FEBS Letters

FEBS PRESS science publishing by scientists

The crystal structure of a *Fusarium oxysporum* feruloyl esterase that belongs to the tannase family

Maria Dimarogona[1,2,*], Evangelos Topakas[2], Paul Christakopoulos[3] and Evangelia D. Chrysina[1]

1 Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece
2 School of Chemical Engineering, National Technical University of Athens, Greece
3 Biochemical Process Engineering, Division of Chemical Engineering, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Sweden

Correspondence
M. Dimarogona, Structural Biology and Biotechnology Laboratory, Department of Chemical Engineering, University of Patras, 26504 Rio, Patras, Greece
Tel: +30 2610997797
E-mail: mdimarog@chemeng.upatras.gr
and
E. D. Chrysina, Institute of Chemical Biology, National Hellenic Research Foundation, 48, Vassileos Constantinou Avenue, Athens 11635, Greece
Tel: +30 2107273851
E-mails: echrysina@eie.gr; echry@tee.gr

Present address
*Department of Chemical Engineering, University of Patras, Patras, Greece

(Received 29 August 2019, revised 13 March 2020, accepted 17 March 2020, available online 16 April 2020)

doi:10.1002/1873-3468.13776

Edited by Stuart Ferguson

Feruloyl esterases are enzymes of industrial interest that catalyse the hydrolysis of the ester bond between hydroxycinnamic acids such as ferulic acid and sugars present in the plant cell wall. Although there are several structures of biochemically characterized feruloyl esterases available, the structural determinants of their substrate specificity are not yet fully understood. Here, we present the crystal structure of a feruloyl esterase from *Fusarium oxysporum* (*Fo*FaeC) at 2.3 Å resolution. Similar to the two other tannase-like feruloyl esterases, *Fo*FaeC features a large lid domain covering the active site with potential regulatory role and a disulphide bond that brings together the serine and histidine of the catalytic triad. Differences are mainly observed in the metal coordination site and the substrate binding pocket.

Enzymes
E.C.3.1.1.73.

Databases
*The sequence of Fo*FaeC has been deposited with UniProt with accession code A0A1D3S5H0_FUSOX and the atomic coordinates of the three-dimensional structure with Protein Data Bank, with PDB code: 6FAT.

Keywords: biocatalysis; feruloyl esterase; *Fusarium oxysporum*; tannase family; X-ray crystallography

Ferulic acid esterases (FAEs) [E.C. 3.1.1.73, also known as feruloyl esterases and cinnamic acid hydrolases] are carboxylic acid esterases that hydrolyse the ester bonds between hydroxycinnamic acids or their dimers, and sugars present in the plant cell wall. They display significant variations in their amino acid sequence and biochemical characteristics, rendering their classification in subfamilies quite complicated. Initially, FAEs were categorized into four types (A, B, C and D), based on the construction of a neighbourhood-joining phylogenetic tree combined with substrate specificity data. The members of each subfamily displayed sequence homology and exhibited common preferences towards mono- and diferulates, as well as substitutions on the phenolic ring and linkage position on the arabinose moiety [1,2]. However, as the number of characterized FAEs increased, the ABCD classification ceased to cover the wealth of putative FAEs encoded in microbial genomes. Consequently, an updated phylogenetic analysis of published

Abbreviation
*Fo*FaeC, *Fusarium oxysporum* feruloyl esterase type C.

fungal genomes resulted initially to the division of FAEs into seven [3] and then to 13 subfamilies (SF1–SF13) [4]. This analysis revealed that FAEs have evolved from highly divergent enzyme families, involving tannases (SF1–4 and SF9–11), acetyl xylan esterases (SF5 and SF6), lipases (SF7) and lipase–choline esterases (SF8, SF12–13). The characterized members of each subfamily display distinct biochemical features, suggesting that this classification could actually reflect the different specificities of each FAE subgroup. In the ESTHER database [5], where α/β hydrolases are classified according to their sequence homology, FAEs are found in five different families, including antigen85c, esterase_phb and tannase from Block X and Lipase_3 from Block L. Over 1000 putative fungal FAEs have been predicted by similarity-based genome mining. Out of these, a set of 27 enzymes were selected to be investigated for their FAE activity, out of which, 20 exhibited similar activities to the ones previously shown [6].

Type A esterase from *Aspergillus niger* (FAEA, FAE-III or *An*FaeA) is one of the most extensively studied fungal FAEs. Its crystal structure has been determined in complex with ferulic acid or feruloylated oligosaccharides, showing that it is the ferulic acid part of the substrate that is oriented towards the core of the catalytic site interacting with the residues in the vicinity [7-9]. The structural features comprising the catalytic triad, that is Ser–His–Asp, the lid region shielding the active site and the oxyanion hole lying in the vicinity have been identified. The 3D structure of a type B esterase from *Aspergillus oryzae* was determined by Suzuki *et al.* [10]. It is the first crystal structure of a feruloyl esterase from the fungal tannase superfamily (Block X), a group that mainly involves tannases, which hydrolyse tannins to release gallic acid, and FAEs of fungal origin. *Ao*FaeB maintains most of the structural characteristics of serine proteases including *An*FaeA. The major differences it exhibits are located at the lid domain. A metal binding site has been revealed (calcium ion), the tight coordination of which seems to enhance the lid domain stability in the closed conformation [10]. A type C feruloyl esterase also from *A. niger* has been identified with a rather broad specificity and catalysed the hydrolysis of methyl 3,4-dimethoxycinnamate, ethyl ferulate, methyl ferulate (MFA), methyl *p*-coumarate (M*p*CA), ethyl coumarate, methyl sinapate (MSA) and methyl caffeate (MCA; Fig. S1) [11]; however, its crystal structure is not yet known. Most recently, the 3D structure of a type D feruloyl esterase from *Streptomyces cinnamoneus* (rR18) was determined at high resolution from a chimeric construct of two homologous strains (R18 and TH2–18) [12].

Type C feruloyl esterase from *Fusarium oxysporum*, *Fo*FaeC, is an enzyme of significant biotechnological interest that shows broad pH stability and considerable synergistic effect when applied together with xylanolytic enzymes on lignocellulosic biomass [13,14]. It is active against MCA, MSA, M*p*CA and to a lesser extent against MSA [13]. Recently, a protein engineering approach was employed to increase activity of *Fo*FaeC against MSA, revealing amino acids that influence its biochemical characteristics [15]. Similar to *Ao*FaeB, *Fo*FaeC is categorized into the fungal tannase superfamily (Block X). In this work, we present the three-dimensional structure of *Fo*FaeC. Comparison of our structure with the other available tannase-like FAE structures revealed alterations in the residues forming the substrate binding cleft, despite the overall conserved fold that could explain their kinetic properties. These differences are discussed, contributing to carving our understanding in structure–function relationships of these enzymes.

