Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure at 1.6 Å resolution of the protein product of the At4g34215 gene from *Arabidopsis thaliana*

Eduard Bitto, Craig A. Bingman, Jason G. McCoy, Simon T. M. Allard, Gary E. Wesenberg and George N. Phillips Jr*

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, USA

Correspondence e-mail: phillips@biochem.wisc.edu

Received 9 September 2005
Accepted 20 October 2005

PDB Reference: At4g34215, 2apj, r2apjsf.

The crystal structure of the At4g34215 protein of *Arabidopsis thaliana* was determined by molecular replacement and refined to an R factor of 14.6% ($R_{\rm free}$ = 18.3%) at 1.6 Å resolution. The crystal structure confirms that At4g34215 belongs to the SGNH-hydrolase superfamily of enzymes. The catalytic triad of the enzyme comprises residues Ser31, His238 and Asp235. In this structure the catalytic serine residue was found to be covalently modified, possibly by phenylmethylsulfonyl fluoride. The structure also reveals a previously undescribed variation within the active site. The conserved asparagine from block III, which provides a hydrogen bond for an oxyanion hole in the SGNH-hydrolase superfamily enzymes, is missing in At4g34215 and is functionally replaced by Gln30 from block I. This residue is positioned in a catalytically competent conformation by nearby residues, including Gln159, Gly160 and Glu161, which are fully conserved in the carbohydrate esterase family 6 enzymes.

1. Introduction

The gene At4g34215 of *Arabidopsis thaliana* encodes a protein with molecular weight 28.3 kDa (residues 1–260; UniProt code Q8L9J9). The biochemical function of At4g34215 is not yet established. Based on a *PSI-BLAST* search (Altschul *et al.*, 1997), the sequence of At4g34215 shows homology to a range of enzymes related to acetylxylan esterases ($E = 1 \times 10^{-60}$, 30% identity over ∼240 aligned amino acids). Many of the enzymes identified by *PSI-BLAST* are members of carbohydrate esterase family 6 (CE6; Coutinho & Henrissat, 1999*a*). Based on a *SUPERFAMILY* server search (Gough *et al.*, 2001), At4g34215 belongs to the SGNH-hydrolase superfamily of enzymes ($E = 4.2 \times 10^{-6}$). The structurally characterized enzymes of this superfamily include serine esterase from *Streptomyces scabies* (Wei *et al.*, 1995), the esterase domain of haemagglutinin-esterase-fusion glycoprotein HEF1 from influenza C virus (Rosenthal *et al.*, 1998), platelet-activating factor acetylhydrolase from *Bos taurus* (Ho *et al.*, 1997), rhamnogalacturonan acetylesterase from *Aspergillus aculeatus* (Molgaard *et al.*, 2000), thioesterase I of *Escherichia coli* (Lo *et al.*, 2003), a putative lipase from *Nostoc* sp. (PDB code 1vjg; Joint Center For Structural Genomics) and the recently elucidated structure of a putative acetylxylan esterase from *Clostridium acetobutylicum* (PDB code 1zmb). The enzymes of the SGNH-hydrolase superfamily facilitate the hydrolysis of ester, thioester and amide bonds in a range of substrates including complex polysaccharides, lysophospholipids, acyl-CoA esters and other compounds (Lo *et al.*, 2003; Molgaard *et al.*, 2000). The enzymes of the superfamily possess a catalytic triad consisting of the catalytic

Table 1
Summary of crystal parameters, data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Hi-Res	Low-Res
Space group	*P*1	
Unit-cell parameters (Å, °)	a = 40.7, b = 71.9, c = 93.2, α = 108.8, β = 93.3, γ = 90.4	
Data-collection and phasing statistics		
Energy (keV)	12.620	12.750
Wavelength (Å)	0.98245	0.97243
Resolution range (Å)	36.06–1.60 (1.64–1.60)	47.26–2.60 (2.69–2.60)
No. of reflections (measured/unique)	469713/125043	107337/29893
Completeness (%)	94.5 (83.7)	93.9 (62.1)
$R_{\rm merge}$†	0.047 (0.240)	0.071 (0.437)
Redundancy	3.8 (3.4)	3.6 (2.3)
Mean $I/\sigma(I)$	17.57 (4.61)	16.66 (2.78)
Molecular-replacement Corr$F$		0.274
Refinement and model statistics		
Resolution range (Å)	35.85–1.60	
Data set used in refinement	Hi-Res	
No. of reflections (total/test)	124841/6226	
$R_{\rm cryst}$‡	0.146	
$R_{\rm free}$§	0.183	
R.m.s.d. bonds (Å)	0.018	
R.m.s.d. angles (°)	1.617	
Average B factor (Å^2)	8.47	
Average solvent B factor (Å^2)	17.8	
No. of water molecules	1611	
Ramachandran plot, residues in		
Most favorable region (%)	90.8	
Additional allowed region (%)	9.2	
Generously allowed region (%)	0.0	
Disallowed region (%)	0.0	

