Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases

Anne Mølgaard[1], Sakari Kauppinen[2] and Sine Larsen[1]*

Background: The complex polysaccharide rhamnogalacturonan constitutes a major part of the hairy region of pectin. It can have different types of carbohydrate sidechains attached to the rhamnose residues in the backbone of alternating rhamnose and galacturonic acid residues; the galacturonic acid residues can be methylated or acetylated. *Aspergillus aculeatus* produces enzymes that are able to perform a synergistic degradation of rhamnogalacturonan. The deacetylation of the backbone by rhamnogalacturonan acetylesterase (RGAE) is an essential prerequisite for the subsequent action of the enzymes that cleave the glycosidic bonds.

Results: The structure of RGAE has been determined at 1.55 Å resolution. RGAE folds into an α/β/α structure. The active site of RGAE is an open cleft containing a serine–histidine–aspartic acid catalytic triad. The position of the three residues relative to the central parallel β sheet and the lack of the nucleophilic elbow motif found in structures possessing the α/β hydrolase fold show that RGAE does not belong to the α/β hydrolase family.

Conclusions: Structural and sequence comparisons have revealed that, despite very low sequence similarities, RGAE is related to seven other proteins. They are all members of a new hydrolase family, the SGNH-hydrolase family, which includes the carbohydrate esterase family 12 as a distinct subfamily. The SGNH-hydrolase family is characterised by having four conserved blocks of residues, each with one completely conserved residue; serine, glycine, asparagine and histidine, respectively. Each of the four residues plays a role in the catalytic function.

Addresses: [1]Centre for Crystallographic Studies, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark and [2]Enzyme Research, Novo Nordisk A/S, Novo Allé, DK-2880 Bagsvaerd, Denmark.

*Corresponding author.
E-mail: sine@ccs.ki.ku.dk

Key words: carbohydrate esterase, catalytic triad, pectin, rhamnogalacturonan, SGNH-hydrolase

Received: **13 December 1999**
Revisions requested: **2 February 2000**
Revisions received: **9 February 2000**
Accepted: **16 February 2000**

Published: **23 March 2000**

Structure 2000, **8**:373–383

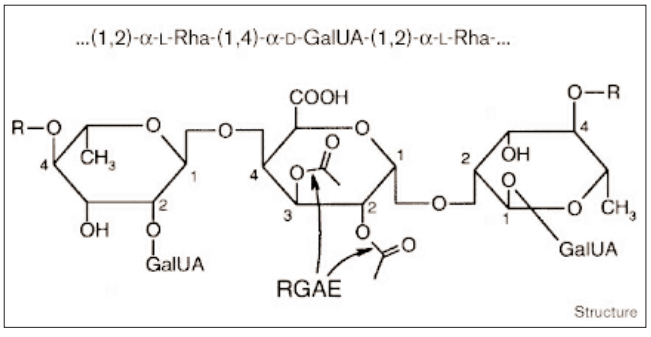
0969-2126/00/$ – see front matter

Introduction

Rhamnogalacturonan is one of the more complex polysaccharides in the pectic substances that constitute part of the middle lamella and the primary cell-wall of higher plants. The backbone of its main component, rhamnogalacturonan I (RG-I), is composed of alternating rhamnose (Rha) and galacturonic acid (GalUA) residues with the dimer repeat unit (1,2)-α-L-Rha-(1,4)-α-D-GalUA shown in Figure 1. RG-I is a heavily branched polysaccharide with carbohydrate sidechains such as arabinan, galactan and arabinogalactan attached to C4 of Rha. The GalUA residues can have methylated carboxylic acid groups and can be acetylated in the C2 and/or C3 position [1] as shown in Figure 1. The complexity of RG-I makes it understandable that a complex set of enzymes is required for its degradation. *Aspergillus aculeatus* can produce enzymes that work in synergy on the degradation of RG-I, each having their own specialized function [2–4].

From Figure 1 it is apparent that the acetylation of the GalUA residues creates a steric hindrance for the cleavage of the glycosyl linkages. One of the first steps in the degradation of RG-I must be the removal of the acetyl groups from the GalUA residues. Rhamnogalacturonan acetylesterase from *Aspergillus aculeatus* (RGAE) has been shown to possess this specificity [2–4]. Its action is

Figure 1



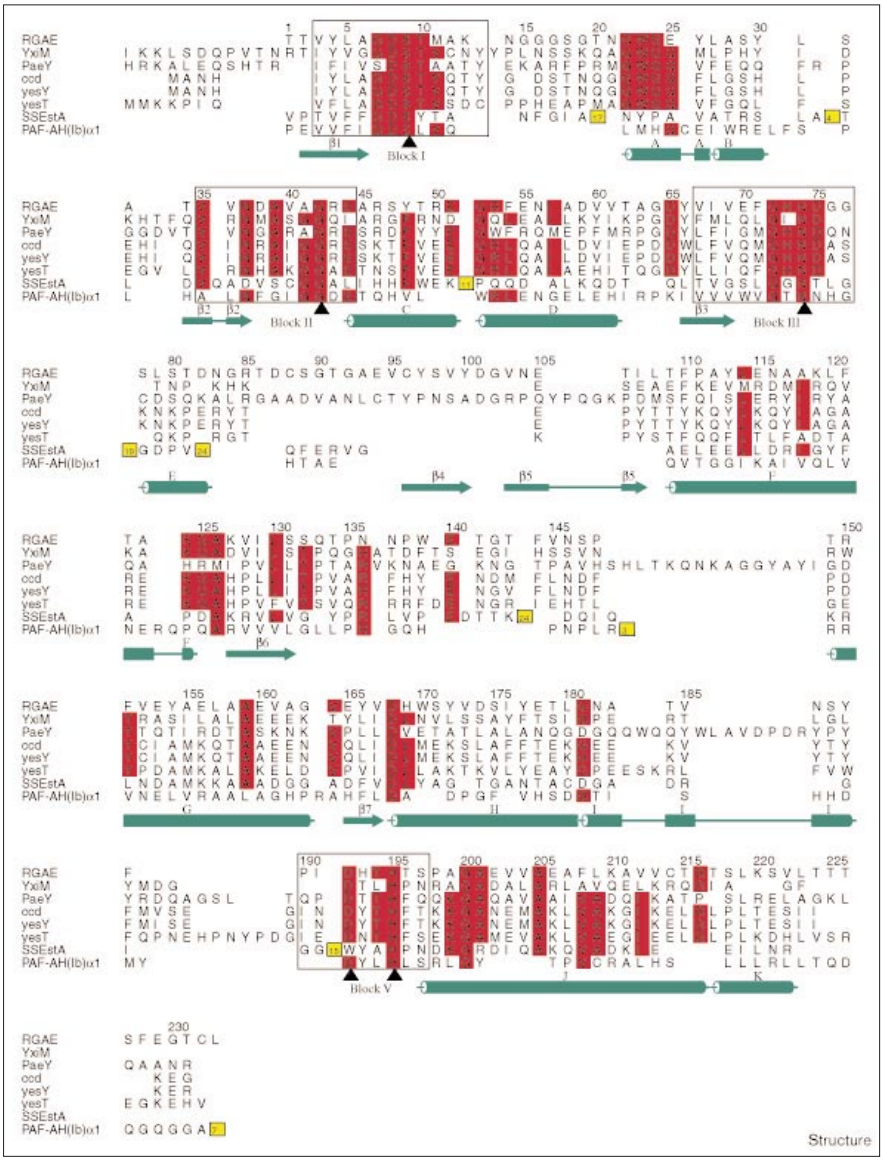
Schematic diagram of a trimer of the backbone of the RG-I substrate. The atoms that are referred to in the text are labeled, and the polysaccharide sidechains are indicated by R. The bonds that are cleaved by RGAE are indicated by arrows. The figure was generated using the program Isis Draw, version 2.1.3d, MDL Information Systems, Inc.

a prerequisite for the lyases and hydrolases acting on the backbone glycosyl bonds of RG-I. At present, the structure is known for one of these enzymes, rhamnogalacturonase A (RGaseA), which hydrolyses the α-(1,2) glycosyl linkages of the backbone [5].

