Multifunctional Xylooligosaccharide/Cephalosporin C Deacetylase Revealed by the Hexameric Structure of the Bacillus subtilis Enzyme at 1.9Å Resolution

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · AUGUST 2003

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(03)00632-6 · Source: PubMed

11 AUTHORS, INCLUDING:

Florence Vincent
French National Centre for Scientific Research
22 PUBLICATIONS **794** CITATIONS
SEE PROFILE

David Jan Scott
University of Nottingham
93 PUBLICATIONS **1,581** CITATIONS
SEE PROFILE

Shirley M Roberts
The University of York
40 PUBLICATIONS **1,142** CITATIONS
SEE PROFILE

James A Brannigan
The University of York
124 PUBLICATIONS **4,045** CITATIONS
SEE PROFILE

doi:10.1016/S0022-2836(03)00632-6 *J. Mol. Biol.* (2003) **330**, 593–606

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Multifunctional Xylooligosaccharide/Cephalosporin C Deacetylase Revealed by the Hexameric Structure of the *Bacillus subtilis* Enzyme at 1.9 Å Resolution

Florence Vincent[1], Simon J. Charnock[1], Koen H. G. Verschueren[1] Johan P. Turkenburg[1], David J. Scott[1], Wendy A. Offen[1] Shirley Roberts[1], Gavin Pell[2], Harry J. Gilbert[2], Gideon J. Davies[1]* and James A. Brannigan[1]

[1]*Structural Biology Laboratory Department of Chemistry University of York, Heslington York YO10 5YW, UK*

[2]*School of Cell and Molecular Biosciences, The Agriculture Building, University of Newcastle upon Tyne Newcastle upon Tyne NE1 7RU, UK*

**Corresponding author*

Esterases and deacetylases active on carbohydrate ligands have been classified into 14 families based upon amino acid sequence similarities. Enzymes from carbohydrate esterase family seven (CE-7) are unusual in that they display activity towards both acetylated xylooligosaccharides and the antibiotic, cephalosporin C. The 1.9 Å structure of the multifunctional CE-7 esterase (hereinafter CAH) from *Bacillus subtilis* 168 reveals a classical α/β hydrolase fold encased within a 32 hexamer. This is the first example of a hexameric α/β hydrolase and is further evidence of the versatility of this particular fold, which is used in a wide variety of biological contexts. A narrow entrance tunnel leads to the centre of the molecule, where the six active-centre catalytic triads point towards the tunnel interior and thus are sequestered away from cytoplasmic contents. By analogy to self-compartmentalising proteases, the tunnel entrance may function to hinder access of large substrates to the poly-specific active centre. This would explain the observation that the enzyme is active on a variety of small, acetylated molecules. The structure of an active site mutant in complex with the reaction product, acetate, reveals details of the putative oxyanion binding site, and suggests that substrates bind predominantly through non-specific contacts with protein hydrophobic residues. Protein residues involved in catalysis are tethered by interactions with protein excursions from the canonical α/β hydrolase fold. These excursions also mediate quaternary structure maintenance, so it would appear that catalytic competence is only achieved on protein multimerisation. We suggest that the acetyl xylan esterase (EC 3.1.1.72) and cephalosporin C deacetylase (EC 3.1.1.41) enzymes of the CE-7 family represent a single class of proteins with a multifunctional deacetylase activity against a range of small substrates.

Keywords: α/β hydrolase; acetylxylan; carbohydrate esterase; cephalosporin; X-ray structure

F.V. and S.J.C. contributed equally to this work.

Present addresses: S. J. Charnock, Megazyme International Ireland Ltd, Bray Business Park, Bray, Co.Wicklow, Ireland; K. H. G. Verschueren, Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands; D. J. Scott, Department of Biochemistry, School of Medical Sciences, University Walk, Bristol, BS8 1TD, UK.

Abbreviations used: AXE, acetylxylan esterase; CAH, cephalosporin acetyl hydrolase; CE, carbohydrate esterase; MAD, multiple-wavelength anomalous dispersion; MPD, 2-methyl-2,4-pentanediol; ORF, open reading frame; 4-NPA, 4-nitrophenylacetate.

E-mail address of the corresponding author: davies@ysbl.york.ac.uk

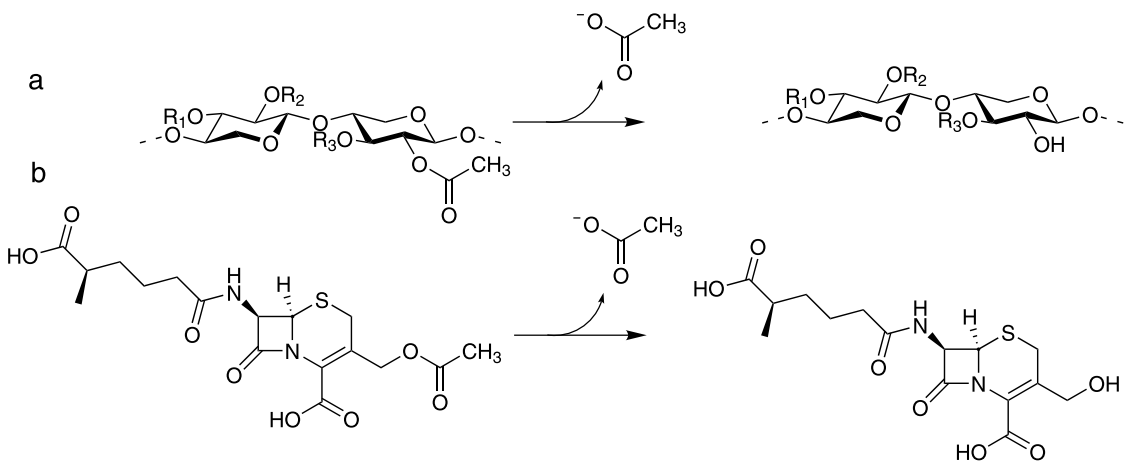

Figure 1. Reactions catalysed by family CE-7 esterases. (a) Deacetylation of short xylooligosaccharides: implicated in the hydrolysis of plant-cell-wall xylans.[71] (b) Deacetylation of cephalosporin C: a deacetylase specific for the acetyl group on position C10 is useful in the semi-synthesis of novel antibiotics.

Introduction

As with the classification of glycoside hydrolases,[1] the Carbohydrate-Active enZYmes Server (CAZy) provides a classification of carbohydrate esterases based upon amino acid sequence similarities†.[2] Currently, 14 such families have been described, ten of which contain enzymes involved in the degradation of plant-cell wall polysaccharides such as xylan, pectin and rhamnogalacturonan. Three-dimensional structures have been reported for ferulate esterases[3] from family CE-1, acetyl-xylan esterases[4–6] from family CE-5, acetyl rhamnogalacturonan esterases[7] from family CE-12 and trehalose transacetylases[8] from family CE-14. All utilise a serine-histidine-carboxylate catalytic triad with catalysis performed *via* the formation and subsequent breakdown of a covalent acyl-enzyme intermediate flanked by tetrahedral transition states.[9] Most are variations on the classic α/β hydrolase fold,[10] although the CE-12 enzyme presents the triad in a modified manner.[7] In addition, the family CE-8 pectin methylesterase displays an altogether different structure with a dual-aspartate catalytic centre located within a right-handed parallel β-helical structure.[11]

Twenty-two sequences are currently (May 2003) classified into CAZy carbohydrate-esterase family CE-7. These enzymes are notable for their unusual specificity for both acetylated xylan and cephalosporin C (Figure 1) and their subsequent industrial application for the chemoenzymatic deacylation of cephalosporins for antibiotic synthesis.[12] The CE-7 family contains some enzymes (EC 3.1.1.72) that are directly implicated in the deacetylation of xylan, including acetylxylan esterases (AXEs) from *Thermoanaerobacterium*[13,14] and *Thermotoga maritima*.[15] The family also includes cephalosporin acetyl hydrolase (CAH) proteins (EC 3.1.1.41) that are known to be deacetylases of the antibiotic cephalosporin C.[12] In contrast to other families of carbohydrate esterases, characterised CE-7 enzymes are known to display an unusually high oligomeric state, with values of four to five, six and eight reported.[14,16–19] The reason, if any, for multimeric assembly remains unclear. The cellular location and substrate specificity of the CE-7 enzymes is also uncertain. Many appear to be more active on short acetylated xylooligosaccharides than xylan, their ORFs display no obvious signal peptide and many occur on operons with known β-xylosidases.[13,14] Together, this has been taken to suggest that in some xylanolytic organisms CE-7 proteins reside in an intracellular location and function in the deacetylation of short acetylated xylooligosaccharides imported from the extracellular degradation of xylan.

