Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a feruloyl esterase from *Aspergillus niger*

Katherine E. McAuley,[a] Allan Svendsen,[b] Shamkant A. Patkar[b] and Keith S. Wilson[c]*

[a]CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England, [b]Department of Protein Chemistry, Novozymes A/S, Novo Alle, 6B-S107 2880 Bagsvaerd, Denmark, and [c]York Structural Biology Laboratory, University of York, York, England

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

Received 4 February 2003
Accepted 3 March 2004

PDB References: feruloyl esterase, 1uza, r1uzasf; feruloyl esterase–ferulic acid complex, 1uwc, r1uwcsf.

The crystallographic structure of feruloyl esterase from *Aspergillus niger* has been determined to a resolution of 1.5 Å by molecular replacement. The protein has an α/β-hydrolase structure with a Ser-His-Asp catalytic triad; the overall fold of the protein is very similar to that of the fungal lipases. The structure of the enzyme–product complex was determined to a resolution of 1.08 Å and reveals dual conformations for the serine and histidine residues at the active site.

1. Introduction

Feruloyl esterases (FAEs) are involved in the degradation of plant cell walls and are found in both bacterial and fungal species (Donaghy & McKay, 1997; de Vries *et al.*, 1997; Blum *et al.*, 2000; Kroon *et al.*, 2000). Plant cell walls contain the polymers lignin, cellulose, hemicellulose and pectin. The most abundant hemicellulose is arabinoxylan, which consists of a polymeric backbone of β-(1,4)-linked xylan units with attached arabinose, acetate or methylglucuronate side chains (Fig. 1). Chains of arabinoxylan are strengthened by cross-linking ferulic acid dimers, which are ester-linked to the arabinose sugars. Ferulic acid is also involved in the linking of hemicellulose to lignin (Ralph *et al.*, 1995).

For a bacterium or a fungus to break down the cell wall, it requires not only a range of hemicellulases to degrade the polysaccharides but also feruloyl esterases to increase the accessibility of the cell-wall hydrolases to their substrates by releasing arabinoxylan from lignin and breaking the cross-links between arabinoxylan chains. Feruloyl esterase breaks the cross-links by hydrolysing the ester bond between ferulic acid and arabinose. The product of the hydrolysis, ferulic acid, can be used in the production of compounds of commercial interest, *e.g. p*-coumaric acid, which is commonly used in sun creams, or vanillin, a flavouring agent widely used in the food industry.

One of the most thoroughly studied FAEs is feruloyl esterase A (FAEA or FAE-III) from *Aspergillus niger*. It has high substrate specificity for esters in which the feruloyl group is attached to the C-5 position of arabinose; it is inactive towards C-2-linked esters (Ralet *et al.*, 1994). Other factors that affect substrate specificity are changes to the substituents on the aromatic ring and changes to the type and number of attached sugar residues (Kroon *et al.*, 1997). FAE-III is also able to catalyze the release of diferulic acids from plant cell walls (Garcia-Conesa *et al.*, 1999).

Sequence analysis of FAE-III indicates that the enzyme is an α/β-hydrolase with a serine, histidine, aspartic acid catalytic triad. The protein with the highest sequence similarity, not including the very similar FAEs with >93% identity from

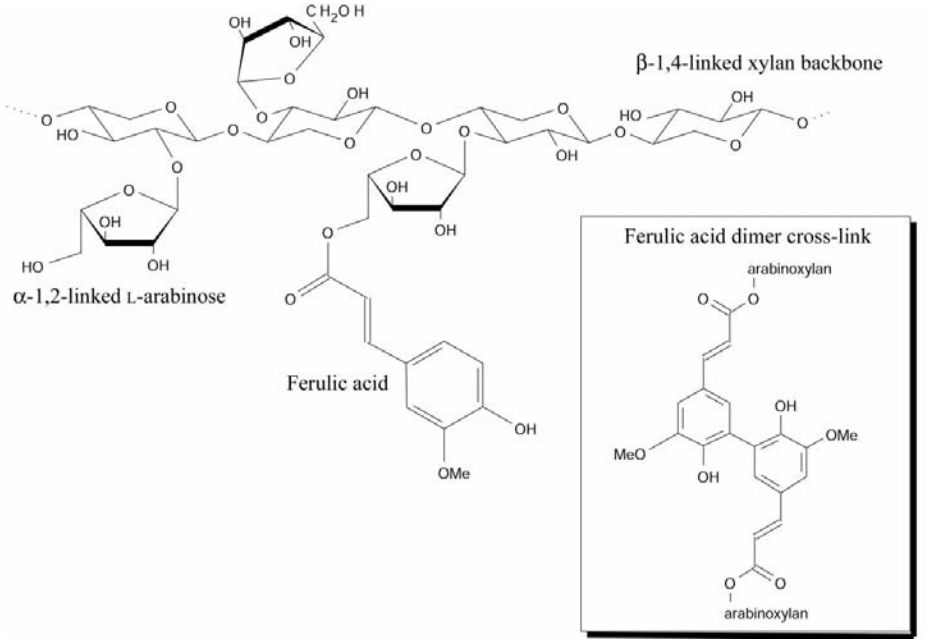


Figure 1
Structure of arabinoxylan. The inset shows one type of ferulic acid dimer that can be formed.

Table 1
Data statistics.

Values in parentheses refer to the highest resolution shell.

	Native	Native	Complex
Data-collection source	In-house Cu $K\alpha$	ESRF ID14-2	SRS 9.6
Wavelength (Å)	1.54	0.933	0.87
Temperature (K)	120	100	110
Space group	*P*1	*P*1	*P*1
Unit-cell parameters			
a (Å)	38.3	38.3	38.8
b (Å)	39.7	39.7	40.2
c (Å)	77.1	77.0	77.9
α (°)	104.9	75.2	75.0
β (°)	91.4	78.8	78.6
γ (°)	109.2	71.3	71.1
V_M (Å^3 Da^{-1})	1.8	1.8	1.8
Solvent content (%)	32	32	32
Resolution (Å)	30–1.86	25–1.5	25–1.08
R_{merge} (%)	8.3 (18.6)	6.4 (17.2)	5.8 (26.0)
Mean $I/\sigma(I)$	18.2 (6.1)	11.5 (2.5)	15.8 (2.5)
$I > 3\sigma(I)$ (%)	85	74	69
Unique reflections	31690	64076	170583
Redundancy	3.6 (2.1)	2.1 (1.8)	2.9 (1.8)
Completeness	93.5 (93.5)	94.4 (95.3)	93.5 (81.9)

A. awamori and *A. tubingensis*, is the lipase from *Rhizomucor miehei* (RmL; Brady *et al.*, 1990), with an overall sequence identity of 32%. FAE-III contains the characteristic lipase serine active-site motif and, given the sequence similarities, it was predicted that FAE-III would have considerable structural homology and a similar catalytic mechanism to the fungal lipases. However, FAE-III has been shown not to possess lipase activity (Aliwan *et al.*, 1999).