RGAE has been cloned and expressed in *Aspergillus oryzae* [6]. The pH optimum of the enzyme is 6.0 and the theoretical pI is 4.46. Two potential *N*-glycosylation sites were identified from the primary sequence at residues Asn104 and Asn182. Accordingly, mass spectrometry has shown that the protein is heterogeneously glycosylated [6,7].

The mature polypeptide chain (233 amino acids) of RGAE shows no obvious sequence homology to proteins with known structure. Recently a bacterial pectin acetylesterase from *Erwinia chrysanthemi* (PaeY) was described that has a low sequence identity (25%) with RGAE. PaeY works in synergy with other pectinases, exhibiting an increased deacetylation rate when the substrate is demethylated prior to the action of PaeY and increasing the subsequent degradation of the substrate by pectate lyases [8]. The most conserved region between RGAE and PaeY comprises 42 residues and has 47% identity. Three proteins of unknown function from *Bacillus subtilis*, YxiM, yesY and yesT, and, recently, cephalosporin C deacetylase from *Bacillus* sp. *KCCM10143* (ccd) were also shown to be homologous to RGAE (Figure 2). These six proteins are classified as family 12 of the carbohydrate esterase families [9]. The structure presented here of

Figure 2



Sequence alignment of RGAE (SwissProt Q00017), YxiM (SwissProt P42304), PaeY (SwissProt O32563), ccd (GenBank AF184175), yesY (SwissProt O31528) and yesT (SwissProt O31523) performed using the program MULTALIGN [38,39]. The sequences of SsEstA (PDB accession number 1esc) and PAF-AH(Ib)α_1 (PDB accession number 1wab) have been aligned based on a structural comparison with RGAE performed by the program DALI [19,20]. In places, where residues have been omitted from the alignment, the number of omitted residues is noted in a yellow box. The sequence of RGAE corresponds to the mature protein without the N-terminal 17-residue signal peptide. The entire sequences of ccd, yesY and yesT are included. The sequence of YxiM starts at residue 167, the sequence of PaeY starts at residue 267, the sequence of SsEstA starts at residue 6 and the sequence of PAF-AH(Ib)α_1 starts at residue 39. Blank spaces are gaps introduced to improve the alignment. Residues conserved in five or more sequences are shown in red. The residue numbering above the alignment and the secondary structure assignments below the alignment correspond to RGAE. Regions of conserved blocks of residues are boxed and the catalytically important residues are indicated by black triangles. The figure was prepared using the program ALSCRIPT [40].

RGAE represents the first known three-dimensional structure in this family.

Dalrymple *et al.* [10] showed that the six proteins in family 12 represent a subfamily of a larger family of hydrolases that are characterized by having three blocks of conserved residues. Outside the blocks there is little or no sequence homology between the members of the larger family, which also contains two proteins with known structures, the serine esterase from *Streptomyces scabies* (SsEst) [11] and the α_1 subunit of the platelet-activating factor acetylhydrolase (PAF-AH(Ib)α_1) from *Bos taurus* (nomenclature as described by Ho *et al.* [12]).

Most esterases are found to adopt a common fold known as the α/β hydrolase fold [13] and contain a catalytic serine–histidine–aspartic acid triad. Proteins with low or no sequence identity and a variety of different functions belong to this structurally defined family which, besides the esterases, comprises lipases, peroxidases, proteases, dehalogenases and epoxide hydrolases. Examples of esterases that appear not to adopt the α/β hydrolase fold were encountered in the structures of SsEst and PAF-AH(Ib)α_1 [14]. Knowledge of the structure of RGAE and its relations to SsEST and PAF-AH(Ib)α_1 allows us not only to elucidate the function of RGAE, but also to characterize the structural relationship within the family of hydrolases identified by its blocks of conserved sequence.

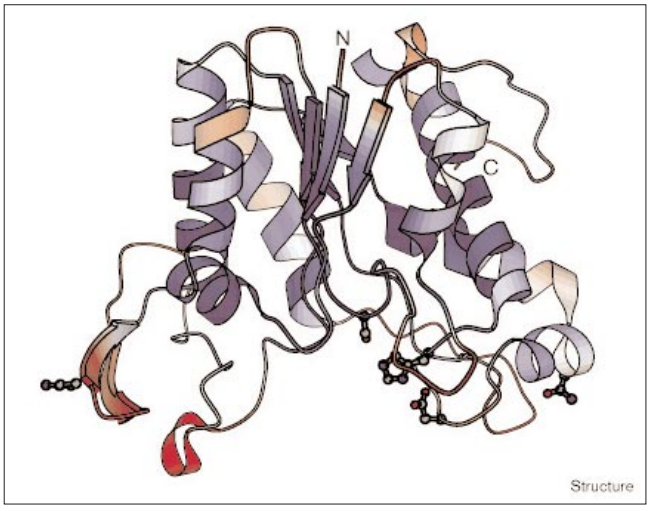
Results and discussion

The overall structure

The first structure of a RGAE is presented here in its orthorhombic crystal form, which could be obtained using either sulfate (Li^+ or NH_4^+) or polyethylene glycol (PEG) as the precipitant [7]. The structure has been determined using the multiple isomorphous replacement (MIR) method and refined using data to 1.55 Å resolution collected on the sulfate-grown crystals. These results provided the basis for the subsequent determination of the structure to a resolution of 1.9 Å for the same crystal form grown with PEG as precipitant. The different precipitants do not affect the structure of RGAE, the two structures superimpose with a root mean squared (rms) difference of 0.12 Å. Out of the 102 water molecules, 75 included in the final model of the structure in the PEG-grown crystals were found in identical positions in the sulfate structure. Because the structure determined for the sulfate-grown crystals is at a higher resolution, most of the subsequent discussion will refer to this structure.

