Structural Basis for the Substrate Specificity of the Feruloyl Esterase Domain of the Cellulosomal Xylanase Z from *Clostridium thermocellum*†

Florian D. Schubot,‡ Irina A. Kataeva,‡ David L. Blum,‡,§ Ashit K. Shah,‡ Lars G. Ljungdahl,‡ John P. Rose,‡ and Bi-Cheng Wang*,‡

Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, Georgia 30602

Received July 3, 2001; Revised Manuscript Received August 21, 2001

ABSTRACT: Feruloyl esterases function in the cleavage of ferulic acid's bonds to arabinoxylan and pectin where the ferulic acid moieties cross-link the layers of polysaccharide chains within hemicellulose. This work presents the crystal structure of FAE_XynZ, the domain of *Clostridium thermocellum's* cellulosomal xylanase Z that displays feruloyl esterase activity. The structure was obtained via multiple isomorphous replacement with anomalous scattering (MIRAS) using three heavy atom derivatives and refined against X-ray diffraction data of up to 1.75 Å resolution. The *R*-value of the final model was 0.187 (R_{free} = 0.21). FAE_XynZ displays an eight-stranded α/*β*-fold with the characteristic "catalytic triad" at the heart of the active site. To define the substrate specificity determinants of the enzyme, the crystal structures of FAE_XynZ and the inactive FAE_XynZ(S172A) mutant were determined in complexes with the feruloyl-arabinoxylans FAXX and FAX_3, respectively. In the complex crystals, the ferulic acid moieties are clearly recognizable and allowed identification of the hydrophobic binding pocket. The carbohydrate part of both substrates is not visible in either structure. The location of the putative carbohydrate binding-pocket was inferred based on the location and orientation of the adjacent ferulic acid molecule. Five of the six residues lining the pocket were found to be conserved in FAE A from *Orpinomyces* sp., which further supports the proposed role of these amino acids.

For structural integrity and as protection from invading organisms, plants have evolved an intricate network of cross-linked polymers that constitute their cell walls. The pore sizes within these walls are too small for even viruses to penetrate (*1*). The main components of plant cell walls can be classified into four groups of polymeric compounds: cellulose, hemicellulose, pectin, and lignin. Hemicellulose in the form of arabinoxylan is the primary component of the cell walls in grasses of the graminae family and consists of *β*(1 → 4)-linked xylose chains that carry acetyl, methyl-glucuronyl, and arabinosyl substituents (*2*, *3*). Dimers of ferulic acid cross-link the xylan chains and also attach them to lignin (*4*, *5*).

A number of bacterial and fungal organisms that exploit plants as nutrient resources have evolved a large array of hydrolytic enzymes, enabling them to disintegrate and solubilize plant cell wall material. As plant cell walls consist of a large variety of cross-linked carbohydrates and other compounds such as ferulic acid, an equal diversity of enzymatic activities for their degradation is required. Invading organisms secrete these cellulolytic and other hydrolytic enzymes either to enter the plant cell or to directly utilize the cell wall material as nutritional resource (*6*). In the bacterium *Clostridium thermocellum* as well as a number of other anaerobic bacteria (*7*) and fungi (*8*, *9*), the secreted enzymes are assembled into large complexes of several millions in molecular weight termed cellulosomes (*10*, *11*). In *C. thermocellum*, the cellulosomal subunits are held together by the noncatalytic cellulosome integrating protein (CipA),[1] a large multidomain protein that consists of a single carbohydrate binding domain (CBD) and multiple "cohesin" domains. The cohesin regions interact with enzymatic subunit's "dockerin" domains that are organized in a modular arrangement with the actual enzymatic domains (*7*). In addition to these two domains, some of the enzymatically active subunits contain a second catalytic domain and other domains such as CBDs, immunoglobin-like domains of currently unknown function, other domains of unknown function, and linker regions rich in serine and threonine or proline residues. The XynZ subunit of the *C. thermocellum* cellulosome, which is the focus of the present work, consists of an N-terminal feruloyl esterase domain that is bridged to

† Support for this research was provided in part by funds to B.-C.W. from the University of Georgia Research Foundation, the Georgia Research Alliance, by the U.S. Department of Energy Grant # DE-FG02-93ER20127/A009 to L.G.L., and by Aureozyme Inc., Atlanta, GA.

* Corresponding author: Dr. Bi-Cheng Wang, Room B204A Life Sciences Building, Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, GA 30602. E-mail: wang@bcl1.bmb.uga.edu. Phone: (706) 542-1747. Fax: (706) 542-3077.

‡ Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, Georgia 30602

§ Current Address: Diversa Corporation, 4955 Directors Place, San Diego, California 92121

[1] Abbreviations: CBD = carbohydrate binding domain, CipA = cellulosome integrating polypeptide A, *C. thermocellum* = *Clostridium thermocellum*, FAE_XynZ = feruloyl esterase domain of XynZ, FAXX = *O*-{5-*O*-[(*E*)-feruloyl]-α-L-arabinofuranosyl}-(1 → 3)-*O*-*β*-D-xylopyranosyl-(1 → 4)-D-xylopyranose, FAX_3 = 5-*O*-[(*E*)-feruloyl]-[*O*-*β*-D-xylopyranosyl-(1 → 2)]-*O*-α-L-arabino- furanosyl-[1 → 3]-*O*-*β*-D-xylopyranosyl-(1→4)-D-xylopyranose, MIRAS = multiple isomorphous replacement with anomalous scattering, PCR = polymerase chain reaction, rmsd = root-mean-square deviation, XynY = Xylanase Y, XynZ = xylanase Z.

10.1021/bi011391c CCC: $20.00
Published on Web 09/28/2001

a xylanase domain via a proline-rich linker, a CBD, and a dockerin domain. The enzyme XynY, also from *C. thermocellum*, displays a similar modular organization featuring the same catalytic domains but in reversed order (*12*).

The xylanase domain of XynZ has been thoroughly studied, and its 1.4 Å crystal structure has been reported. The catalytic core was assigned to the glycoside hydrolase family 10 xylanases with a typical β-glycanase fold (*13*), which is characterized by an eight-stranded α/β-barrel. This fold is also called the TIM-barrel and forms the active site cleft at the carboxy-terminal end of the sheet. Enzymes of this family hydrolyze their substrate via a double displacement mechanism that involves two acidic residues and leads to a retention of the absolute configuration at the anomeric carbon atom (*14*). In the case of the XynZ's xylanase domain, Glu754 acts as the nucleophile and Glu645 serves as the acid/base catalyst.

The presence of the feruloyl esterase domain in XynZ (FAE_XynZ) was only recently discovered when a primary sequence analysis revealed more than 30% sequence identity of this region with feruloyl esterase FAE-A from strain PC-2 of the anaerobic fungus *Orpinomyces* sp. (*12*). A subsequent substrate specificity analysis of the cloned FAE_XynZ domain confirmed the predicted enzymatic function (*12*). Feruloyl esterases facilitate plant cell wall degradation by cleaving the cross-linking ester bond between the ferulic acid and the carbohydrate moieties. Primary structure analysis allowed the categorization of FAE_XynZ as an α/β-hydrolase with a Asp/His/Ser catalytic triad at its active site (*12*) and suggested that substrate cleavage is achieved by a well-studied mechanism that is commonly conserved among carboxylic acid hydrolases (*15*−*17*). As also observed for other feruloyl esterases, FAE_XynZ requires both the ferulic acid as well as the carbohydrate part of the substrate for optimal enzymatic activity (*6*, *12*). The protein is active over a pH range from 4 to 7, has an optimal reaction temperature of 50−60 °C, and is inhibited by ferulic acid but not by arabinose or xylanose (*12*).