Recently, the structures of two bacterial feruloyl esterases have been published: FAE_XynZ from *Clostridium thermocellum* (Schubot *et al.*, 2001) and FAE_Xyn10B (Prates *et al.*, 2001) also from *C. thermocellum*. Although these enzymes have similar functionality to FAE-III, there is no apparent sequence homology. Therefore, in order to examine the structure of the enzyme and its substrate specificity, FAE-III has been crystallized and the crystallographic structures solved for the native enzyme and for the enzyme in complex with ferulic acid.

2. Experiment

2.1. Protein preparation

FAE-III was expressed in *A. oryzae* and secreted in the extracellular media. The fermentation supernatant was purified by anion-exchange chromatography using fast-flow Q-Sepharose (Pharmacia-Amersham) as described in Andersen *et al.* (2002).

2.2. Crystallization

FAE-III was crystallized by vapour diffusion from hanging drops. 1 µl of 15 mg ml^{-1} protein solution was mixed with 1 µl precipitant solution containing 1.0 *M* ammonium sulfate, 0.1 *M* sodium acetate pH 4.5 and the drop was equilibrated against 500 µl precipitant solution. Crystals grew after one to two weeks. The crystals were cryoprotected by brief soaking in a solution of mother liquor supplemented with 35% glucose. The space group is *P*1 and the unit-cell parameters for the native enzyme are a = 38.3, b = 39.7, c = 77.1 Å, α = 75.2, β = 78.8, γ = 71.3°.

The complex with ferulic acid was obtained by soaking a crystal in 1.5 *M* ammonium sulfate, 0.1 *M* sodium acetate pH 4.5, 10 m*M* ferulic acid and 5% ethanol for approximately 2 h.

2.3. Data collection

The first native data set was collected in the home laboratory using a rotating-anode source and a MAR 345 image-plate detector. The resolution of the data extended to 1.86 Å and the data were 93.5% complete, with an overall R_{merge} of 8.3%. The data were processed with *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994).

A second, higher resolution native data set was collected at the European Synchrotron Radiation Facility (ESRF) and a data set for the complex was collected at the Synchrotron Radiation Source (Darebury, UK). Both data sets were processed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). The resolution of the native data set was 1.5 Å, with a completeness of 94.4% and an overall R_{merge} of 6.4%. The resolution of the data set of the complex was 1.08 Å, with a completeness of 94% and an overall R_{merge} of 5.8%. The statistics of all the data collections are given in Table 1.

Table 2
ARP/wARP results.

Resolution (Å)	No. residues built (%)	*R* value	Free *R* value†
1.9	281 (54%)	19.5	50.3 ↑
1.8	275 (53%)	20.5	49.2 ↑
1.7	331 (64%)	20.4	44.5 ↓
1.6	430 (83%)	19.6	34.8 ↓
1.5	487 (94%)	18.4	27.0 ↓

† The arrow indicates an increase or decrease in the free *R* value during the *ARP/wARP* procedure.

2.4. Molecular replacement

The lipase from *R. miehei* was chosen as the search model for the molecular-replacement trial. The Matthews coefficient (Matthews, 1968) was consistent with one or two molecules in the asymmetric unit, giving solvent contents of 66 or 32%, respectively. After performing a rotation search in *AMoRe* (Navaza, 1994) two solutions were clearly better than the others, with correlation coefficients (CC_F) of 10.7 and 9.1% for data in the resolution range 10–3.5 Å. The next highest solution had a correlation coefficient of 5.8%. Following a translation search to locate the second molecule in relation to the first and a rigid-body refinement in *AMoRe*, the correlation coefficient increased to 31% and the *R* value was 49%. There were a significant number of close contacts between C^{α} atoms of symmetry-related molecules; however, on inspection it was observed that the clashes only affected surface loops.

2.5. Refinement

The refinement strategy is discussed in more detail below but, in summary, the packages used were *REFMAC*5 or *REFMAC*5.1 (Murshudov *et al.*, 1997, 1999), *ARP/wARP* 5.1 (Perrakis *et al.*, 1999) and *DM* (Cowtan, 1994). Model rebuilding was carried out in *XtalView* (McRee, 1999).

2.5.1. Refinement of in-house data. A rigid-body refinement in *REFMAC* of the molecular-replacement model with data in the resolution range 20–1.86 Å led to an *R* value and a free *R* value of 50.5 and 49.3%, respectively. The residues involved in steric clashes were removed at this stage and refinement was continued using the NCS-phased refinement task of the ccp4i GUI (Collaborative Computational Project, Number 4, 1994). This task combines *REFMAC* refinement with solvent flattening, histogram matching and NCS averaging in *DM*. The electron-density maps produced from this procedure were used to continue rebuilding the model, but the maps were difficult to interpret. After several rounds of refinement and model rebuilding, the free *R* value had decreased to 46%, but it failed to fall any further.

A program that has proven useful in the refinement of poor molecular-replacement solutions is *ARP/wARP* (Perrakis, 2001). The autobuilding option has proved to be generally useful when the resolution is better than 2.0 Å. However, in this case, the poor quality of the phases from the molecular-replacement model resulted in very few residues being built by *ARP/wARP* and in fact the free *R* value increased and the maps deteriorated. The autobuilding was tried both before and after the refinement steps described above.

The failure of the 1.86 Å data to refine further led to the decision to collect higher resolution data at a synchrotron.

2.5.2. Refinement of synchrotron data. The increased information from the higher resolution data had an immediate impact on the refinement. The rebuilt model (obtained from refinement of the in-house data) was positioned in the new unit cell using *AMoRe* and the warpNtrace mode of *ARP/wARP* was then used to automatically build a new model. The first building cycle of warpNtrace found 243 residues out of the expected 520 and resulted in a free *R* value of 44%. During the iterative refinement, model updating and building cycles the free *R* decreased steadily until it reached 26% and 488 residues had been built. The electron-density maps were now of excellent quality and the remaining residues were easily built into the density.

However, the process of obtaining the model that was successfully used in *ARP/wARP* involved many days of work, so some alternative protocols were investigated. To start with, the molecular-replacement solution was input into *ARP/wARP* without prior refinement. This failed to work. Secondly, the molecular-replacement solution was converted to a poly-alanine model and put through two cycles of *REFMAC* refinement combined with *ARP*. The parameters in *ARP* were set so that up to 200 atoms below 1σ density could be removed in each cycle and 20 new atoms could be added. The aim of this was to remove all the atoms involved in clashes with neighbouring molecules and in addition to remove regions of the lipase that are different in the feruloyl esterase. 231 atoms were actually removed from the model. The new 'trimmed' model was input into warpNtrace and the automated building was able to continue successfully from this point.

The effects of increasing resolution were further investigated by running multiple *ARP/wARP* procedures with the same parameters but with different high-resolution limits between 1.9 and 1.5 Å. After 100 refinement cycles and 33 building cycles, the number of residues that had been built by *ARP/wARP* varied depending on the resolution (see Table 2). These results demonstrate that in this particular case and with the default parameters as set up some months ago, a resolution better than 1.6 Å is required for automated building of the majority of the protein model. We plan to use the FAE-III data as a test case for improving the refinement of poor molecular-replacement solutions.