In order to investigate the structural basis for this unusual specificity and to dissect the oligomeric state of the protein complex we have expressed the multi-functional "cephalosporin C deactylase" CAH from *Bacillus subtilis* 168 and characterised its kinetics on a range of acetylated substrates. The native 3-D structure has been determined by multiple-wavelength anomalous dispersion (MAD) methods at 1.9 Å. The structure presents an unusual "doughnut-shaped" 32 hexamer with the six active centres disposed towards the central pore. Complex structures reveal catalytic centre specificity for acetate with a large hydrophobic cavity that is likely to be the non-specific binding site for xylobiose or cephalosporins. The hexameric array may function to hinder access to large acetylated ligands and presumably plays a role in conferring some degree of specificity upon an otherwise promiscuous active centre.

† Available at URL http://afmb.cnrs-mrs.fr/CAZY/index.html

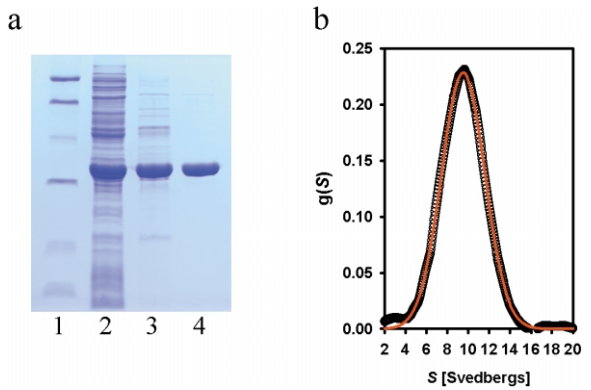

Figure 2. CAH is a hexamer in solution. (a) Purification of CAH. Lane 1, Molecular mass markers; lane 2, soluble cell extract; lane 3, 70–80% AS cut; lane 4, CAH after gel filtration. The molecular masses of the markers are (in descending order) 97.4 kDa, 66.2 kDa, 45.0 kDa, 31.0 kDa and 21.5 kDa. S-200 gel filtration estimated the molecular mass of CAH to be in the range 220–290 kDa. (b) A single Gaussian peak in the distribution of sedimentation coefficients, centred at 9.3 S, suggests that the protein is highly homogenous, with no evidence of other multimeric forms at the concentrations assayed. Sedimentation equilibrium experiments estimate a native molecular mass of 220,380(±6800) Da, compared with 216,966 Da calculated for a hexamer.

Results

Protein characterisation and structure solution

The purified protein (Figure 2(a)) is a hexamer in solution, as revealed by ultra-centrifugation (Figure 2(b)). The structure was solved by the MAD method (as described in Materials and Methods; Table 1), which confirmed a hexameric quaternary structure formed by a trimer of dimers. In the *R*3 crystal form the asymmetric unit contains one complete hexamer together with a dimer which itself forms a second, equivalent hexamer by virtue of a crystallographic 3-fold axis. The *R*32 crystal form contains a dimer, which again forms the equivalent hexamer around a crystallographic 3-fold axis.

Secondary structure elements were used to analyse a primary sequence alignment of the protein family (Figure 3), highlighting the sequence motif GxSxG common to esterase/lipase active sites, and a catalytic triad in the order Ser, Asp, His characteristic of α/β hydrolases.[20,21] Despite low sequence similarity, alignment based on secondary structures identifies features common to the wider family, including conserved amino acids at the edges of secondary structure elements, mostly Gly and Pro residues that are structurally important, mediating the tight turns required for fold maintenance. There is a conserved salt-bridge between an aspartate residue in helix αB and an arginine after strand β4 (Arg118–Asp156 in CAH) common to the α/β hydrolase family.[22] The active site serine has backbone angles ($\psi = -116.7$, $\varphi = 62.4$) outside of the favoured region of the Ramachandran plot,[23] exposing the side-chain proud from a strand-turn-helix substructure coined the "nucleophile elbow". The position of the catalytic triad residues Ser181, Asp269 and His298 are completely consistent with the fold family, but sequence variations on the motifs surrounding these residues act as a "signature" for the CE-7 family. The backbone

Table 1. Crystal, data and refinement statistics

	Peak	Inflection	Remote	Native	Mutant + acetate
A. *Crystal parameters*					
Space group	*R*32	*R*32	*R*32	*R*3	*R*32
Cell dimensions					
$a = b$ (Å)	156.7	156.7	156.7	315.2	157.2
c (Å)	132.7	132.7	132.7	68.5	132.1
No. mols./AU	2	2	2	8	2
B. *Data quality*					
Wavelength (Å)	0.9786	0.9787	0.9184	0.908	1.5418
Resolution of data (Å)	30–2.5	30–2.5	30–2.5	15–1.9	22–1.69
(outer shell)	2.56–2.5	2.56–2.5	2.56–2.5	1.97–1.9	1.74–1.69
Unique reflections	20,577	20,093	20,953	199,895	63,610
R_{merge} (outer shell)[a]	0.06 (0.138)	0.077 (0.192)	0.054 (0.133)	0.045 (0.221)	0.048 (0.393)
Mean $I/\sigma I$ (outer shell)	19.9 (5.5)	16.0 (3.9)	25.3 (11.0)	25.8 (4.4)	14.9 (2.7)
Completeness (outer shell) (%)	97.7 (95.5)	99.6 (94.2)	99.9 (100)	99.9 (98.8)	97.6 (95.9)
Multiplicity (outer shell)	3.8 (1.9)	3.8 (1.9)	4.0 (4.0)	3.3 (3.0)	3.1 (2.8)
C. *Refinement*					
Protein atoms	–	–	–	20168	5636
Solvent water molecule s	–	–	–	2064	562
Ions				8Cl + 2Mg	
R_{cryst}	–	–	–	15.4	15.5
R_{free}	–	–	–	18.6	19.1
r.m.s. deviation 1–2 bonds (Å)	–	–	–	0.011	0.016
r.m.s. deviation 1–3 angles (deg.)	–	–	–	1.32	1.58

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - \langle I_{hkl}\rangle|/\Sigma_{hkl}\langle I_{hkl}\rangle$.