Because of a lack of structural data for any member of the feruloyl esterase subclass of carboxylic acid hydrolases, little is known about the residues that provide substrate specificity to these enzymes. The present work reports the solution of the three-dimensional crystal structure of FAE_XynZ determined at 1.75 Å resolution. Furthermore, substrate complex structures of FAE_XynZ with FAXX **(***O*-{5-*O*-[(*E*)-feruloyl]-α-L-arabinofuranosyl}-(1 → 3)-*O*-β-D-xylopyranosyl-(1 → 4)-D-xylopyranose) at 2.0 Å as well as that of mutant FAE_XynZ (S172A) with FAX_3 **(**5-*O*-[(*E*)-feruloyl]-[*O*-β-D-xylopyranosyl-(1 → 2)]-*O*-α-L-arabino- furanosyl-[1 → 3]-*O*-β-D-xylopyranosyl-(1 → 4)-D-xylopyranose**)** at 1.8 Å resolution have been determined to explore the structure of the enzyme and the mechanism of substrate binding and to identify crucial residues beyond those directly involved in catalysis. The results support a conserved mechanism of hydrolysis as outlined in Figure 1.

MATERIALS AND METHODS

Purification of Native FAE_XynZ. The FAE_XynZ domain from *C. thermocellum* was expressed in *Escherichia coli* and purified as reported previously (*12*) by heat treatment at 70 °C for 30 min followed by gel filtration with a TSK3000SW column (TosoHaas, Montgomeryville, PA). The expressed protein released ferulic acid from FAXX and other substrates, demonstrating that it is enzymatically active when assayed by the method described in Borneman et al. (*2*). The substrates FAXX and FAX_3 were purified from wheat bran as described in Borneman et al. (*2*).

Site-Directed Mutagenesis. Serine in position 172 of FAE_XynZ was selectively changed to alanine using the following oligonucleotide primers:

5′-CGGGCGATTGCAGGACTT**GCA**ATGGGTG-GAGGACAATCG-3′ and 5′-CGATTGTCCTCCACCCAT-**TGC**AAGTCCTGCAATCGCCCG-3′. Plasmid pET-21b(+) containing the DNA fragment encoding the FAE_XynZ served as a template for PCR. PCR with mutagenesis primers was carried out using the QuickChange Site-Directed Mutagenesis kit (Stratagene Cloning Systems, La Jolla, Calif.). The PCR product was used to transform BL21(DE3) competent cells (Stratagene). Plasmid DNA was isolated and sequenced to confirm the mutation.

Purification of FAE_XynZ(S172A). FAE_XynZ(S172A) was expressed in *E. coli* and purified via heat treatment according to the same protocol as the native enzyme (*12*). Denatured protein was removed by centrifugation. The supernatant was precipitated with ammonium sulfate (80% saturation) on ice and dialyzed against a buffer containing 20 mM Tris-HCl buffer (pH 7.5) and 0.1 M NaCl. The concentrated sample was subjected to gel filtration chromatography on a Superdex-200 (Pharmacia Biotech. Inc., Cotati, Calif.) equilibrated with the same buffer. Eluted FAE_XynZ-(S172A) was concentrated and dialyzed against 50 mM Tris-HCl buffer (pH 8.0) containing 0.2 M NaCl and 5% glycerol. The purified FAE_XynZ(S172A) was assayed using FAXX and FAX_3 as substrates (*2*) and found inactive.

Crystallization. The crystallization of the FAE_XynZ domain has been described elsewhere (*18*). The native enzyme−FAXX complex was crystallized by the hanging-drop vapor-diffusion technique at 273 K using 3 μL drops. A setup contained 1 μL of protein solution (15 mg/mL) in 50 mM Tris-HCl buffer (pH 8.5) containing 0.2 M NaCl and 5% glycerol, 1 μL of 100 mg/mL FAXX, and 1 μL of the reservoir solution containing 30% (w/v) polyethylene glycol 8000, 0.2 M ammonium sulfate, and 0.1 M sodium cacodylate pH 6.5 (Hampton Research). Crystals of the FAE_XynZ(S172A) complex with FAX_3 were obtained from the same setup with the exception that the protein buffer contained 0.2 M sodium chloride, 5% glycerol, and 50 mM Tris-HCl, pH 8.0, and that 100 mg/mL FAX_3 was used instead of FAXX. In both cases, hexagonal crystals with dimensions of 50 × 50 × 100 μm where obtained.

Data Collection. Data sets for the native crystals, the FAX and FAX_3 complex crystals, and the K_2PtCl_4 derivative were collected on a Rigaku R-Axis IV image plate detector on loop-mounted (*20*) and flash-frozen (93K) crystals using 5kW Cu Kα radiation focused with Rigaku/MSC Blue confocal optics. The data were indexed, integrated, and scaled using HKL 1.9.1 (*19*). Data sets for the Xenon and $K_2Pt(CN)_4$ derivatives were collected at ID17 (IMCA-CAT), Advanced Photon Source, Argonne National Laboratory, on a Mar-Research 165 CCD detector on loop mounted (*20*) flash-cooled (100 K) crystals using X-ray wavelengths of 1.54 and 1.00 Å respectively. The data were indexed, integreted,

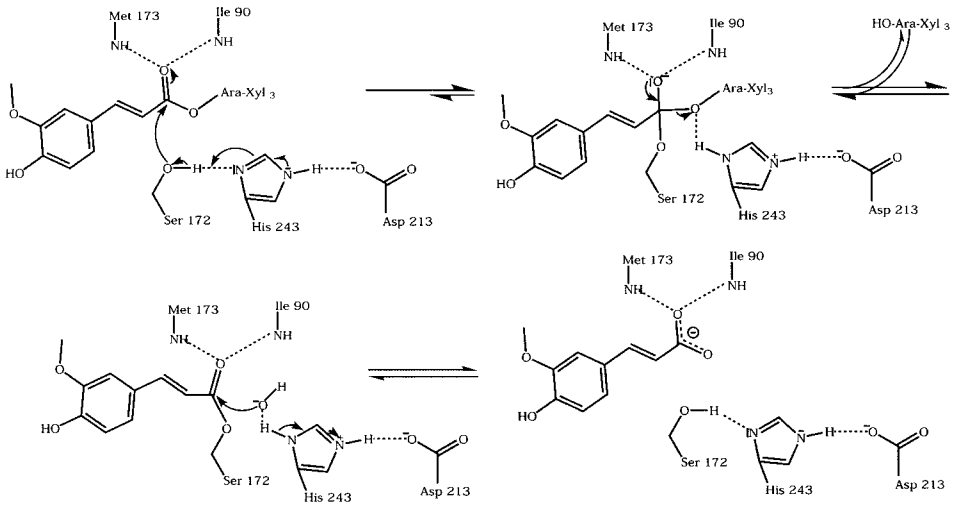

FIGURE 1: Mechanism of substrate hydrolysis for esterases featuring a "catalytic triad" (shown here for FAX_3). After substrate binding, the serine is activated by the histidine, which allows the nucleophilic attack of the substrate's carbonyl carbon atom. The resulting tetrahydral intermidiate is stabilized through interactions with two main chain NH groups in the "oxyanion hole". The histidine donates a proton to the substrate's forming alcohol component. This leads to the formation of a second covalent intermediate: the "acyl enzyme". An incoming water molecule is activated by the histidine residue, resulting in a nucleophilic attack on the carbonyl carbon and release of the substrate's acid component.