2.5.3. Phasing and refinement of FAE–ferulic acid complex. The program *ACORN* (Foadi *et al.*, 2000) was used to obtain phases for the complex data in order to avoid model bias that may have arisen from phasing by molecular replacement. A 50-residue fragment (residues A17–A67) of the native structure was used as a starting point for *ACORN*. The initial phases calculated from the fragment were refined by Patterson superposition, two cycles of Sayre-equation refinement and 47 cycles of dynamic density modification. The correlation coefficient for the small *E* values increased from 0.0360 to 0.269. The phase-refinement process was completed within 3 min (Pentium III, 1 GHz PC).

Table 3
Refinement statistics.

	Native	Complex
Resolution range (Å)	30–1.5	25–1.08
R value (%)	15.6	11.5
Free *R* value (%)	18.8	13.8
R.m.s. deviation, bond distance (Å)	0.019	0.021
R.m.s. deviation, bond angle (°)	1.8	1.9
Total non-H atoms	4476	4922
Water molecules	399	713
Sulfate ions	2	2
Ferulic acid molecules	0	3
Glycosylation sites	NAG, Asn*A*79 and Asn*B*79	NAG, Asn*A*79 and Asn*B*79
Mean *B* factors (Å^2)		
All atoms	15.9	11.8
Main chain (*A* chain, *B* chain)	12.8, 14.3	8.5, 9.0
Side chain (*A* chain, *B* chain)	15.4, 16.6	10.6, 11.1
Water molecules	26.9	23.8
Sulfate ions	35.6	17.5
Ferulic acid molecules	—	11.3
NAG units	10.9, 12.2	7.3, 7.4
Ramachandran statistics† (*A* chain, *B* chain)		
Most favoured regionss	89.8, 90.2	91.6, 90.3
Additional allowed regions	8.9, 8.4	7.0, 8.4
Generously allowed regions	1.3, 1.3	0.9, 1.3
Disallowed regions	0, 0	0.4, 0
PDB code	1uza	1uwc
PDB structure-factor code	r1uzasf	r1uwcsf

† The different regions of the Ramachandran plot are described in Morris *et al.* (1992).

ARP/*wARP*5.1 was used to build the protein model using the phases from *ACORN*. After 12 cycles of refinement and four building cycles, 501 residues were built. At this stage, the *R* value was 25% and the free *R* value was 28%. The remaining protein residues and two ferulic acid molecules were added manually. The *R* value and free *R* value after refinement in *REFMAC*5.1 and addition of water molecules using *ARP*/*wARP* were 13.7 and 15.9%, respectively. Another ferulic acid molecule and two sulfate ions were then located and added to the model. The final *R* value and free *R* value following anisotropic temperature-factor refinement using all the data to 1.08 Å were 11.5 and 13.8%, respectively.

We retrospectively tested the ability of *ACORN* to solve the structure of the complex using the RmL model positioned in the unit cell by *AMoRe*. As stated above, this model was challenging to refine with *REFMAC* and *ARP*/*wARP* in a straightforward manner. With the 93.5% complete 1.08 Å spacing data, *ACORN* provided a set of phases after 7 min computational time. The resulting map revealed essentially all of the atoms and could easily be rebuilt using *ARP*/*wARP* with little user intervention. In retrospect, this would have been the optimum solution to the phase problem for FAE-III and demonstrates further the power of *ACORN* to give an unbiased set of phases for atomic resolution data from a relatively poor molecular-replacement solution. The importance of data completeness is shown by the failure to produce a satisfactory development of phases with *ACORN* starting from a higher resolution 1.05 Å data set which was 85% complete (these data are not described in detail as they were not used in the final refinement).

3. Results and discussion

3.1. Native structure and active site

The final model consists of two molecules of FAE-III, two sulfate ions and 399 water molecules. The protein is glycosylated at Asp79 and a single *N*-acetyl glucosamine (NAG) residue was built at each glycosylation site. The electron-density maps in this region are clear and well ordered and there is no indication in the maps of additional carbohydrate residues attached to the NAG residue. The N-terminal alanine residue is missing from the electron-density maps for both molecules in the asymmetric unit, but all other residues are present. The crystallographic *R* value for the model is 15.6% (free *R* value 18.8%) for all data in the 25–1.5 Å resolution range. Table 3 gives details of the refinement statistics and model quality.

FAE-III has an α/β-hydrolase fold (Ollis *et al.*, 1992) consisting of a nine-stranded β-sheet core surrounded by five α-helices and two additional β-strands. The active-site residues of FAE-III are Ser133, Asp194 and His247. His247 occupies a position where atom N^{ε} forms a hydrogen bond to the O^{γ} atom of Ser133 (distance of 2.8 Å) and atom N^{δ} hydrogen bonds to Asp194 (distances of 2.8 Å to $O^{\delta1}$ and 3.0 Å to $O^{\delta2}$). This arrangement of residues is characteristic of enzymes that cleave amide or ester bonds by nucleophilic attack and is known as a catalytic triad (shown in Fig. 4*a*). The histidine acts as a general base by deprotonating the serine, which allows nucleophilic attack of the carbonyl C atom of the substrate by the serine to occur, yielding a tetrahedral transition state. The resulting positive charge on the histidine is stabilized by an electrostatic interaction with the negatively charged aspartic acid residue. The tetrahedral transition state collapses to give an acyl-enzyme intermediate. The next step is attack on the acyl-enzyme by a nucleophilic water molecule, again activated by the histidine residue, leading to release of the acid product *via* a second tetrahedral intermediate transition state. The negative charge on the carbonyl O atom of the tetrahedral intermediates is stabilized by the oxyanion hole formed by backbone N atoms of, in this case, residues Thr68 and Leu134. The serine-protease mechanism has recently been reviewed by Hedstrom (2002).

Both molecules of FAE-III in the native crystal structure have well ordered active-site residues with single conformations corresponding to an active catalytic triad with the expected pattern of hydrogen bonds between the residues.

3.2. Comparison with lipases and bacterial FAEs

The tertiary structure of FAE-III is most similar to that of the fungal lipases, as expected from the sequence homologies. The root-mean-square (r.m.s.) deviations between FAE-III and RmL and the *Thermomyces lanuginosa* lipase (TlL) are 1.0 Å over 217 aligned C^{α} atoms and 1.3 Å over 205 aligned C^{α} atoms, respectively.

Lipases can exist in two conformational states: an active form where the catalytic residues are exposed and an inactive form where a helical lid involving residues 82–96 (TlL numbering; Brzozowski *et al.*, 2000) covers the active site. Fig. 2

shows a comparison of FAE-III with an open-form and a closed-form lipase. The helix in FAE-III (residues 71–77), corresponding to the lid of TlL, does not shield the active-site residues and corresponds roughly to the open form of the lipase. Since FAEs do not demonstrate lipase activity (Aliwan *et al.*, 1999), it is unlikely that this helix needs to adopt two conformations.

Although the overall structures are similar, the active-site regions have to be different in order to accommodate the differences in substrates. This can be seen in the surface representations shown in Fig. 2: FAE-III has a greater number of polar residues exposed in the region around the active site compared with the lipase structures, which have more hydrophobic surfaces.