The location of the secondary structural elements in the sequence, assigned using the program PROMOTIF [15], is indicated in Figure 2. The structure adopts a three-layer sandwich α/β/α fold, with a central five-stranded parallel β sheet surrounded by α helices, as shown in Figure 3. The topology of the central sheet is –1X 2X 1X 1X. An insertion of a two-stranded antiparallel β sheet after the third β strand in the central sheet makes the order of the β strands in the central sheet 2–1–3–6–7, as shown in the topology diagram in Figure 4a. Of the 11 helices in the structure, four are 3_{10} helices: helix A, B, E and K. The assignment of A and B as two helices is because of the lack of the Glu25–Ala28 and Tyr26–Ser29 backbone hydrogen bonds. The four cysteine residues in the sequence form two disulfide bridges, linking residues 88–96 and 214–232. The latter disulfide bridge anchors the C-terminal loop to α helix J. The electron-density maps do not show any evidence of *O*-glycosylation, but show that both potential *N*-glycosylation sites, Asn104 and Asn182, are glycosylated, allowing a total of seven carbohydrate residues to be fitted into the electron density.

Figure 3

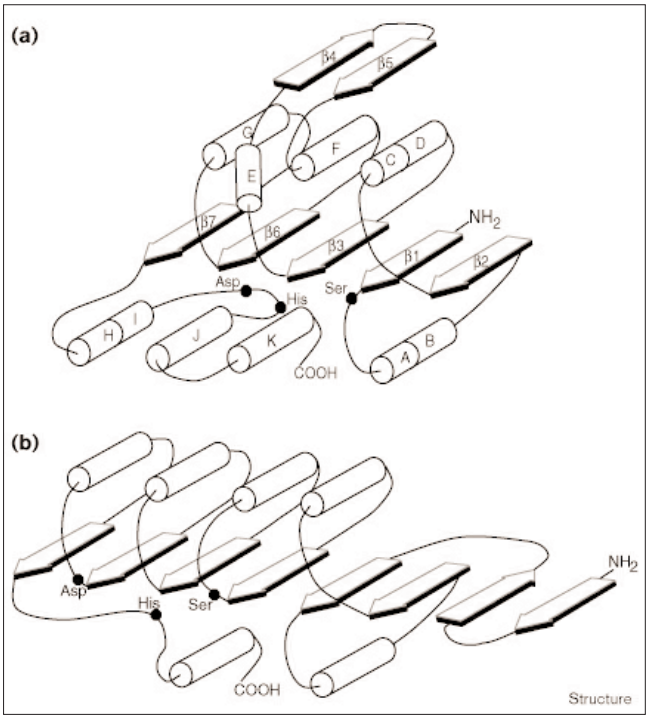


Schematic representation of the overall structure of RGAE. The secondary structure elements are colored according to B factor, with regions with B factors less than 10 Å^2 in deep blue and regions with B factors greater then 20 Å^2 in dark red. The two *N*-glycosylation sites and the catalytic triad residues are shown in ball-and-stick representation. The figure was prepared using the program Bobscript [41].

The active site

The location of the active site was inferred from the structure. Ser9, located at a topological switchpoint at the end of β strand 1, forms the characteristic hydrogen-bonding pattern of a catalytic triad with the nearby His195 and Asp192 (Figure 3). A sulfate ion, which is very well defined in the electron-density maps, is bound in the active site. One of its oxygen atoms occupies the oxyanion hole, making hydrogen bonds to the mainchain NH groups of Ser9 and Gly42 and the sidechain amide of Asn74 (Figure 5). These three residues are located in loop regions adjacent to the carboxy ends of the first three strands in the central β sheet. Another sulfate oxygen atom is hydrogen bonded to the imidazole ring of His195, and the sulfate ion thus connects four different regions in the protein structure. With its tetrahedral configuration,

Figure 4



Topology diagrams. The distribution of secondary structure elements in **(a)** RGAE and **(b)** a classical α/β hydrolase fold. The figure was prepared using the Microsoft program PowerPoint 97.

the sulfate ion mimics the transition state, suggesting that it might be a competitive inhibitor for the enzyme, although this has not been tested experimentally.

The active site in the structure of the PEG-grown crystals also contains residual density, which suggests that the enzyme has a high affinity for binding substrate analogs. It has not been possible to identify the density, which is rather diffuse. The strongest part of the density was therefore modeled as two water molecules.

The nucleophilic Ser9 residue is part of a type I β turn, and the two remaining members of the catalytic triad, Asp192 and His195, are in a loop region connecting two helices near the C terminus. The active site is at the bottom of an open cleft making it easily accessible to a large substrate. Several arginine residues are positioned in suitable positions to interact with the carboxylate groups of the negatively charged substrate (Figure 1).

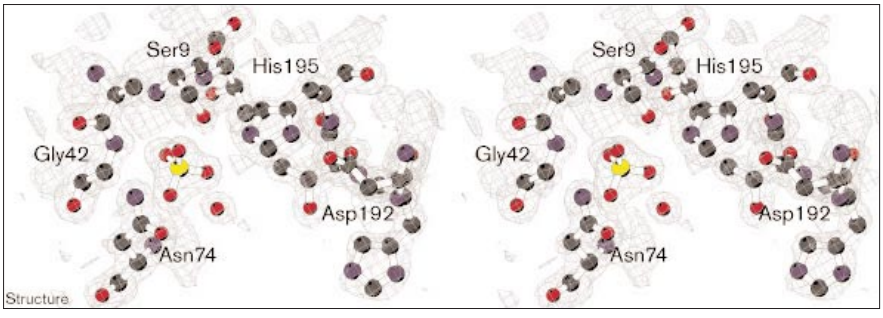
The fold is different from the α/β hydrolase fold

Most structures of neutral lipases and esterases have been found to adopt the common α/β hydrolase fold, which comprises a central, mostly parallel eight-stranded β sheet surrounded by α helices [13]. The sequential order of the catalytic triad residues in the α/β hydrolase family enzymes is invariably nucleophile–acid–histidine. The catalytic residues occur at similar places in the topology in most members of the family. The nucleophile is normally located at the end of β strand 5, the acid at the end of β strand 7 and the histidine in a loop region following strand 8 (Figure 4b). A characteristic motif of the α/β hydrolase fold is the so-called nucleophilic elbow [13,16], in which the nucleophile is located in a sharp turn between a β strand and an α helix, giving the backbone of the nucleophile a strained ε conformation [13,16]. The consensus sequence for this motif is Sm–X–Nu–X–Sm–Sm (Sm, small residue; X, any residue; Nu, nucleophilic residue).

The methionine in the Nu+2 position in RGAE violates this consensus sequence. The large sidechain of Met11 makes it sterically impossible to form the sharp turn that is necessary for the nucleophilic elbow motif. In RGAE the nucleophile is located on a type I β turn followed by a loop region, leaving the backbone conformation of the nucleophile completely unstrained. The highly conserved α helix C in the α/β hydrolase fold [17] is not conserved in the structure of RGAE.

The sequential order of the catalytic residues in RGAE is Ser–Asp–His, as in the α/β hydrolases, but whereas the last two residues in the α/β hydrolases are quite far from each other in the sequence, they are separated by only two residues in RGAE.

Figure 5



The active site in the structure of the Li_2SO_4-grown crystals showing the three catalytic residues, Ser9, His195 and Asp192. The $2F_o–F_c$ electron density is contoured at the 1σ level. A sulfate ion is bound with one of its oxygen atoms occupying the oxyanion hole making hydrogen bonds to backbone amides of Ser9, Gly42 and the sidechain nitrogen in Asn74. His195 of the catalytic triad is also within hydrogen-bonding distance of an oxygen atom in the sulfate ion. The figure was prepared using the program MOLSCRIPT [42].

