer gives rise to the intra-trimer domain-swapping fold as defined by Hubbard *et al.* (2005) that is shared by ECH, dienoyl-CoA isomerase, 4-chlorobenzoyl dehalogenase and the human AUH protein. The trimeric substructure is at odds with the previous description of HCHL as a functional dimer as determined by calibrated gel filtration (Mitra *et al.*, 1999). In crotonases with the intra-trimer domain-swapping fold, the trimer is stabilized by extensive interactions caused by the swapping of the

C-terminal domain of one monomer with its neighbour, whereas crotonases with the self-association fold, wherein the C-terminal helix loops back over the N-terminal domain, have been shown to have a metal ion at the threefold axis that may contribute to trimer stability (Whittingham *et al.*, 2003; Benning *et al.*, 2000). The nature of the monomer fold in HCHL may be determined by the turn between HCHL *A*Glu228 and *A*Glu232. The first of these residues forms a salt bridge with an arginine (HCHL *A*Arg103) in the N-terminal domain and is conserved in ECH, but does not form an equivalent salt bridge. As the conservation of amino-acid sequence throughout the length of the C-terminal helix is poor, it is difficult to attribute the formation of the discrete C-terminal domain to the primary sequence as such, but the question remains as to the primary factors that result in the recruitment of either the intra-trimer or the self-associating fold for the catalytic chemistry of different crotonases. This does not seem to be merely dependent on coenzyme A binding, as crotonases that are both dependent on and independent of coenzyme A for activity, such as methylmalonyl-CoA decarboxylase (Benning *et al.*, 2000) and 6-oxo camphor hydrolase (Whittingham *et al.*, 2003), respectively, have been shown to possess the self-association fold. Amongst other interactions that are involved in monomer *A*–monomer *B* interactions in the trimer, there is a deeply embedded tryptophan residue *B*Trp231 that hydrogen bonds through its side chain to *A*Asp160. *B*Tyr239 is hydrogen bonded to *A*Asn152; *B*Glu235 forms a salt bridge with *A*Lys100 and is hydrogen bonded to *A*Ser95. It has been suggested that three residues thought to be conserved amongst all members of the LSI/H superfamily (Gly115, Asp129 and Gly186 in HCHL) serve a function in either substrate binding or trimer integrity (Hubbard *et al.*, 2005), although in the HCHL structure none of these residues makes close contacts with a neighbouring monomer.



3.3. Oligomer structure

HCHL is a homohexamer consisting of a dimer of trimers as described for the majority of previously solved crotonase structures. 12 molecules were observed in the asymmetric unit (chains *A*–*F* and *G*–*L*), giving rise to two hexamers. A side view of the hexamer is illustrated in Fig. 2(*b*). The hexamer

Figure 2
(*a*) Structure of HCHL monomer. The N-terminal domain is shown in red and the C-terminal domain in blue. Residue side chains that form salt bridges that stabilize the domain interactions between Asp129 and Lys220 and between Arg227 and Asp160 are shown as cylinders. (*b*) Side view of the HCHL hexamer illustrating monomers *A* (red), *B* (blue), *C* (green), *D* (cyan), *E* (yellow) and *F* (magenta). (*c*) Semi-transparent surface model of HCHL with underlying ribbons of monomer *A* (light blue) and *B* (pink) showing modelled coenzyme A ferulic acid ester [3-(4-hydroxy-3-methoxy-phenyl)prop-2-enoic acid, yellow) in the active-site cleft at the monomer interface.

has dimensions of approximately 77 × 80 × 74 Å, occupying a space of 455 840 Å^3. Close inter-trimer contacts were observed between *D*Arg92 and *B*Glu49, and between *D*Asp236, *C*Lys241 and *C*Tyr237. A comparison of the architecture of the trimers and hexamers of HCHL homologues is summarized in Table 2. These data support and explain the efficiency of *T. thermophilus* ECH (ECH*Tt*; PDB code 1uiy) as the model in the molecular-replacement strategy. The best superposition of HCHL and ECH*Tt* trimers requires only ∼1.8° rotation, indicative of their similar organization. However, the assemblies of these trimers into hexamers are quite different and a rotation of ECH*Tt* trimers by ∼2 × 4.8° with respect to each other was required for the best superposition of both enzymes. The combination of a wide range of rotational differences in quaternary structures of trimers (∼1.8–3.8°) and hexamers (∼0.7–21.6°) within the crotonase superfamily illustrates the plasticity of the trimer/hexamer architecture that is adopted to support efficient catalysis of a particular type of chemical process.

Table 2
Comparison of molecular architecture within the LSI/H superfamily.