† $R_{\rm merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection. ‡ $R_{\rm cryst} = \sum_h ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_h |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ was calculated as $R_{\rm cryst}$ using 5.0% of the randomly selected unique reflections that were omitted from structure refinement.

serine, histidine and aspartate residues. In all available structures of SGNH-hydrolase superfamily enzymes, the catalytic triad residues line up approximately perpendicular to the central β-sheet of the protein. The catalytic serine residue is positioned in a loop at the carboxy end of the first β-strand, the histidine of the catalytic triad is located in a loop preceding the carboxy-terminal helix and the aspartate residue resides on the loop just three amino acids downstream of the histidine. In addition, conserved glycine and asparagine residues are involved in formation of the active site by providing hydrogen bonds to the oxyanion hole of the enzymes. Furthermore, a unique hydrogen-bond network which stabilizes the catalytic site and is conserved in the structurally characterized members of the SGNH-hydrolase superfamily has recently been described (Lo *et al.*, 2003).

Here, we report the three-dimensional structure of At4g34215 protein at 1.6 Å. We show that the fold of At4g34215 is similar to that of the SGNH-hydrolase superfamily enzymes. Also, we discuss how the active site of At4g34215 (and likely all other carbohydrate esterase family 6 enzymes, which are closely related to At4g34215 by sequence) differs from the active site described in structurally characterized members of the SGNH-hydrolase superfamily. The structure was determined under the National Institutes of Health NIGMS Protein Structure Initiative.

2. Materials and methods

The *A. thaliana* gene At4g34215 was cloned, and native and selenomethionyl proteins were purified following the standard Center for Eukaryotic Structural Genomics (CESG) pipeline protocol for cloning (Thao *et al.*, 2004), protein expression (Sreenath *et al.*, 2005), protein purification (Jeon *et al.*, 2005) and overall information management (Zolnai *et al.*, 2003). Crystals of At4g34215 were grown by the hanging-drop method from 10 mg ml^{-1} protein solution in buffer (50 m*M* NaCl, 3 m*M* NaN_3, 0.3 m*M* TCEP, 5 m*M* Bis-Tris pH 6.0) mixed with an equal amount of well solution containing 17% PEG 4000, 20 m*M* KNO_3, 50 m*M* MES, 50 m*M* sodium acetate pH 5.5 at 277 K. Crystals grew as thin plates with significant morphological defects, including twisting. The native crystals of At4g34215 belong to space group *P*1, with unit-cell parameters a = 40.7, b = 71.9, c = 93.2 Å, α = 108.2, β = 93.3, γ = 90.4°. These crystals were cryoprotected by soaking in a solution containing 24% PEG 4000, 20 m*M* KNO_3, 50 m*M* MES, 50 m*M* sodium acetate pH 5.5 at 277 K supplemented with increasing concentrations of glycerol up to a final concentration of 18%. Selenomethionyl crystals grew from conditions similar to those of the native crystals; however, no diffraction-quality selenomethionyl crystals were obtained. X-ray diffraction data of the native crystals extending to 2.6 and 1.6 Å were collected at the SER-CAT and GM/CA-CAT beamlines at the Advanced Photon Source of Argonne National Laboratory, respectively. The diffraction images were integrated and scaled using *HKL*2000 (Otwinowski & Minor, 1997). The crystal structure of At4g34215 was solved by molecular replacement using *MOLREP* from the *CCP*4 suite (Vagin & Teplyakov, 1997; Collaborative Computational Project, Number 4, 1994) against the 2.6 Å diffraction data set using a homology model created by the *Phyre* server (http://www.sbg.bio.ic.ac.uk/~phyre/) based on a crystal structure of a putative acetylxylan esterase from *C. acetobutylicum* (PDB code 1zmb). The homology model showed a higher signal-to-noise ratio during rotational search than the original acetylxylan esterase model. Four monomers of At4g34215 were identified and placed in the asymmetric unit of the crystal. The structure was completed using alternate cycles of manual building in *Xfit* and refinement in *REFMAC*5 (McRee, 1999; Murshudov *et al.*, 1997). Tight positional and thermal restraints between the four monomers in the asymmetric unit were applied during the refinement. Also, a TLS refinement of four groups corresponding to the individual monomers was used. All refinement steps were monitored using an $R_{\rm free}$ value based on 4.9% of the independent reflections. After the structure refinement against the 2.6 Å data set was completed, a data set extending to 1.6 Å was obtained. Refinement against this data set was completed following the protocol established for the lower resolution data set. The main difference in the refinement protocol was the use of medium positional and thermal NCS restraints instead of strict restraints. Water

molecules were placed automatically using *ARP_waters* in peaks greater than 3.0σ in difference maps and within hydrogen-bonding distance of N or O atoms of the protein and other solvent molecules. The stereochemical quality of the final model was assessed using *PROCHECK* and *MolProbity* (Laskowski *et al.*, 1993; Lovell *et al.*, 2003). Refined coordinates were deposited in the RCSB Protein Data Bank (Berman *et al.*, 2000) with accession number 2apj. The figures were prepared using *PyMol* (DeLano, 2002) and *TopDraw* (Bond, 2003) based on a topology analysis of At4g34215 structure by the *TOPS* server (Westhead *et al.*, 1999).