Another prominent difference between the α/β hydrolase fold and the fold of RGAE is the orientation of the catalytic residues with respect to the remaining part of the structure. In the α/β hydrolase fold, the residues in the catalytic triad align parallel to the central β sheet, but in RGAE, the line of the catalytic triad is almost perpendicular to the central β sheet (Figure 6). This does not seem to have importance for the catalytic function of the enzymes, as the line of attack of the nucleophile on the carbonylic carbon is conserved, favoring attack on the *si* face of the substrate ester [18].

Comparison with related structures

A structural-homology search was performed using the program DALI, which is based on a distance criterion and does not take sequence information into account [19,20]. Using RGAE as search model, two structurally similar proteins were identified, namely SsEst [11] and PAF-AH(Ib)α_1 [12] (Figure 7).

The sequence identity between RGAE and SsEst after structural alignment is 17%, and the two structures are aligned with a positional rms deviation of 3.1 Å over 173 residues. For PAF-AH(Ib)α_1 the sequence identity to RGAE after structural alignment is 19% and the structures are superimposed with an rms deviation of 3.6 Å over 161 residues. The superposition of the three structures shown in Figure 7 reveals that their central parallel β sheet superimposes well with the catalytic residues lined up perpendicular to the central β sheet. It is noteworthy that SsEst and PAF-AH(1b)α_1, as well as RGAE, lack the nucleophilic elbow motif.

The most prominent differences between the fold of RGAE and SsEst encompass the long loop regions in SsEst that cover the active site (Figure 7), and the location of the two-stranded antiparallel β sheet, which in RGAE comes after the third β strand in the central parallel sheet, but in SsEst comes after the second β strand in the central parallel sheet. The structural alignment reveals that the location of the antiparallel β sheet in SsEst corresponds to the location of the kink that separates helices C and D in RGAE (Figure 7), suggesting that there might have been a similiar insertion in the ancestral versions of RGAE.

The structures of RGAE and PAF-AH(Ib)α_1 are more similar, and the size and overall shape of the molecules are very alike. The structure of PAF-AH(Ib)α_1 features an additional α helix prior to the first β strand, which is involved in the dimer-forming interface with the α_2 subunit, and lacks the antiparallel β sheet that is found in RGAE. The similarities between the overall folds of RGAE, SsEst and PAF-AH(Ib)α_1 leads us to conclude that the three enzymes belong to the same structural family.

The SGNH-hydrolase family

RGAE and SsEst have been previously classified as members of a new family of hydrolases characterized by its three blocks of conserved sequence [10], identified as the blocks I, III and V in the five blocks of conserved sequence used by Upton and Buckley to define a new family of lipolytic enzymes [21]. This new family of hydrolases contains proteins from a large variety of organisms, and we predict that they will have an overall fold that resembles the one described for RGAE, SsEst and PAF-AH(Ib)α_1. The carbohydrate esterase family 12 defined by Coutinho and Henrissat [9] constitutes a prominent subgroup of this new hydrolase family. We have conducted sequence alignments for its members RGAE, Yxim, PaeY, ccd, yesY and yesT. SsEst and

Figure 6

The orientation of the catalytic triad residues with respect to the overall fold. **(a)** Schematic diagram of the structure of dienelactone hydrolase (PDB accession number 1din) [43], an α/β hydrolase, viewed from the N-terminal side of the central β sheet. The three catalytic residues align roughly parallel to the central β sheet. **(b)** RGAE viewed from a similar orientation. The catalytic residues align almost perpendicular to the central β sheet. The figure was prepared using the program MOLSCRIPT [42].

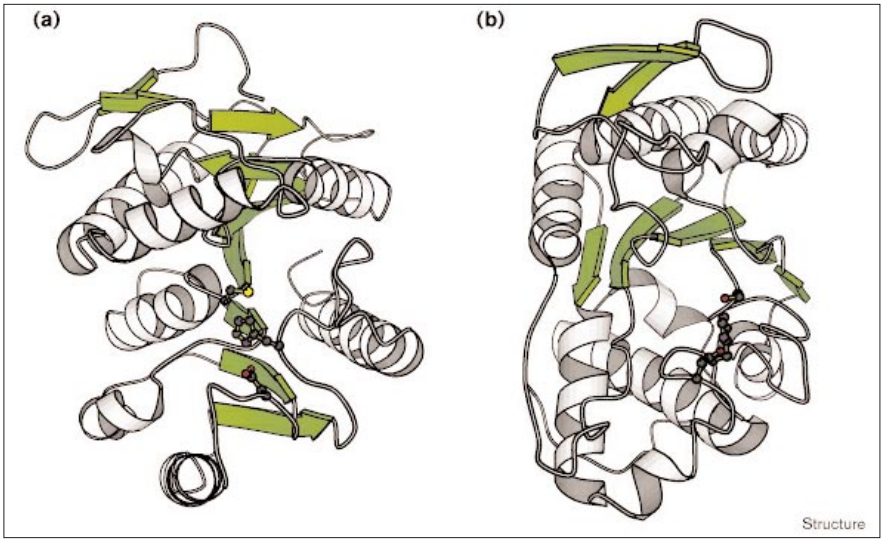
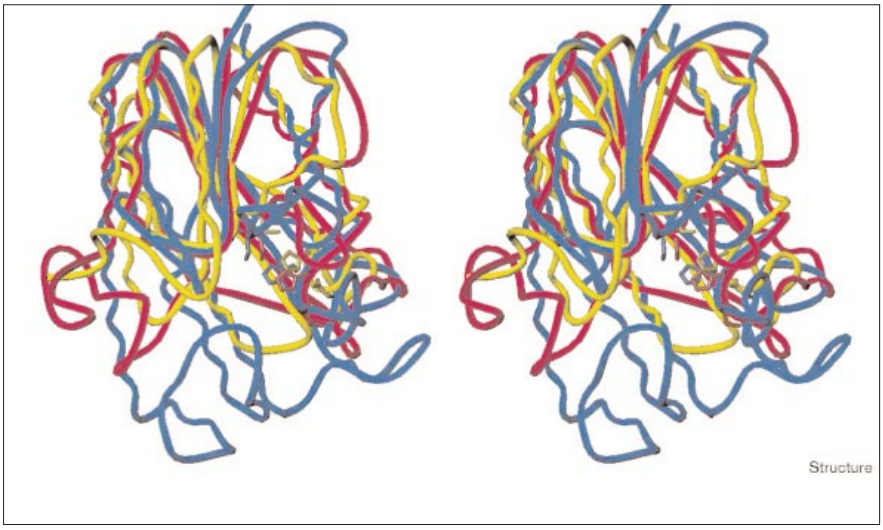


Figure 7



The structures of RGAE (pink), SsEst (blue) and PAF-AH(1b)α_1 (yellow) aligned using the orientation matrices from DALI. The catalytic residues are shown as sticks. The figure was prepared using the program TURBO [44].