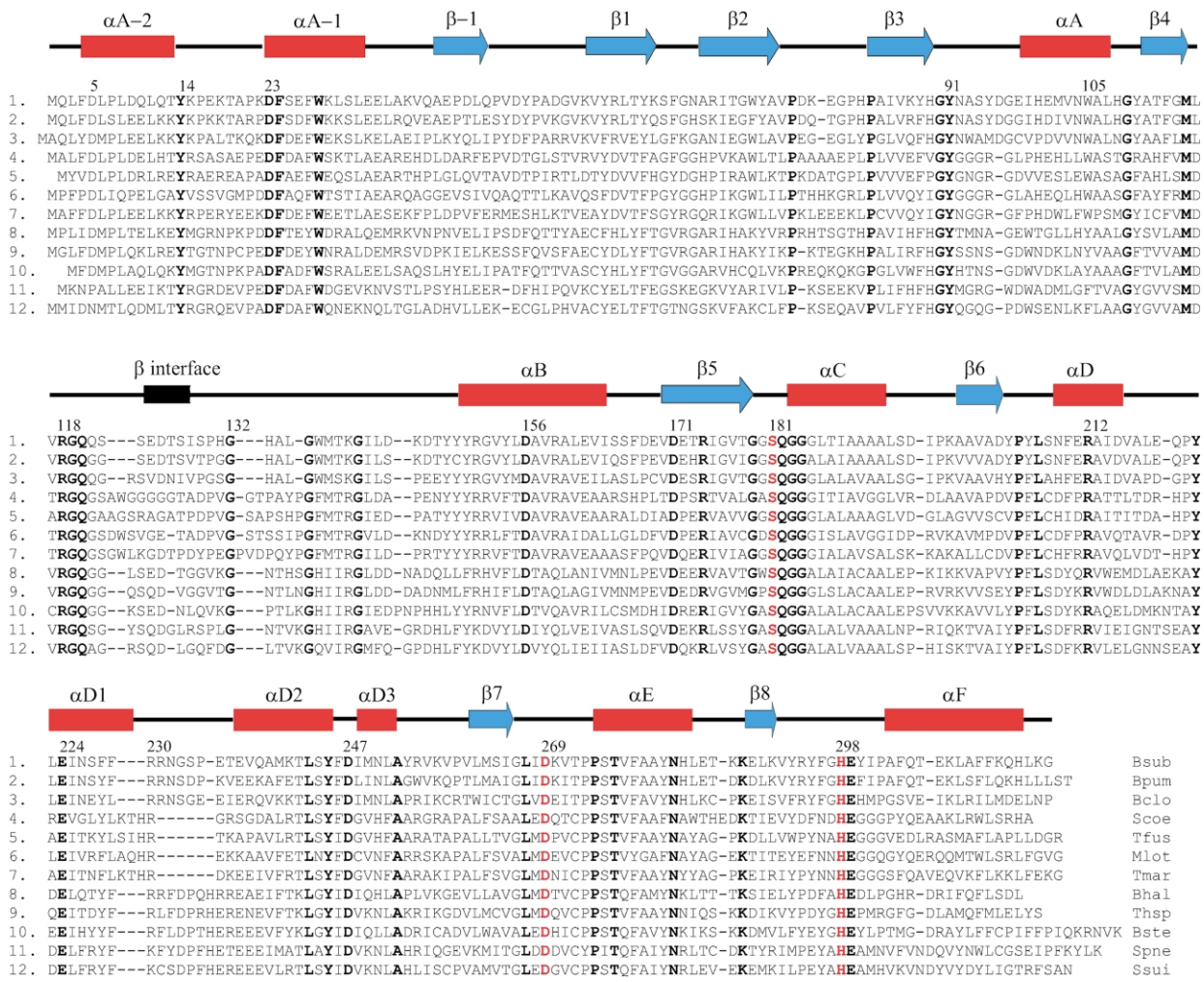

Figure 3. Structure-based sequence alignment of CE-7 esterases. Secondary structure elements of CAH (sequence 1): α-helices, red boxes; β-strands, blue arrows; numbered as for Figure 4(b). The β-sheet-like interface between adjacent subunits in the hexamer is depicted as a black box. Strictly conserved residue code letters are emboldened, with the active site catalytic triad highlighted in red. Amino acid residues of CAH mentioned in the text are numbered. In addition, Asp171 is numbered as this residue makes a salt bridge interaction with Arg174. Both residues are strictly conserved on the strand β5 that positions the catalytic serine181. Codes: 1, Bsub, *Bacillus subtilis*; 2, Bpum *B. pumilus* (TrEMBL q9k5f2); 3, Cthe, *Clostridium thermocellum* (NCBI zp_00062045); 4, Scoe, *Streptomyces coelicolor* (q9fc27); 5, Tfus, *Thermobifida fusca* (residues 80–402, zp_00057353); 6, Mlot, *Mesorhizobium loti* (q987m4); 7, Tmar, *Thermatoga maritima* (q9wxt2); 8, Bhal, *B. halodurans* (q9k7n5); 9, Thsp, *Thermoanaerobacter* sp.JW/SL YS485 (o30361); 10, Bste, *B. stearothermophilus* (q9jp86); 11, Spne, *Streptococcus pneumoniae* (q97pe0); 12, Ssui, *S. suis* (q8rr40).

amide of the residue following the active site nucleophile contributes to the oxyanion hole in α/β hydrolases. In CE-7 proteins, the invariant glutamine (Gln182) at this position interacts with a residue (Gln120) in the strictly conserved motif RGQ at the end of strand β4. The acidic glutamate side-chain (Glu299) following the active site His makes a bidentate salt bridge with Arg230 in the αD helical bundle insertion from an adjacent subunit. In other members of the family, a Glu residue positioned near the beginning of αF is speculated to coordinate hydrolytic water molecules. The sequence around the equivalent RGQ region has been used to identify epoxide hydrolases.[24] We suggest that the presence of the motifs RGQ…GxSQG…HE at the appropriate spacing, define a signature sequence motif for the CE-7 family.

The core of each monomer (Figure 4(a)) corresponds to the canonical α/β hydrolase fold,[10] consisting of a central, twisted β-pleated sheet flanked by α-helices packing above and below the sheet, with the catalytic triad residues Ser at the end of β5, Asp at the end of β7 and His contributed from a loop between β8 and αF (Figure 4(b)). This helix αF packs against the concave side of the sheet, along with αA, which is implicated in substrate binding. There is an insertion of a three-helix bundle between αD and β7. It is common for α/β hydrolases to have excursions from the canonical fold at this position.[25,26] For instance, α-helical insertions between β6 and β7

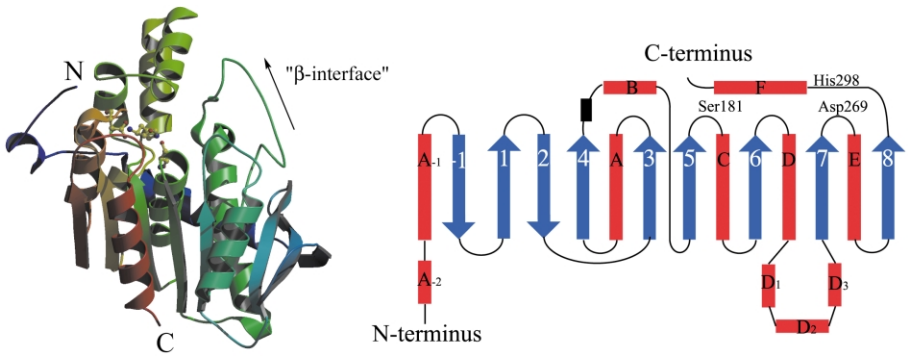

Figure 4. The fold of a CAH monomer. (a) Ribbon diagram, with secondary structure elements from the N to the C terminus colour ramped blue to red. The active site catalytic triad residues are depicted as ball-and-stick models, with oxygen atoms coloured red and nitrogen atoms blue. The extended portion of the sequence on the top right of the molecule in this orientation forms the β-sheet-like interface between adjacent subunits. This and subsequent Figures were drawn with BOBSCRIPT[72]/MOLSCRIPT[73] and raster3D.[74] (b) Topology diagram. The α-helices and β-strands are coloured and labelled as for Figure 3. The β-sheet-like interface-region is shaded in black. The catalytic triad of Ser181, Asp269 and His298 is indicated.

form the "cap" domains of epoxide hydrolase[24] and cocaine esterase.[27] There is also an extension of a β-strand and two α-helices at the N terminus. The strand β-1 extends the central sheet, giving a topology[28] of +1, +1, +2, −1x, +2x, $(+1x)_3$, with only strands β-1 and β2 antiparallel. The nine-strand extended sheet is highly twisted, with the first and last strands crossing at ~130° relative to each other. In a manner similar to brefeldin A esterase,[29] the insertions to the canonical fold interact, capping the active site and shielding it from solvent, much the same as the "lid" which mediates interfacial activation of lipases.[30] In CAH, the role of these insertions is much more profound, as the N terminus helix αA-2 can interact with the αD-insertion helices, notably by a salt bridge between Glu5 in αA-2 and Arg212 in αD1, thanks to the long N-terminal extension which wraps around each monomer.

These α-helical arrays interact with those from neighbouring subunits, to form both the entrance and cavity walls of the hexamer (Figures 5(a) and 6). They also pack tightly against residues in the helix αE, which displays an unusually high degree of sequence conservation (Figure 3). As well as α-helical interfaces between monomers in the hexamer, there is a short antiparallel β-strand-like interaction between residues 125–129 of each monomer, which take up an extended peptide conformation. The tight turns required for this interaction are mediated by Ser122, which has backbone dihedrals in the non-favoured region of the Ramachandran plot ($\psi = -46.8$, $\varphi = 67.2$) and Gly132.

The molecular surface of the hexamer has two openings, one on either face of the "doughnut", separated by approximately 36 Å. These entrances, of approximately 104 Å^2 (Figure 5(b)), open into a large chamber whose walls are approximately 60 Å apart at the widest point. The six active site residues point into this cavity. There appears to be no other way apart from the axial entrances for substrate and reaction products to enter and egress the catalytic chamber. A molecular surface based on electrostatic potential (Figure 5(b)) suggests that the cavity is predominately negatively charged. The reason for this is unknown, but it may aid expulsion of the reaction product, acetate, which is a weak competitive inhibitor.[31]

Substrate binding and the oxyanion hole

In order to facilitate determination of a structure of CAH in complex with ligand, an inactive mutant of the catalytic serine (Ser181Ala) was harnessed for co-crystallisation studies. Despite catalytic inactivity, complex data (not shown) did not reveal binding sites for cephalosporin or xylobiose, consistent with subsequent kinetics which show that these compounds display elevated K_M values. This mutant did reveal the binding site for the reaction product, acetate. Acetate expels three water molecules seen in the native structure. It binds in the catalytic site and is involved in hydrogen bonds with the $N^{\epsilon 2}$ atom of the active site His298, and the backbone amide hydrogen atoms of Tyr91 and of Gln182, which lie immediately adjacent to the active site nucleophile (Figure 7). This arrangement within the oxyanion hole is typical of α/β hydrolases. Here, however, the side-chains of the residues contributing their backbone amides are strictly conserved in CE-7 family enzymes. Tyr91 interacts with the strictly conserved residue Glu224, which caps the first helix in the αD helical insertion. Gln182 adjacent to the active site Ser bonds to the invariant Gln120 in the signature RGQ motif at the end of strand β4 and preceding the β-sheet-like interface. The oxyanion hole of CAH is therefore extremely well tethered, by interactions that are also intergral to hexameric maintenance.