Materials and methods

Experimental procedures

Expression and purification of recombinant *Fo*FaeC

Recombinant *Fo*FaeC was prepared in *Pichia pastoris* as described previously [13]. Purification was performed using a metal ion affinity chromatography column (Talon; Clontech Laboratories Inc., a Takara Bio Company, Mountain View, CA, USA) equilibrated with 20 mM Tris/HCl buffer containing 300 mM NaCl (pH 8.0). The column was first washed with 60 mL buffer, and then, a linear gradient from 0 to 100 mM imidazole in 20 mM Tris/HCl buffer containing 300 mM NaCl (60 mL, pH 8.0) was applied at a flow rate of 2 mL·min^{-1}. The purified samples were assessed by SDS/PAGE. A single band corresponding to a molecular mass of 62 kDa was observed, indicating that the protein was suitable for crystallization trials.

Crystallization and data collection

Purified *Fo*FaeC was concentrated to 25 mg·mL^{-1} in 20 mM Tris/HCl pH 8.0 buffer and submitted to crystallization trials using the sitting drop vapour diffusion technique. A large number of crystallization conditions were screened in 96-well MRC crystallization plates (Molecular Dimensions, Cambridge, UK) with the aid of an OryxNano crystallization robot (Douglas Instruments Ltd, Hungerford Berkshire, UK) installed at NHRF, using commercially available crystallization kits. Crystals of *Fo*FaeC appeared in various PEG-based conditions and diffracting crystals were obtained in 30% (v/v) PEG400, 0.1 M Tris/HCl pH 8.5 buffer at 16 °C. X-ray diffraction data were collected at 100 K using the synchrotron

radiation source at PETRA III, EMBL-Hamburg beamline P14 (λ = 1.2395 Å, oscillation range 0.2°). Prior to data collection, the crystals were flash-cooled to 100 K in the nitrogen stream. A complete data set was collected at 2.3 Å from a single crystal. Data processing was performed with *XDS* [16] followed by data integration and scaling with *AIMLESS* [17] as implemented in the *CCP4* programme suite [18]. X-ray diffraction data analysis showed that *Fo*FaeC crystals were grown in spacegroup $P2_1$ with unit-cell dimensions a = 67.5 Å, b = 87.5 Å, c = 106.6 Å, β = 106.2°, and two molecules in the asymmetric unit. Data collection statistics for *Fo*FaeC are presented in Table 1.

Structure determination, refinement and analysis

The structure of *Fo*FaeC was determined by molecular replacement with PHASER [19] using the crystal structure of *Ao*FaeB determined at 1.5 Å resolution (PDB code 3WMT) [10] as starting model. *Fo*FaeC and *Ao*FaeB share 49% sequence identity for 89% coverage. Iterative cycles of model building and refinement were performed with *COOT* [20] and *PHENIX* [21]. Validation of the final refined model, the metal binding sites and the glycosylation sites was performed with *MolProbity* [22], *CheckMyMetal* (*CMM*) [23], *NetNGlyc* [24] and *pdb-care* web servers [25], respectively. The *CATH* database of domain structures server [26] was used for the structural classification of *Fo*FaeC folding. The secondary structure elements were identified with *DSSP* [27,28], and the topology diagram of the structure was extracted using *PDBsum* (http://www.ebi.ac.uk) [29] on the EBI server (Fig. S2). All structural figures were generated with *Chimera* [30]. The refined model and structure factors are deposited to the Protein Data Bank with accession code 6FAT.

Results and Discussion

Sequence analysis

Recombinant *Fo*FaeC is composed of 563 amino acids. According to the *NetNGlyc* 1.0 [24] and *NetOGlyc* 4.0 [31] servers, there are six predicted *N*-glycosylation sites, involving residues Asn101, Asn151 and Asn362, followed by Asn66, Asn111 and Asn324, and four possible *O*-glycosylation sites, involving Thr68, Thr75, Thr505 and Thr506. *Fo*FaeC is a type C feruloyl esterase [13] and belongs to subfamily SF2 of the phylogenetic classification [4], which is related to tannases. In addition, it belongs to the fungal tannase family of *ESTHER* database [5]. A *BLAST* search against the PDB database [32] showed that the closest homologues with known structure were the *A. oryzae* feruloyl esterase C, *Ao*FaeC (PDB code 6G21) sharing 53% sequence identity for 89% sequence coverage and *Ao*FaeB (PDB code 3WMT) sharing 49% sequence identity for 89% sequence coverage (Fig. 1).

Table 1. Diffraction data and refinement statistics for *Fo*FaeC structure. Values in parentheses are for the outermost shell.

Data collection and processing statistics	
Crystallization condition	0.1 M Tris/HCl pH 8.5, 30% (v/v) PEG400
Beamline	P14 (EMBL-Hamburg, PETRA III)
Wavelength (Å)	1.23951
Space group	$P2_1$
Unit-cell parameters (Å)	a = 67.5, b = 87.5, c = 106.6 Å, α = γ = 90°, β = 106.2°
No. of molecules per asymmetric unit	2
Resolution (Å; outermost shell)	87.47–2.30 (2.37–2.30)
No. of observations	202 531 (17 193)
No. of unique reflections	52 149 (4515)
Completeness (%)	98.4 (98.2)
R_{merge}[a] (%)	9.6 (82.4)
$< I/\sigma(I) >$	9.5 (1.7)
$CC_{1/2}$[b]	99.6 (63.5)
Multiplicity	3.9 (3.8)
Wilson B value (Å^2)	29.1
Refinement statistics and model quality	
No. of reflections	49 493
Residues included	Chain A: (36–543), Chain B: (36–543)
No. of protein atoms	7913
No. of heteroatoms	
Solvent molecules	357 (HOH)
Calcium ions	2 (CA)
Glycan chain A	98 (7 NAG),33 (3 BMA), 22 (2 MAN)
Glycan chain B	98 (7 NAG), 22 (2 BMA), 143 (13 MAN)
R/R_{free}	0.207/ 0.235
R.m.s. deviation in	
Bond lengths (Å)	0.007
Bond angles (°)	1.072
MolProbity analysis[c]	
Ramachandran favoured/outliers (%)	97.7/0
Poor rotamer outliers (%)	0.1
Average B (Å^2) for protein residues	
Overall	35.5
Backbone atoms	35.3
Side chain atoms	35.6
Average B (Å^2) for heteroatoms	
Solvent molecules	32.0
Calcium ions	31.8
Glycan chain A	53.6 (NAG), 60.7 (BMA), 59.7 (MAN)
Glycan chain *B*	44.9 (NAG), 38.9 (BMA), 41.5 (MAN)

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, [b] $CC_{1/2}$ is the correlation coefficient between two random half data sets, [c] calculated using *MolProbity* (http://molprobity.biochem.duke.edu/).