PAF-AH(Ib)α_1 are not members of the carbohydrate esterase family 12, but belong to the block-defined hydrolase family. The structural alignments by DALI described above were employed for their alignment with the other six sequences (Figure 2). This sequence comparison revealed an additional block of conserved residues, which seems to correspond to block II in the family of lipolytic enzymes identified by Upton and Buckley [21]. The position of the blocks of conserved residues in the sequence are marked on Figure 2. When this alignment is compared with the larger alignments performed by Dalrymple [10] and Upton and Buckley [21], there is only one single residue that is completely conserved in each of the four blocks: serine, glycine, asparagine and histidine. Not surprisingly, these four residues are all catalytically important. Given that they must represent the utmost exponent for the relationships in this family we found it appropriate to name it the SGNH-hydrolase family.

Knowledge of the structure of RGAE allows us to explain the role of the blocks of conserved residues with respect to structure and function of the SGNH-hydrolase family. The four blocks are marked with different colors on the structure of RGAE in Figure 8a, which illustrates how they are concentrated at the centre of the structure and contain the active-site residues. In the following text we will analyse and describe the functional role of these four blocks in greater detail.

Block I contains the nucleophile and a hydrogen-bond donor to the oxyanion hole

The characteristic sequence motif in this block is GDS, the nucleophilic serine residue is completely conserved [10]. In RGAE, Ser9 is located in a type I β turn from Asp8 to Met11 (Figure 8b). The conformation of the loop is stabilized by hydrogen bonds between the carbonyl group of Nu–1 and backbone amide groups of Nu+1 and Nu+2, leaving only room for a small residue, such as glycine, as residue 7 (Nu–2). The Asp8 residue is conserved in all but one of the sequences comprising the family of hydrolases, where it is replaced with a glutamine residue. It appears that the main function of this aspartic acid is to stabilize the most flexible part of the β turn containing the catalytic serine residue by forming a hydrogen bond between OD1 and the Nu+1 backbone amide group, as shown in Figure 8b.

Interestingly, a corresponding hydrogen bond is found in the active site of the otherwise structurally unrelated serine protease subtilisin [22]. As in RGAE, the nucleophile in subtilisin is located on a type I β turn, but the hydrogen bond from the Nu+1 backbone amide group in subtilisin is accepted by the Nu–1 backbone carbonyl oxygen instead of an aspartic acid sidechain. The net effect is the same as in RGAE, stabilization of the flexible part of the β turn.

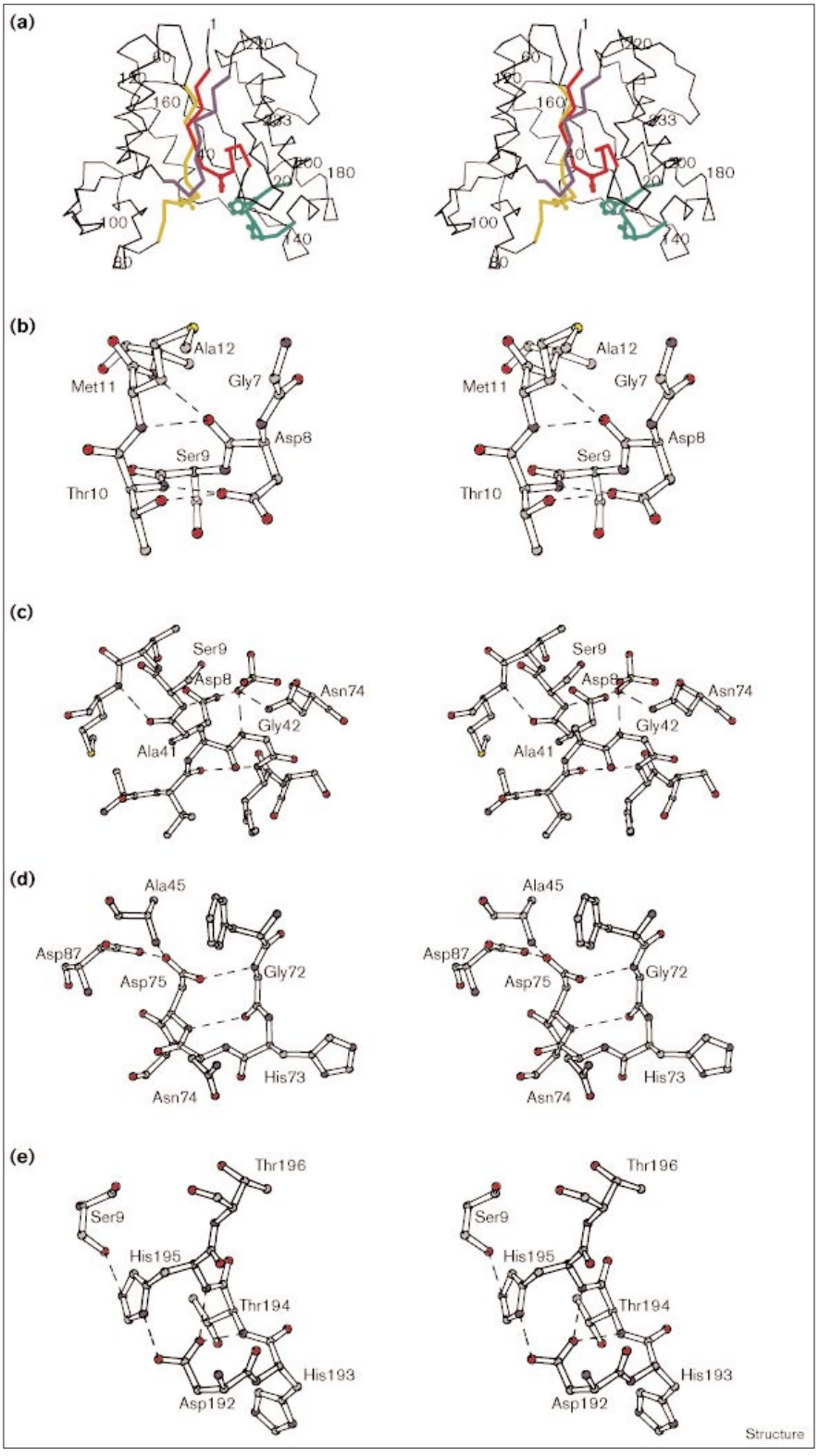
In the structure of PAF-AH(1b)α_1, acetate was bound in the active site, with the methyl moiety in a hydrophobic pocket, created by methyl groups of Leu48, Thr103 and Leu194 [12]. In a recent study, these residues have been mutated to smaller residues, and the role and significance of this specificity pocket has been confirmed [23]. A similar hydrophobic pocket is also found in RGAE, where Cγ2 in Thr10, Cβ in His73 and Cγ2 in Thr194 from blocks I, III and V delineate the pocket. SsEst is not an acetylesterase, and in accordance with its different substrate specificity, this hydrophobic pocket is lacking in the structure of SsEst.

Block II contains a hydrogen-bond donor to the oxyanion hole

This structurally conserved block was not included in the description of the hydrolase family by Dalrymple *et al.*

Figure 8

Position and role of the four blocks of conserved residues in the SGNH-hydrolase family. **(a)** The location of the conserved blocks of residues in the structure of RGAE: block I (red), block II (blue), block III (gold) and block V (green). The catalytically important residues are shown as ball-and-stick models. **(b–e)** Stereoview close-ups of blocks I, II, III and V, respectively. The figure was prepared using the program MOLSCRIPT [42].