Perhaps the best-characterised class of acetylxylan esterases is the *Penicillium purpurogenum* AXEII, a secreted protein, which is

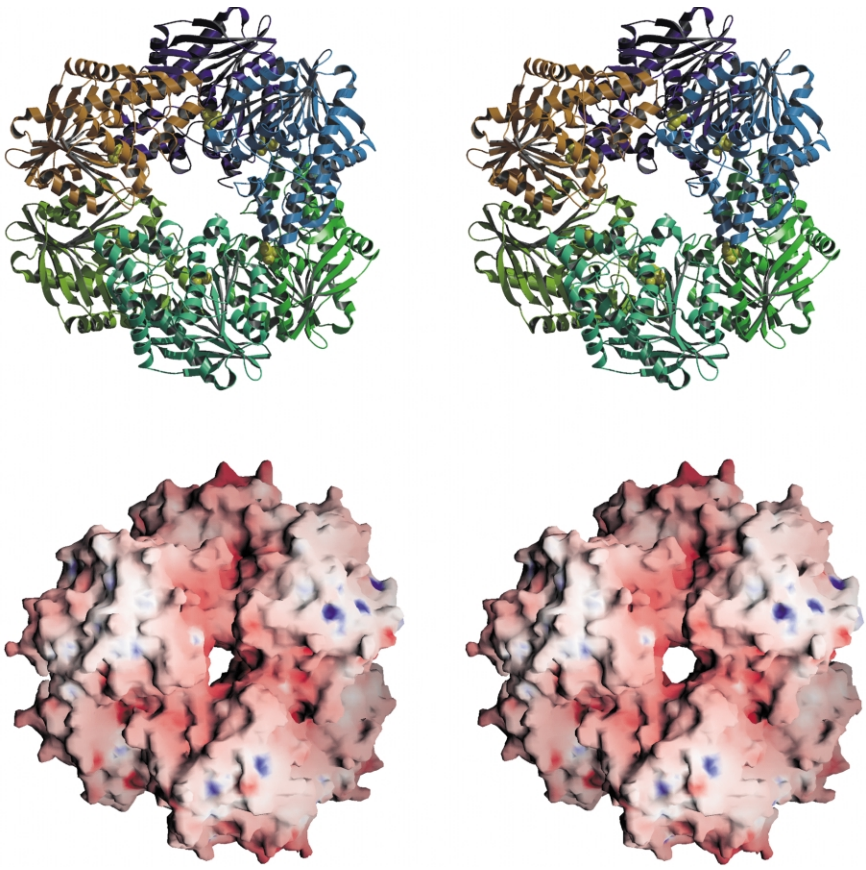

Figure 5. Quaternary structure of CAH. (a) A cartoon (divergent "wall-eyed" stereo) of the hexamer of CAH with each monomer in a different colour. The catalytic serine is shown with van der Waals' surface. (b) Electrostatic surface of the CAH hexamer, coloured from a GRASP[75] "surface potential" of colour property scale −38 (red, negatively charged) to +18 (blue, positive charge).

also a member of the α/β hydrolase superfamily.[5,6] Superposition of CAH with the core domain of AXEII[6] yields an r.m.s. deviation of 1.6 Å over 103 residues encompassing the β-sheet core and associated helices. The catalytic site of AXEII shows two conformations: in conformation A the serine has the same position as the CAH catalytic serine and the catalytic site is filled with water molecules; in conformation B, the serine rotates through 65° and the catalytic site contains a SO_4^{2-} molecule which replaces three of the resident water molecules and interacts with the catalytic His187 $N^{\epsilon 2}$, Thr13 O^{γ} and backbone amides of Thr13 and Gln91. When this structure of AXEII is superposed with the CAH–acetate complex, the acetate takes the same place and orientation as the sulphate ion, with two oxygen atoms of the acetate superimposing with two of the sulphate ion (Figure 8). It is not known if the CAH serine also takes up alternate conformations, as the structure is both a Ser to Ala

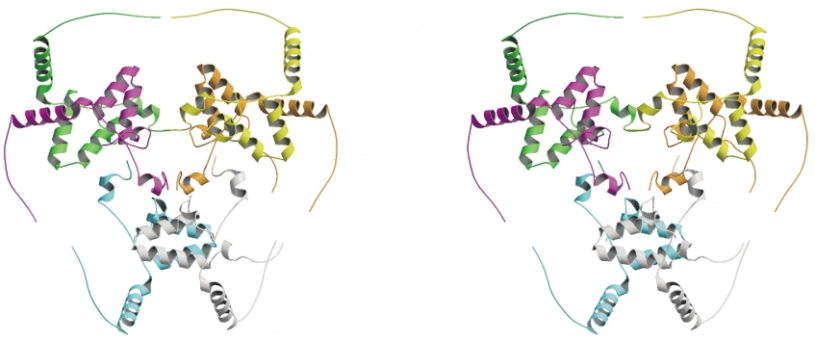

Figure 6. Insertion elements. The α-helices of the N-terminal extension and the αD excursion to the canonical α/β hydrolase fold are coloured for each monomer in the hexameric CAH assembly. This representation emphasises the role of these secondary structure elements in the overall architecture of the hexamer, by forming inter-subunit interfaces, as well as defining the entrance and walls of the catalytic cavity.

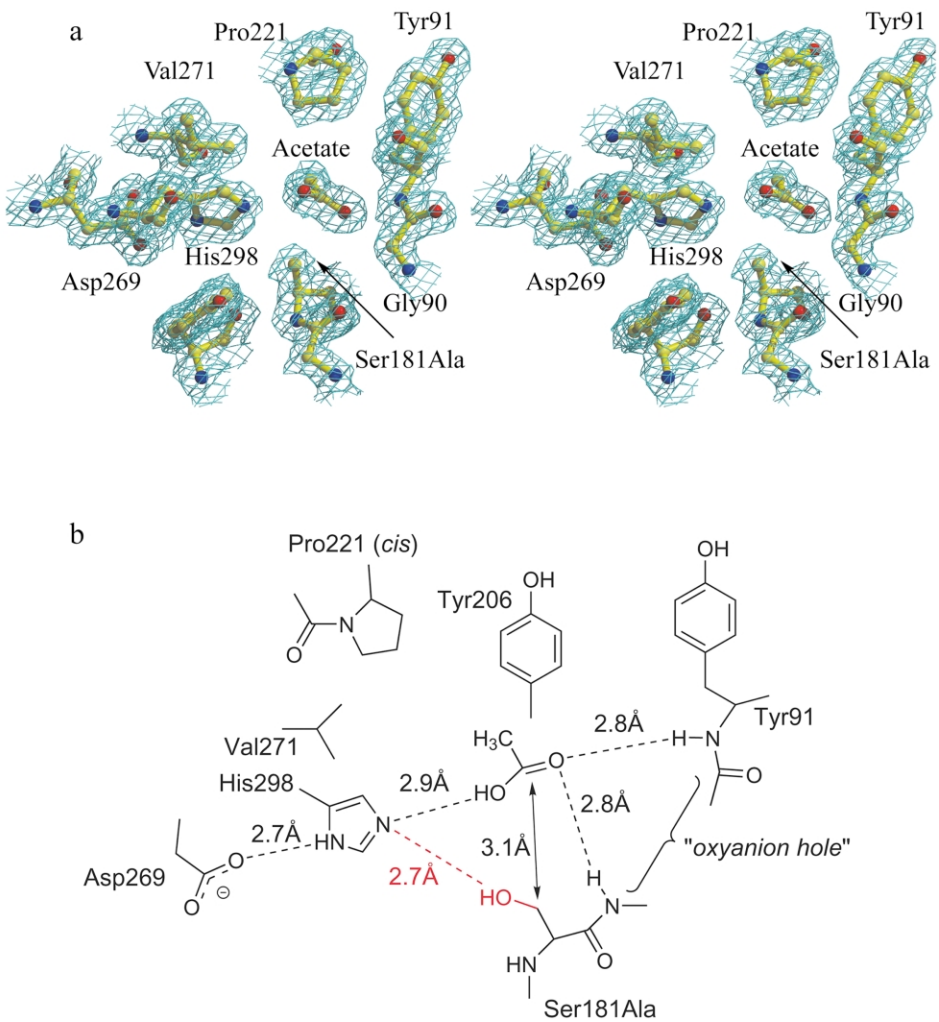

Figure 7. Acetate binding. (a) Observed electron density for the complex CAH with acetate. The density shown in blue mesh is the $2F_o - F_c$ map contoured at the 1 sigma level, and atoms are represented in ball-and-stick. The Figure is in divergent (wall-eyed) stereo and was drawn with BOBSCRIPT.[72] (b) Schematic showing the interactions of acetate in the active centre of (Ser181Ala) CAH. Positions and distances associated with the wild-type structure are included and coloured red.

mutant and was not crystallised with sulphate, which may induce this change. The multiple conformations seen in AXEII, cutinase[32] and a *Bacillus* extracellular lipase[33] may be biologically relevant because these enzymes lack any lid or cap over their active sites, thus providing an alternate mechanism whereby the protein can be held in an inactive state.