3. Results and discussion

The structure of At4g34215 has been refined to a resolution of 1.6 Å. Data-collection, refinement and model statistics are summarized in Table 1. The final model describes four monomers, containing residues 17–260 in molecule *A*, residues 18–260 in molecule *B*, residues 20–260 in molecule *C* and residues 20–106 and 110–260 in molecule *D*. In addition, 1611 water molecules were built into the final model. In each monomer, Ser31 has been covalently modified by phenylmethylsulfonyl fluoride (PMSF) to yield *O*-benzylsulfonylserine. Monomers *A* and *B* are essentially identical. Monomer *C* shows a slight variation in the local conformation of the loop harboring Cys148 and displacement of the loop that spans residues 40–48. Finally, in monomer *D* the loop that spans residues 40–48 shows a displacement in a direction opposite to that seen in monomer *C*. The main-chain C$^{\alpha}$ atoms of His44 in the monomers *C* and *D* are separated by 6 Å in the overlapped structures.

The three-dimensional structure of At4g34215 revealed that this protein belongs to the α/β-class of proteins with a three-layer $\alpha\beta\alpha$-sandwich architecture (see Fig. 1*b*) and Rossmann fold topology (CATH code 3.40.50; Pearl *et al.*, 2005). The central feature of the structure is a seven-stranded β-sheet with *CBDAEFG* topology formed by six parallel β-strands (*B*, *D*, *A*, *E*, *F* and *G*) and strand *C*, which is antiparallel to strand *B*. On one side of the central β-sheet are α-helices *H*4, *H*6 and *H*7, which run in an approximately parallel direction to the β-strands (see Fig. 1*b*), and two single-turn 3_{10}-helices *H*3 and *H*5. On the other side are α-helices *H*2, *H*8 and a single-turn 3_{10}-helix *H*1. The α-helices *H*2 and *H*8 are in close contact with each other and are located at the concave surface formed by the parallel strands of the central β-sheet. Also located on this face of the central β-sheet is a two-stranded β-sheet formed by strands *a* and *b*, which extends from the otherwise globular At4g34215. Several longer loops that do not adopt a regular secondary structure connect helices *H*2, *H*8, *H*1 and β-strands *a* and *b* to the central β-sheet.

To classify the fold of At4g34215, a structural homology search was conducted using the *DALI* and *VAST* servers (Holm & Sander, 1993; Madej *et al.*, 1995). Both *DALI* and *VAST* identified a range of structural homologs of At4g34215. *DALI* identified 286 structural homologs with a structural similarity score $Z > 2$, three of which had a score $Z > 10$. *VAST*