The only residue conserved in the three structures in this block is a glycine, the backbone amide of which is engaged in a hydrogen bond to the atom in the oxyanion hole, as shown in Figure 8c. It is located in the R3 position

of a type II β turn. A survey of type II β turns has shown that for this position, glycine is preferred in order to avoid steric hindrance between the β carbon of R3 and the carbonyl oxygen of the previous residue [24].

Block III contains a hydrogen-bond donor to the oxyanion hole
One of the oxygen atoms of the sulfate ion that occupies the oxyanion hole is the acceptor for three hydrogen bonds each coming from different blocks of conserved sequence, as illustrated in Figure 8d. In addition to the mainchain amides of Ser9 (block I) and Gly42 (block II), it involves the sidechain amide group of Asn74. This residue is part of the characteristic motif of the conserved residues in block III, GXND [10], and is completely conserved in all the sequences in the hydrolase family. In all three known structures, it is placed on a type I β turn. The glycine and aspartic acid located at the –2 and +1 position relative to Asn74 are also highly conserved in the hydrolase family. In RGAE, Asp75 accepts a hydrogen bond from the mainchain amide of Gly72, thus stabilizing the β turn, and participates in hydrogen bonds to the sidechain carboxyl group of Asp87 and the mainchain amide of Ala45. It is buried within the structure and not exposed to the solvent.

Block V contains the catalytic aspartic acid and histidine
The residues that identify this block correspond to the sequence DXXHP [10]. The only totally conserved residue is His195, the second member of the catalytic triad. The last member, Asp192, is also located in block V. Whereas blocks I, II and III are found at the C-terminal ends of the central parallel β sheet, block V is located in a variable loop region. RGAE, SsEst and PAF-AH(Ib)α_1 share the last β strand in the central parallel β sheet and the final α helix, but differ in the loop regions that connect these structural elements.

Previous studies on the identification of the catalytic residues in this family of hydrolases have correctly identified the serine in block I and the histidine in block V as two of the three members of the catalytic triad [11,12,25]. In the structure of PAF-AH(Ib)α_1, the third member of the triad was identified as the aspartate in block V [12]. Before this structure was known, sequence comparison studies within the hydrolase family and kinetic studies of mutants of the lipase/acyltransferase from *Aeromonas hydrophila*, identified the aspartate in block III, and not the aspartate in block V, as the third member of the catalytic triad [25]. One of the arguments for excluding the aspartate in block V as the third member of the catalytic triad was that in all lipases of known structures at the time the three catalytic residues were located in separate loops far from each other in the sequence. The distantly spaced aspartic acid and histidine residues is a characteristic of the α/β hydrolase family, so the fact that those residues are close together in the sequences of the members of the SGNH-hydrolase family is one of the features that distinguishes the two families. Another argument for excluding the aspartic acid in block V was that one of the proteins in the family, *Vibrio parahaemolyticus* lipase, lacked the aspartate in block V, whereas the one in block III was conserved. In the study by Dalrymple *et al.* [10], more proteins have been included in the hydrolase family, and in this larger alignment neither of the two aspartates are completely conserved. ScEst was shown to be catalytically active without the third member of the triad [11], and the *Vibrio parahaemolyticus* lipase could be expected to posses a similar catalytic dyad. The kinetic studies of *Aeromonas hydrophila* lipase/acyltransferase mutants, however, showed both aspartates to be important for catalysis [25], and the structural analysis of the blocks of conserved sequence presented here can now explain these results.

Biological implications

The complex pectic polysaccharides are located predominantly in the middle lamella and primary cell wall of dicotyledonous plants [26]. The main backbone in pectins can be divided into linear homogalacturonan (smooth) regions of up to 200 residues of (1,4)-linked α-D-galacturonic acid (GalUA), and highly branched rhamnogalacturonan (hairy) regions consisting of repeating α-(1,2)-L-Rha-α-(1,4)-D-GalUA disaccharide units [26–28]. In general, about half of the rhamnose residues in the ramified regions are substituted with neutral oligosaccharides such as arabinans, galactans and arabinogalactans. Furthermore, most pectic substances are furthermore esterified with acetyl or methyl groups at some of the GalUA residues in the backbone [26,29]. Recently, several enzymes implicated in the degradation of pectic hairy regions, have been characterized [30]. Rhamnogalacturonan acetylesterase (RGAE) from *Aspergillus aculeatus* has a key role in the enzymatic breakdown of the ramified regions by specifically removing the acetyl groups from the rhamnogalacturonan backbone. Thus, by acting in synergy RGAE provides deacetylated substrate for further degradation by rhamnogalacturonan lyases and rhamnogalacturonan hydrolases [2–4].

The structure of RGAE presented here represents the first structure of a carbohydrate esterase family 12 enzyme. Comparisons with the structures of SsEst and PAF-AH(Ib)α_1 from the larger hydrolase family defined by Dalrymple *et al.* [10] allowed the characterization of their common fold. From sequence alignments of the hydrolase family and analysis of the three structures, we find that there are four blocks of conserved sequence, with one completely conserved residue in each of the four blocks: serine, glycine, asparagine and histidine, respectively. This led us to propose the name the SGNH-hydrolase family for this new hydrolase family, which contain

the carbohydrate esterase family 12 as a significant subfamily.

A number of characteristics of the common fold of the three known three-dimensional structures of SGNH-hydrolase enzymes can be recognized. The structures are α/β structures with a central five-stranded parallel β-sheet with the topology –1X 2X 1X 1X. They possess a catalytic serine positioned in a type I β turn at the carboxy end of β strand 1 and contain a glycine after β strand 2 and an asparagine after β strand 3 as hydrogen bond donors to the oxyanion hole. In the three structures, the histidine in the catalytic triad is located in a loop region after the last β strand in the central sheet, and in two of the structures the aspartate member of the catalytic triad is only three residues upstream from the histidine. SsEst was the first enzyme shown to possess a catalytic dyad, lacking the carboxylic acid member of the conventional triad [11]. In all three structures, the catalytic triad residues line up perpendicular to the central β sheet. Some of these features were also recognized in a recent review on the α/β hydrolase fold by Heikinheimo *et al*. [14], namely the overall topology as a five-stranded flavodoxin-like molecule, the location of the catalytic triad residues in the topology and the lack of the nucleophilic elbow motif.

Catalytic triads constitute the catalytic machinery in enzymes with diverse three-dimensional folds, such as the subtilisins, the serine proteases, the cysteine proteases, the human cytomegalovirus protease [31] and the α/β hydrolases. Of these folds, the α/β hydrolase fold is the one that most closely resembles the SGNH-hydrolase fold. The differences in topology and the location of the catalytic residues in the framework, however, suggest that they have evolved independently or that they diverged before developing their current catalytic mechanism.