For comparison of trimers, each trimer was firstly globally superimposed with the HCHL trimer. This was the subsequent starting position for the best superposition of corresponding monomers of those trimers: it resulted in the rotation by a certain angle that is quoted here as angle (1↔3). Hexamers were also initially globally superimposed and then one trimer of the relevant protein was fitted onto the corresponding trimer of HCHL, giving the rotation angle that is quoted here as angle (3↔6). (The relative rotation of trimers required for the best fit of hexamers is twice as large.) R.m.s.d.s were calculated for three-dimensionally aligned C$^\alpha$ atoms. All superpositions and three-dimensional alignments were performed using the program *O* (Jones *et al.*, 1991).

PDB code	1dub	1hnu	1nzy	1uiy
Angle (1↔3) (°)	3.84	8.65	2.42	1.84
Angle (3↔6) (°)	0.72	21.57	14.93	4.77
R.m.s.d. (monomer) (Å)	1.62	2.12	1.63	1.60
R.m.s.d. (trimer) (Å)	2.08	2.83	1.78	1.80
R.m.s.d. (hexamer) (Å)	2.17	9.98	6.43	2.70

3.4. Active site: modelling feruloyl-coenzyme A into the active site using the structure of enoyl-CoA hydratase bound to 4-(*N*,*N*-dimethylamino)-cinnamoyl-CoA as a basis

The active sites of HCHL, in common with other members of the subgroup of crotonases possessing the intra-trimer association fold, are located at the trimer interface between, for example, helix 3 of the *A* subunit and helix 9 of the *B* subunit, located in the C-terminus of the subunit, by virtue of the domain-swapping phenomenon. A channel at the trimer interface, as revealed by a surface map, houses the putative active site of HCHL (Fig. 2*c*). The substrate feruloyl-CoA has extensive structural similarities with the 4-(*N*,*N*)-dimethylaminocinnamoyl-CoA that has been crystallized in the active site of ECH (Bahnson *et al.*, 2002; PDB code 1ey3), a protein to which HCHL bears significant structural homology throughout the length of the polypeptide chain. A simple three-dimensional overlap of the HCHL and 1ey3 structures revealed that many of the expected binding interactions of 4-(*N*,*N*-dimethylamino)-cinnamoyl-CoA with ECH were conserved with HCHL. As such, this structure (1ey3) supplied a useful and relevant basis for the construction of an energy-minimized model of feruloyl-CoA bound in the active-site cleft of HCHL.

A simple graphical superimposition of 1ey3 bound to 4-(*N*,*N*-dimethyl)-cinnamoyl-CoA with the HCHL coordinates suggested that some of the structural determinants of coenzyme-A binding are well conserved between ECH and HCHL. 1ey3 thus provided a sound basis for *CHARMM*-based modelling of feruloyl-CoA into the structure of HCHL. The model suggests conservation of hydrogen bonds that secure the amide bonds in coenzyme A(HCHL Ala68, ECH Ala96) and the exocyclic adenine NH_2 to the peptide backbone (HCHL Met70, ECH Ala98).

In the model, Tyr239 from the *B* subunit makes a strong hydrogen bond with the phenolic hydroxyl of feruloyl-CoA (Fig. 3). The structural determinants which dictate that feruolyl-CoA, but not cinnamoyl-CoA, which lacks the phenolic group, is a substrate for HCHL (Mitra *et al.*, 1999) are not known, but the model suggests that Tyr239 would be an excellent candidate for one of these determinants. The aromatic ring of the substrate also makes a close contact with the side chain of Met70 and

