The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst

Maria-Despoina Charavgi,[a,b] Maria Dimarogona,[a,b] Evangelos Topakas,[b] Paul Christakopoulos[b] and Evangelia D. Chrysina[a]*

[a]Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, 48 Vassileos Constantinou Avenue, 11635 Athens, Greece, and [b]School of Chemical Engineering, National Technical University of Athens, Iroon Polytechniou Street, Zografou Campus, 15700 Athens, Greece

Correspondence e-mail: echrysina@eie.gr, echry@tee.gr

The increasing demand for the development of efficient biocatalysts is a consequence of their broad industrial applications. Typical difficulties that are encountered during their exploitation in a variety of processes are interconnected with factors such as temperature, pH, product inhibitors *etc.* To eliminate these, research has been directed towards the identification of new enzymes that would comply with the required standards. To this end, the recently discovered glucuronoyl esterases (GEs) are an enigmatic family within the carbohydrate esterase (CE) family. Structures of the thermophilic *St*GE2 esterase from *Myceliophthora thermophila* (synonym *Sporotrichum thermophile*), a member of the CE15 family, and its S213A mutant were determined at 1.55 and 1.9 Å resolution, respectively. The first crystal structure of the S213A mutant in complex with a substrate analogue, methyl 4-*O*-methyl-β-D-glucopyranuronate, was determined at 2.35 Å resolution. All of the three-dimensional protein structures have an α/β-hydrolase fold with a three-layer $\alpha\beta\alpha$-sandwich architecture and a Rossmann topology and comprise one molecule per asymmetric unit. These are the first crystal structures of a thermophilic GE both in an unliganded form and bound to a substrate analogue, thus unravelling the organization of the catalytic triad residues and their neighbours lining the active site. The knowledge derived offers novel insights into the key structural elements that drive the hydrolysis of glucuronic acid esters.

1. Introduction

Lignocellulose in vascular plant cell walls is composed of cellulose, hemicellulose and lignin in relative proportions that vary according to the plant origin (Reddy & Yang, 2005). A selection of diverse processes, ranging from simple burning to advanced bioconversion, has been applied to access the energy stored in the cell-wall polymers. One of the challenges that researchers face today is to make this process cost-competitive in the biofuels market (Himmel *et al.*, 2007), mainly by overcoming the recalcitrance of biomass. This effort becomes even more demanding owing to the complex structure of the plant cell wall, leading to an increased cost of lignocellulosic conversion (Weng *et al.*, 2008).

Carbohydrate esterases (CEs) are employed as potent biocatalysts for the reduction of the protein load required for the breakdown of lignocellulose to fermentable sugars. One of the recently described CEs, glucuronoyl esterase (GE), has been suggested to play an important role in the dissociation of lignin from hemicellulose and cellulose by cleaving the ester bonds between the aromatic alcohols of lignin and the

63

carboxyl groups of 4-*O*-methyl-D-glucuronic acid residues in glucuronoxylan (Špániková & Biely, 2006; Ďuranová *et al.*, 2009). GE was first discovered in the wood-rot fungus *Schizophyllum commune* (Špániková & Biely, 2006), while the first reported amino-acid sequence was from the *Hypocrea jecorina* GE Cip2_GE (Li *et al.*, 2007). The latter launched the emerging CE15 family deposited in the continuously updated Carbohydrate-Active Enzymes database (CAZy; http://www.cazy.org/; Cantarel *et al.*, 2009). To date, six members of this family have been purified and characterized using a series of new synthetic substrates comprising methyl esters of uronic acid and their glycosides. The data obtained showed that the methyl ester of 4-*O*-methyl-D-glucuronic acid was hydrolyzed more efficiently by all GEs examined compared with the corresponding ester without the 4-*O*-methyl group (Fig. 1). This finding further supported the potential significance of the methoxy group in substrate recognition (Ďuranová *et al.*, 2009). To date, the only three-dimensional structure of the CE15 family available is that of Cip2_GE, which confirms the triad arrangement of the putative catalytic residues Ser-His-Glu (Pokkuluri *et al.*, 2011). Here, we report the three-dimensional structures of a recombinant thermophilic GE from *Myceliophthora thermophila* (synonym *Sporotrichum thermophile*; *St*GE2) and its S213A mutant determined at 1.55 and 1.90 Å resolution, respectively. We also present for the first time the crystal structure of a GE in complex with a substrate analogue, methyl 4-*O*-methyl-β-D-glucopyranuronate (MCU), bearing the methoxy group of interest, at 2.35 Å resolution.