Materials and methods

Crystallization and data collection

RGAE was expressed and purified as described previously [6]. Crystals used for structure determination were grown in 1.4 M Li_2SO_4, 0.1 M Na acetate pH 5.0 with a protein concentration of 40 OD_{280}. The crystals belong to the space group $P2_12_12_1$ with a = 52.14 Å, b = 56.87 Å and c = 71.89 Å and one molecule in the asymmetric unit. Their diffraction limit was 1.55 Å with the in-house equipment. A

Table 1

Data collection, phasing and refinement statistics.

	Native, Li_2SO_4	Native, PEG	Hg_1	Hg_2	Pt_1	Pt_2	Pt_3
Cell parameters (Å)	52.14, 56.87, 71.89	52.55, 57.08, 71.86					
Resolution range (Å)	30–1.55	30–1.9	30–2	30–2	30–2.45	30–2	30–2
Resolution range outermost shell (Å)	1.63–1.55	2.0–1.9	2.1–2.0	2.1–2.0	2.58–2.45	2.1–2.0	2.1–2.0
No. of unique reflections	28,693	15,662	15,038	14,308	7819	14,221	14,437
Multiplicity	7.7	4.1	10.2	6.9	6.7	6.7	6.8
Completeness overall (%)	91.0	89.7	100.0	96.6	95.9	95.4	96.7
Completeness outermost shell (%)	77.9	64.1	100.0	94.1	73.8	94.2	95.4
R_{merge}* overall (%)	3.5	6.2	6.1	5.6	4.4	4.1	3.4
R_{merge} outermost shell (%)	29.3	25.8	17.7	16.9	10.9	8.9	7.5
R_{iso}† (%)			6.8	9.2	7.2	8.6	10.8
Heavy atom concentration (mM)			0.5	0.5 + 1.25	3	5	7
Soaking time (days)			1	22 + 2	2	4	5
R_{cullis}‡ centric (%)			71	66	67	75	68
R_{cullis} acentric (%)			71	68	62	72	67
Phasing power§ centric			1.07	1.21	1.22	1.01	1.18
Phasing power acentric			1.59	1.76	1.95	1.51	1.78
R factor/R_{free}	0.164/0.200	0.162/0.218					
Rms deviations							
bond lengths (Å)	0.014	0.014					
angles (°)	1.60	1.62					
dihedrals (°)	24.3	24.5					
Number of non-hydrogen atoms							
protein	1735	1735					
carbohydrate, sulfate	86, 10	86, 0					
water	153	102					
⟨B⟩ (Å²)							
protein mainchain, sidechain	14.6, 16.5	18.8, 20.6					
carbohydrate, sulfate	35.6, 39.3	37.5, –					
water	30.1	28.7					

*$R_{merge} = \Sigma_{hkl} \Sigma_i | I(hkl)i - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_i I(hkl)i$, where i represents the number of batches. †$R_{iso} = \Sigma_{hkl} || F_{PH} | - | F_P || / \Sigma_{hkl} | F_P |$. ‡$R_{cullis} = \Sigma_{hkl} | | F_{PH} \pm F_P | - F_H(calc) | / \Sigma_{hkl} | F_{PH} - F_P |$. §Phasing $= \Sigma_{hkl} F_H / \Sigma_{hkl} | F_{PH}(obs) - F_{PH}(calc) |$.

native data set was collected for the orthorhombic crystal. Heavy-atom derivatives were prepared by soaking experiments, where the heavy-atom solution was diluted with reservoir solution prior to soaking. Two heavy-atom compounds gave isomorphous heavy-atom derivatives, parahydroxymercuri benzene sulfonate (PHMBS) and chloro(2,2′:6′,2″-terpyridine)platinum(II)chloride ('orange platinum'). The crystals were quite sensitive to PHMBS, and cracked when the heavy-atom solution was added to the drop. For this reason, low concentrations of PHMBS were used, and in one case the concentration was increased after the crystal had been exposed to a low concentration of PHMBS for 22 days. A total of five data sets were collected (Table 1). All derivative data sets and the native PEG data set were collected at room temperature on an in-house R axis II imaging plate system with a Rigaku rotating anode operating at 50 kV and 180 mA. Cu-K_α radiation was employed. The exposure time was 20 min and the oscillation range was 2° for all data sets. Integrated intensities were obtained using the HKL package [32] and subsequent data processing was performed using the CCP4 programs [33].

Structure determination and refinement

Both the Pt and the Hg derivatives contained two heavy-atom sites, which were refined using the CCP4 program MLPHARE [33]. After ten cycles of heavy-atom refinement using phases to a maximum resolution of 2.0 Å, the figure of merit was 0.53. The initial phases were improved using solvent flattening, histogram matching and Sayre's equation as implemented in the CCP4 program DM [33]. The solvent content was set to 35%, corresponding to one molecule of RGAE per asymmetric unit.

An electron-density map was calculated using the phases obtained by the density modification procedures, and a bones skeleton was constructed for the initial model building. Using the program O [34], the skeleton was manually modified until the entire polypeptide chain had been built. The initial electron-density map was very clear, and the model building proceeded without complications. The sulfate ion in the active site was easily identified in the initial map, but was not included in the starting model. A subset of the reflections from the native data set were kept out of the subsequent refinements and used to calculate the R_{free} [35], to avoid over-parameterization of the model. The working set hereafter comprised 25,800 reflections and the test set, which consisted of 10% of the reflections from the original native data set, chosen randomly from 30–1.55 Å, comprised 2847 reflections. The initial model was refined using X-PLOR [36], with a standard slow-cool protocol followed by restrained individual B-factor refinement. As the electron density improved after refinement, carbohydrate residues were included in the model at the two *N*-glycosylation sites, where there was evidence for it in both the $2F_o–F_c$ and the $F_o–F_c$ maps. Water molecules were included in the model, where good density was present in both the $2F_o–F_c$ and the $F_o–F_c$ maps, and where there was sensible hydrogen-bonding geometry. Bulk-solvent correction was performed using the protocol described by Jiang and Brünger [37], using all data from 30.0–1.55 Å. The final model includes all 233 amino acid residues, seven carbohydrate residues, two SO_4^{2-} ions and 153 water molecules. The final R factor is 0.164 and the R_{free} is 0.200. The mean B factors for protein, carbohydrate and solvent atoms are 15.4 Å^2, 35.6 Å^2 and 30.1 Å^2, respectively. The rms deviations from ideality are 0.014 Å for bond lengths, 1.60° for bond angles and 24.3° for dihedrals.

The model from the structure of the SO_4^{2-} grown crystals was used as a starting point for refinement of the PEG structure, not including carbohydrate residues and solvent molecules. The model was refined against the native PEG data using reflections from 30.0–1.9 Å. The refinement protocol was similar to the one used for the SO_4^{2-} structure. The final R factor is 0.162 and the R_{free} is 0.218. The mean B factors are 19.6 Å^2, 37.5 Å^2 and 28.7 Å^2 for protein, carbohydrate and solvent atoms, respectively. The rms deviations from ideality are 0.014 Å for bond lengths, 1.62° for bond angles and 24.5° for dihedrals.

Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 1deo and 1dex.

Acknowledgements

The authors are grateful for the funding from the Danish National Research Foundation which enabled this research.