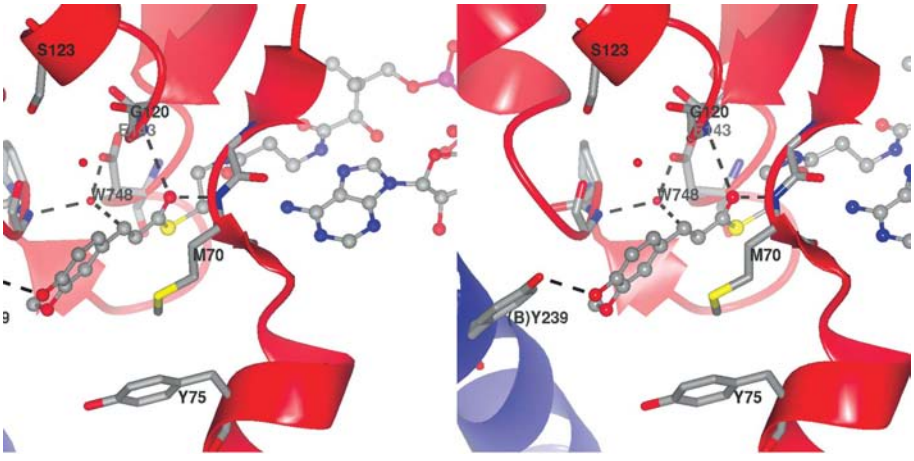


Figure 3
Active site of HCHL, into which the natural substrate feruloyl-CoA has been modelled. The active site is at the interface between monomer *A* (red) and monomer *B* (blue). In the foreground, interactions between the side chains of *A*Met70, *A*Tyr75 and *B*Tyr239 and the aromatic ring of feruloyl-CoA can be seen. *B*Tyr239 is shown making a hydrogen bond (dashed line) to the phenolic hydroxyl of the substrate. The carbonyl group of the thioester of coenzyme A is bound in the oxyanion hole formed by the peptidic NH groups of Met70 and Gly120 (indicated by dashed lines). A structural water, W748, is shown coordinated to both Glu143 and the NH group of Gly151 (indicated by dashed lines) and was at a distance of 3.3 Å from the benzylic C atom in the model.

a stacking interaction with *A*Tyr75 at a distance of approximately 6 Å. The thioester carbonyl, which is spatially equivalent to the negatively charged oxygen of a putative enolate intermediate in the reaction coordinate, is within hydrogen-bonding distance of the peptidic NH groups of Met70 and Gly120, sequence homologues of Ala98 and Gly141 in ECH, and illustrates the structural conservation of mechanism in the crotonase superfamily (Fig. 3). The NH of the mercaptoethylamine/pantothenate amide bond is hydrogen bonded to the backbone peptide carbonyl of Ala68 and the adjacent carbonyl is hydrogen bonded to the side chains of both Trp146 and Ser142. The geminal dimethyl group of coenzyme A is orientated towards the plane of the aromatic ring of *A*Phe118 at a distance of approximately 4 Å. The α-phosphate of coenzyme A interacts with the side chain of *A*Arg30. The adenine NH_2 moiety is bonded to the peptide carbonyls of both *A*Ala68 and *A*Met70. Further interactions of the coenzyme A molecule with the most C-terminal region of HCHL are not observed in the model because of the C-terminal truncation discussed above.

In the model, the benzylic C atom of feruloyl-CoA, which would undergo attack by an activated water molecule in the first step of the hydration reaction, is 3.3 Å distant from a structural water molecule, W748, which is in turn hydrogen bonded to *A*Glu143 (the homologue of the catalytic acid Glu164 in ECH) and to the peptidic NH group of *A*Gly151 (Fig. 3). In the active site of HCHL modelled with feruloyl-CoA, we observed that the putative catalytic water W748 that is hydrogen bonded with Glu143 is also bound not by Ser123, as might be expected, given that this is the homologous residue of Glu144 in ECH, but rather by the backbone NH of Gly151. Ser123 is in fact 7.7 Å distant from W748 in the HCHL structure. In the concerted E1cb acid–base mechanism of ECH-catalysed hydration proposed by Bahnson *et al.* (2002), the catalytic water is coordinated to two carboxylate residues and the peptidic NH of Gly172 (the sequence homologue of HCHL Gly151) and all three atoms of the water molecule are incorporated into the hydrated product. In the absence of a second suitably positioned carboxylate residue in HCHL, it is difficult at this stage to suggest an equivalent concerted mechanism in HCHL. Bahnson and coworkers eliminated the possibility of proton donation by any species other than the water as there were no candidate-residue side chains and the hydrogen-bonding pattern in the substrate-binding site suggested that both glutamates are ionized in the presence of substrate. The origin of the proton required to complete the hydration step in HCHL has not yet been ascertained, although Ser123 is in the active-site region and is hydrogen bonded to W505, a water molecule that is approximately 4 Å from the benzylic C atom of feruloyl-CoA in the model. However, it is possible that the shifts in protein tertiary structure that accompany actual substrate binding may bring Ser123 into closer contact with the substrate and associated water molecules.

If the hydration step catalysed by HCHL is analogous to that observed in ECH, the reaction mechanism would require both general base-catalysed activation of a water molecule for attack at the benzylic carbon of the substrate and an oxyanion hole for stabilization of the kinetically unstable enolate intermediate in hydration. It is difficult at this stage to speculate on the mechanism of the retro-aldol half-reaction, yet the model suggests that the only reactive residue side chains that are within a suitable distance, *via* a structural water, of the relevant benzylic C atom of feruloyl-CoA are Glu143 and Ser123 and the latter residue appears to be somewhat distant. Therefore, the actual structural bases for the mechanisms of each half-reaction catalysed by HCHL await elucidation by appropriate mutational analyses.