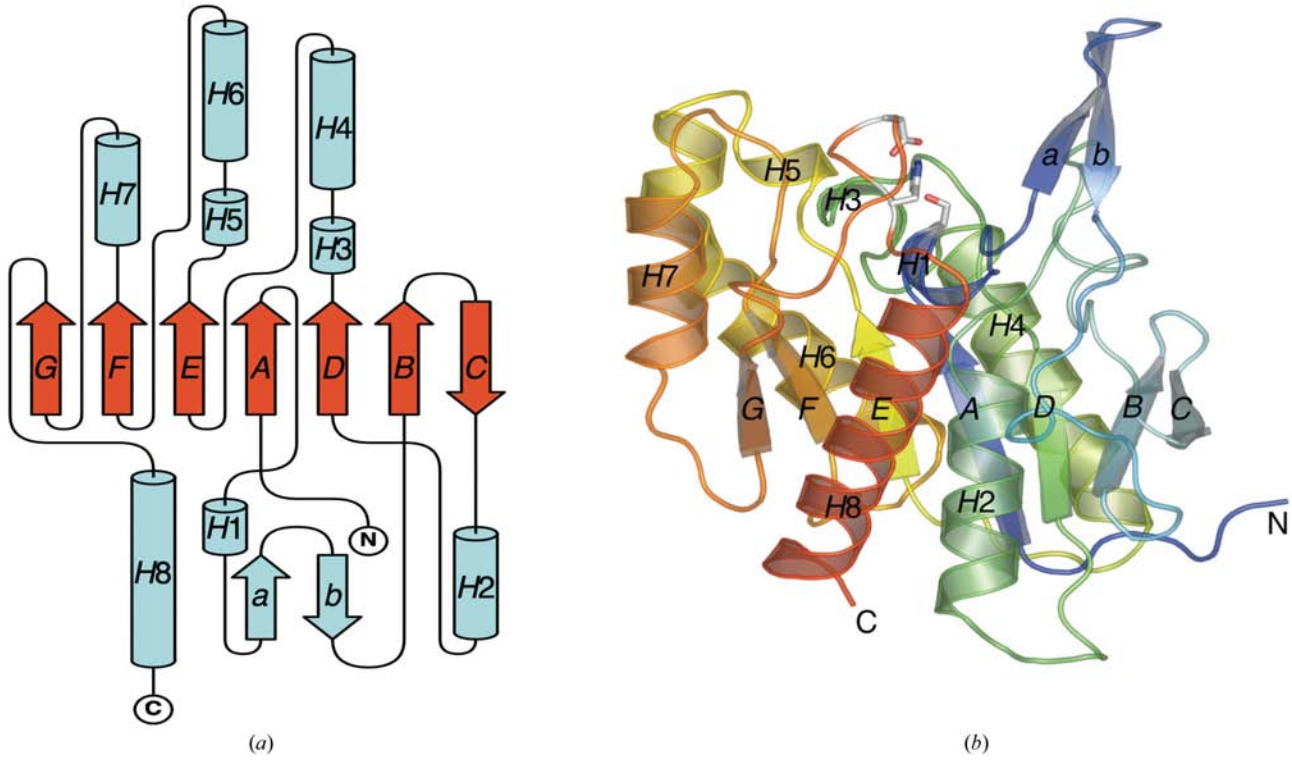

Figure 1
(*a*) A topology diagram of the At4g34215 structure (PDB code 1apj). The central seven-stranded β-sheet (red arrows) is surrounded by several helices (cyan cylinders) and the auxiliary two-stranded β-sheet comprising β-strands *ab* (cyan arrows). (*b*) A ribbon diagram of the At4g34215 structure with rainbow coloring from amino-terminus (blue) to carboxy-terminus (red). The catalytic triad residues Ser31, His238 and Asp235 are depicted in stick representation. The structure is labeled to match the topology diagram.

identified 1299 structural neighbors with non-identical sequences. Both servers identified the structure of a putative acetylxylan esterase from *C. acetobutylicum* (PDB code 1zmb) as the closest homolog of At4g34215. Specifically, *DALI* calculated for this protein pair a similarity score Z of 28.6, an r.m.s.d. of 1.6 Å and 30% sequence identity over 208 aligned C^{α} residues. *VAST* aligned this pair with the following statistics: a *VAST* score of 19.1, an r.m.s.d. of 1.3 Å and 29.7% sequence identity over 209 aligned residues. Additional significant structural homologs identified by *DALI* (with $Z > 10$) were *S. scabies* esterase with a similarity score Z of 13.6, an r.m.s.d. of 2.6 Å and 10% sequence identity over 176 aligned C^{α} residues (PDB code 1esc; Wei *et al.*, 1995) and platelet-activating factor acetylhydrolase from *B. taurus* with a similarity score Z of 11.0, an r.m.s.d. of 2.9 Å and 13% sequence identity over 152 aligned C^{α} residues (PDB code 1wab; Ho *et al.*, 1997). *VAST* identified a large number of additional weaker structural homologs with similar *VAST* scores. The top two of them are *E. coli* thioesterase I with *VAST* score 14.4, r.m.s.d. 2.9 Å and 11% sequence identity over 163 aligned residues (PDB code 1ivn; Lo *et al.*, 2003) and carbonyl reductase Sniffer from *Drosophila melanogaster* with *VAST* score 14.3, r.m.s.d. 3.5 Å and 12% sequence identity over 142 aligned residues (PDB code 1sny; Sgraja *et al.*, 2004). In summary, the several closest structural homologs of At4g34215 identified by *DALI* and *VAST* are members of the SGNH-hydrolase superfamily of enzymes.

The overall folds of At4g34215 and of a putative acetylxylan esterase from *C. acetobutylicum* (CAC0529) are very similar. In fact, this allowed us to successfully solve the structure of At4g34215 by molecular replacement. However, there are several significant differences between the two structures (see Fig. 2). The amino-terminus of CAC0529 is 20 residues shorter than that of At4g34215. Residues 1–16 of At4g34215 may be highly flexible as no electron density corresponding to this segment could be located in the crystallographic maps. CAC0529 has a significantly longer carboxy-terminus; residues 227–284 form an extension wrapping approximately a third of the way around the protein core and then form a two-helix bundle that runs parallel to the core. The secondary-structural elements of the two proteins align very well. The loops connecting these elements show slight to large variations. Specifically, the three most notable differences are (i) the At4g34215 loop that spans residues 33–52 contains a unique two-stranded β-sheet *ab* (see Fig. 1*b*) while the structurally equivalent area of CAC0529 is shorter, formed by residues 18–27, (ii) the CAC0529 loop spanning residues 159–170, located next to the catalytic triad aspartate, is much longer than that in At4g34215 (residues 201–204) and (iii) the At4g34215 loop spanning residues 82–89 located close to the catalytic serine contains an additional three residues compared with the corresponding loop spanning residues 47–51 in CAC0529.