2. Materials and methods

2.1. Production and purification of recombinant *St*GE2

Recombinant *St*GE2 and its S213A mutant were produced in *Pichia pastoris* and subsequently purified using immobilized metal-ion affinity chromatography (IMAC) as described previously. The S213A mutant exhibited a complete loss of enzyme activity towards methyl 4-*O*-methyl-D-glucopyranuronate (Topakas *et al.*, 2010). The homogeneity of the purified samples were assessed by SDS–PAGE using a 12.5% polyacrylamide gel. A single band corresponding to a molecular mass of 43 kDa was observed, indicating that the protein samples were suitable for crystallization trials.

2.2. Sequence analysis

Acknowledging the scarcity of clear crystallographic evidence regarding the putative catalytic site of *St*GE2 among
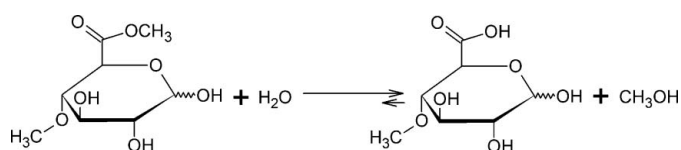
Figure 1
Hydrolysis of methyl 4-*O*-methyl-D-glucopyranuronate catalyzed by *St*GE2.

CE15 family members, a broad sequence analysis was performed using *Basic Local Alignment Search Tool* (*BLAST*) from the National Center for Biotechnology Information (NCBI) data bank (Altschul *et al.*, 1997). Multiple sequence alignment of homologous enzymes was performed with *ClustalW*2 (Larkin *et al.*, 2007) on the EBI server (http://www.ebi.ac.uk/Tools/clustalw2/) and the results were visualized using *ESPript*2.2 (Gouet *et al.*, 1999; Fig. 2). Secondary-structure assignment and analysis were performed with *PROMOTIF* (Hutchinson & Thornton, 1996) as implemented in *PDBsum* (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/; Laskowski *et al.*, 1997) on the EBI server (the results are presented as Supplementary Material[1]).

2.3. Crystallization and data collection

The protein samples were concentrated to 20 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0 buffer prior to crystallization. Initial crystallization conditions for *St*GE2 were established from the commercially available screen The PEGs Suite (Qiagen) using a protein concentration of 10 mg ml^{-1} in the drop. The crystallization experiments were carried out with the aid of an OryxNano crystallization robot (Douglas Instruments Ltd, UK) using the sitting-drop vapour-diffusion method and 96-well SWISSCI MRC crystallization plates (SWISSCI AG, Zug, Switzerland). Needle-shaped crystals grew at 289 K and at a pH ranging from 7.5 to 8.5 in the presence of PEG 2000 MME, 3350, 4000 or 6000 and 25, 30 or 35%(*w*/*v*) concentrations of the precipitating agent. Further optimization of both crystal size and quality was carried out and diffracting crystals were grown in 30%(*w*/*v*) PEG 3350, 0.1 *M* Tris–HCl pH 8.0 to average dimensions of ∼10 × 10 × 350 μm in an orthorhombic habit within one month. X-ray diffraction data were collected at 100 K using the synchrotron-radiation source at EMBL Hamburg Outstation beamline X13 (λ = 0.8123 Å). Prior to data collection, the crystals were transferred into 20%(*v*/*v*) glycerol, which was used as a cryoprotectant. A complete data set was collected to 1.55 Å resolution from a single crystal on a 225 mm MAR CCD detector using the *DNA* software package available at beamline X13. Data processing was performed with the *XDS* package (Kabsch, 2010), followed by data integration and scaling with *SCALA* from the *CCP*4 program suite (Evans, 2006). X-ray diffraction data analysis showed that the *St*GE2 crystal symmetry was consistent with space group $P2_12_12_1$, with unit-cell parameters $a = 46.0$, $b = 58.7$, $c = 136.2$ Å, $\alpha = \beta = \gamma = 90°$ and one molecule per asymmetric unit. Data-collection statistics for *St*GE2 are summarized in Table 1.