The acetate molecule lies in a surface depression into which it fits nicely, and defines one end of the substrate-binding pocket (Figure 9). This substrate-binding gorge has sides 9 Å apart, formed by helices αF and αA, and a depth of 14 Å whose base is formed by the central β-strands β3, β4 and β5. The isolation of a Bacillus CAH with enhanced cephalosporin hydrolytic rates displays four sequence differences to this CAH.[12] Three of these positions are completely exposed to solvent. The Thr112 to Ala replacement, however, occurs on strand β4, near one end of the substrate-binding channel, and is likely to affect substrate binding. The gorge length is 21 Å and is delimited by the helix αD1 at the catalytic site end and the C-terminal residues of helix αF and helix αA. In particular, Trp105 from αA closes the end of the gorge, and may put length limitations on which substrates would be able to bind with good affinity. This substrate-binding pocket is similar to that of AXEII, except that the acetylxylan esterase lacks the Trp closing gate. The overall environment around the CAH pocket is highly hydrophobic, with an extended aromatic patch consisting of tyrosine residues from positions 91, 148, 149, 150, 154, 204, 206 and 222, plus Phe210 and Phe248.

Enzyme activity of CAH

The data presented in Table 2 show that CAH is able to remove acetyl groups from a range of *O*-acetylated small molecules, although the enzyme was unable to attack *N*-acetyl-D-glucosamine, indicating that the enzyme is specific for ester bonds and is unable to cleave amide linkages. For all

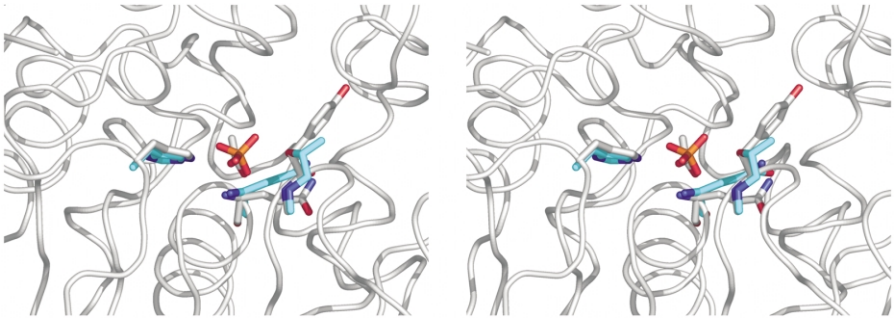

Figure 8. Comparison of the oxyanion hole of CAH and acetyl-xylan esterase AXEII. Divergent stereographic view of the superposition of the residues involved in the interaction of CAH with acetate and AXEII[6] with sulphate. AXEII residues Thr13 and Gln91, which interact with the sulphate ion are in cyan whilst the equivalent CAH residues (Tyr91 and Gln182) interacting with the acetate are in white. This Figure was drawn using PyMOL[76] (http://www.pymol.org).

substrates, except 4-nitrophenylacetate (4-NPA), the activity of the enzyme was directly proportional to substrate concentration [*S*], indicating that the assays were carried out at [*S*] values considerably below K_M, thus precluding the determination of individual kinetic parameters, although catalytic efficiency could be calculated from the slope (Table 2). The activity of CAH against acetylated xylose, xylobiose and deoxyadenosine was broadly similar, but the enzyme displayed lower activity against cephalosporin C. The esterase removed all the acetyl groups from tetra-acetyl-xylose and penta-acetyl-xylobiose. CAH displayed no activity against fully or partially acetylated xylan. Treatment of these decorated polysaccharides with the xylanase pc Xyn10A (see Materials and Methods) converted 5% and 100%, respectively, of the fully and partially acetylated xylan to small oligosaccharides, primarily the mono and disaccharide. Subsequent addition of CAH removed 86% and 5% of the acetyl groups from xylanase treated partially and fully acetylated xylan, respectively.

4-NPA was sufficiently soluble to allow determination of both kinetic parameters yielding a k_{cat} value of 4712(±198) min^{-1}, K_M 0.288(±0.012) mM and hence k_{cat}/K_M of 1.6×10^7 min^{-1} M^{-1}. CAH is more active against 4-NPA than the other substrates with a K_M value >tenfold lower than the other acetylated molecules. This may reflect an unusually high affinity of CAH for 4-NPA, presumably because the enzyme is better able to accommodate the aryl moiety than the organic alcohol component of the other substrates. Esterases, however, cleave substrates *via* a two-stage process. In the initial phase the catalytic nucleophile attacks the carbonyl group of the ester bond generating a covalent enzyme-acetyl intermediate (k_2), which is then hydrolysed by an activated water molecule (k_3). For most substrates, acetylation is rate limiting, but for substrates that contain good leaving groups, k_2 is rapid and k_3 becomes rate-limiting. When deacetylation is the rate-limiting step the acetyl-enzyme intermediate accumulates at low substrate concentration and thus the apparent K_M value will be much lower than for substrates in which k_2 is rate-limiting. As the leaving group of 4-NPA has a much lower pK_a value (7.18)[34] than sugar hydroxyl groups (~12), it is likely that the rate-limiting step for 4-NPA cleavage will be k_3, reflected in the relatively low K_M value for this substrate.

Discussion

CAH displayed no activity against acetylated or even partially acetylated (polymeric) xylan.

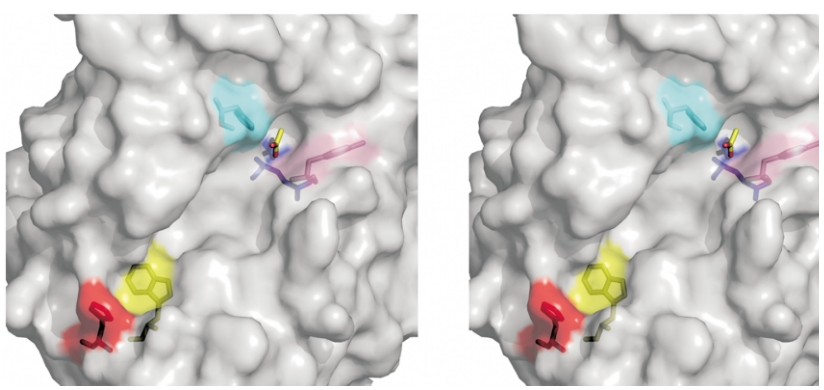

Figure 9. Substrate binding gorge. Surface representation of the CAH catalytic gorge. The acetate is represented in yellow liquorice and residues Tyr91 (pink), His298 (cyan) and Gln182 (dark blue) are highlighted behind the grey molecular surface. The substrate-binding channel extends to the bottom left of the diagram, where one end is closed by Trp105 (yellow). The Figure was drawn with PyMOL[76] (http://www.pymol.org)

Table 2. Enzyme kinetics

Substrate	Substrate concentration range (mM)	Enzyme concentration[a] (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)	Standard error
Cephalosporin C	0.4–1.7	0.05	7.1×10^4	$\pm 2.9 \times 10^3$
2,3,2′,3′,4′-Penta-acetyl-xylobiose	1.2–3.7	0.1	2.7×10^5	$\pm 4.8 \times 10^3$
4-Nitrophenyacetate	0.1–0.45	0.0125	1.6×10^7	$\pm 6.8 \times 10^5$
3′-*O*-Acetyl-2′-deoxyadenosine	0.7–2.7	0.05	5.4×10^5	$\pm 3.9 \times 10^4$
1,2,3,4-Tetra-*O*-acetyl-xylopyranose	0.3–1.3	0.05	7.6×10^5	$\pm 7.3 \times 10^3$
N-Acetyl-D-glucosamine	2.3-13.6	2.0	No activity	NA
100% Acetylated xylan	34.8	1.0	No activity	NA
57% Acetylated xylan	19.8	1.0	No activity	NA
100% Acetylated xylan + Xyn10A[b]	34.8	1.0	1.7 mM[c]	NA
57% Acetylated xylan + Xyn10A[b]	19.8	1.0	17.0 mM[c]	NA

NA, not applicable.
[a] Concentration of CAH monomer.
[b] Pre-treatment with 1 μM Xyn10A for 18 hours (unchanged HPLC from two hours) followed by enzyme inactivation.
[c] Acetate released after 24 hours hydrolysis (unchanged from two hours).