The next hits in the *BLAST* output displayed no homology to *Fo*FaeC, demonstrating the uniqueness of the overall fold of the three esterases. Since *Ao*FaeC

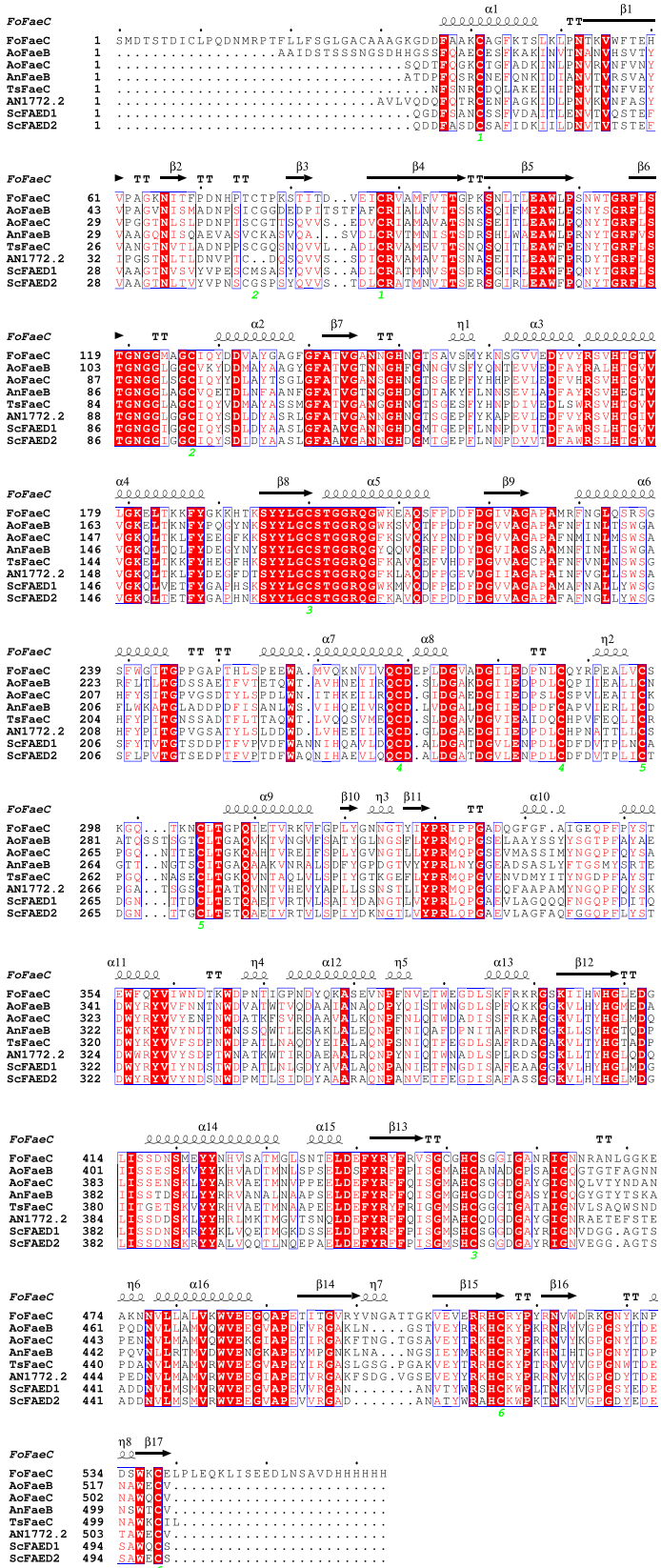

Fig. 1. Sequence alignment of characterized FAEs that belong to the fungal tannase family of *ESTHER* database: *Fo*FaeC, *Ao*FaeB (XP_001818628) and *Ao*FaeC (XP_001819091) from *Aspergillus oryzae*; *An*FaeB (UniProt Q8WZI8) from *Aspergillus niger*; *Ts*FaeC from *Talaromyces stipitatus* (AJ_505939); AN1772.2 from *Aspergillus nidulans*; and *Sc*FaeD1 (XP_003027549.1) and *Sc*FaeD2 (XP_003027556.1) from *Schizophyllum commune*. The secondary structure elements are shown for *Fo*FaeC structure, with α-helices, 3_{10}-helices and β-strands being denoted *α*, *η* and *β*, respectively. The figure was prepared using *ESPRIPT* [33].

structure was not available when the diffraction data for *Fo*FaeC were collected, we used the *Ao*FaeB structure as starting model and solved the structure of *Fo*FaeC by molecular replacement.

Analysis of the *Fo*FaeC crystal structure

The crystal structure of *Fo*FaeC was determined at 2.3 Å resolution using the *Ao*FaeB structure (PDB code 3WMT, 1.5 Å resolution) [10] as starting model from crystals grown in spacegroup $P2_1$, with unit-cell dimensions $a = 67.5$ Å, $b = 87.5$ Å, $c = 106.6$ Å, $\beta = 106.2°$ and two molecules in the asymmetric unit (Table 1, Fig. 2). Out of the 563 amino acids, a total of 508 were modelled in the density, while the *N*- and *C*-termini ones, comprising residues (1–35) and residues (544–564) were excluded from the structure due to insufficient density. The geometry of the residues was checked by *MolProbity* [22]; all lay in allowed regions of the Ramachandran plot except Lys527, which is located in a flexible loop with rather poor density.

*Fo*FaeC is a serine-type hydrolase, and its three-dimensional structure has an α/β-hydrolase fold with a three-layer ($\beta\alpha\beta$) sandwich architecture and a Rossmann topology, bearing a twisted β-sheet, intercalated between two layers of α-helices (Fig. 2). This folding deviates from the canonical α/β hydrolase fold that comprises a β-sheet of eight strands and two clusters of α-helices bilaterally (four and two α-helices) [34]. The secondary structure elements are derived by *PROMOTIF* [35] as implemented in *PDBsum* [29], which is presented in detail in Figs S2 and S3. The overall structure is stabilized by six disulphide bonds, that is Cys41–Cys88, Cys76–Cys127, Cys200–Cys453, Cys269–Cys287, Cys296–Cys304 and Cys516–Cys538. *Fo*FaeC is monomeric in solution, as confirmed with gel filtration (data not shown). In the crystal, however, it appears as a dimer (Fig. 2), the packing of which in the unit cell is shown in Fig. S4.