FAE-III and the FAEs from *C. thermocellum* have no significant sequence homology. Sequence alignment using *ClustalX* (Thompson *et al.*, 1997) fails to align the catalytic residues and sequence-database searches with the FAE-III sequence do not identify FAE_XynZ or FAE_Xyn10B as homologues. The structures of FAE-III and the *C. thermocellum* FAEs were superimposed using the program *CE* (Shindyalov & Bourne, 1998) and the resulting structure-based sequence alignment is shown in Fig. 3(*a*). In this alignment, the catalytic residues superimpose and there is overlap of several of the secondary-structure elements. The r.m.s. deviation between FAE-III and FAE_XynZ is 4.0 Å for 146 aligned C^{α} atoms and for FAE_Xyn10B the r.m.s. deviation is 4.4 Å for 153 aligned C^{α} atoms. These deviations are higher than those found for the fungal lipases described above, confirming that FAE-III is indeed substantially more similar to the fungal lipases than to the bacterial FAEs.

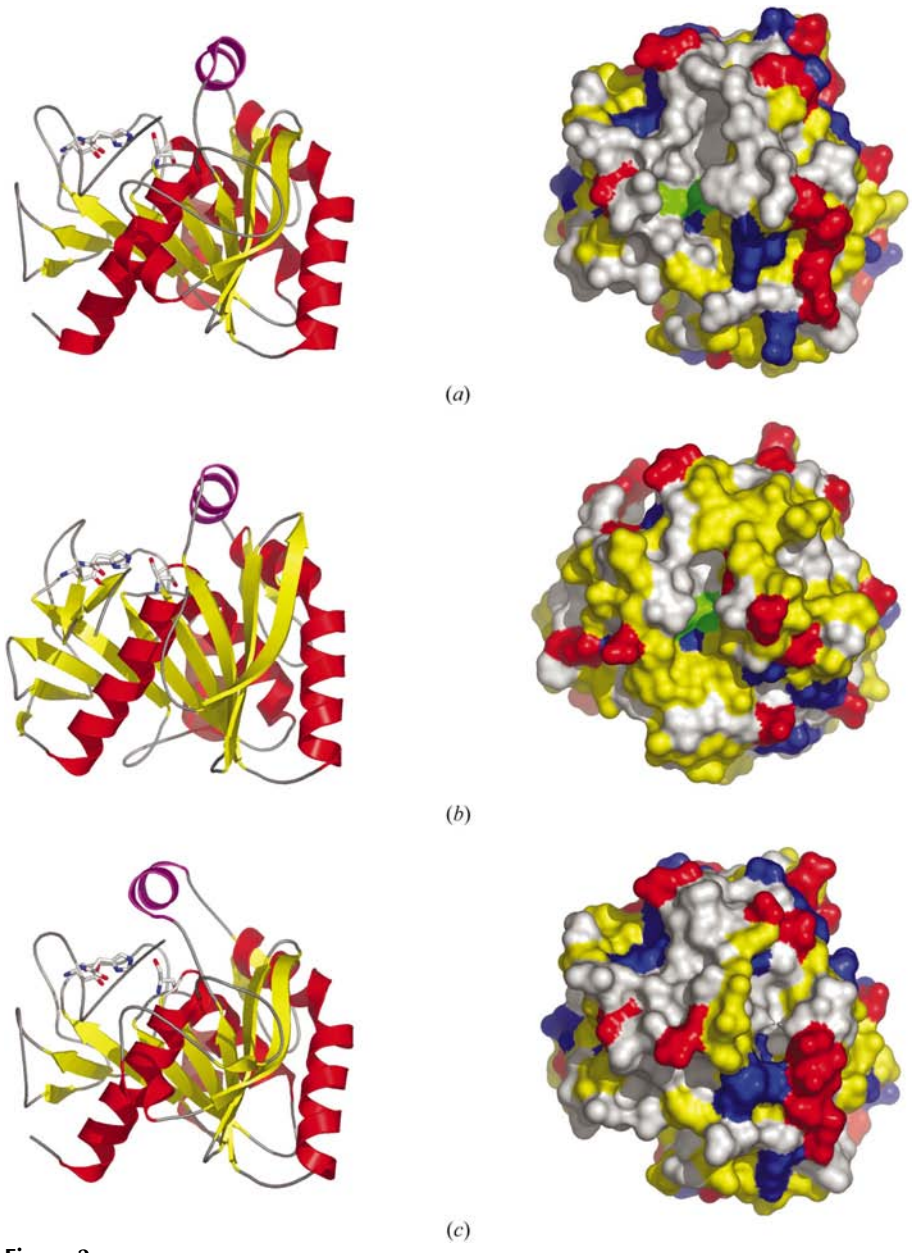


Figure 2
Diagrams of (*a*) open-form TlL, (*b*) FAE-III and (*c*) closed-form TlL. The ribbon diagrams on the left are coloured yellow for β-sheet, red for α-helices and violet for the lid helix. On the right are the solvent-accessible surface representations. The surface is coloured by residue type: blue, basic; red, acidic; yellow, polar; grey, hydrophobic; green, catalytic triad. Figs. 2, 4, 5(*b*) and 6 were produced using *PyMOL* (DeLano, 2002).

FAE-III, the *Clostridium* FAEs and the lipases are all members of the α/β-hydrolase family and can be assumed to have originated from a common ancestor. All the FAEs share the serine motif common to serine hydrolases, *i.e.* G*X*S*X*G, and they all have a catalytic triad and central β-sheet core surrounded by α-helices. However, the sequences and topologies (Fig. 3*b*) suggest that the FAE-III and lipases diverged at a later stage than the divergence of the bacterial FAEs from the common ancestor, *i.e.* the ferulic acid esterase activity evolved twice: once for the bacterial FAEs and once for the fungal FAEs.

3.3. Changes in the conformations of residues in the catalytic triad

The structure of FAE-III in complex with its product, ferulic acid, was determined at atomic resolution and the refinement statistics are given in Table 3. Overall, there is little difference between the complexed and the free protein structures. The r.m.s. deviations for the C^{α} atoms were calculated to be 0.24 Å for molecule *A* and 0.32 Å for molecule *B*. However, there are differences in the conformations of the residues at the active site.

3.3.1. Ser133. The catalytic serine residue Ser133 is in a single active conformation in both molecules of the native structure (Fig. 4*a*). However, we observe two conformations of the serine residue when ferulic acid is bound at the active site (Figs. 4*b* and 4*c*). One conformation is the same as in the native structure, with a χ_1 torsion angle of approximately -155°; this corresponds to the active conformation. In this conformation, the O^γ atom forms a hydrogen bond to the carboxylate group of ferulic acid. There is an additional hydrogen bond in molecule *B*, also seen in the uncomplexed enzyme, between Ser133 and the N^ε atom of one of the His247 conformations (shown in Fig. 4*c*). The χ_1 angles of the second conformation of the serine residue are -49° for molecule *A* and -54° for molecule *B*. In this conformation, the O^γ atom of the serine makes an hydrogen bond to a water molecule. Alternate conformations of the catalytic serine have been observed in previous high-resolution esterase structures, *e.g.* the *Fusarium solani* cutinase (Longhi *et al.*, 1997), the *Penicillium purpurogenum* acetylxylan esterase (Ghosh *et al.*, 2001) and the *Bacillus subtilis* lipase (Kawasaki *et al.*, 2002).