Decorated oligosaccharides, derived from these polysaccharides by the action of a xylanase, however, acted as substrates, indicating that the active site is unable to accommodate large substrates. CAH shows absolute specificity for acetylated substrates reflected in its inability to hydrolyse cinnamoyl-esters, for example (data not shown). These biochemical properties are similar to those reported for other esterases of the CE-7 family, although the equivalent enzyme from *B. pumilus* appears to display low activity (less than 1% of the specific activity for xylose tetraacetate) against acetylated xylan.[19] All published reports of CE-7 activity suggest that xylose tetraacetate is the best substrate tested, and that the cephem nucleus (7-ACA) is a better substrate than cephalosporin C, with K_M values around 10 mM. Given the broad specificity of CAH it is difficult to define its role in *B. subtilis*. It is an intracellular protein, evidenced by the lack of signal peptide and cysteine residues in the sequence. Few clues can be gleaned from the gene organisation on the Bacillus chromosome. There is a good consensus −10 sequence for binding of the primary vegetative sigma factor[35] centred at position −24 from the start codon (TcTTATAAa) but the promoter seems to lack a strong −35 sequence. There is a good match for the minor sigma factor σ^H at position −43 (ATAGGATaC), suggesting that this gene is under transcriptional control and may require activation by ancillary factors. There is no genetic linkage to known carbohydrate processing enzymes, unlike the organisation seen around some CE-7 encoding genes, such as in *B. stearothermophilus*, where it is flanked by genes for β-mannanase and α-galactosidase,[36] or linked to β-xylosidase in *Thermoanaerobacter.*[13,14] In Thermatogae, the gene is similarly linked to a mannanase and xylosidase in a hemi-cellulose catabolic gene cluster, and in *Streptococcus suis*, it is flanked by a putative glucose kinase and *N*-acetyl-mannosamine-6-P epimerase.[37] The presence of known CE-7 genes is very sparsely distributed in the microbial phylogenetic sub-divisions, being limited to the alpha proteobacteria, thermotogae, actinomycetes and firmicutes. It is possible that its main function in *B. subtilis* is to deacetylate decorated xylooligosaccharides that are transported into the cytoplasm of the bacterium. This conclusion, however, must be viewed with some caution, as there is an extensive repertoire of small organic molecules that contain acetyl groups, any one of which could act as a substrate for this enzyme, and it is known that α/β hydrolases mediate a large number of reactivities.[38]

Shielding of active sites in α/β hydrolases occurs in a number of ways. Lipases use α-helical lids, which move on contact with lipids, leading to interfacial activation.[30] The active site of prolyl oligopeptidase is accessed through the central tunnel of a seven-bladed β-propeller domain by partial opening of the circular structure.[39] Of special relevance to discussion of CE-7 family structure organisation are the structures of brefeldin A esterase,[29] proline iminopeptidase[40] and X-prolyl dipeptidyl aminopeptidase.[41] All display α-helical insertions after strand β6. In brefeldin A esterase the αD insertion interacts with N-terminal α-helices, much like CAH. Proline iminopeptidase has an insertion of six helices in the topologically and spatially equivalent location to the three-helix bundle of CAH. This feature is shared by serine carboxypeptidases, where they not only contribute to substrate binding, but also mediate dimerisation. CAH seems to have taken this one step further, utilizing the α-helical interactions as both a cap and dimer interface, but added further interactions between them by the hexameric assembly, leading to complete shielding of the active sites, with access limited by the entrance mouth formed by the αD helices.

Such sequestration of active sites by multimeric assembly is a strategy shared by self-compartmentalising proteases, a number of which are structurally characterised. The best-known examples are barrel-shaped with associated ATPase or chaperone activities. The proteasome, FtsH, HlsVU, ClpP, Lon and DegP/HtrA[42–46] are

thought to act by progressively unfolding and degrading proteins. Other peptidases, which lack ATPase domains, share the strategy of restricting access to catalytic sites by forming a chamber, entered by narrow axial or lateral pores, e.g. Gal6/bleomycin hydrolase, tricorn protease, DppA and leucine aminopeptidase.[47–50] Many use a post-translational processing event as an additional measure to ensure that no functional activity is possible until the active sites are safely hidden away from cellular components, which must be protected from such potent, non-specific proteolytic activities. A protease from *Pyrococcus horikoshii* has an elaboration on this theme of ensuring minimal proteolytic activity before assembly, the active sites are formed at the interfaces of monomer pairs in the hexamer, with the acid (Glu) residue of the triad donated by an adjacent monomer.[51] By analogy, we suggest that CAH is the first example of a self-compartmentalising esterase, where access is initially limited by the sieving effect of the tunnel entrance, then further by a size limitation in the rather non-specific substrate-binding groove. The precise conformation of the active centre, can only be formed after assembly into the trimer of dimers, as the position of the oxyanion hole is tightly tethered by residues intimately associated with dimer interfaces, and the crucial active site histidine residue is positioned by a salt-bridge from its adjacent glutamic acid residue to a neighbouring subunit.

Materials and Methods

Protein production and crystallisation

The *Bacillus subtilis cah* (cepahalosporin acetyl hydrolase) gene was cloned into the T7-promoter based expression vector pET26b (Novagen) and expressed in *Escherichia coli* BL21(DE3), a strain with an inducible T7 RNA polymerase gene.[52] PCR with a high-fidelity DNA polymerase (Pfu; Stratagene) was used to amplify *cah* from *B. subtilis* IG20 (168*trpC2*) chromosomal DNA, using primers designed to introduce convenient restriction sites for cloning (*Nde* I and *Bam* HI). DNA sequencing of independent isolates revealed a single amino acid alteration from the predicted protein sequence (Asp255 GAC to Tyr TAC). Site-directed mutagenesis of the active site serine AGC codon to an alanine codon (GCT, simultaneously incorporating a new *Sac* I restriction site to aid screening) was mediated by complementary 41-mer oligonucleotides (MWG-Biotech) using a QuikChange kit (Stratagene). For protein production, cell cultures were grown in LB media containing 30 μg/ml kanamycin to an absorbance of 0.8 at 600 nm before induction of protein expression by the addition of IPTG to a final concentration of 1mM. After four hours further growth, the cells were harvested by centrifugation and disrupted by sonication. The lysate was clarified by centrifugation and fractionated using ammonium sulphate. Protein precipitated at 70–80% saturation was dissolved in 6 ml of 1.7 M ammonium sulphate, 50 mM Tris (pH 8.0), applied to a phenyl Sepharose column and eluted using a linear gradient of decreasing ammonium sulphate. CAH-containing fractions were concentrated and further purified by gel filtration (Superdex200 HR10/30, Pharmacia) in a 50 mM Tris (pH 8.0), 250 mM NaCl buffer, with apparent molecular mass estimated by comparison to standard protein markers. Seleno-L-methionyl (SeMet) labelled protein was produced in the *E. coli* methionine auxotroph B834 (DE3) grown in a defined media using a standard protocol,[53] and purified as described above but in the presence of 10 mM DTT and omitting the phenyl Sepharose column step. Precipitated SeMet-CAH was dissolved and buffer exchanged into gel filtration buffer containing 10 mM DTT using a PD-10 desalting column (Pharmacia), and then concentrated to 25 mg/ml using a 30 kDa centrifugation membrane (Pall Filtron). Pure fractions of SeMet-CAH were pooled, washed into fresh reducing agent and concentrated to 30 mg/ml using centrifugation membranes as described above. Final protein yield was ~50 mg/l of cell culture, which ran as a single band on both native PAGE and isoelectric focussing (pI 5.2). MALDI-TOF mass spectrometric analysis of both SeMet- and native CAH confirmed the protein masses and indicated that selenium incorporation had been successful (data not shown). Protein crystals were grown by vapour-phase diffusion using the hanging drop method. For native protein, a rhombohedral crystal form in space group *R*3 was obtained when equal volumes of protein (20 mg/ml) were mixed with 30% (v/v) MPD, 0.1 M $MgCl_2$, 0.1 M Mops (pH 6.5). The unit cell, dimensions $a = b = 315.2$ Å, $c = 68.5$ Å, contain eight molecules in the asymmetric unit, corresponding to a solvent content[54] of 46%. SeMet-CAH protein crystallises in the related space group *R*32, with unit cell dimensions $a = b = 156.74$ Å, $c = 132.7$ Å and two molecules in the asymmetric unit, corresponding to 44% solvent content. This crystal form grew over a period of 24 hours from 25% (v/v) MPD, 0.1 M Tris–HCl (pH 7.5), 0.2 M $CaCl_2$ and 10 mM DTT. Crystals of the mutant protein in complex with acetate were grown from the same solution containing 0.2 M sodium acetate.

X-ray data collection and processing

Native data were collected at a temperature of 100K on a frozen crystal after transfer into a stabilising solution of 30% (v/v) MPD for 15 seconds, on beamline X11 at the EMBL, Hamburg, using a MAR Research 345 image-plate as detector, collecting 200 images of 0.5° per image. For SeMet-CAH, a three-wavelength MAD experiment was conducted on beamline BM14 at the European Synchrotron Radiation Facility (ESRF, Grenoble) at a temperature of 100 K using a MAR CCD detector. Datasets were collected at the wavelengths corresponding to the minimum f', the maximum f'' and a reference wavelength at an energy above the absorption edge to maximise dispersive differences, as chosen by scanning through the absorption edge of the crystal (Table 1). X-ray data for the mutant-acetate complex were collected at 100 K using a Cu Kα rotating anode, long-focusing mirror optics, with a MAR-research imaging-plate system. Data were integrated, scaled and reduced using DENZO and SCALEPACK[55,56] as part of HKL2000, with all further computation using the CCP4 suite of programs.[57] Details of the data statistics are given in Table 1.