Residues Ser201, His452 and Asp 412 form the catalytic triad. Ser201 lies at the interface of two secondary structure elements, a β-strand and an α-helix (S8 and H7) forming the characteristic 'nucleophile elbow' of the α/β hydrolase fold with Ramachandran angles in the allowed region [34]. His452 belongs to the 'histidine loop' region (connecting β-strand S13 and 3_{10} helix, h21) that ensures the residue is oriented in a suitable position to facilitate the hydrogen bond interactions formed with both Ser201 and Asp412, through its dual action as an acid/base. Asp412 is located between a β-strand and an α-helix (S12 and H19) and complements the organization of the active site (Figs 2 and 3 and Fig. S3).

Two calcium ions with an octahedral geometry were included in the structure as suggested by both the $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ electron density maps in the lid domains of the dimer and confirmed by the server *CheckMyMetal* [23] (Fig. S5). Although calcium was not part of the crystallization conditions, it appears as a natural ligand of the enzyme, in accordance with previous studies [10]. The identity of the bound ion was further confirmed by ICP-SFMS (Fig. S6). The two calcium ions are coordinated by six oxygen atoms in each monomer, out of which, three belong to aspartic residues (Asp270 OD1; Asp274 OD1; Asp278 OD1), two come from the backbone oxygen atom of Val276 and Ile280, and one from a water molecule (Wat30 O and Wat87 O for monomers A and B, respectively; Fig. 4).

Additional density was observed next to Asn66, Asn101, Asn151 and Asn362, indicating that these residues were indeed glycosylated in both monomers, in accordance with what was suggested by *NetNGlyc* [24]. Details on the glycosylation sites are shown in Fig. 5.

Comparison of *Fo*FaeC with other known feruloyl esterase structures

A search for *Fo*FaeC structural homologues using *DALI* server [36] indicated that the closest structural homologue of *Fo*FaeC (chain A) is *Ao*FaeC (PDB code 6Q21, chain A), with a *Z*-score of 61.1 and r.m.s.d 1.2 Å, followed by *Ao*FaeB (chain A) with a *Z*-score of 56.3 and r.m.s.d. 1.1 Å. The third closest structural homologue is a mono-(2-hydroxyethyl) terephthalate (MHET) esterase (PDB code 6QG9, chain A), with a Z-score of 41.2 and r.m.s.d. 2.6 Å. The rest of the structures that were enlisted in the catalogue generated by *DALI* server had a significantly lower *Z*-score (< 18.5). Although all structures follow the same α/β hydrolase fold, only the three ones exhibiting the highest structural similarity share a rather extended 'lid domain'. The 'lid' lies at the entrance of the catalytic site limiting the access of the substrate (Fig. 7b), unlike the rest of the known FAEs deposited to Protein Data Bank, where the corresponding site is fully or partially exposed (forming a groove) to the solvent.

Characterized FAEs that belong to the fungal tannase family of the *ESTHER* database involve *Ao*FaeC (XP_001819091) and *Ao*FaeB (XP_001818628) from *A. oryzae* [37], *An*FaeB (UniProt Q8WZI8) from *A. niger* [38], *Ts*FaeC from *Talaromyces stipitatus* (AJ_505939) [39] and AN1772.2 from *Aspergillus nidulans* [40] (Fig. 1). Even though these esterases belong to the same family and share sequence homology, they

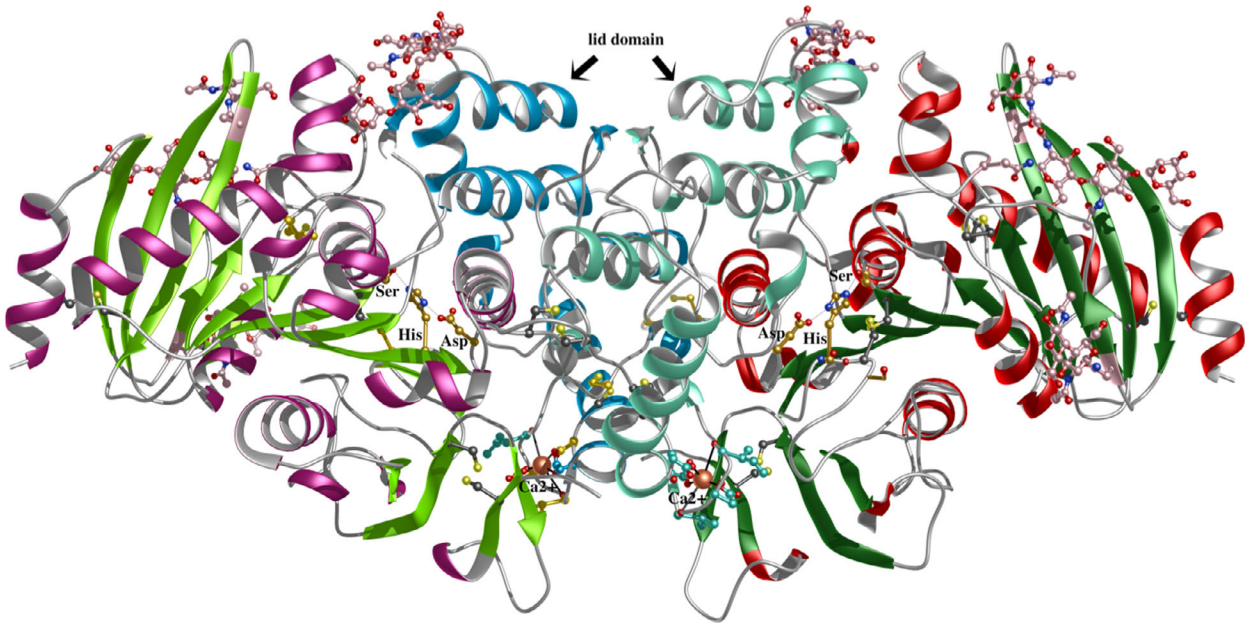

Fig. 2. The overall structure of *Fo*FaeC determined at 2.3 Å resolution. The two monomers follow the α/β hydrolase fold with three layers of ($\beta\alpha\beta$) sandwich architecture folded in a twisted β-sheet (shown in light and dark green for chains A and B) surrounded by two layers of α-helices (shown in purple and red). Each monomer has its own lid domain depicted in blue and turquoise that shields the active site of the enzyme. The catalytic triad residues (shown in gold) and the disulphide bonds formed are indicated. Two calcium ions, which were identified in the structure, and the glycosylation sites of the enzyme are also presented.