Table 1: Data Collection, Phasing, and Refinement Statistics for the Native FAE_XynZ Structure

crystal	native	Xe-derivative	$PtCl_4$-derivative	$Pt(CN)_4$-derivative
X-ray source	RU-200	17ID APS	RU-200	17ID APS
wavelength (Å)	1.54	1.54	1.54	1.00
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
cell dimensions	a=43.22 Å; *b*=63.32 Å; *c*=79.41 Å	a=43.22 Å; *b*=63.45 Å; *c*=79.51 Å	a=43.46 Å; *b*=63.50 Å; *c*=79.81 Å	a=43.29 Å; *b*=63.65 Å; *c*=79.78 Å
maximum resolution (Å)	2.3	2.4	2.6	1.75
oscillation range (°)	0.5	0.5	0.5	0.5
rotation range (°)	203	253	663	180
total reflections	222 837	222 936	578 231	260 585
unique reflections	9858	8168	6966	21940
completeness (%) [last shell]	91.8 [84.2]	90.4 [79.7]	96.3 [86.9]	95.3 [83.4]
I/σ [last shell]	21.7 [7.5]	27 [7.3]	27.6 [6.7]	23.3 [2.5]
R_{sym} [last shell]	0.065 [0.16]	0.082 [0.27]	0.10 [0.34]	0.07 [0.33]
phasing statistics				
no. of heavy atom sites		1	2	1
overall figure of merit (up to 3.0 Å)	0.7			
refinement statistics				
resolution (Å)	35−1.75			
reflections	21 940			
no. of atoms in model	2259			
no. of solvent molecules	293			
R-factor (%)	18.7			
free *R*-factor (%)	21.0			
RMS bond length (Å)	0.005			
RMS bond angle (°)	1.18			

and scaled using HKL 2000 (*19*). Details of the data collection and processing for all data sets are collected in Table 1.

Structure Solution and Refinement. For phasing, a Xenon derivative was prepared by incubating a FAE_XynZ crystal under high pressure (200 psi) Xenon for 10 min and flash-cooled in liquid Freon using the Molecular Structure Corporation CryoXe-Siter. The platinum derivatives were generated by soaking native crystals overnight in preequilibrated hanging drops containing 5 mM $K_2Pt(CN)_4$ and K_2PtCl_4, respectively. All three heavy atom derivatives were analyzed using SOLVE (*21*). The Xenon and $K_2Pt(CN)_4$ derivatives gave a single heavy atom site, while the K_2PtCl_4 derivative gave an additional site. Phasing statistics are provided in Table 1. After the heavy atom positions had been refined with MLPHARE (*22*), new phases were calculated. The improved MIRAS phases were then modified using the solvent flipping protocol implemented in SOLOMON (*23*). The resulting electron density map (Figure 2c) was of reasonable quality, exhibiting clear protein/solvent boundaries and recognizable features of secondary structure. The density modified map allowed the fitting (*24*) of a polyalanine chain for 150 of the 266 residues, which was carried out in the molecular modeling program O (*24*). After several rounds of phase combination using SIGMAA (*22*) and model building, the remaining residues and side chains could be fit into the electron density map. The model was refined using the simulated annealing (SA) protocol of X-PLOR (*25*) followed by manual adjustment of the model using SA omit maps. The data obtained for the $K_2Pt(CN)_4$ derivative at the