Analytical ultracentrifugation and enzyme assays

Analytical ultracentrifuge experiments were carried

out on a Beckman XL-A machine using an AN60Ti rotor. Sedimentation velocity experiments were performed at 40,000 rpm using Epon two-channel centrepieces. Data were obtained by radial scans at 280 nm every five minutes during the run. Sedimentation coefficient distributions were determined using the program DCDT.[58] Sedimentation equilibrium was performed in Epon six-channel centrepieces. Samples at 0.7 mg/ml, 0.35 mg/ml and 0.15 mg/ml were centrifuged to equilibrium (10–12 hours) at 7000 rpm, 9000 rpm, and 11,000 rpm at 20 °C. A true optical baseline was obtained by increasing the rotor speed to 36,000 rpm for eight hours. Data were fitted using NONLIN[59] to obtain a molecular mass, with the best-fit model obtained from inspection of the residuals and the fit statistics.

Enzyme assays were carried out in 50 mM sodium phosphate/citrate buffer(pH 6.5) at 37 °C over a two hour period in a 1 ml reaction volume, unless otherwise stated. Purified CAH at a concentration of 12 nM–2 μM was incubated with substrate at various concentrations, and at regular time intervals a 100 μl aliquot was removed and the release of acetate quantified using the Boehringer Mannheim (R-Biopharm Rhone Ltd) acetic acid test kit following the manufacture's instructions. Activity against 4-nitrophenylacetate (4-NPA) was determined by continuous measurement of 4-nitrophenolate at 410 nm, using a molar extinction coefficient of 3658 $M^{-1}cm^{-1}$, which was determined experimentally, while the hydrolysis of 4-methylumbelliferyl *p*-trimethylammoniocinnamate chloride was evaluated by the appearance of the fluorometric molecule 4-methylumbelliferone. The substrates used in the enzyme assays were obtained as follows: cephalosporin C, *N*-acetyl-D-glucosamine, 3′-*O*-acetyl-2′-deoxyadenosine, 4-methylumbelliferyl *p*-trimethylammoniocinnamate chloride and 4-NPA were purchased from Sigma Chem. Co. (Poole, UK) while 1,2,3,4-tetra-acetyl-xylose was obtained from CMC Chemicals Ltd (Abingdon, UK), and 2,3,2′,3′,4′-penta-acetyl-xylobiose was prepared by the method of Charnock *et al*.[60] Maximum substrate concentrations used in the assays were 3.7 mM due to the sparsely soluble nature of these acetylated molecules. Fully acetylated xylan (100%) was synthesised from the soluble portion of oat spelt xylan (Sigma Chem. Co., UK) as described.[61] The substrate was treated with 100 mM NaOH at 0 °C for 15 minutes to generate xylan that was 57% acetylated equating to approximately one acetyl group per xylose residue in the polysaccharide. To generate acetylated xylooligosaccharides from both forms of acetylxylan, the polysaccharides, at a concentration of 5 mg/ml, were incubated with 1 μM of *Cellvibrio japonicus* Xyn10A[62] in PC buffer at 37 °C for 18 hours. The extent of xylan degradation and the nature of the products generated were assessed by HPLC.[60]

Phasing, model building and refinement

The anomalous differences at the wavelength corresponding to the f'' maximum were used as input to the direct methods program SnB.[63] Of 1000 initial phase sets, 86 had a significantly lower Minimal Function value (0.21–0.24) than the noise (peaking around 0.7). The top phase sets yielded a set of 24 peaks, in agreement with the predicted two molecules per asymmetric unit. The Se positions were refined and phases calculated using the CCP4 program MLPHARE, giving an overall figure-of-merit (FOM) of 0.58. The resulting phases were used as a starting set for phase improvement by solvent flattening in DM[64] giving an overall FOM of 0.85. No averaging was used in those calculations. An electron density map calculated with the coefficients from DM was of sufficient quality to allow the tracing of the entire molecule using the X-AUTOFIT[65] module in Quanta (Accelrys Inc. San Diego, USA). The model was refined using the CCP4 program REFMAC,[66] initially with the phases from DM included as experimental restraints. A partially refined monomer model in this space group (*R*32) with a crystallographic *R*-factor[67] of 27% (R_{free} 32%) was used as a search model to solve the structure in *R*3 using the program AMoRe.[68] Four molecules were confidently placed in the asymmetric unit and provided a starting refinement model for REFMAC/ARP/WARP.[66,69] The warpNtrace mode produced a model, after 100 cycles, consisting of many fragments, which could be grouped into eight molecules. In retrospect, all eight molecules were present in the initial AMoRe peak-lists. The side-chains were placed using the side_dock option in ARP/WARP. Additional refinement using REFMAC and Quanta yielded a final model consisting of 20,168 non-hydrogen protein atoms with 2054 water molecules and ten ions. The crystallographic *R*-factor and R_{free} factor are, respectively, 15.4% and 18.6% (see Table 1). One dimer of the final model was used to find the solution of the complex (CAH mutant + acetate) structure with the program MOLREP.[70] The Correlation coefficient and the *R*-factor were 0.72 and 0.32, respectively. Several refinement cycles of REFMAC have been applied and the final *R* and R_{free} values are 15.5% and 19.1%, respectively (see Table 1).

Data Bank accession codes

The atomic coordinates and structure factors for the native and complex protein structures have been deposited in the European Bioinformatics Institute Protein Databank with PDB codes 1ODS and 1ODT, respectively.

Acknowledgements

The authors thank Marek Brzozowski for expert guidance in crystallisation, and the ESRF (Grenoble), and EMBL Hamburg for excellent beam-line facilities. This work was funded by The Wellcome Trust (grant 063963) and by the BBSRC through a Centre award (87/SB/09829). Support for data collection at the EMBL Hamburg synchrotron outstation was provided under the TMR/LSF program, reference ERBFMGECT980134. G.J.D. is a Royal Society University Research Fellow.

References

1. Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.*, 695–696.
2. Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated approach. In *Recent Advances in Carbohydrate Engineering* (Gilbert, H. J., Davies, G. J., Svensson, B. & Henrissat, B., eds), pp. 3–12, Royal Society of Chemistry, Cambridge, UK.
3. Prates, J. A. M., Tarbouriech, N., Charnock, S. J.,