3.3.2. His247. His247 in the native enzyme is present in a single conformation corresponding to the active orientation for a catalytic triad histidine residue, *i.e.* within hydrogen-bonding distance of both the serine and the aspartic acid residues. However, this is not the case in the enzyme–product complex, where we observe differences between the active sites of molecule *A* and molecule *B*. In molecule *A*, His247 has moved away from Asp194 and the hydrogen bond between His N^δ and the carboxylate group of the aspartate is broken (inactive form). Instead, the histidine forms hydrogen bonds to two water molecules as shown in Fig. 4(*b*). This conformation of the histidine of a catalytic triad is unusual, but in some serine-protease inhibitor complexes the histidine has rotated out of the active site in a similar manner (*e.g.* James *et al.*, 1980; Bone *et al.*, 1991; Koszelak *et al.*, 1997). His247 of molecule *B* is present in two conformations (Fig. 4*c*). One

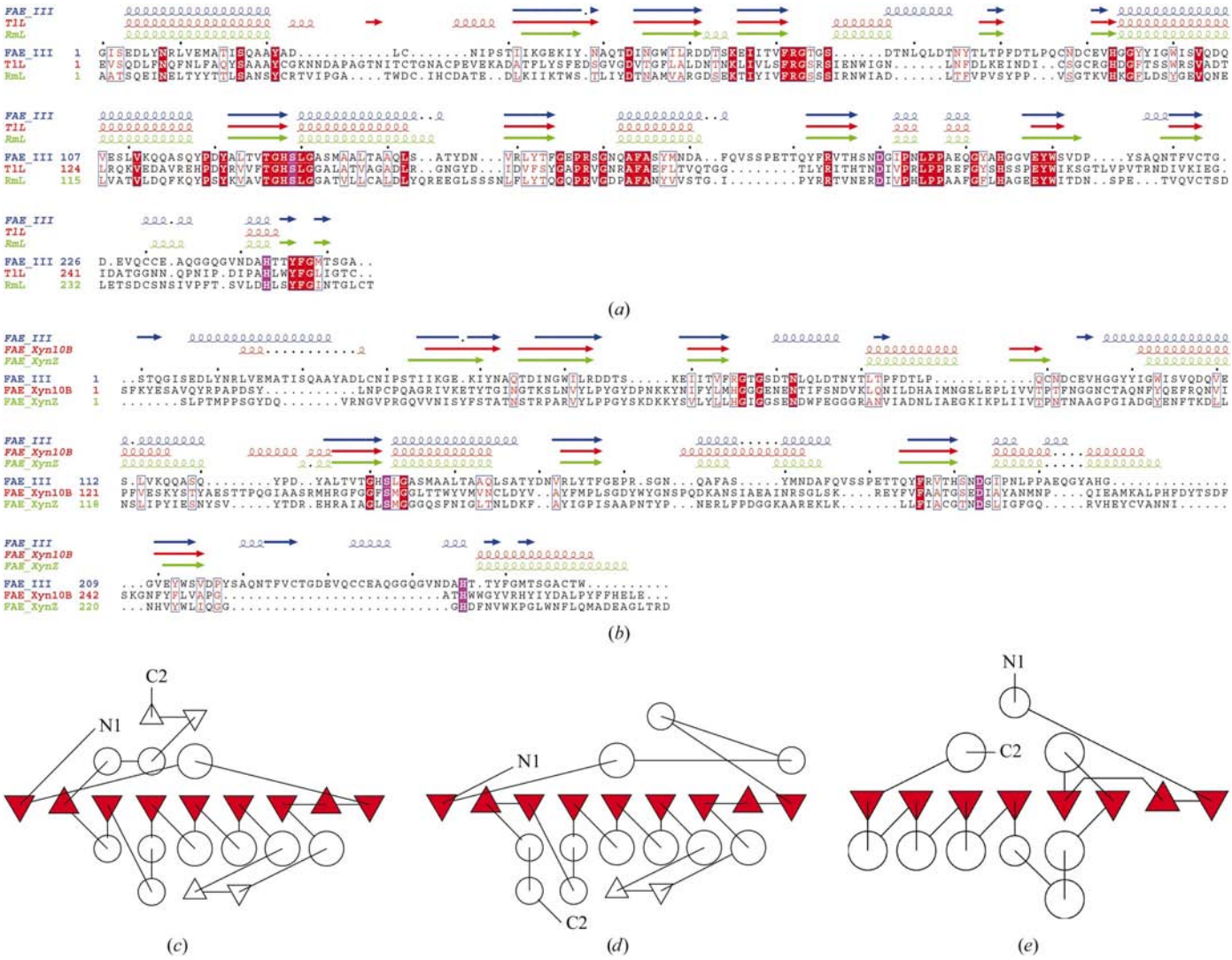


Figure 3
Structure-based sequence alignments of (*a*) FAE-III with TlL and RmL lipases and (*b*) FAE-III with FAE_Xyn10B and FAE_XynZ. Secondary-structure elements are shown above the sequence alignments. The catalytic residues are highlighted in pink (*ESPript*; Gouet *et al.*, 1999). Topology diagrams of (*c*) FAE-III, (*d*) RmL and (*e*) FAE_Xyn10B created by *TOPS* (Westhead *et al.*, 1999). The circles represent helices and the triangles represent β-strands. The strands forming the central β-sheet are coloured red.

conformation is as described for molecule *A* and the other is in the position observed in the native structure, *i.e.* the catalytically active form.

Why do we observe two conformations for His247 in molecule *B* and only one conformation in molecule *A*? This is not because of differences in the occupancies of the ferulic acid molecules, as *B*-value comparisons between the ferulic acid molecules and surrounding protein residues indicate a similar occupancy for both molecules. The average *B* factors for protein residues within 10 Å of the binding site are 8.3 Å^3 for molecule *A* and 7.9 Å^3 for molecule *B* and the average *B* factor for ferulic acid is 10.5 Å^3 for both the *A* and *B* molecules. This implies that the difference in free energy between the two histidine conformations is close to zero and so the two forms can easily interconvert.

3.3.3. Asp194. Asp194 remains unchanged when ferulic acid is bound to FAE-III. Asp194 has a single conformation that is the same in both molecules of the complex and in both molecules of the uncomplexed native structure.

3.4. A broken catalytic triad

The dual conformations of the serine and the histidine residues were not observed in the native structure despite having sufficiently high-resolution data to allow easy identification of alternate conformations. The only difference between the native and the complexed crystals is that the latter have been soaked in a stabilizing solution containing ferulic acid. The lack of interaction between the histidine and the serine and aspartic acid residues in molecule *A* of the FAE-III–product complex means that the enzyme cannot be active and so we must consider the cause of this inactivation. A shift in the position of the histidine side chain was also seen in the structure of FAE_XynZ–FAXX when crystallized at pH 8.5 (Schubot *et al.*, 2001). Once the FAE_XynZ–FAX$_3$ complex was crystallized at pH 8.0, the histidine returned to the position it adopts in the free-enzyme structure. The authors concluded that the change in orientation of the histidine arose from a pH effect. However, for FAE-III the free and complexed structures were both determined at pH 4.5, which is close to the enzyme's optimal pH for activity of pH 5.0 (de Vries *et al.*, 1997). Therefore, it is unlikely that the pH is solely responsible for the changes in the catalytic triad of FAE-III.