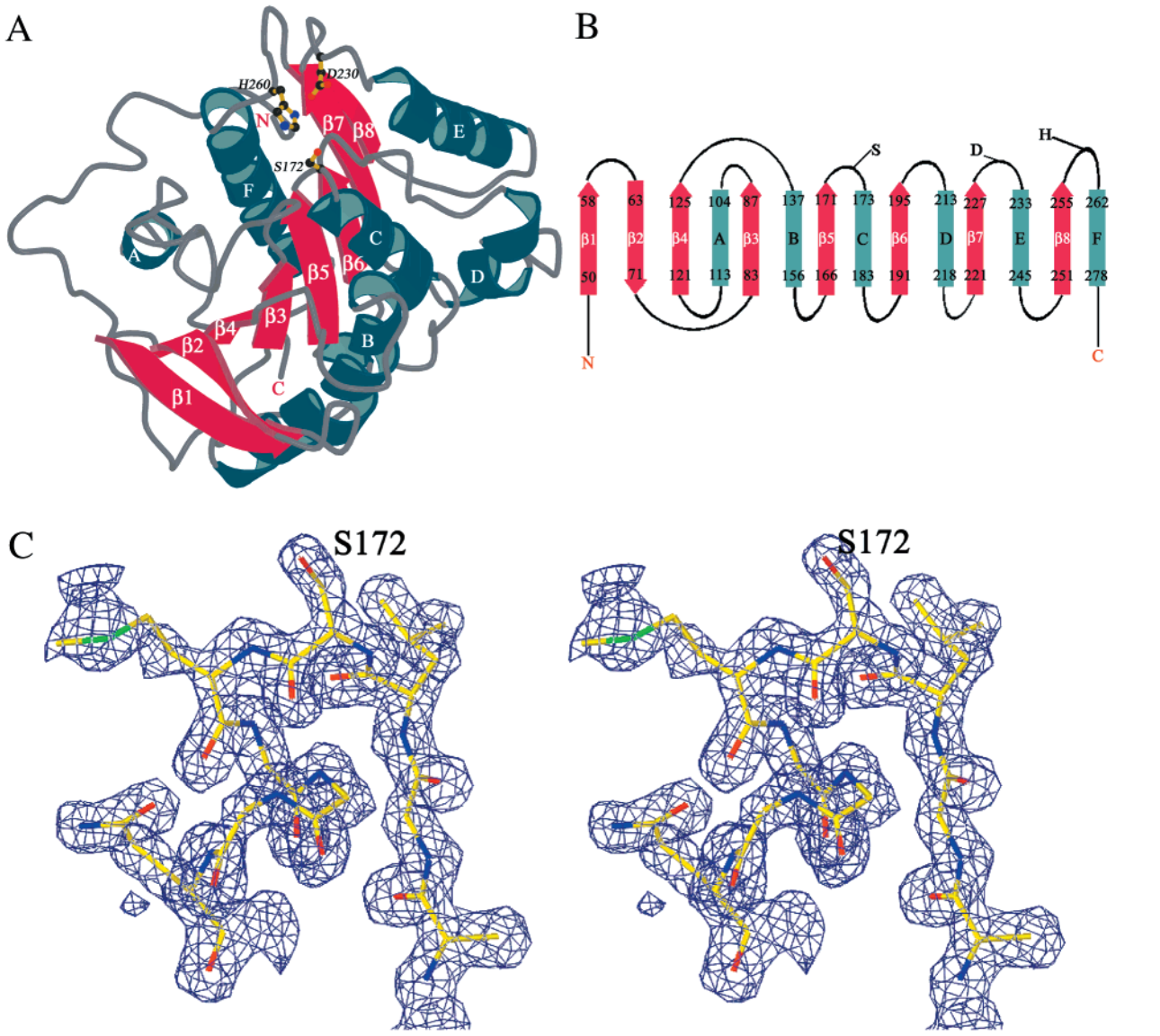

FIGURE 2: (A) Ribbon drawing of the FAE_XynZ backbone structure showing the α/β domain with the twisted sheet at the center. Also shown are the residues that constitute the "catalytic triad". The image was generated with MOLSCRIPT (*36*). (B) Topology plot of FAE_XynZ with the helices displayed as blue cylinders and the strands as pink arrows. The positions of the catalytic residues at the C-terminal end of the β-sheet are provided. (C) Stereoview of the "nucleophilic elbow" region together with the associated $2F_o - F_c$ electron density map at 1.75 Å resolution, using phases calculated from the final refined model and contoured at 1σ.

synchrotron beamline had the highest resolution at 1.75 Å and the best data statistics. Therefore, this data set was used for model refinement and solvent building. The final refinement statistics are given in Table 1. The structures of the FAE_XynZ−FAXX complex and that of FAE_XynZ-(S172A) with FAX_3 were solved via molecular replacement (MR) using the native enzyme structure as a search model and the MR routines available in X-PLOR (*25*). At this point, the space group was determined to be $P6_122$ rather than $P6_5$-22. Both models were refined with the SA protocol of X-PLOR (*25*) and manually adjusted in O. The final refinement statistics for all three models are given in Table 2. Coordinates have been submitted to the Protein Data Bank (*26*) (PDB code 1JJF).

RESULTS

Structure of the Feruloyl Esterase Domain of XynZ (FAE_XynZ). The crystal structure of the XynZ feruloyl esterase domain was determined at 1.75 Å resolution employing the MIRAS procedure after numerous failed attempts to obtain a structure solution via molecular replacement. The refined model had an *R*-value of 18.7% (R_{free} = 21%). The geometry of the final model (Figure 2a) has been evaluated with PROCHECK (*27*) and was found to be excellent, with 92% of the residues being located in the most favorable region of the Ramachandran Plot (*28*). Only Ser172 at the center of the "nucleophilic elbow" (Figure 2c) and Phe98 displayed a strained geometry. While the cloned construct contained amino acids 20 to 286 of XynZ enzyme, the N-terminal 10 residues as well as the two C-terminal residues could not be identified in the crystal structure. FAE_XynZ displays an α/β-topology (Figure 2b) with the central eight-stranded sheet assuming an open conformation and the helices being positioned on both sites (Figure 2a). The sheet carries a superhelical twist, creating approximately a 90° angle between the first and the last strand of the protein. The observed fold allows classification of FAE_XynZ as a member of the α/β-hydrolase super family of enzymes as

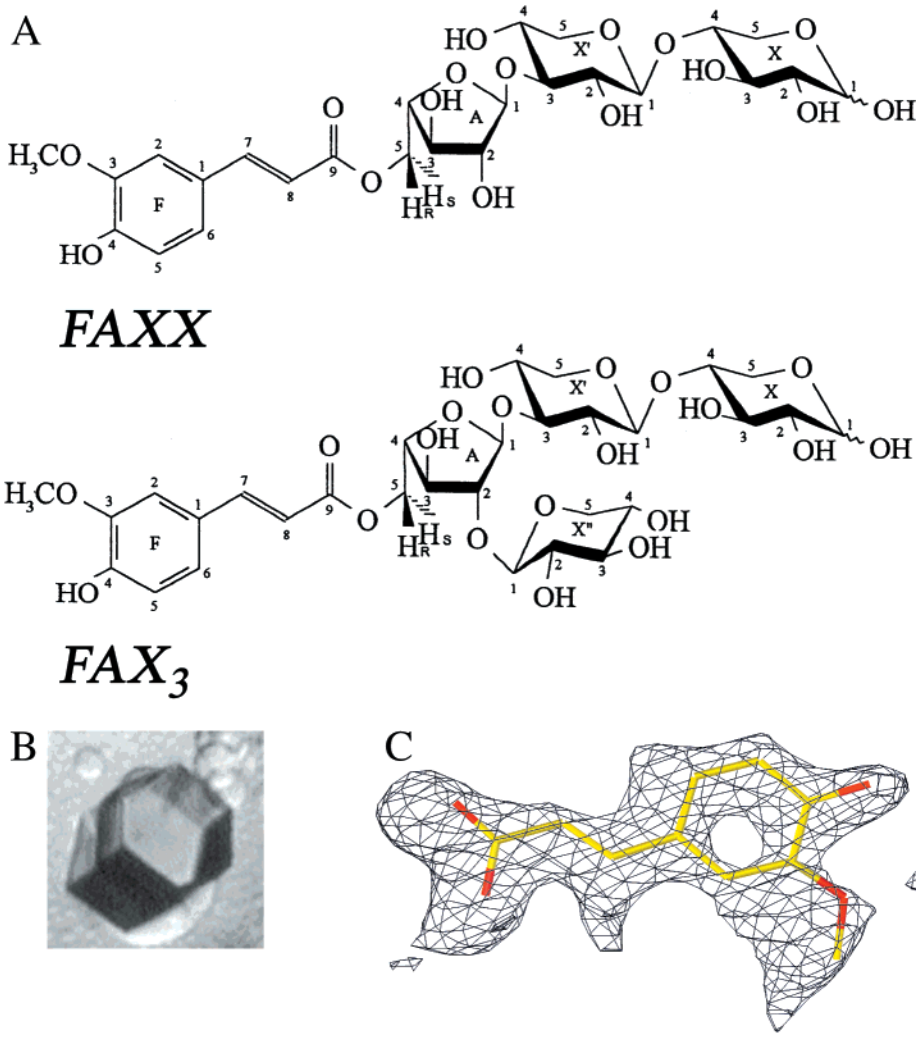

FIGURE 3: (A) Schematics of the substrate molecules FAXX and FAX_3. (B) Picture of the FAE_XynZ−FAXX crystal measuring 50 × 50 × 100 μm. (C) Electron density for ferulic acid in the FAE_XynZ−FAXX complex obtained from a $F_o - F_c$ map at a 2σ contour level and 2.0 Å resolution. The associated phases were calculated from the refined protein model that had been derived via molecular replacement.