Fontes, C. M. G. A., Ferreira, L. M. A. & Davies, G. J. (2001). The structure of the feruloyl esterase module of xylanase 10B from *Clostridium thermocellum* provides insights into substrate recognition. *Structure*, **9**, 1183–1190.
4. Hakulinen, N., Tenkanen, M. & Rouvinen, J. (2000). Three-dimensional structure of the catalytic core of acetylxylan esterase from *Trichoderma reesei*: insights into the deacetylation mechanism. *J. Struct. Biol.* **132**, 180–190.
5. Ghosh, D., Erman, M., Sawicki, M., Lala, P., Weels, D. R., Li, N. *et al.* (1999). Determination of a protein structure by iodination: the structure of iodinated acetylxylan esterase. *Acta Crystallog. sect. D*, **55**, 779–784.
6. Ghosh, D., Sawicki, M., Lala, P., Erman, M., Pangborn, W., Eyzaguirre, J. & Gutiérrez, R. (2001). Multiple conformations of catalytic serine and histidine in acetylxylan esterase at 0.90 Å. *J. Biol. Chem.* **276**, 11159–11166.
7. Molgaard, A. & Larsen, S. (2002). A branched N-linked glycan at atomic resolution in the 1.12 Å structure of rhamnogalacturonan acetylesterase. *Acta Crystallog. sect. D*, **58**, 111–119.
8. Ronning, D. R., Klabunde, T., Besra, G. S., Vissa, V. D., Besisle, J. T. & Sacchettini, J. C. (2000). Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nature Struct. Biol.* **7**, 141–146.
9. Dodson, G. & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347–352.
10. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M. *et al.* (1992). The α/β hydrolase fold. *Protein Eng.* **5**, 197–211.
11. Jenkins, J., Mayans, O., Smith, D., Worboys, K. & Pickersgill, R. W. (2001). Three-dimensional structure of *Erwinia chrysanthemi* pectin methylesterase reveals a novel esterase active site. *J. Mol. Biol.* **305**, 951–960.
12. Mitsushima, K., Takimoto, A., Sonoyama, T. & Yagi, S. (1995). Gene cloning, nucleotide sequence, and expression of a cephalosporin-C deacetylase from *Bacillus subtilis*. *Appl. Environ. Microbiol.* **61**, 2224–2229.
13. Lee, Y.-E. & Zeikus, J. G. (1993). Genetic organisation, sequence and biochemical characterization of recombinant β-xylosidase from *Thermoanaerobacterium saccharolyticum* strain B6A-RI. *J. Gen. Microbiol.* **139**, 1235–1243.
14. Lorenz, W. W. & Wiegel, J. (1997). Isolation, analysis, and expression of two genes from *Thermoanaerobacterium* sp. strain JW/SL YS485: a β-xylosidase and a novel acetyl xylan esterase with cephalosporin C deacetylase activity. *J. Bacteriol.* **179**, 5436–5441.
15. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H. *et al.* (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermatoga maritima*. *Nature*, **399**, 323–329.
16. Takimoto, A., Mitsushima, K., Yagi, S. & Sonoyama, T. (1994). Purification, characterization and partial amino acid sequences of a novel cephalosporin-C deacetylase from *Bacillus subtilis*. *J. Ferment. Bioeng.* **77**, 17–22.
17. Shao, W. & Wiegel, J. (1995). Purification and characterization of two thermostable acetyl xylan esterases from *Thermoanaerobacterium* sp. strain JW/SL-YS485. *Appl. Environ. Microbiol.* **61**, 729–733.
18. Degrassi, G., Okeke, B. C., Bruschi, C. V. & Venturi, V. (1998). Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus*. *Appl. Environ. Microbiol.* **64**, 789–792.
19. Degrassi, G., Kojic, M., Ljubijankic, G. & Venturi, V. (2000). The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. *Microbiology*, **146**, 1585–1591.
20. Brenner, S. (1988). The molecular evolution of genes and proteins—a tale of two serines. *Nature*, **334**, 528–530.
21. Drablos, F. & Petersen, S. B. (1997). Identification of conserved residues in family of esterase and lipase sequences. *Methods Enzymol.* **284**, 28–61.
22. Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K. & Doctor, B. P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterase, lipases, and related proteins. *Protein Sci.* **2**, 366–382.
23. Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical conformations for polypeptide and protein conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.
24. Zou, J., Hallberg, B. M., Bergfors, T., Oesch, F., Arand, M., Mowbray, S. L. & Jones, T. A. (2000). Structure of *Aspergillus niger* epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure*, **8**, 111–122.
25. 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.
26. Nardini, M. & Dijkstra, B. W. (1999). α/β hydrolase fold enzymes: the family keeps growing. *Structure*, **9**, 732–737.
27. Larsen, N. A., Turner, J. M., Stevens, J., Rosser, S. J., Basran, A., Lerner, R. A. *et al.* (2002). Crystal structure of a bacterial cocaine esterase. *Nature Struct. Biol.* **9**, 17–21.
28. Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 168–339.
29. Wei, Y., Contreras, J. A., Sheffield, P., Osterlund, T., Derewenda, U., Kneusel, R. E. *et al.* (1999). Crystal structure of brefeldin A esterase, a bacterial homolog of the mammalian hormone-sensitive lipase. *Nature Struct. Biol.* **6**, 340–345.
30. Brzozowski, A. M., Derewenda, U., Derewenda, Z. S., Dodson, G. G., Lawson, D. M., Turkenburg, J. P. *et al.* (1991). A model for interfacial activation of lipases from the structure of a fungal lipase–inhibitor complex. *Nature*, **351**, 491–497.
31. Zhu, X., Larsen, N. A., Basran, A., Bruce, N. C. & Wilson, I. A. (2003). Observation of an arsenic adduct in an acetyl esterase crystal structure. *J. Biol. Chem.* **278**, 2008–2014.
32. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997). Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis. *J. Mol. Biol.* **268**, 779–799.
33. Kawasaki, K., Kondo, H., Suzuki, M., Ohgiya, S. & Tsuda, S. (2002). Alternate conformations observed in catalytic serine of *Bacillus subtilis* lipase determined at 1.3 Å resolution. *Acta Crystallog. sect. D*, **58**, 1168–1174.
34. Tull, D. & Withers, S. G. (1994). Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry*, **33**, 6363–6370.
35. Helmann, J. D. & Moran, C. P. Jr (2002). RNA

polymerase and sigma factors. In Bacillus subtilis *and its Closest Relatives: from Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 289–312, ASM Press, Washington DC.
36. Ethier, N., Talbot, G. & Sygusch, J. (1998). Gene cloning, DNA sequencing, and expression of thermostable β-mannanase from *Bacillus stearothermophilus*. *Appl. Environ. Microbiol.* **64**, 4428–4432.
37. Takamatsu, D., Osaki, M. & Sekizaki, T. (2002). Evidence for lateral transfer of the suilysin gene region of *Streptococcus suis*. *J. Bacteriol.* **184**, 2050–2057.
38. Cousin, X., Hotelier, T., Giles, K., Toutant, J. P. & Chatonnet, A. (1998). aCHEdb: the database system for ESTHER, the α/β fold family of proteins and the cholinesterase gene server. *Nucl. Acids Res.* **26**, 226–228.
39. Fülöp, V., Böcskei, Z. & Polgár, L. (1998). Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. *Cell*, **94**, 161–170.
40. Medrano, F. J., Alonso, J., Garcia, J. L., Romero, A., Bode, W. & Gomis-Ruth, F. X. (1998). Structure of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*: a prototype for the prolyl oligopeptidase family. *EMBO J.* **17**, 1–9.
41. Rigolet, P., Mechin, I., Delage, M-M. & Chich, J-F. (2002). The structural basis for catalysis and specificity of the X-prolyl dipeptidyl aminopeptidase from *Lactococcus lactis*. *Structure*, **10**, 1383–1394.
42. Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). Crystal structure of the 20 S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science*, **268**, 533–539.
43. Bochtler, M., Ditzel, L., Groll, M. & Huber, R. (1997). Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **94**, 6070–6074.
44. Wang, J. M., Hartling, J. A. & Flanagan, J. M. (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell*, **91**, 447–456.
45. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. (2002). Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature*, **416**, 455–459.
46. Li, W., Srinivasula, S. M., Chai, J., Li, P., Wu, J-W., Zhang, Z., Alnemri, E. S. & Shi, Y. (2002). Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nature Struct. Biol.* **9**, 436–441.
47. O'Farrell, P. A., Gonzalez, F., Zheng, W., Johnston, S. A. & Joshua-Tor, L. (1999). Crystal structure of human bleomycin hydrolase, a self-compartmentalising cysteine protease. *Structure*, **7**, 619–627.
48. Brandsetter, H., Kim, J. S., Groll, M. & Huber, R. (2001). Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature*, **414**, 466–470.
49. Remaut, H., Bompard-Gilles, C., Goffin, C., Frere, J-M. & Van Beeumen, J. (2001). Structure of the *Bacillus subtilis* D-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nature Struct. Biol.* **8**, 674–678.
50. Burley, S. K., David, P. R., Taylor, A. & Lipscomb, W. N. (1990). Molecular structure of leucine aminopeptidase at 2.7-Å resolution. *Proc. Natl Acad. Sci. USA*, **87**, 6878–6882.
51. Du, X., Choi, I.-G., Kim, R., Wang, W., Jancarik, J., Yokota, H. & Kim, S.-H. (2000). Crystal structure of an intracellular protease from *Pyrococcus horikoshii* at 2-Å resolution. *Proc. Natl Acad. Sci. USA*, **97**, 14079–14084.
52. Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130.
53. Ducros, V. M.-A., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P. *et al.* (2001). Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in *Bacillus subtilis*. *J. Mol. Biol.* **306**, 759–771.
54. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
55. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing: Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N. & Bailey, S., eds), Science and Engineering Research Council, Daresbury, UK.
56. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, 307–326.
57. Collaborative Computational Project No 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
58. Stafford, W. F. (1992). Methods for obtaining sedimentation coefficient distributions. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 359–393, Royal Society of Chemistry, Cambridge, UK.
59. Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical centrifuge by non-linear least-squares techniques. *Biophys. J.* **36**, 575–588.
60. Charnock, S. J., Spurway, T. D., Xie, H., Beylot, M. H., Virden, R., Warren, R. A. *et al.* (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *J. Biol. Chem.* **273**, 32187–32199.
61. Johnson, K. G., Fontana, J. D. & MacKenzie, C. R. (1988). Measurement of acetylxylan esterase in Streptomyces. *Methods Enzymol.* **160**, 551–560.
62. Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J. & Gilbert, H. J. (1989). Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies cellulosa: internal signal sequence and unusual protein processing. *Mol. Microbiol.* **3**, 1211–1219.
63. Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB v2.0. *J. Appl. Crystallog.* **32**, 120–124.
64. Cowtan, K. & Main, P. (1996). Improvement of macromolecular electron density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallog. sect. D*, **49**, 148–157.
65. Oldfield, T. J. (1996). Real space refinement as a tool for model building. In *Macromolecular Refinement. Proceedings of the CCP4 Study Weekend* (Dodson, E., Moore, M., Ralph, A. & Bailey, S., eds), pp. 67–74, SERC Daresbury Laboratory, Warrington, UK.
66. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
67. Brünger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
68. Navaza, J. (1994). AMoRe: an automated package for

molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
69. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D*, **49**, 129–147.
70. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.
71. Christov, L. P. & Prior, B. A. (1993). Esterases of xylan-degrading microorganisms—production, properties and significance. *Enzyme Microb. Technol.* **15**, 460–475.
72. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
73. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
74. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
75. Sharp, K., Fine, R. & Honig, B. (1987). Computer-simulations of the diffusion of a substrate into the active-site of an enzyme. *Science*, **236**, 1460–1463.
76. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.

Edited by R. Huber

(Received 25 February 2003; received in revised form 5 May 2003; accepted 8 May 2003)