There is some debate in the literature as to whether or not movement of the catalytic histidine is required for the serine esterase catalysis to succeed. Since the histidine is favourably positioned to remove a proton from the serine, it would subsequently appear to be favourably positioned to donate the proton back to the serine during breakdown of the tetrahedral intermediate, leading to regeneration of the substrate rather than the reaction proceeding to product. A similar argument would apply to the collapse of the second tetrahedral intermediate. Several proposals involving movement of the histidine have been put forward to resolve this problem. ^{15}N NMR work suggested that the Asp-His hydrogen bond is broken when the tetrahedral intermediate is formed and that the positively charged histidine moves into a position to donate a proton to the leaving group (Bachovchin, 1986). Another theory is that the imidazole of the histidine undergoes a 'ring

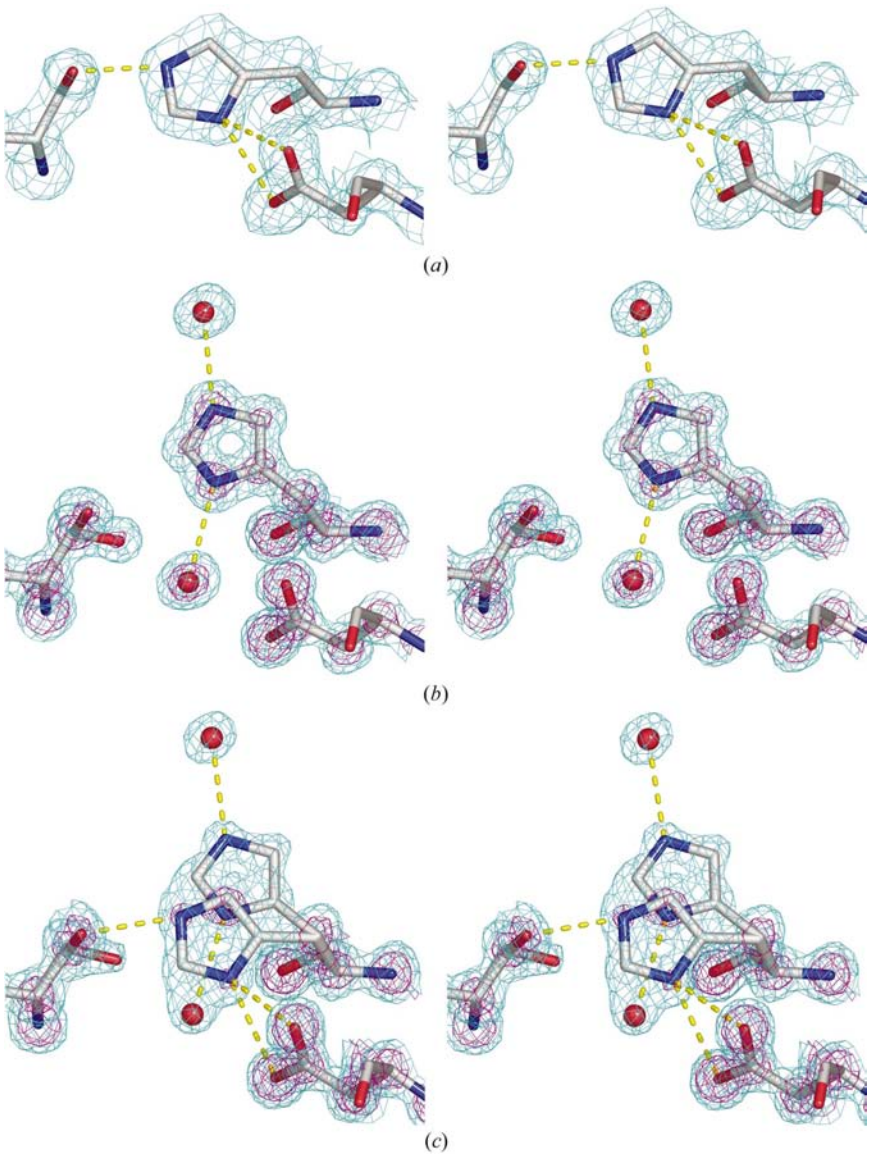
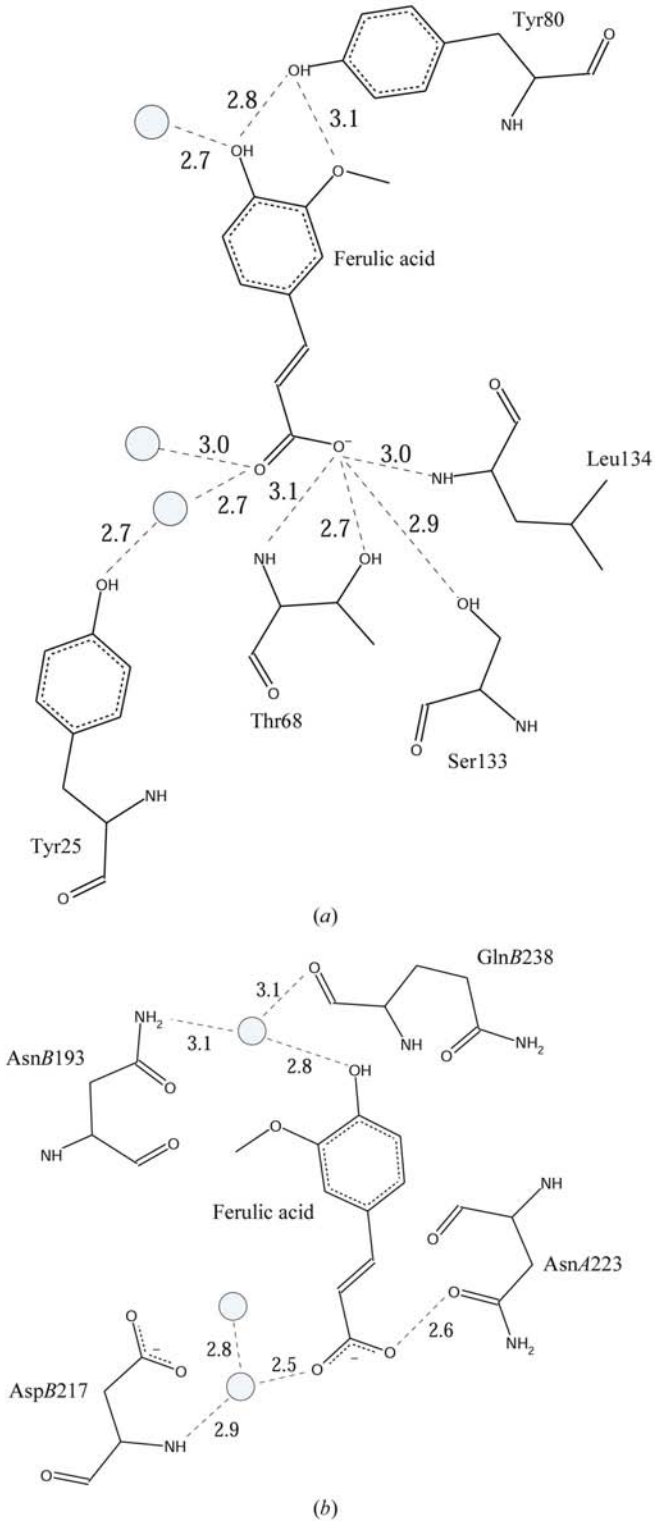