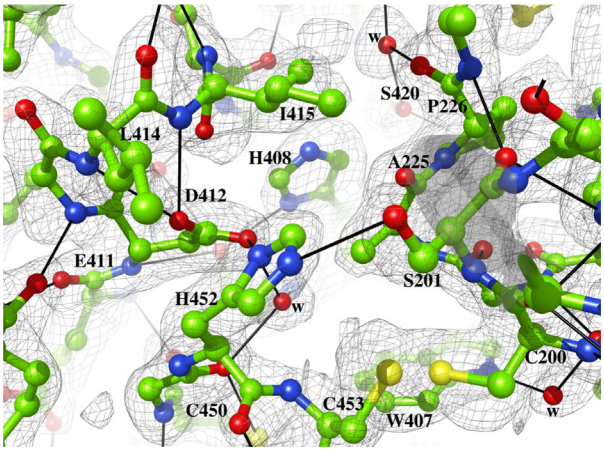

Fig. 3. Close view of the *Fo*FaeC active site with the $2F_{obs} - F_{calc}$ electron density map contoured at 1σ level. The catalytic triad residues Ser201, His452 and Asp412 and their hydrogen bond interactions (depicted as solid black lines) formed are indicated.

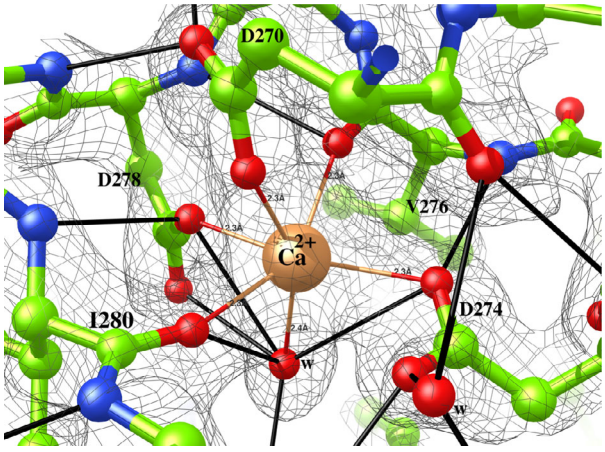

Fig. 4. The coordination of the calcium binding site *Fo*FaeC determined at 2.3 Å resolution is shown along with the $2F_{obs} - F_{calc}$ contoured at 1σ level.

seem to differ as far as their specificity on methyl esters of cinnamic acids is concerned. More specifically, *Ao*FaeB, *An*FaeB and *An*1772.2 cannot hydrolyse MSA. The other members of the family are active on all esters; however, *Ao*FaeC and *Fo*FaeC display significantly reduced activity against MSA. *Ts*FaeC, a true type C feruloyl esterase, displays similar activity against MFA, MpCA, MSA and slightly lower against MCA. As indicated by Suzuki *et al.* [10], most of the residues forming the substrate binding pocket are highly conserved among the aforementioned enzymes, and small differences could be the reason for this observed variation. Sinapinic acid is the bulkiest

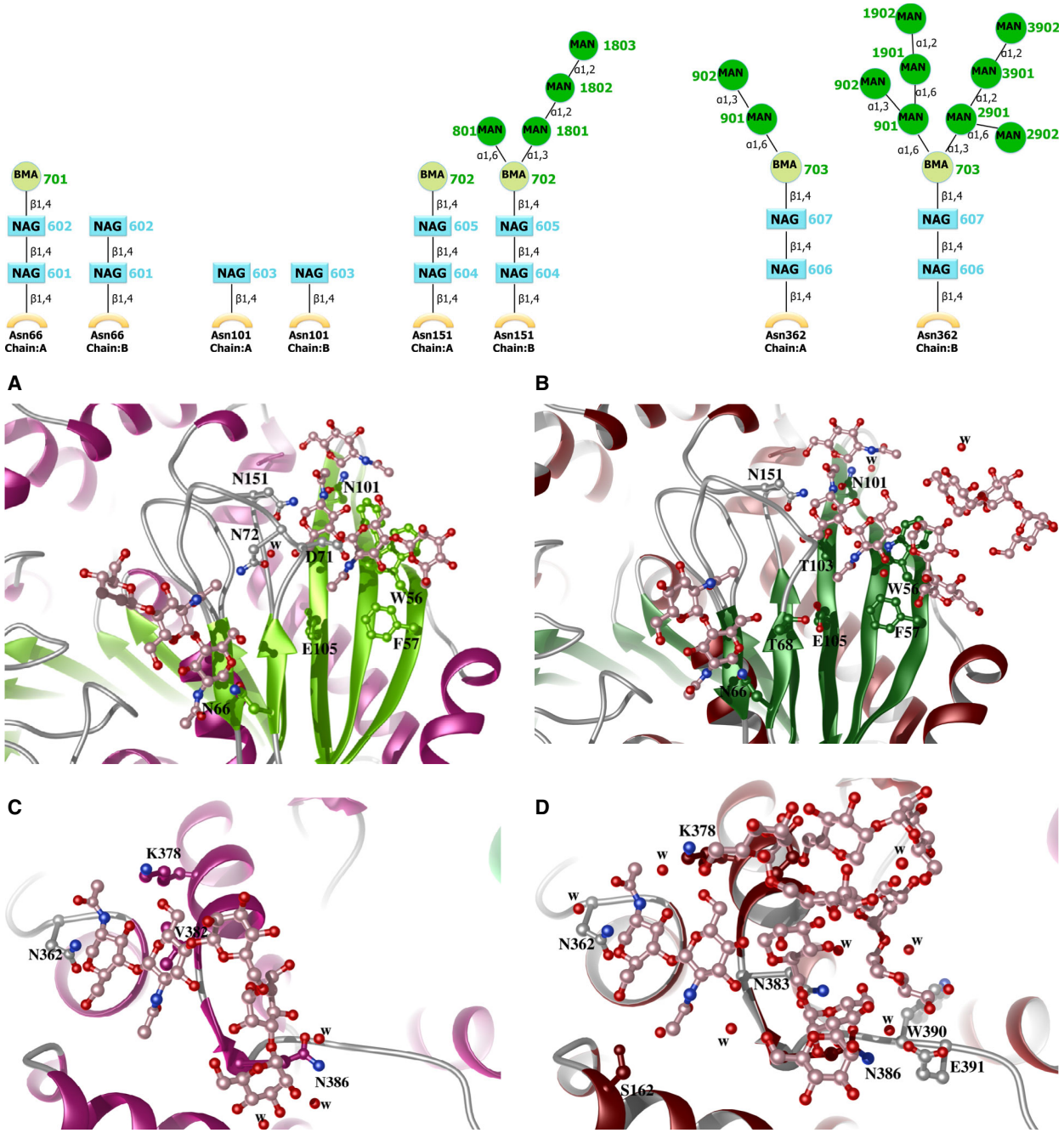

Fig. 5. (Top) Detailed view of *Fo*FaeC *N*-glycosylation sites presenting the type of the links formed between the sugars. (Bottom) The carbohydrate atoms are depicted in ball-and-stick representation and for chain A (A, C) and chain B (B, D) following the colour code from Fig. 2 for the two monomers. The glycans are attached to the amide nitrogen of Asn66, Asn101, Asn151 (A, B) and of Asn362 (C, D).

cinnamic acid with two methoxy and one hydroxyl groups on its phenolic ring, and a more congested active site could impede the accommodation of its side chain.