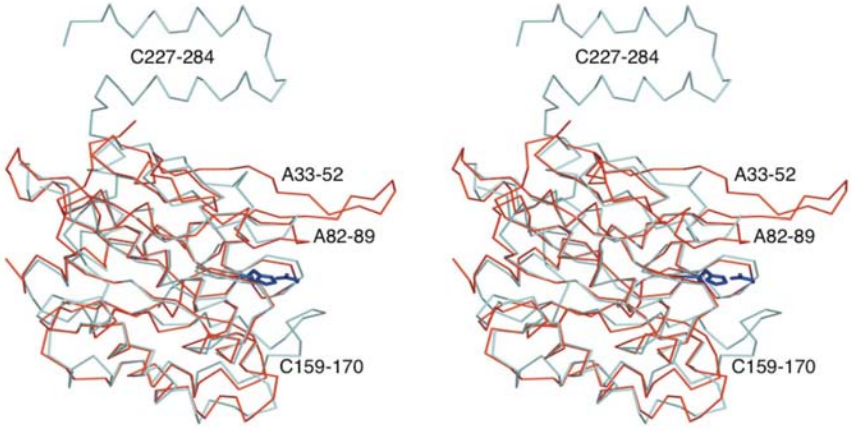

Figure 2
Structural superposition of *A. thaliana* At4g34215 (red; PDB code 2apj) and CAC0529 (cyan; PDB code 1zmb). The catalytic triad residues Ser31, His238 and Asp235 of At4g34215 are depicted in stick representation (blue). The carboxy-terminus of CAC0529 and surface loops showing the most notable differences are labeled for convenience.

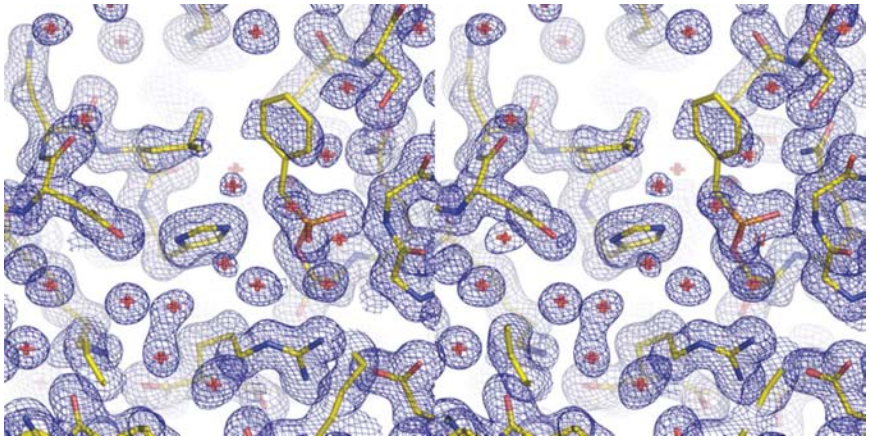

Figure 3
A stereo representation of the *A. thaliana* At4g34215 electron-density map depicted at the 1σ contour level and the final refined model (yellow sticks). The active-site catalytic triad consists of Ser31 (right of His238), His238 (center) and Asp235 (left of His238). Ser31 is covalently modified by the protease inhibitor PMSF; the sulfonate portion of the inhibitor is very well defined in the electron density. The sulfonate O atom that mimics an oxyanion of the transition state is stabilized by three hydrogen bonds from the main-chain amide N atoms of the catalytic Ser31 and the conserved Gly122 (right of sulfonate) and the amide N atom of Gln31.

The catalytic triad of At4g3215 was identified in the structure and contains residues Ser31, His238 and Asp235. The spatial arrangement of the catalytic triad and the position of individual residues of the triad in the sequence of At4g34215 is consistent with the pattern expected for SGNH-hydrolase superfamily enzymes (Molgaard *et al.*, 2000). The catalytic Ser31 was identified as interesting early in the

protein refinement because this residue was covalently modified in every At4g34215 monomer in the asymmetric unit. Our best hypothesis is that the protease inhibitor phenylmethylsulfonyl fluoride used during the protein purification trapped Ser31, the catalytic residue of At4g34215. The sulfonate group is very well defined in the electron-density map; the density for the benzyl group is less convincing as this group is probably rotationally disordered (see Fig. 3).

Another important feature of enzymes in the SGNH-hydrolase superfamily is the conservation of residues involved in the formation of an oxyanion hole. The conserved glycine from the so-called block II of conserved residues defined for members of the SGNH-hydrolase superfamily (Upton & Buckley, 1995; Lo *et al.*, 2003) and asparagine from block III provide hydrogen bonds to the oxyanion hole (see Fig. 4*b*). In fact, the crucial role of these two residues is reflected in the name of the SGNH-hydrolase superfamily: S stands for the catalytic serine, G and N stand for the conserved glycine and asparagine involved in the formation of the oxyanion hole and H represents the catalytic triad histidine. The order of these residues reflects their consecutive positions in the sequence of the SGNH-hydrolase superfamily enzymes. The third residue of the catalytic triad, Asp, is not represented in the abbreviation because it is not universally conserved in the superfamily. *S. scabies* esterase (PDB code 1esc; Wei *et al.*, 1995) is an example of an enzyme from the SGNH-hydrolase superfamily with a catalytic dyad Ser-His, instead of a catalytic triad. In this case, the proper orientation of the active-site imidazole was found to be maintained by a hydrogen bond between the N^{δ}—H group of the catalytic histidine and a main-chain O atom of a residue in a position usually occupied by the side chain of the third member of the catalytic triad.