References

1. Ishii, T. (1997). *O*-acetylated oligosaccharides from pectins of potato tuber cell walls. *Plant Physiol.* **113**, 1265-1272.
2. Schols, H.A., Geraeds, C.C.J.M., Searle-van Leeuwen, M.F., Kormelink, F.J.M. & Voragen, A.G.J. (1990). Rhamnogalacturonase: a novel enzyme that degrades the hairy regions of pectin. *Carbohydr. Res.* **206**, 105-115.
3. Kofod, L., *et al*., & Dalbøge, H. (1994). Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*. *J. Biol. Chem.* **269**, 29182-29189.
4. Searle-van Leeuwen, M.J.F., van den Broek, L.A.M., Schols, H.A., Beldman, G. & Voragen, A.G.J. (1992). Rhamnogalacturonan acetylesterase: a novel enzyme from *Aspergillus aculeatus*, specific for the deacetylation of hairy (ramified) regions of pectins. *Appl. Microbiol. Biotechnol.* **38**, 347-349.
5. Petersen, T., Kauppinen, S. & Larsen, S. (1997). The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed paralled β helix. *Structure* **5**, 533-544.
6. Kauppinen, S., Christgau, S., Kofod, L.V., Halkier, T. Dörreich, K. & Dalbøge, H. (1995). Molecular cloning and characterization of a rhamnogalacturonan acetylesterase from *Aspergillus aculeatus*. *J. Biol. Chem.* **270**, 27172-27178.
7. Mølgaard, A., *et al*., & Larsen, S. (1998). Crystallization and preliminary X-ray diffraction studies of the heterogeneously glycosylated enzyme rhamnogalacturonan acetylesterase from *Aspergillus aculeatus*. *Acta Crystallogr. D* **54**, 1026-1029.
8. Shevchik, V. & Hugouvieux-Cotte-Pattat, N. (1997). Identification of a bacterial pectin acetyl esterase in *Erwinia chrysanthemi*. *Mol. Microbiol.* **24**, 1285-1301.
9. Coutinho, P. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge.
10. Dalrymple, B., *et al*., & Lowry, J. (1997). Three *Neocallimastix patriciarum* esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. *Microbiology* **143**, 2605-2614.
11. Wei, Y., Schottel, J., Derewenda, U., Swenson, L., Patkar, S. & Derewenda, Z. (1995). A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nat. Struct. Biol.* **2**, 218-223.
12. Ho, Y., *et al*., & Derewenda, Z. (1997). Brain acetylhydrolase that inactivates platelet-activating factor is a γ-protein-like trimer. *Nature* **385**, 89-93.
13. Ollis, D., *et al*., & Goldman, A. (1992). The α/β hydrolase fold. *Protein Eng.* **5**, 197-211.
14. Heikinheimo, P., Goldman, A., Jeffries, C. & Ollis, D.L. (1999). Of barn owls and bankers: a lush variety of α/β hydrolases. *Structure* **7**, R141-R146.
15. Hutchinson, E.G. & Thornton, J.M. (1996). Promotif – a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212-220.
16. Derewenda, Z. & Derewenda, U. (1991). Relationships among serine hydrolases: evidence for a common structural motif in triacylglyceride lipases and esterases. *Biochem. Cell Biol.* **69**, 842-851.
17. Schrag, J. & Cygler, M. (1997). Lipases and the α/β hydrolase fold. *Methods Enzymol.* **284**, 85-107.
18. Derewenda, Z. & Wei, Y. (1995). Molecular mechanism of enantiorecognition by esterases. *J. Am. Chem. Soc.* **117**, 2104-2105.
19. Holm, L., Ouzounis, C., Sander, C., Tuparev, G. & Vriend, G. (1992). A database of protein structure families with common folding motifs. *Protein Sci.* **1**, 1691-1698.
20. Holm, L. & Sander, C. (1994). The fssp database of structurally aligned protein fold families. *Nucleic Acids Res.* **22**, 3600-3609.
21. Upton, C. & Buckley, J. (1995). A new family of lipolytic enzymes?

Trends Biochem. Sci. **20**, 178-179.
22. Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M. & Bott, R. (1998). The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry* **37**, 13446-13452.
23. Ho, Y.S., *et al*., & Derewenda, Z.S. (1999). Probing the substrate specificity of the intracellular brain platelet-activating factor acetylhydrolase. *Protein Eng*. **12**, 693-700.
24. Chou, P. & Fasman, G. (1977). β-Turns in proteins. *J. Mol. Biol.* **115**, 135-175.
25. Brumlik, M. & Buckley, J. (1996). Identification of the catalytic triad of the lipase/acyltransferase from *Aeromonas hydrophila. J. Bacteriol.* **178**, 2060-2064.
26. Carpita, N.C. & Gibeaut, D.M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J*. **3**, 1-30.
27. O'Neill, M., Albersheim, P. & Darvill, A. (1990). The pectic polysaccharides of primary cell walls. *Methods Plant Biochem*. **2**, 415-441.
28. Thibault, J.-F., Renard, C.M.G.C., Axelos, M.A.V., Roger, P. & Crepeau, M.-J. (1993). Studies of the length of homogalacturonic regions in pectins by acid hydrolosis. *Carbohydr. Res.* **238**, 271-286.
29. Schols, H.A., Posthumus, M.A. & Voragen, A.G.J. (1990). Structural features of hairy regions of pectins isolated from apple juice produced by the liquefaction process. *Carbohydr. Res.* **206**, 117-129.
30. Mutter, M., Colquhoun, I.J., Schols, H.A., Beldman, G. & Voragen, A.G. (1996). Rhamnogalacturonase B from *Aspergillus aculeatus* is a rhamnogalacturonan α-L-rhamnopyranosyl-(1-4)-α-D-galactopyranosyluronide lyase. *Plant Physiol*. **110**, 73-79.
31. Tong, L., Qian, C., Massariol, M.-J., Bonneau, P., Cordingley, M. & Lagace, L. (1996). A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature* **383**, 272-275.
32. Gewirth, D. *The HKL-Manual: An Oscillation Data Processing Suite for Macromolecular Crystallography.* Yale University, New Haven, CT.
33. CCP4. (1994). The ccp4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
34. Jones, T., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and location of the errors in these models. *Acta Crystallogr. A* **47**, 110-119.
35. Brünger, A. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
36. Brünger, A.T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
37. Jiang, J.-S. & Brünger, A. (1994). Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsion and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100-115.
38. Barton, G.J. & Sternberg, M.J.E. (1987). A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J. Mol. Biol.* **198**, 327-337.
39. Barton, G. (1990). Protein multiple sequence alignment and flexible pattern matching. *Methods Enzymol.* **183**, 403-428.
40. Barton, G.J. (1993). Alscript a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.
41. Esnouf, R. (1997). An extensively modified version of molscript that includes greatly enhanced coloring capacities. *J. Mol. Graph.* **15**, 132-134.
42. Kraulis, P. (1991). Molscript: a program to produce detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
43. Pathak, D. & Ollis, D. (1990). Refined structure of dienelactone hydrolase at 1.8 Å. *J. Mol. Biol*. **214**, 497-525.
44. TURBO-FRODO (1992). Biographics and AFMB (Architecture et fonction des Macromolecules Biologiques). Marseille, France.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.