A more detailed structural comparison was performed between *Fo*FaeC and *Ao*FaeB structures, the latter being the first tannase-like feruloyl esterase that its three-dimensional structure revealed the presence of the novel lid domain and the disulphide bond in the vicinity that brings together Ser and His residues of the catalytic triad [10]. Overall, the crystal structures of both *Fo*FaeC and *Ao*FaeB are quite similar with

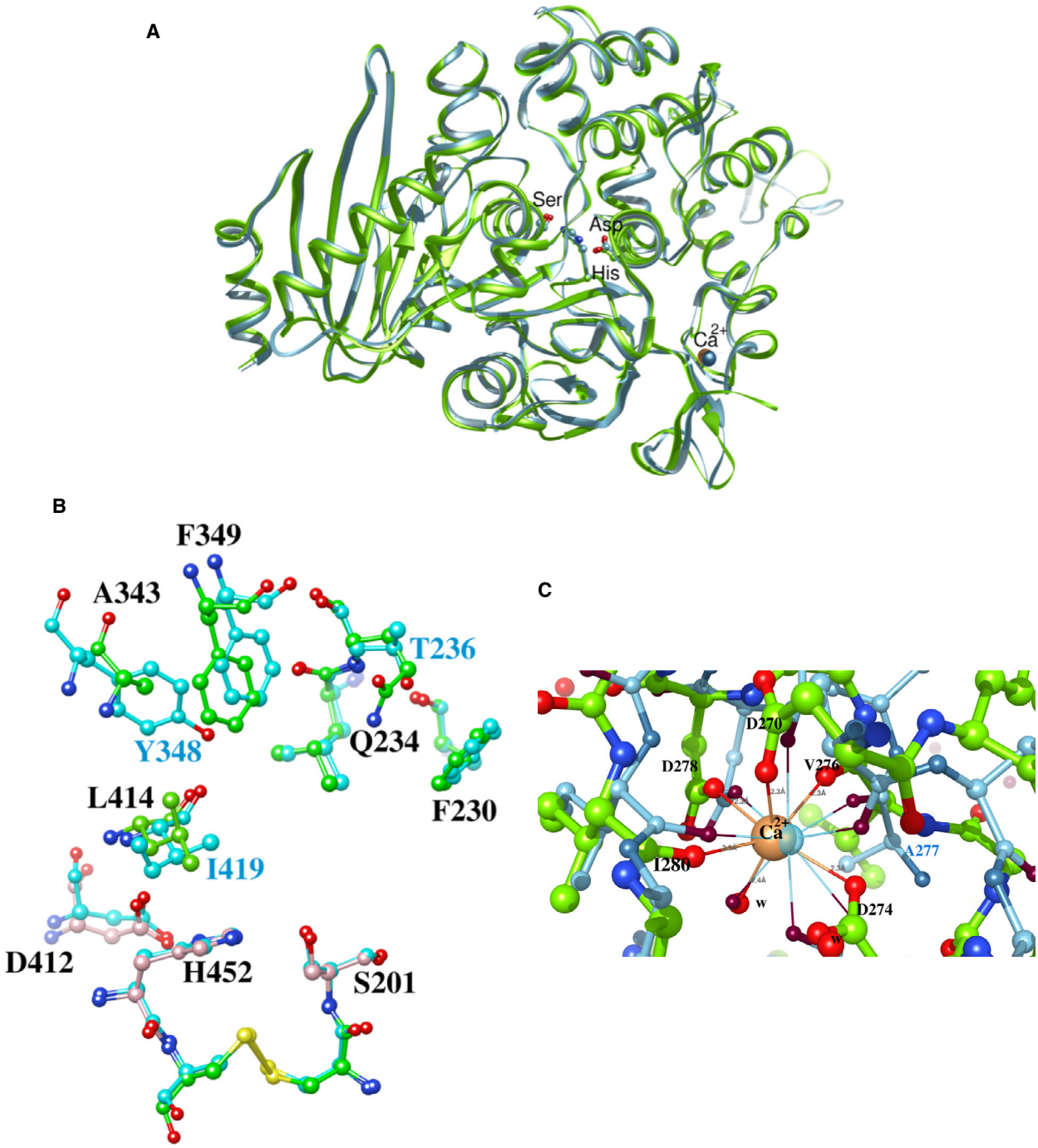

Fig. 6. (A) Superposition of *Fo*FaeC crystal structure (shown in green) on *Ao*FaeB (cyan) using the secondary structure elements. The catalytic site and the calcium binding sites are indicated. Close-up view of (B) substrate binding pocket (the catalytic triad residues are highlighted in pink for *Fo*FaeC) and (C) the calcium binding site.

minor changes mainly observed in loop regions away from the catalytic sites of the two enzymes (Fig. 6a). Changes were also observed in the amino acids forming the substrate binding pocket, as estimated by Suzuki *et al.* [10]. In specific, docking studies showed that the phenolic group of the substrate is held in a hydrophobic pocket surrounded by residues Phe232, Leu235, Thr236, Tyr348, Phe354, Tyr356 and Ile419 of *Ao*FaeB [10]. In the case of *Fo*FaeC structure, there is an alanine (343) in the place of Tyr348 and a