The structure of At4g34215 reveals an important variation in the active-site residues responsible for the formation of the oxyanion hole. While the conserved glycine from block II is present in At4g34215 (Gly122), the conserved asparagine found in block III of SGNH-hydrolase superfamily enzymes is missing. Instead, Ser162 is located at the equivalent position in At4g34215. Upon closer inspection of the active site, it became clear that Gln30 functionally substitutes for the missing asparagine and provides a hydrogen bond to the oxyanion hole of At4g34215 (see Fig. 4*a*). The amide N atom of Gln30 is found 2.85 Å from the sulfonate O atom of *O*-benzylsulfonyl serine that mimics the transition-state oxyanion. A correct spatial positioning of Gln30 is achieved through a network of hydrogen bonds that may involve hydrogen bonds between the amide O atom of Gln30 and any of the backbone amides of residues Gly160, Glu161 or Ser162 or the amide N atom of Gln159 (2.93, 3.02, 3.01 or 3.12 Å, respectively). Also, the hydroxyl O atom of Ser162 can form a hydrogen bond with the amide N atom of Gln30 (2.89 Å, see Fig. 4*a*).

At4g34215 is a member of carbohydrate esterase family 6 (CE6; Coutinho & Henrissat, 1999*a*). A multiple sequence alignment of 15 members of this family revealed 14 fully conserved residues. Fig. 5 provides a representative portion of this alignment. Only six members of the CE6 family are shown; however, the consensus sequence is based on the alignment of all 15 members of the family. Immediately obvious from the alignment are the two stretches of fully conserved residues: GQSNM-G and QGE-(D/N). The former segment contains the catalytic Ser31 and the above-mentioned Gln30, which is involved in the formation of the oxyanion hole. The full conservation of the latter segment, containing residues Gln159, Gly160 and Glu161, lends further credence to our hypothesis that these residues are crucial for the proper positioning of Gln30 through a hydrogen-bond network. Inspection of the recently determined structure of another member of the CE6 family, CAC0529, confirmed that the local conformation and the stabilizing hydrogen-bond network formed by these residues is conserved. Also, a residue corresponding to Ser162 in At4g34215 is highly conserved in the

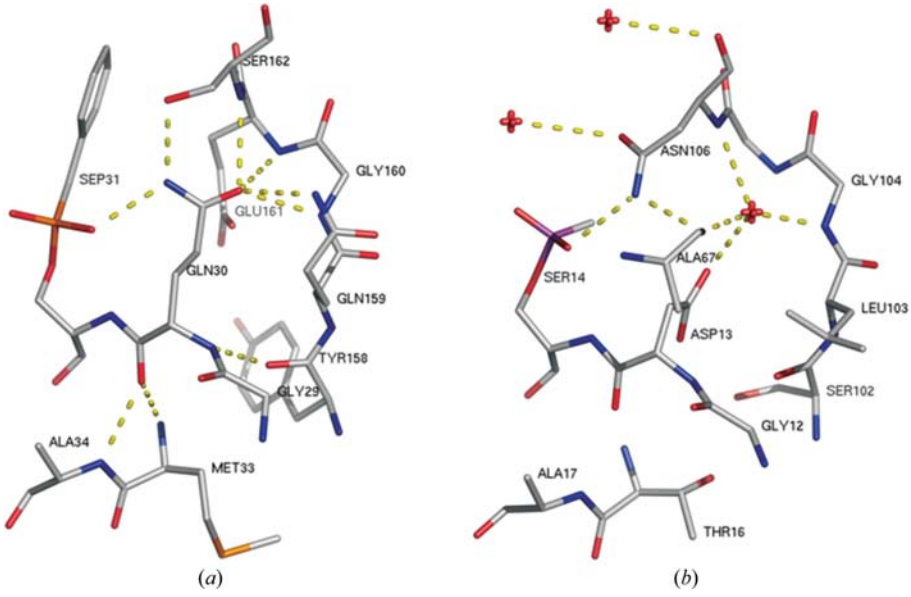

Figure 4
(*a*) A portion of the active site of At4g34215 involved in formation of the oxyanion hole. Gln30, which is fully conserved among proteins in the carbohydrate esterase 6 family (see Fig. 5), forms the crucial hydrogen bond to the oxygen of *O*-benzylsulfonyl serine (Sep31) that mimics the transition-state oxyanion. The network of (possible) hydrogen bonds (yellow dashed lines) to nearby residues including Gln159, Gly160, Glu161 and Ser162 positions Gln30 in the catalytically competent conformation. (*b*) A portion of the active site of a representative SGNH-superfamily enzyme involved in a formation of the oxyanion hole. The figure is based on the structure of serine esterase from *S. scabies* (PDB code 1esd; Wei *et al.*, 1995). The conserved Asn106 makes the crucial hydrogen bond to the O atom of methylphosphic serine (Ser14), which is mimicking the transition-state oxyanion. The local conformation in the vicinity of Asp106 is stabilized by a hydrogen-bond network (yellow dashed lines) that involves residues Asn106, Gly104, Asp14 and Ala67 and a conserved water molecule.