In the case of the S213A mutant, The PEGs Suite crystallization screen (Qiagen) was again used to identify crystallization conditions. Crystals grew to average dimensions of ∼30 × 30 × 90 μm within one week in the presence of 0.1 *M* sodium acetate pH 4.6, 25%(*v*/*v*) PEG 550 MME at 289 K. Diffraction data were collected using a SuperNova in-house

---

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: RR5026). Services for accessing this material are described at the back of the journal.

X-ray generator (Agilent Technologies) equipped with a 135 mm ATLAS CCD detector and a four-circle kappa goniometer at the Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation (NHRF; Cu high-intensity X-ray microfocus source, λ = 1.5418 Å) at 100 K using 15%(v/v) glycerol as cryoprotectant. Preliminary characterization showed that the S213A mutant crystals belonged to space group $P2_122_1$, with unit-cell parameters a = 52.0, b = 69.6, c = 103.8 Å, $\alpha = \beta = \gamma = 90°$ and one molecule per asymmetric unit. Complete data were collected to 1.9 Å resolution (Table 1) and were processed using the $CrysAlis^{Pro}$ software system (Agilent Technologies) followed by $SCALA$ (Evans, 2006), employing standard protocols for indexing, integration and scaling.

Structural studies of StGE2 were performed with the S213A mutant in complex with methyl 4-O-methyl-D-glucopyranuronate, which was kindly provided by Dr Peter Biely, Institute of Chemistry of the Slovak Academy of Sciences, Bratislava, Slovakia. This substrate analogue exhibited affinity for StGE2, as shown previously (Topakas $et~al.$, 2010). Crystals of the S213A mutant were soaked in 5 mM methyl 4-O-methyl-D-glucopyranuronate dissolved in the mother liquor (Figs. 3a and 3b) for 1 h. A longer soaking time resulted in crystal deterioration, preventing data collection. The soaked crystals were tested at 100 K and complete diffraction data were collected to 2.35 Å resolution using the in-house X-ray source at NHRF. The crystal of the complex remained isomorphous to the crystal of the unliganded S213A mutant after soaking, with unit-cell parameters a = 52.1, b = 69.8, c = 103.9 Å, $\alpha = \beta = \gamma = 90°$. Data processing and scaling was conducted as described for the S213A mutant and the statistics are outlined in Table 1.