The structure of HCHL and the model of the enzyme bound to the substrate HCHL provide the first basis for rational engineering of the enzyme to investigate both mechanism and substrate specificity, both of which will aid in the application of the enzyme in improved industrial processes for the production of natural substances for human consumption.

We thank the Biotechnology and Biosciences Research Council for funding (Project No. 87B/B17227) and the National Institutes of Health for the award of grant NIH R0234601 to AL.

References

Bahnson, B. J., Anderson, V. E. & Petsko, G. A. (2002). *Biochemistry*, **41**, 2621–2629.
Benning, M. M., Haller, T., Gerlt, J. A. & Holden, H. M. (2000). *Biochemistry*, **39**, 4630–4639.
Benning, M. M., Taylor, K. L., Liu, R.-Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D. & Holden, H. M. (1996). *Biochemistry*, **35**, 8103–8109.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.
Brünger, A. T. & Karplus, M. (1988). *Proteins*, **4**, 148–156.
Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Eberhard, E. D. & Gerlt, J. A. (2004). *J. Am. Chem. Soc.* **126**, 7188–7189.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engel, C. K., Mathieu, M., Zeelen, J. P., Hiltunen, J. K. & Wierenga, R. K. (1996). *EMBO J.* **15**, 5135–5145.
Gasson, M. J., Kitamura, Y., McLauchlan, W. R., Narbad, A., Parr, A. J., Parsons, E. H. L., Payne, J., Rhodes, M. J. C. & Walton, N. J. (1998). *J. Biol. Chem.* **273**, 4163–4170.
Gerlt, J. A. & Babbitt, P. C. (2001). *Annu. Rev. Biochem.* **70**, 209–246.
Grogan, G., Roberts, G. A., Bougioukou, D., Turner, N. J. & Flitsch, S. L. (2001). *J. Biol. Chem.* **276**, 12565–12572.
Holden, H. M., Benning, M. M., Haller, T. & Gerlt, J. A. (2001). *Acc. Chem. Res.* **34**, 145–157.
Hubbard, P. A., Yu, W., Sculz, H. & Kim, J.-J. P. (2005). *Protein. Sci.* **14**, 1545–1555.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kurimoto, K., Fukai, S., Nureki, O., Muto, Y. & Yokoyama, S. (2001). *Structure*, **9**, 1253–1263.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leonard, P. M. & Grogan, G. (2004). *J. Biol. Chem.* **279**, 31312–31317.
Leonard, P. M., Marshall, C. M., Dodson, E. J., Walton, N. J. & Grogan, G. (2004). *Acta Cryst.* D**60**, 2343–2345.

Lomascolo, A., Stentelaire, C., Asther, M. & Lesage-Meessen, L. (1999). *Trends. Biotechnol.* **17**, 282–289.
MacKerell, A. D. *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.
Mitra, A., Kitamura, Y., Gasson, M. J., Narbad, A., Parr, A. J., Payne, J., Rhodes, M. J. C., Sewter, C. & Walton, N. J. (1999). *Arch. Biochem. Biophys.* **365**, 10–16.
Modis, Y., Filppula, S. A., Novikov, D. K., Norledge, B., Hiltunen, J. K. & Wierenga, R. K. (1998). *Structure*, **6**, 957–970.
Muheim, A. & Lerch, K. (1999). *Appl. Microbiol. Biotechnol.* **51**, 456–461.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Mursula, A. M., van Aalten, D. M., Hiltunen, J. K. & Wierenga, R. K. (2001). *J. Mol. Biol.* **309**, 845–853.
Narbad, A. & Gasson, M. J. (1998). *Microbiology*, **144**, 1397–1405.
Pelletier, D. A. & Harwood, C. S. (1998). *J. Bacteriol.* **180**, 2330–2336.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Priefert, H., Rabenhorst, J. & Steinbüchel, A. (1997). *J. Bacteriol.* **179**, 2595–2607.
Rabenhorst, J. (1996). *Appl. Microbiol. Biotechnol.* **46**, 470–474.
Rabenhorst, J. & Hopp, R. (1997). European Patent Application EP 0 761 817 A2.
Truglio, J. J., Theis, K., Feng, Y., Gajda, R., Machutta, C., Tonge, P. J. & Kisker, C. (2003). *J. Biol. Chem.* **278**, 42352–42360.
Vagin, A & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Walton, N. J., Mayer, M. J. & Narbad, A. (2003). *Phytochemistry*, **63**, 505–515.
Whittingham, J. L., Turkenburg, J. P., Verma, C. S., Walsh, M. A. & Grogan, G. (2003). *J. Biol. Chem.* **278**, 1744–1750.