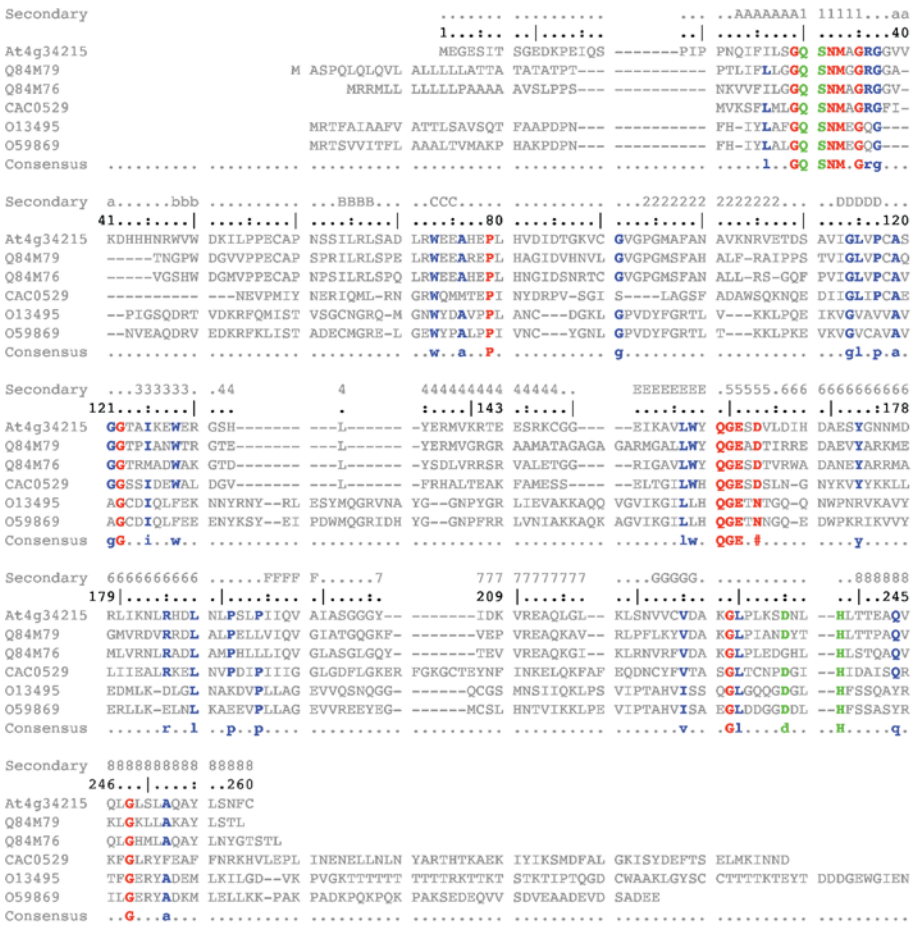

Figure 5
A sequence alignment of representative members of carbohydrate esterase family 6 (CE6). Fully conserved residues (>95% conservation) are shown in red; conserved residues (70–95% conservation) are highlighted in blue. The residues of the catalytic triad and Gln30 are depicted in green. The alignment is numbered according to the At4g34215 sequence. The top line presents the structure-based secondary-structure assignment of individual amino acids of At4g34215 in a notation consistent with that introduced in Fig. 1. The consensus sequence is based on the alignment of 15 currently annotated members of the CE6 family (Coutinho & Henrissat, 1999*b*; Bairoch *et al.*, 2005), including At4g34215 of *A. thaliana* (UniProt code Q8L9J9), ABC0603 of *Bacillus clausii* KSM-K16 (GenPept code BAD63143.1), BLi02801 of *B. licheniformis* DSM 13ATCC 14580 (GenPept code AAU41672.1), BT4180 of *Bacteroides thetaiotaomicron* VPI-5482 (GenPept code AAO79285.1), CAC0529 of *Clostridium acetobutylicum* ATCC 824 (UniProt code Q97LM8), acetylxylan esterase A of *Fibrobacter succinogenes* S85 (GenPept code AAG36766.1), acetylxylan esterase BnaI of *Neocallimastix patriciarum* (UniProt code O13495), acetylxylan esterase A of *Orpinomyces* sp. PC-2 (UniProt code O59869), OSJNBa0009K15.9 of *Oryza sativa* (GenPept code CAE05089.3), OSJNBa0027H06.22 of *O. sativa* (GenPept code CAD39440.1), OSJNBa0059G06.14 of *O. sativa* (UniProt code Q84M79), OSJNBa0059G06.16 of *O. sativa* (UniProt code Q84M76), RB5007 of *Rhodopirellula baltica* SH 1 (GenPept code CAD78234.1), RB763 of *R. baltica* SH 1 (GenPept code CAD71736.1) and SSP2193 of *Staphylococcus saprophyticus* ATCC 15305 (GenPept code BAE19338.1).

other members of the CE6 family. We hypothesize that the hydrogen bond between the hydroxyl O atom of serine or threonine at this position and the amide N atom of Gln30 provides further stabilization of the portion of the active site involved in the formation of the oxyanion hole.