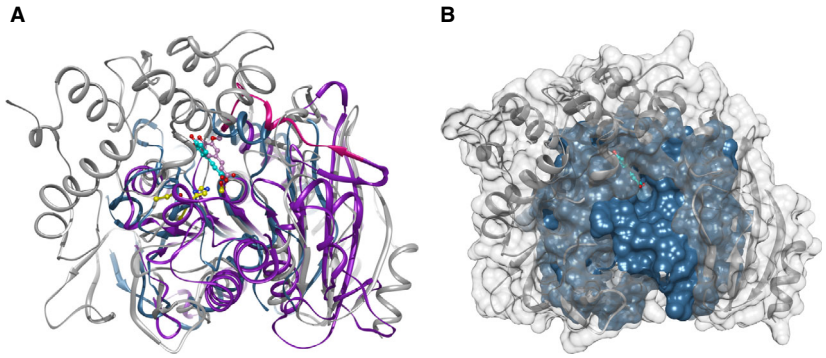

Fig. 7. (A) Superposition of the rR18 (shown in purple) and *Aspergillus niger* (shown in blue) feruloyl esterase crystal structures onto *Fo*FaeC (shown in grey). The catalytic site residues for *Fo*FaeC are indicated (in yellow) and ferulic acid bound at the active site for rR18 and *A. niger* (shown in mauve and cyan, respectively). The ligand coordinates for rR18 have been obtained as described in [12]). Residues (140–154) lining the R18 loop are highlighted in pink. (B) Surface representation of *Fo*FaeC structure (shown in grey) and *A. niger* (shown in silver blue). The molecular surface of *Fo*FaeC highlights the additional domains present in *Fo*FaeC structure.

glutamine (234) instead of Thr236 (Fig. 6b) Antonopoulou *et al.* [15] managed to increase *Fo*FaeC activity against MSA by employing a protein engineering approach; their double-mutant F230H/T202V opened up the binding pocket, allowing the accommodation of MSA and resulting in a fivefold improvement in catalytic efficiency compared to wild-type. Changes between *Fo*FaeC and *Ao*FaeB structures were also observed in the coordination of the calcium ion and in residues in the vicinity (Fig. 6c). One of the residues participating in the metal coordination through its backbone oxygen varied from alanine to valine in *Fo*FaeC, while the interaction with Asp272 O is missing. Interestingly, the geometry of the calcium ion is not pentagonal bipyramid but octahedral, implying a tighter packing of the surrounding oxygen ligands. This type of geometry was also observed in two of the calcium ions bound in the lower structural homologue of *Fo*FaeC, MHET esterase; the identity of the metal ion in the latter was also confirmed by X-ray fluorescence spectroscopy [41]. Although calcium ion is known to exhibit variable coordination numbers, without directionality [42], its importance for the structural integrity of *Fo*FaeC is yet to be examined by targeted studies, for example employing chelating agents during the preparation stage and protein engineering studies. Insight on the changes observed in the vicinity of the metal binding site, including the solvent structure, requires also further investigation through mutational studies and high-resolution structures that may reveal the structural determinants that dictate the geometry of the active site and the catalytic activity of the enzyme.

The crystal structure *Fo*FaeC was further compared with rR18, a type D feruloyl esterase from *S. cinnamoneus*, the structure of which was recently determined (PDB code: 5YAL) [12] and a complex of *A. niger* feruloyl esterase with ferulic acid (PDB code: 1UWC) [7] (Fig. 7). Superposition was performed using the secondary structure elements with reference structure the *Fo*FaeC. The results showed that all three structures followed the α/β-hydrolase fold and the most profound differences were observed in the environment of the active site. More specifically, the 'lid' domain of *Fo*FaeC, present also in *Ao*FaeB, is missing from *A. niger* structure and rR18. The loop region of rR18 that was suggested to be crucial for the release of ferulic acid [12] is not present either and does not seem to match with any other structural features that *Fo*FaeC, *Ao*FaeB or *A. niger* bear.

A comparison with the feruloyl esterases deposited to the protein data bank showed that the characteristic α/β-hydrolase fold of the family is retained, and the most remarkable differences occur in the 'lid domain', the role of which appears to be regulatory. The 'lid' domain as a unique structural feature of *Fo*FaeC, *Ao*FaeC and *Ao*FaeB until present may have a pivotal role in their function as compared to the rest of feruloyl esterases. Structural studies of more fungal esterase structures are essential to reveal its importance and allow interpretation of the structure–function relationships underlying the regulatory role of the 'lid' domain in enzymic activity. To this end, further mutational studies are required to fully unwind the potential of FAEs as biocatalysts. Metagenomic libraries are an endless source of information to be examined by consortia that make the most of the currently available research infrastructures for structural biology. Determination of more feruloyl esterase structures at high resolution could shed light on the function of these

enzymes and their evolution pathway bridging the gap with the industrial sector and allowing better results.

Acknowledgements

We thank the European Molecular Biology Laboratory (Hamburg) for the allocation of synchrotron beam time and EMBL P13/P14 PETRA III scientists for assistance. The work was supported by the European Commission's Seventh Framework Program (FP7) Research Potential of Convergence Regions REGPOT-2009-1 No. 245866 'ARCADE'; the research leading to these results also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under 'BioStruct-X' (FP7 Research Infrastructures grant agreement No. 283570) providing access to PETRA III synchrotron (projects numbers 1196, 3976 and 7054 to EDC). We acknowledge EMBL P13/P14 PETRA III scientists for their help and especially Dr Gleb Bourenkov for his assistance in collecting the FoFaeC data set in the frame of the joint EMBL-CCP4 training course 'European School for Macromolecular Crystallography (ESMAX)' that was held at EMBL-Hamburg, Germany (19–26 November 2012), and Drs Michelle Cianci and Guillaume Pompidor for their help with the efforts on the derivatization of *Fo*FaeC crystals. We would also like to thank Dr Io Antonopoulou (Luleå University of Technology) for comments on the manuscript about current research on feruloyl esterases. Last, we would like to thank Dr Anastasia Zerva for producing and purifying the sample sent for ICP-SFMS analysis. We also acknowledge support of this work by the project 'INSPIRED-The National Research Infrastructures on Integrated Structural Biology, Drug Screening Efforts and Drug target functional characterization' (MIS 5002550), which is implemented under the Action 'Reinforcement of the Research and Innovation Infrastructure', funded by the Operational Programme 'Competitiveness, Entrepreneurship and Innovation' (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Author contributions

MD and ET performed expression and purification of the protein. MD carried out the structural biology experiments, processed, analysed the data and contributed in writing up of the paper. PC and ET have performed the initial expression and purification of the enzyme. EDC supervised the structural biology experiments, analysed the data, wrote and edited the paper. All authors reviewed the results and approved the final version of the manuscript.