2.4. Structure determination and refinement

The three-dimensional structure of StGE2 was determined by molecular replacement using the structure of Cip2_GE (Pokkuluri $et~al.$, 2011; PDB entry 3pic) as a starting model and the $Phaser$ crystallographic software (McCoy $et~al.$, 2007) as implemented in CCP4 (Winn $et~al.$, 2011). The final model was obtained by alternating rounds of anisotropic temperature-factor refinement of all atoms with $REFMAC$ (Murshudov $et~al.$, 2011) and manual model building with $Coot$ (Emsley $et~al.$, 2010). Water molecules that fulfilled the criteria of forming direct or water-mediated hydrogen-bond interactions with the protein were incorporated into the model also
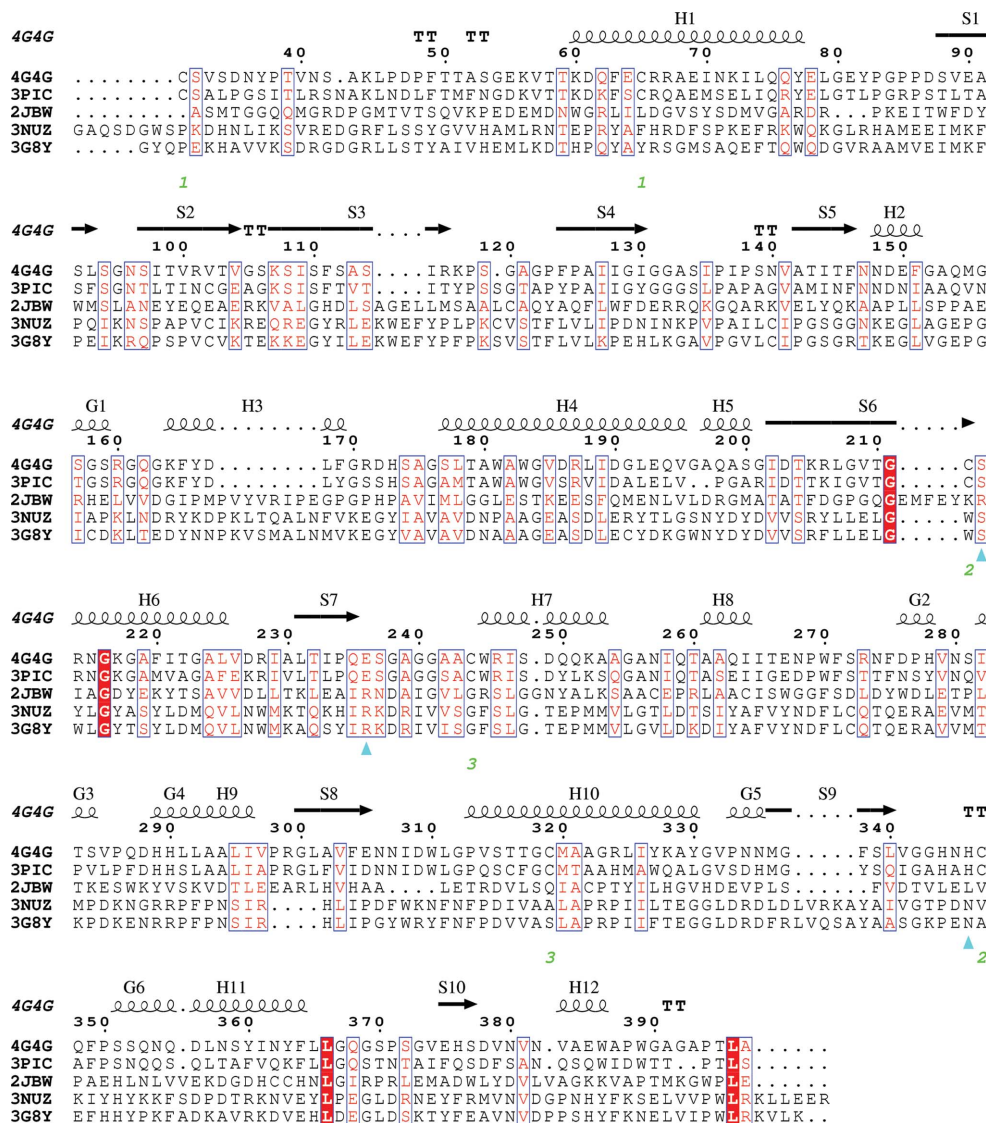
Figure 2
Sequence alignment of StGE2 (PDB entry 4g4g) with the best sequence and structural homologues: Cip2_GE (the closest sequence and structural homologue, with 56% identity over 97% sequence coverage and a Z-score of 59.1; PDB entry 3pic; Pokkuluri $et~al.$, 2011), 2,6-dihydroxy-pseudo-oxynicotine hydrolase (the second closest sequence homologue, with 28% identity over 37% sequence coverage and a Z-score of 16.2; PDB entry 2jbw; Schleberger $et~al.$, 2007), putative acetylxylan esterase (a structural homologue with a Z-score of 18.8; PDB entry 3nuz; Joint Centre for Structural Genomics, unpublished work) and SusD/RagB-associated esterase-like protein (a structural homologue with a Z-score of 18.8; PDB entry 3g8y; Joint Centre for Structural Genomics, unpublished work). Identical and similar residues are shown in white on a red background and in red on a white background, respectively. The disulfide bonds (numbers in green) and the residues of interest (cyan triangles) are also indicated on the same line. The secondary-structure elements α-helices, 3_{10}-helices, β-strands and strict β-turns are denoted H, G, S and TT, respectively.

Table 1
Diffraction data and refinement statistics for *St*GE2, the S213A mutant and the S213A mutant in complex with MCU.

Values in parentheses are for the outermost shell.