In conclusion, the crystal structure of At4g34215 confirmed that this protein belongs to the SGNH-hydrolase superfamily of enzymes. The catalytic triad of this enzyme consists of residues Ser31, His238 and Asp235. An important variation in the formation of the oxyanion hole involving Gln30 and nearby conserved residues Gln159, Gly160, Glu161 and Ser162 was found in this enzyme and a related CAC0529 protein. This structural feature is likely to be conserved in all members of the carbohydrate esterase family 6.

We acknowledge financial support from NIH National Institute for General Medical Sciences grants P50 GM64598 and U54 GM074901. Use of the Advanced Photon Source was supported by the US Departmentof Energy, Basic Energy Sciences, Office of Science under contract No. W-31-109-ENG-38. Data were collected at GM/CA CAT, which has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Data were also collected at Southeast Regional Collaborative Access Team (SERCAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org/members.html. Special thanks goes to all members of the

CESG team, including Todd Kimball, John Kunert, Nicholas Dillon, Rachel Schiesher, Juhyung Chin, Megan Riters, Andrew C. Olson, Jason M. Ellefson, Janet E. McCombs, Brendan T. Burns, Blake W. Buchan, Holalkere V. Geetha, Zhaohui Sun, Ip Kei Sam, Eldon L. Ulrich, Janelle Warrick, Bryan Ramirez, Zsolt Zolnai, Peter T. Lee, Jianhua Zhang, David J. Aceti, Russell L. Wrobel, Ronnie O. Frederick, Hassan Sreenath, Frank C. Vojtik, Won Bae Jeon, Craig S. Newman, John Primm, Michael R. Sussman, Brian G. Fox and John L. Markley.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L. S. (2005). *Nucleic Acids Res.* **33**, D154–D159.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coutinho, P. & Henrissat, B. (1999*a*). *Recent Advances in Carbohydrate Bioengineering*, edited by H. Gilbert, G. Davies, B. Henrissat & B. Svensson, pp. 3–12. Cambridge: Royal Society of Chemistry.

Coutinho, P. & Henrissat, B. (1999*b*). *Carbohydrate-Active Enzymes Server.* http://afmb.cnrs-mrs.fr/CAZY/.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

Gough, J., Karplus, K., Hughey, R. & Chothia, C. (2001). *J. Mol. Biol.* **313**, 903–919.

Ho, Y. S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., Inoue, K. & Derewenda, Z. S. (1997). *Nature (London)*, **385**, 89–93.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Jeon, W., Aceti, D. J., Bingman, C., Vojtik, F., Olson, A., Ellefson, J., McCombs, J., Sreenath, H., Blommel, P., Seder, K., Buchan, B., Burns, B., Geetha, H., Harms, A., Sabat, G., Sussman, M., Fox, B. & Phillips, G. (2005). *J. Struct. Funct. Genomics*, **6**, 143–147.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lo, Y. C., Lin, S. C., Shaw, J. F. & Liaw, Y. C. (2003). *J. Mol. Biol.* **330**, 539–551.

Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Madej, T., Gibrat, J. F. & Bryant, S. H. (1995). *Proteins*, **23**, 356–369.

Molgaard, A., Kauppinen, S. & Larsen, S. (2000). *Structure*, **8**, 373–383.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pearl, F. *et al.* (2005). *Nucleic Acids Res.* **33**, D247–D251.

Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C. H., Meier-Ewert, H., Skehel, J. J. & Wiley, D. C. (1998). *Nature (London)*, **396**, 92–96.

Sgraja, T., Ulschmid, J., Becker, K., Schneuwly, S., Klebe, G., Reuter, K. & Heine, A. (2004). *J. Mol. Biol.* **342**, 1613–1624.

Sreenath, H. K., Bingman, C. A., Buchan, B. W., Seder, K. D., Burns, B. T., Geetha, H. V., Jeon, W. B., Vojtik, F. C., Aceti, D. J., Frederick, R. O., Phillips, G. N. Jr & Fox, B. G. (2005). *Protein Expr. Purif.* **40**, 256–267.

Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P. G., Riters, M., Newman, C. S., Fox, B. G. & Wrobel, R. L.(2004). *J. Struct. Funct. Genomics*, **5**, 267–276.

Upton, C. & Buckley, J. T. (1995). *Trends Biochem. Sci.* **20**, 178–179.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Wei, Y., Schottel, J. L., Derewenda, U., Swenson, L., Patkar, S. & Derewenda, Z. S. (1995). *Nature Struct. Biol.* **2**, 218–223.

Westhead, D. R., Slidel, T. W., Flores, T. P. & Thornton, J. M. (1999). *Protein Sci.* **8**, 897–904.

Zolnai, Z., Lee, P. T., Li, J., Chapman, M. R., Newman, C. S., Phillips, G. N. Jr, Rayment, I., Ulrich, E. L., Volkman, B. F. & Markley, J. L. (2003). *J. Struct. Funct. Genomics*, **4**, 11–23.