References

1 Crepin VF, Faulds CB and Connerton IF (2004) Functional classification of the microbial feruloyl esterases. *Appl Microbiol Biotechnol* **63**, 647–652.
2 Topakas E, Vafiadi C and Christakopoulos P (2007) Microbial production, characterization and applications of feruloyl esterases. *Process Biochem* **42**, 497–509.
3 Benoit I, Danchin EGJ, Bleichrodt RJ and De Vries RP (2008) Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnol Lett* **30**, 387–396.
4 Dilokpimol A, Mäkelä MR, Aguilar-Pontes MV, Benoit-Gelber I, Hildén KS and de Vries RP (2016) Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol Biofuels* **9**, 231.
5 Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P and Chatonnet A (2013) ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res* **41**: D423-9.
6 Dilokpimol A, Mäkelä MR, Varriale S, Zhou M, Cerullo G, Gidijala L, Hinkka H, Brás JLA, Jütten P, Piechot A *et al.* (2018) Fungal feruloyl esterases: Functional validation of genome mining based enzyme discovery including uncharacterized subfamilies. *N Biotechnol* **41**, 9–14.
7 McAuley KE, Svendsen A, Patkar SA and Wilson KS (2004) Structure of a feruloyl esterase from *Aspergillus niger*. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 878–887.
8 Faulds CB, Molina R, Gonzalez R, Husband F, Juge N, Sanz-Aparicio J and Hermoso JA (2005) Probing the determinants of substrate specificity of a feruloyl esterase, AnFaeA, from *Aspergillus niger*. *FEBS J* **272**, 4362–4371.
9 Hermoso JA, Sanz-Aparicio J, Molina R, Juge N, González R and Faulds CB (2004) The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *J Mol Biol* **338**, 495–506.
10 Suzuki K, Hori A, Kawamoto K, Thangudu RR, Ishida T, Igarashi K, Samejima M, Yamada C, Arakawa T, Wakagi T *et al.* (2014) Crystal structure of a feruloyl esterase belonging to the tannase family: a disulfide bond near a catalytic triad. *Proteins* **82**, 2857–2867.
11 Dilokpimol A, Mäkelä MR, Mansouri S, Belova O, Waterstraat M, Bunzel M, de Vries RP and Hildén KS (2017) Expanding the feruloyl esterase gene family of

Aspergillus niger by characterization of a feruloyl esterase, FaeC. *N Biotechnol* **37**, 200–209.
12 Uraji M, Tamura H, Mizohata E, Arima J, Wan K, Ogawa K, Inoue T and Hatanaka T (2017) Loop of Streptomyces feruloyl esterase plays an important role in its activity of releasing ferulic acid from biomass. *Appl Environ Microbiol* **84**, e02300-17.
13 Moukouli M, Topakas E and Christakopoulos P (2008) Cloning, characterization and functional expression of an alkalitolerant type C feruloyl esterase from *Fusarium oxysporum*. *Appl Microbiol Biotechnol* **79**, 245–254.
14 Xiros C, Moukouli M, Topakas E and Christakopoulos P (2009) Factors affecting ferulic acid release from Brewer's spent grain by *Fusarium oxysporum* enzymatic system. *Bioresour Technol* **100**, 5917–5921.
15 Antonopoulou I, Hunt C, Cerullo G, Varriale S, Gerogianni A, Faraco V, Rova U and Christakopoulos P (2018) Tailoring the specificity of the type C feruloyl esterase FoFaeC from *Fusarium oxysporum* towards methyl sinapate by rational redesign based on small molecule docking simulations. *PLoS ONE* **13**, e0198127.
16 Kabsch W (2010) XDS. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 125–132.
17 Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr Sect D Biol Crystallogr* **67**, 282–292.
18 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* **67**, 235–242.
19 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
20 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 486–501.
21 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 213–221.
22 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 12–21.
23 Zheng H, Chordia MD, Cooper DR, Chruszcz M, Muller P, Sheldrick GM and Minor W (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat Protoc* **9**, 156–170.
24 Gupta R, Jung E and Brunak S (2002) Prediction of N-glycosylation sites in human proteins. *Pacific Symp Biocomput* **7**, 310–322.
25 Lütteke T and von der Lieth C-W (2004) pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* **5**, 69.
26 Cuff AI, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton J and Orengo CA (2009) The CATH classification revisited--architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res* **37** (Database), D310–D314
27 Kabsch W and Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
28 Touw WG, Baakman C, Black J, Te Beek TAH, Krieger E, Joosten RP and Vriend G (2015) A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* **43**, D364–D368.
29 Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML and Thornton JM (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci* **22**, 488–490.
30 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
31 Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT-BG, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**, 1478–1488.
32 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
33 Robert X and Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.
34 Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J *et al.* (1992) The alpha/beta-hydrolase fold. *Protein Eng* **5**, 197–211.
35 Hutchinson EG and Thornton JM (1996) PROMOTIF-A program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212–220.
36 Holm L and Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38** (suppl_2), W545–W549.

37 Koseki T, Hori A, Seki S, Murayama T and Shiono Y (2009) Characterization of two distinct feruloyl esterases, AoFaeB and AoFaeC, from *Aspergillus oryzae*. *Appl Microbiol Biotechnol* **83**, 689–696.
38 Kroon PA, Faulds CB and Williamson G (1996) Purification and characterization of a novel esterase induced by growth of *Aspergillus niger* on sugar-beet pulp. *Biotechnol Appl Biochem* **23** (Pt 3), 255–262.
39 Garcia-Conesa MT, Crepin VF, Goldson AJ, Williamson G, Cummings NJ, Connerton IF, Faulds CB and Kroon PA (2004) The feruloyl esterase system of *Talaromyces stipitatus*: production of three discrete feruloyl esterases, including a novel enzyme, TsFaeC, with a broad substrate specificity. *J Biotechnol* **108**, 227–241.
40 Shin HD and Chen RR (2007) A type B feruloyl esterase from *Aspergillus nidulans* with broad pH applicability. *Appl Microbiol Biotechnol* **73**, 1323–1330.
41 Palm GJ, Reisky L, Böttcher D, Müller H, Michels EAP, Walczk MC, Berndt L, Weiss MS, Bornscheur UT and Weber G (2019) Structure of the plastic-degrading *s* MHETase bound to a substrate. *Nat Commun* **10**, 1717.
42 Kretsinger RH, Uversky VN and Permyakov EA (Eds.) (2013) Encyclopedia of Metalloproteins. Springer, New Yor, NY.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. FAEs substrates.

Fig. S2. The secondary structure elements derived by *PROMOTIF* [35] as implemented in *PDBsum* [29].

Fig. S3. The secondary structure elements identified with *DSSP* [27,28] and the topology diagram of the structure as extracted using *PDBsum* (http://www.ebi.ac.uk) [29] on the EBI server.

Fig. S4. The packing of the *Fo*FaeC dimer in the unit cell.

Fig. S5. Output from *CheckMyMetal* server [23].

Fig. S6. ICP-SFMS analysis output from *Fo*FaeC (3.76 mg·mL^{-1} in 20mM Tris/HCl, pH 8.0).