Figure 4
Views of the active-site residues of FAE-III: Ser133, Asp194 and His247. Electron density is shown at 1σ in cyan and 3σ in purple. (*a*) The native enzyme with an active catalytic triad. (*b*) The FAE-III–ferulic acid complex as seen in molecule *A*. His247 has changed position, leaving an inactive triad. (*c*) Molecule *B* of the FAE-III–ferulic acid complex. His247 is seen in both the active and inactive conformations. (*b*) and (*c*) also show the two conformations of the catalytic Ser133 residue.

flip' (Ash *et al.*, 2000). After the tetrahedral intermediate has formed, the imidazole ring flips, *i.e.* a rotation of approximately 180° about the $C^{\beta}-C^{\gamma}$ bond occurs, so that the proton on $N^{\delta1}$ is now positioned near the leaving group. There are counter arguments to this proposal: an *ab initio* QM/MM dynamics simulation of the tetrahedral intermediate found that the histidine moved between the two O atoms of the tetrahedral structure but no ring-flip occurred during the simulation (Topf *et al.*, 2002). Although there are numerous high-resolution crystallographic structures of serine-protease complexes (*e.g.* Kuhn *et al.*, 1998; Katona *et al.*, 2002), the catalytic histidine has been observed in the flipped orientation in only one instance, *i.e.* subtilisin in the presence of 50% dimethylformamide (Kidd *et al.*, 1999).

However, neither of the two orientations of His247 in the FAE-III–product complex correspond to the hypothetical 'flipped' orientation of the catalytic histidine. Instead, we observe two orientations for His247. The two positions are related to each other by an approximately 38° rotation about the $C^{\alpha}-C^{\beta}$ bond. The effect of this movement is to break the histidine–aspartic acid hydrogen bond and so our structure instead lends support to the moving-histidine mechanism proposed by Bachovchin (1986).

3.4.1. Binding interactions of ferulic acid. The electron-density maps clearly show a ferulic acid molecule bound at the active site of the esterase for both molecules in the asymmetric unit. The ferulic acid interacts through its OH and OCH_3 groups with the hydroxyl group of Tyr80. The carboxylate group is located in the region of the oxyanion hole, close to the main-chain N atom of Leu134, the backbone N atom and OH group of Thr68 and the active-site serine residue (Fig. 5*a*). Leu199 and Ile196 provide a hydrophobic environment. The amino acids directly involved in binding ferulic acid are unchanged in the native and the complex structures. The high resolution of the electron-density maps reveals that the ferulic acid molecule is not completely planar as modelled in the previous FAE–ferulic acid complexes (Schubot *et al.*, 2001; Prates *et al.*, 2001). The carboxylate group in particular is twisted away from a planar configuration (see Fig. 5*a*). The



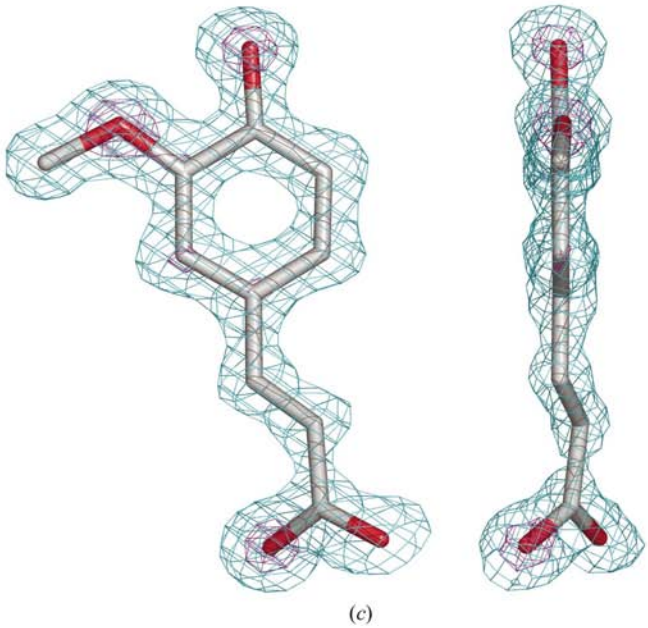


Figure 5
Direct and water-mediated enzyme–product interactions shown for ferulic acid binding at the active site (*a*) and at a site distant to the active site (*b*). (*c*) Orthogonal views of ferulic acid, with electron density contoured at 1.5σ in cyan and 5σ in violet.

atomic resolution of the data set of the complex allows estimation of the charge distribution of the carboxylate group of ferulic acid. The negative charge appears to be localized on the O atom (O2) that is closest to Ser133, judging by the bond lengths and the electron-density levels at the O-atom positions. Electron density of more than 6σ is observed at this atom, but is only observed at the 4σ level for the other O atom (O1). The C—O1 bond distance is 1.26 Å and the C—O2 distance is 1.34 Å. This would support a protonated state of the carboxylate group with the hydrogen attached to O2. However, there is no significant difference density to support this.

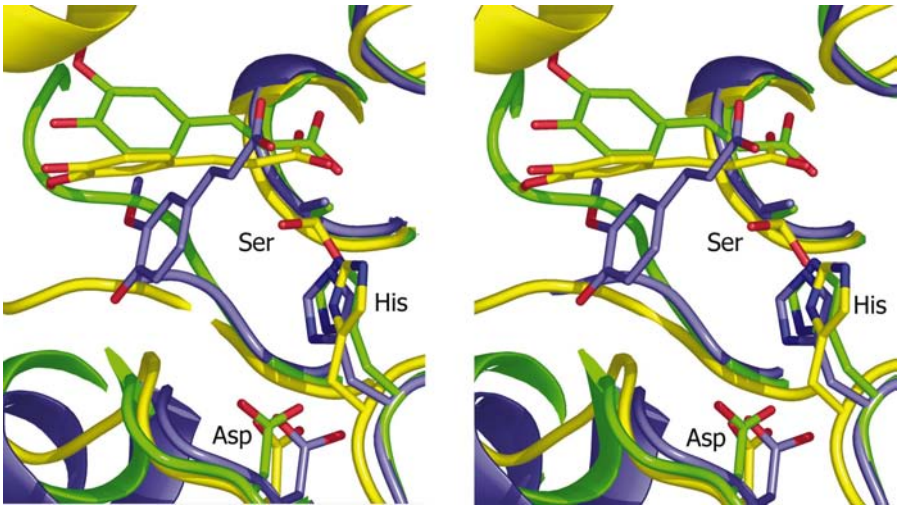


Figure 6
Stereoview of the active sites of FAE-III (C atoms and ribbon in yellow), FAE_Xyn10B (blue) and FAE_XynZ (green). The residues of the catalytic triads are labelled; however, FAE_Xyn10B and FAE_XynZ have Ser to Ala mutations.