Table 2: Data Collection, Phasing, and Refinement Statistics for the Native FAE_XynZ Structure

crystal	FAE-FAXX	FAE-(S172A)-FAX_3
X-ray source	RU-200	RU-200
wavelength (Å)	1.54	1.54
space group	$P6_122$	$P6_122$
cell dimensions	$a = b = 64.74$ Å; $c = 222.36$ Å	$a = b = 64.53$ Å; $c = 221.82$ Å
maximum resolution (Å)	2.0	1.8
oscillation range (°)	0.5	0.5
rotation range (°)	188	152
total reflections	534,076	348,106
unique reflections	18679	24740
completeness (%) [last shell]	94.8 [70.4]	91.6 [74.4]
I/σ [last shell]	16.4 [2.7]	13.2 [2.7]
R_{sym} [last shell]	0.11 [0.32]	0.07 [0.3]
refinement statistics		
resolution (Å)	20−2.0	20−1.8
reflections	18836	24740
no. of atoms in model	2280	2291
no. of solvent molecules	301	309
R-factor (%)	19.11	18.7
free *R*-factor (%)	23.7	21.4
RMS bond length (Å)	0.005	0.006
RMS bond angle (°)	1.2	1.2

had already been suspected based on the primary structure of this domain (*12*).

Crystal Structure of the FAE_XynZ−FAXX Complex. To explore the structural basis for the observed substrate specificity, the feruloyl esterase was cocrystallized with its natural substrate FAXX (Figure 3a). As the enzyme's optimal reaction temperature lies between 50 and 60 °C, the experiment was set up at 4 °C to prevent substrate hydrolysis. Hexagonal crystals were obtained after 1 day and used for X-ray diffraction analysis. Utilizing the native model as the search template, the substrate−complex structure was determined via molecular replacement and refined to 2.0 Å resolution (Table 2). The overall geometry of the refined model is even better than that of the free enzyme, with 93% of the residues lying in the most favorable region of the Ramachandran Plot (*28*). Again, the ten N-terminal and two C-terminal residues could not be observed in the electron density map. The conformation of the already discussed Ser172 has slightly changed and it is now positioned in the normally disallowed region of the plot. The molecular arrangement of the enzyme in this crystal is quite different as compared to that of the free FAE_XynZ molecules, which crystallized in the space group $P2_12_12_1$ and had a solvent content of 33%. The substrate complex crystals, on the other hand, belong to hexagonal space group $P6_122$ and have a significantly larger solvent content at 45% (Figure 3b). These shifts in the molecular crystal packing can be attributed to the formation of a biochemically insignificant 3.02 Å hydrogen-bond between O4 of the ferulic acid moiety that

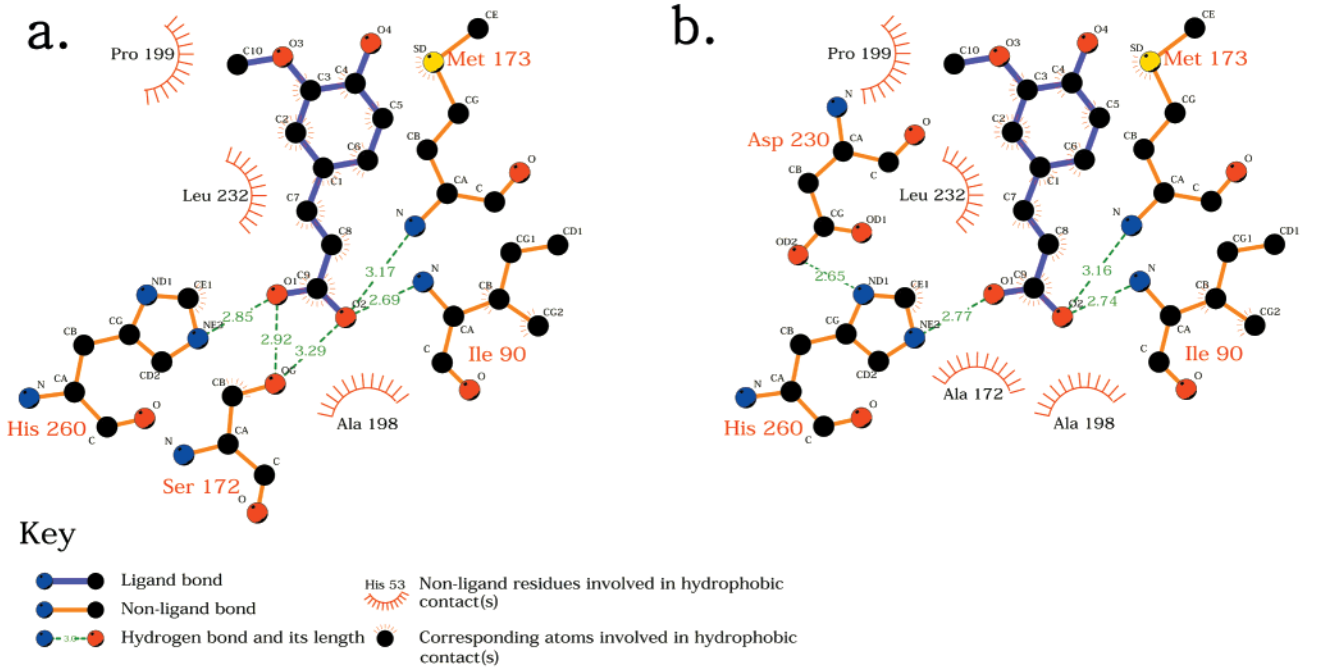

FIGURE 4: (a) LIGPLOT (*37*) drawing showing the hydrogen bonds formed as well as the hydrophobic interactions of the ferulic acid moiety with FAE_XynZ in the FAXX complex. The carboxyl group of the ligand forms the anticipated hydrogen bonds with the NH-groups of Met173 and Ile90 as well as with Ser172 and His260. The ferulic acid binding site consists of exclusively hydrophobic residues. (b) LIGPLOT drawing of the hydrophobic and hydrophilic protein−substrate interactions in the FAE_XynZ(S172A)−FAX_3 complex.

is bound in the active site of one enzyme molecule and the NH group of Asn53 in a neighboring molecule. The interpretation of the substrate's electron density was only possible for the ferulic acid part (Figure 3c), while little density was visible in the region where the position of the carbohydrate part had been expected. The overall structure of the FAE_XynZ−FAXX complex differs little from that of the free protein. The root-mean-square deviation (rmsd) for the main chain residues was calculated at 0.44 Å. The largest deviation between the two structures was observed in the loop region spanning residues 112 to 119. In this region, which is only poorly defined in the free protein structure, rmsd rose to 0.7 Å for the main chain atoms. All other structural differences primarily involve amino acid side chains and will be discussed below if they are the result of substrate binding interactions.