When the structure of the FAE-III–ferulic acid complex was superimposed with the *C. thermocellum* FAE–ferulic acid complexes, it was found that the binding orientations of the ferulic acid are very different (see Fig. 6). Of the three different complexes, the orientations and positions of the ferulic acid from FAE-III and FAE_XynZ are the most similar, with deviations in atom positions ranging from a minimum of 0.85 Å at the carboxylate group to a maximum of 4.9 Å at the opposite end of the molecule. It is feasible that the ferulic acid in FAE_Xyn10B is displaced from its 'usual' binding site owing to the presence of a glycerol molecule at the active site and the mutation of the active-site serine to an alanine. If the glycerol is deleted from the model then the ferulic acid can be moved so that the carboxylate group is in a similar position to that observed in the other enzyme–ferulic acid structures and only a small rotation of the molecule is required to eliminate steric clashes. Alternatively, the different binding mode could reflect independent evolution of FAE activity from the ancestral protein as suggested at the end of §3.2.

There is supporting evidence that ferulic acid can bind to proteins in a non-productive manner since there is a third molecule of ferulic acid found in the FAE-III complex at a distance of approximately 20 Å from the active site. This ferulic acid interacts with molecule *A* at residue Asn*A*223 and with a molecule *B* from a different asymmetric unit *via* a water molecule to Asp*B*217; these interactions are shown in Fig. 5(*b*).

4. Conclusions

The high-resolution crystallographic data obtained for FAE-III have been useful in a number of ways. Firstly, improving the resolution to 1.5 Å allowed us to refine the molecular-replacement solution, which we were unable to do with the original 1.86 Å data. By collecting a 1.08 Å data set for the enzyme complex, we were able to use *ACORN* to perform phase refinement starting from just a small fragment of the model. In this way, we could avoid model bias. Secondly, high-resolution data allows the protein model and water structure to be constructed using automated methods with little user intervention. Finally, the high resolution allows us to examine the residues involved in substrate binding and hydrolysis in detail. For example, we can see that the ferulic acid is non-planar and we can identify dual conformations of the active-site residues, which would not have not been possible at a lower resolution.

References

Aliwan, F. O., Kroon, P. A., Faulds, C. B., Pickersgill, R. & Williamson, G. (1999). *J. Sci. Food Agr.* **79**, 457–459.
Andersen, A., Svendsen, A., Vind, J., Lassen, S. F., Hjort, C., Borch, K. & Patkar, S. A. (2002). *Colloids. Surf. B*, **26**, 47–55.
Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V., Haddad, K. C., Bradshaw, E. M., Sanford, D. G. & Bachovchin, W. W. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 10371–10376.
Bachovchin, W. W. (1986). *Biochemistry*, **25**, 7751–7759.
Blum, D. L., Kataeva, I. A., Li, X.-L. & Ljungdahl, L. G. (2000). *J. Bacteriol.* **182**, 1346–1351.
Bone, R., Sampson, N. S., Bartlett, P. A. & Agard, D. A. (1991). *Biochemistry*, **30**, 2263–2272.
Brady, L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J., Dodson, G., Tolley, S., Turkenburg, J. P., Christiansen, L., Hugejensen, B., Norskov, L., Thim, L. & Menge, U. (1990). *Nature (London)*, **343**, 767–770.
Brzozowski, A. M., Savage, H., Verma C. S., Turkenburg, J. P., Lawson, D. M, Svendsen, A. & Patkar, S. (2000). *Biochemistry*, **39**, 15071–15082.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Donaghy, J. & McKay, A. M. (1997). *J. Appl. Microbiol.* **83**, 718–726.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J.-X. & Zheng, C.-D. (2000). *Acta Cryst.* D**56**, 1137–1147.

Garcia-Conesa, M.-T., Kroon, P. A., Ralph, J., Mellon, F. A., Colquhoun, I. J., Saulnier, L., Thibault, J.-F. & Williamson, G. (1999). *Eur. J. Biochem.* **266**, 644–652.
Ghosh, D., Sawicki, M., Lala, P., Erman, M., Pangborn, W., Eyzaguirre, J., Gutierrez, R., Jornvall, H. & Thiel, D. J. (2001). *J. Biol. Chem.* **276**, 11159–11166.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hedstrom, L. (2002). *Chem. Rev.* **102**, 4501–4523.
James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C.-A. (1980). *J. Mol. Biol.* **144**, 43–88.
Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R. & Schofield, C. J. (2002). *J. Biol. Chem.* **277**, 21962–21970.
Kawasaki, K., Kondo, H., Suzuki, M., Ohgiya, S. & Tsuda, S. (2002). *Acta Cryst.* D**58**, 1168–1174.
Kidd, R. D., Sears, P., Huang, D.-H., Witte, K., Wong, C.-H. & Farber, G. K. (1999). *Protein Sci.* **8**, 410–417.
Koszelak, S., Ng, J. D., Day, J., Ko, T.-P., Greenwood, A. & McPherson, A. (1997). *Biochemistry*, **36**, 6597–6604.
Kroon, P. A., Faulds, C. B., Brezillon, C. & Williamson, G. (1997). *Eur. J. Biochem.* **248**, 245–251.
Kroon, P. A., Williamson, G., Fish, N. M., Archer, D. B. & Belshaw, N. J. (2000). *Eur. J. Biochem.* **267**, 6740–6752.
Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**, 27–33.
Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997). *J. Mol. Biol.* **268**, 779–799.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 283–291.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). *Acta Cryst.* D**57**, 1445–1450.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Prates, J. A. M., Tarbouriech, N., Charnock, S. J., Fontes, C. M. G. A., Ferreira, L. M. A. & Davies, G. J. (2001). *Structure*, **9**, 1183–1190.
Ralet, M.-C., Faulds, C. B., Williamson, G. & Thibault, J.-F. (1994). *Carbohydr. Res.* **263**, 257–269.
Ralph, J., Grabber, J. H. & Hatfield, R. D. (1995). *Carbohydr. Res.* **275**, 167–178.
Schubot, F. D., Kataeva, I. A., Blum, D. L., Shah, A. K., Ljungdahl, L. G., Rose, J. P. & Wang, B.-C. (2001). *Biochemistry*, **40**, 12524–12532.
Shindyalov, I. N. & Bourne, P. E. (1998). *Protein Eng.* **11**, 739–747.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). *Nucleic Acids Res.* **25**, 4876–4882.
Topf, M., Varnai, P. & Richards, W. G. (2002). *J. Am. Chem. Soc.* **124**, 14780–14788.
Vries, R. P. de, Michelsen, B., Poulsen, C. H., Kroon, P. A., van den Heuvel, R. H. H., Faulds, C. B., Williamson, G., van den Hombergh, J. P. T. W. & Visser, J. (1997). *Appl. Environ. Microbiol.* **63**, 4638–4644.
Westhead, D. R., Slidel, T. W., Flores, T. P. & Thornton, J. M. (1999). *Protein Sci.* **8**, 897–904.