Structure of the Mutant FAE_XynZ(S172A) Complex with FAX_3. The native enzyme−FAXX structure gave a noninterpretable electron density in the projected location of arabinosyl moiety within the active site. Assuming that this was caused by residual activity of the enzyme even at 4 °C, the inactive FAE_XynZ(S172A) mutant was constructed and cocrystallized with substrates FAXX and FAX_3. FAXX and FAX_3 are equally good substrates (*12*), and while the FAE_XynZ(S172A)−FAXX crystals obtained from the slightly modified crystallization conditions were too small for X-ray diffraction studies the larger FAE_XynZ(S172A)−FAX_3 crystals gave X-ray diffraction up to 1.8 Å resolution (Table 2). These crystals belonged to the same space group as those of the native protein−substrate complex, and the refined model derived from these data deviated only slightly from that of the native protein substrate complex. As in the latter case, the ferulic acid moiety was readily visible in the

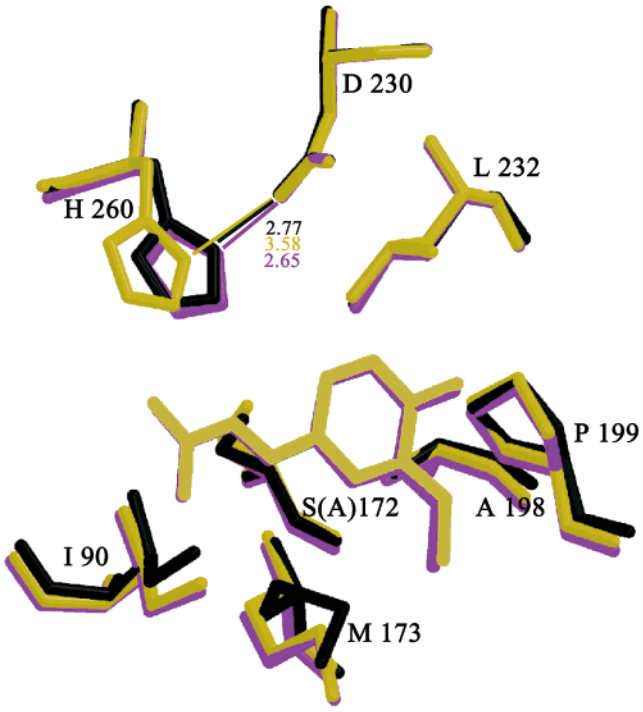

FIGURE 5: Superposition of those residues that are involved in catalysis and binding of the ferulic acid part of the substrate in FAE_XynZ (black), FAE-XynZ−FAXX (gold), and FAE_XynZ-(S172A)−FAX_3 (purple). Overall, there is little movement in the protein when accommodating the substrate. The shift in the position of His260's side chain in the FAE_XynZ−FAXX structure is likely due to the high pH used in the crystallization experiment (see text). The apparent movement in the side chain of Met 173 is due to the poor definition of the methionine's side chain in the native structure caused by the binding of the platinum ion. The figure was generated with MOLSCRIPT (*36*)

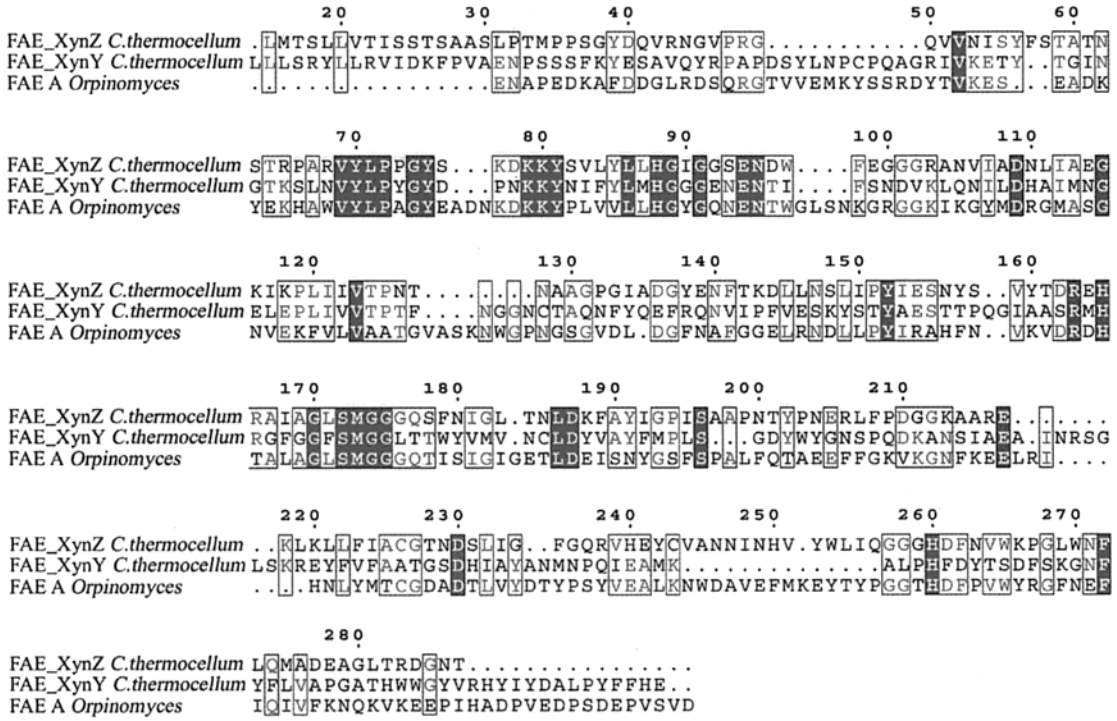

FIGURE 6: ESPript (*38*) sequence alignment of FAE_XynZ with FAE_XynY from *C. thermocellum* and FAE A from *Orpinomyces* sp. Even though FAE_XynY originates from the same organism as FAE_XynZ, its sequence diverges significantly also in those residues, Asp96, Trp97, Leu171, Phe262, and Trp265 in FAE_XynZ, that form the putative carbohydrate binding site and that are conserved between FAE_XynZ and FAE A.

$F_o - F_c$ electron density map. However, again none of the carbohydrate moieties were clearly discernible from the map.

DISCUSSION

In the α/β-hydrolase superfamily, the catalytic site residues are arranged in a Ser/His/Asp "catalytic triad" located at the C-terminal end of a beta sheet (*29*). Even though different members of this group target a large variety of substrates, the general mechanism of hydrolysis, involving the serine nucleophile, an activating histidine, and a catalytic acid, is believed to be conserved (Figure 1). After activation through interaction with a neighboring histidine, the serine "attacks" the substrate carbonyl carbon, leading to the formation of a transient tetrahedral intermediate. Concurrently, a proton is transferred from the serine to the histidine. At this stage of the reaction, the negatively charged carbonyl oxygen atom ("oxyanion") is stabilized by interactions with two main chain NH-groups ("oxyanion hole"), while the positively charged histidine is stabilized by a hydrogen bond to the catalytic acid. In the next step, the formed alcohol is released from the substrate and the acid part forms an ester bond with the serine oxygen. This bond, in turn, is hydrolyzed in a two-step mechanism involving a water molecule, and the enzyme is returned to the starting point (*30*). The serine nucleophile is located at the center of a universally conserved pentapeptide with the consensus sequence G−X−S−X−G. This pentapeptide segment constitutes the so-called "nucleophilic elbow," an unusually tight turn with the serine sitting at the top (Figure 2c). Since the sequence homology among α/β hydrolases, despite the shared structural elements, is generally low, the nucleophilic elbow has become the fingerprint feature commonly used to identify proteins of this enzyme family based on their primary structure alone.

Comparison of FAE_XynZ with its closest structural homologue a carboxyesterase from *Pseudomonas fluorescens* (*31*) revealed that the "catalytic triad" is formed by Ser172, Asp230, and His260 (Figure 2a), while the "oxyanion hole" is created by the NH groups of residues Met173 and Ile90. Even though the two proteins share only 10% of their primary sequences, more than 80% of the carboxyesterase's 218 residues align structurally with residues in the feruloyl esterase at an rmsd of 3.0 Å. While the carboxyesterase from *P. fluorescens* displays a broad substrate specificity (*31*), FAE_XynZ is, as discussed earlier, very sensitive to substitutions of the ferulic acid or the arabinosyl moieties in the potential substrate, suggesting that the crucial interactions are formed with both parts during substrate binding (*12*). The selectivity with regard to the carbohydrate part of the substrate is shared by other feruloyl esterases. For instance, the extracellular feruloyl esterases FAE-I and FAE-II produced by the anaerobic fungus *Neocallimastix* readily cleave FAXX but display a 1000 times lower activity on ethylferulate (*32*). Similarly, FAE-III from *Aspergillus niger* did not only display a 100 times larger K_m for methylferulate as compared to that for FAXX (*33*) but also proved sensitive to the positional attachment of the feruloyl group to the arabinose moiety (*33*, *34*). FAE-III efficiently cleaved the C-2 linked substrate while it was unable to hydrolyze the C-5 linked molecule. The enzyme CinnAE from the same organism is less particular and efficiently cleaves both substrates (*35*).

Substrate Binding and Structural Basis for Substrate Specificity. The FAE_XynZ−FAXX complex structure allows the identification of the specific protein−substrate interactions. The substrate's electron density was obtained from inspection of a $F_o - F_c$ map at a contour level of 2σ.

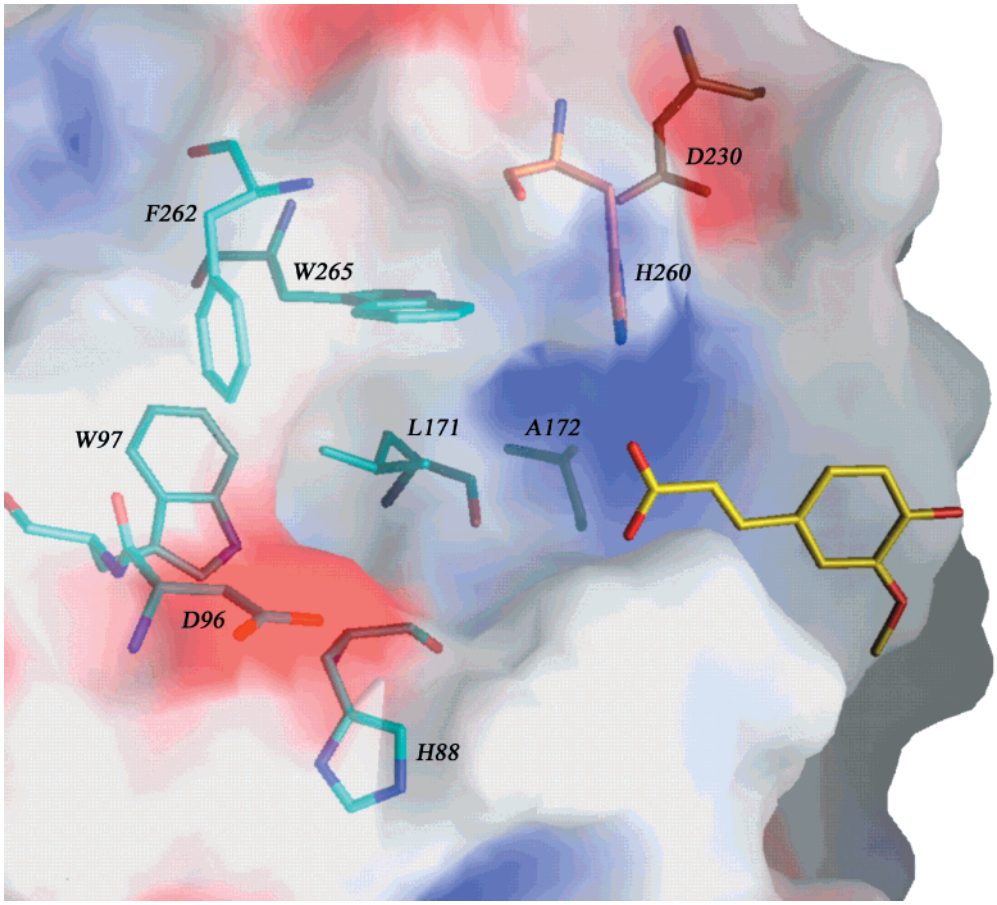

FIGURE 7: Electrostatic surface potential plot of FAE_XynZ's active site in the FAE_XynZ(S172A)−FAX_3 complex generated with GRASP (*39*). Regions with positive potential are blue, while regions of negative potential are red. Also displayed are the residues that form the catalytic triad and the potential carbohydrate binding site.

The density of the planar ferulic acid moiety was clearly visible in the active site and could be modeled with ease (Figure 3c). The remainder of the substrate was not recognizable. At the arabinosyl group's expected position, a circular arrangement of five peaks with heights all larger than 2σ could be observed in the $F_o - F_c$ map. However, the overall density was too weak to permit reasonable modeling of the arabinosyl unit, suggesting either that even at 4 °C hydrolysis had proceeded to a large degree or that protein−carbohydrate interactions are too weak to allow precise definition of the sugar molecules' conformation and position.

The aromatic acid part of the ligand forms the anticipated hydrogen bonds with nucleophile Ser172 and the catalytic base His260 as well as the NH-groups of residues Ile90 and Met173 that constitute the "oxyanion hole". The ferulic acid binding pocket in FAE_XynZ is created by the hydrophobic side chains of residues Ile90, Met173, Ala198, Pro199, and Leu232 (Figure 3). The aromatic ring of the ligand is stacked almost parallel with that of Pro199. The open and solvent-exposed ferulic acid binding site suggests that the enzyme belongs to the type-A ferulic acid esterases that are able to accommodate not only mono- but also diferulates in their active site (*35*). Figure 6 depicts a sequence alignment of FAE_XynZ with two feruloyl esterases that display enough sequence homology to FAE_XynZ to allow a direct comparison: FAE_XynY also from *C. thermocellum* and FAE A from the fungus *Orpinomyces* sp. While the first 200 residues of all three proteins match well, the remaining amino acids do not align as unambiguously. Here the positions of the amino acids that complete the catalytic triad in FAE_XynZ Asp230 and His260 were used to align the other amino acids in the two unknown structures with those in FAE_XynZ. Of the residues that form the ferulic acid binding pocket merely Met173 is identical and Leu232 is conserved in all three enzymes, suggesting a certain degree of flexibility in the mode of ferulic acid binding by the various enzymes. This observation could prove valuable in achieving the goal of engineering a more efficient feruloyl esterase.

Of note was also a shift in the side chain position of His260 in the FAE_XynZ−FAXX complex as compared to the free FAE_XynZ structure. While it was originally hydrogen-bonded to Ser172 as well as an oxygen atom of Asp230 in the complex structure, the N−O distance has risen to 3.5 Å. This is probably the result of the relatively high pH value of 8.5 at which these crystals were grown in an attempt to avoid substrate hydrolysis and resulting in the deprotonation of His260. This was addressed when crystallizing the FAE_XynZ(S172A)−FAX_3 complex by reducing the pH to 8.0, and the expected Asp230−His260 hydrogen bond was observed in that structure (Figure 5).

To ensure that no substrate degradation occurred in the mildly alkaline conditions, the drops from the crystallization setups for the FAE_XynZ(S172A)−FAX_3 complex were analyzed via HPLC. Only the characteristic peak for FAX_3 was observed. Therefore, the poorly defined electron density for the carbohydrate moiety of FAX_3 in the FAE_XynZ(S172A)−FAX_3 complex can be attributed to a presence of

only weak interactions between the protein and the carbohydrate part of the substrate. This finding is somewhat unexpected since substrate specificity studies for FAE_XynZ have shown it unable to efficiently cleave ethylferulate (*12*), and thus implying crucial interactions of the protein with the carbohydrate part of the ligand. Even though the carbohydrate moieties are not visible in the electron density map, the potentially involved amino acid can be extrapolated from the position of the ferulic acid group. Adjacent to the ferulic acid binding site, FAE_XynZ forms a hydrophobic pocket consisting of primarily aromatic residues (Figure 7). Of these residues His88, Trp97, Leu171, Phe262, and Trp265 are all strictly conserved between FAE_XynZ and FAE A from *Orpinomyces* sp. underlining their importance and suggesting similar binding interactions of both enzymes with the carbohydrate part of the substrate. FAE_XynY, on the other hand, although stemming from the same organism as FAE_XynZ, seems to possess a differently designed binding pocket that points toward a potentially distinct substrate specificity for this enzyme and making it an interesting target for future structural and specificity studies. Alternatively, the cause for the low enzymatic activity for ethylferulate could be that ethanol is a poor leaving group as compared to the carbohydrates in FAX_3 rather than specific protein−substrate interactions. However, this would not explain why different feruloyl esterases display such a range of activity on methyl- or ethylferulate (*6*, *12*).

In summary, the crystal structure of native FAE_XynZ and its substrate complexes have allowed the identification of the residues involved in substrate binding and hydrolysis. The "catalytic triad" is formed by Ser172, Asp230, and His260, while the ferulic acid moiety interacts with residues Ile90, Met173, Ala198, Pro199, and Leu232. Only the methionine and the leucine are conserved in FAE_XynY and *Orpinomyces* sp.'s FAE A. The putative binding site for the carbohydrate part of the substrate is formed by residues His88, Trp97, Leu171, Phe262, and Trp265, which are also conserved in FAE A from *Orpinomyces* sp.

ACKNOWLEDGMENT

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-ENG-38.

REFERENCES

1. Borneman, W. S., Akin, D. E., and VanEseltine, W. P. (1986) *Appl. Environ. Microbiol. 52*, 1331−9.
2. Borneman, W. S., Hartley, R. D., Himmelsbach, D. S., and Ljungdahl, L. G. (1990) *Anal. Biochem. 190*, 129−33.
3. Puls, J., and Schuseil, J. (1993) in *Hemicellulose and hemicellulases* (Michael P. Coughlan, G. P. H., Ed.) pp 1−28, Portland Press, London; Chapel Hill, NC.
4. Ralph, J., Grabber, J. H., and Hatfield, R. D. (1995) *Carbohydr. Res. 275*, 167−178.
5. Fry, S. C. (1982) *Biochem. J. 203*, 493−504.
6. Williamson, G., Faulds, C. B., and Kroon, P. A. (1998) *Biochem. Soc. Trans. 26*, 205−9.
7. Bèguin, P., and Lemaire, M. (1996) *Crit. Rev. Biochem. Mol. Biol. 31*, 201−36.
8. Li, X. L., Chen, H., and Ljungdahl, L. G. (1997) *Appl. Environ. Microbiol. 63*, 4721−8.
9. Fanutti, C., Ponyi, T., Black, G. W., Hazlewood, G. P., and Gilbert, H. J. (1995) *J. Biol. Chem. 270*, 29314−22.
10. Bayer, E. A., Shimon, L. J., Shoham, Y., and Lamed, R. (1998) *J. Struct. Biol. 124*, 221−34.
11. Lamed, R., Setter, E., and Bayer, E. A. (1983) *J. Bacteriol. 156*, 828−36.
12. Blum, D. L., Kataeva, I. A., Li, X. L., and Ljungdahl, L. G. (2000) *J. Bacteriol. 182*, 1346−51.
13. Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Bèguin, P., and Alzari, P. M. (1995) *Nat. Struct. Biol. 2*, 569−76.
14. Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Bèguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., and Withers, S. G. (1992) *J. Biol. Chem. 267*, 12559−61.
15. Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H. Y., and Poulos, T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3743−5.
16. Bachovchin, W. W. (1986) *Biochemistry 25*, 7751−9.
17. Kraut, J. (1977) *Annu. Rev. Biochem. 46*, 331−58.
18. Blum, D. L., Schubot, F. D., Ljungdahl, L. G., Rose, J. P., and Wang, B. C. (2000) *Acta Crystallogr. D Biol. Crystallogr. 56*, 1027−9.
19. Teng, T. Y. (1990) *J. Appl. Crystallogr. 23*, 387−391.
20. Otwinowski, Z. M., W. (1997) *Methods Enzymol. 276*, *307*−*326*.
21. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D55*, 849−861.
22. Laboratory, S. D. (1986), CCP4. A Suite of Programs for Protein Crystallography, Warrington, England.
23. Abrahams, J. P., and Leslie, A. G. (1996) *Acta Crystallogr. D 52*, 30−42.
24. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
25. Brunger, A. T., Karplus, M., and Petsko, G. A. (1989) *Acta Crystallogr. A45*, 50−61.
26. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzel, T. F., and Weng, J. (1987) in *Crystallographic Databases − Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., and Sievers, R., Eds.) pp 107−132, Data Commission of the International Union of Crystallography, Bonn.
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
28. Ramakrishnan, C., and Ramachandran, G. N. (1965) *Biophys. J. **5***, 909−933.
29. Schrag, J. D., and Cygler, M. (1997) *Methods Enzymol. 284*, 85−107.
30. Jaeger, K. E., Dijkstra, B. W., and Reetz, M. T. (1999) *Annu. Rev. Microbiol. 53*, 315−51.
31. Kim, K. K., Song, H. K., Shin, D. H., Hwang, K. Y., Choe, S., Yoo, O. J., and Suh, S. W. (1997) *Structure 5*, 1571−84.
32. Borneman, W. S., Ljungdahl, L. G., Hartley, R. D., and Akin, D. E. (1992) *Appl. Environ. Microbiol. 58*, 3762−6.
33. de Vries, R. P., Michelsen, B., Poulsen, C. H., Kroon, P. A., van den Heuvel, R. H., Faulds, C. B., Williamson, G., van den Hombergh, J. P., and Visser, J. (1997) *Appl. Environ. Microbiol. 63*, 4638−44.
34. Faulds, C. B., and Williamson, G. (1994) *Microbiology 140*, 779−787.
35. Kroon, P. A., Faulds, C. B., Brezillon, C., and Williamson, G. (1997) *Eur. J. Biochem. 248*, 245−51.
36. Esnouf, R. M. (1997) *J. Mol. Graph. Model. 15*, 132−4, 112−3.
37. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng. 8*, 127−34.
38. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics 15*, 305−8.
39. Nicholls, A. (1992), GRASP: Graphical Representation and Analysis of Surface Properties, Department of Biochemistry and Molecular Biophysics, Columbia University, New York.

BI011391C