	*St*GE2	S213A mutant	S213A mutant complex
Data-collection and processing statistics			
Source	MX beamline X13, EMBL/DESY	In-house, SuperNova, Agilent Technologies, NHRF/IBMCB	In-house, SuperNova, Agilent Technologies, NHRF/IBMCB
Wavelength (Å)	0.8123	1.5418	1.5418
No. of images	299	210	184
Oscillation range (°)	0.3	0.5	0.5
Space group	$P2_12_12_1$	$P2_122_1$	$P2_122_1$
Unit-cell parameters (Å)	$a = 46.0$, $b = 58.7$, $c = 136.2$	$a = 52.0$, $b = 69.6$, $c = 103.8$	$a = 52.1$, $b = 69.8$, $c = 103.9$
No. of molecules in asymmetric unit	1	1	1
No. of observations	190519 (25955)	97087 (8291)	50703 (6741)
No. of unique reflections	51271 (7498)	30113 (4161)	15099 (2119)
Resolution range (Å)	19.45–1.55 (1.63–1.55)	13.97–1.90 (2.00–1.90)	13.48–2.35 (2.48–2.35)
Completeness (%)	94.5 (95.9)	99.0 (95.6)	92.8 (90.4)
R_{merge}†	0.069 (0.498)	0.080 (0.210)	0.110 (0.150)
$\langle I/\sigma(I) \rangle$	9.5 (2.5)	11.4 (3.0)	7.7 (5.3)
Multiplicity	3.7 (3.5)	3.2 (2.0)	3.4 (3.2)
B factor (Å^2) from Wilson plot	13.5	8.5	14.8
Refinement statistics and model quality			
Resolution range (Å)	19.45–1.55	13.97–1.90	13.48–2.35
No. of reflections	48582	28557	14330
Residues included	31–397	29–397	29–397
No. of protein atoms	2751	2760	2736
No. of heteroatoms			
Solvent molecules	412	437	212
Ethylene glycol molecules	14	8	20
Glycerol molecules	2	3	2
MCU	—	—	1
R_{free}‡	0.254 (0.333)	0.208 (0.254)	0.245 (0.298)
R‡	0.204 (0.247)	0.169 (0.212)	0.186 (0.230)
R.m.s. deviation in			
Bond lengths (Å)	0.005	0.006	0.006
Bond angles (°)	0.967	0.997	1.006
MolProbity analysis			
Ramachandran favoured/outliers (%)	96.4/0	96.5/0	96.2/0
Poor rotamers (%)	0.7	0.35	0.4
Average B (Å^2) for protein residues			
Overall	15.5	7.6	8.7
Backbone atoms	15.0	7.5	8.7
Side-chain atoms	16.1	7.7	8.7
Average B (Å^2) for heteroatoms			
Water molecules	26.2	18.0	12.4
Ethylene glycol molecules	28.5	22.5	15.0
Glycerol molecules	31.9	19.9	16.6
MCU	—	—	13.5

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ and $I_i(hkl)$ are the mean and the *i*th measurement of intensity for reflection *hkl*, respectively. ‡ Crystallographic $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. R_{free} is the corresponding R value for a randomly chosen 5% of the reflections that were not included in the refinement.

using *Coot*. Visual inspection of the $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ electron-density maps towards the final stages of refinement revealed additional electron density that was attributed to two glycerol and 14 ethylene glycol molecules originating from the cryoprotectant solution and the crystallization medium, respectively. The structure was refined to a final R factor of 0.204 and a final R_{free} of 0.254. The refinement statistics are presented in Table 1.

The structure of *St*GE2 was in turn employed as a starting model for molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the data for the S213A mutant. Isotropic temperature-factor refinement of all atoms was carried out against experimental data using the same protocol as applied for the native structure. Eight ethylene glycol molecules and

three glycerol molecules were included in the model as suggested by both the $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ electron-density maps. The structure was refined to a final R factor of 0.169 and a final R_{free} of 0.208 (Fig. 4, Table 1).

Similarly, the structure of the S213A mutant in complex with the synthetic substrate analogue methyl 4-*O*-methyl-D-glucopyranuronate was determined by employing the mutant structure as a starting model and following a standard protocol for refinement and model building as described above for the S213A *St*GE2 structure. During the final stages of optimization a portion of extra density was observed at the putative catalytic site of S213A *St*GE2, suggesting binding of only the β anomer of MCU. The MCU model was prepared using the Dundee *PRODRG* server (http://davapc1.bioch.dundee.ac.uk/

prodrg/) and was fitted to the electron density by adjustment of its torsion angles. The model structure was subjected to further refinement (Fig. 3c). In addition, a total of 20 ethylene glycol molecules and two glycerol molecules were incorporated in the final model, which was refined to an R factor of 0.186 and an R_{free} of 0.245 (Table 1).

2.5. Structure analysis

The stereochemistry of the protein residues was validated using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010). The potential hydrogen-bond and van der Waals interactions formed upon the binding of MCU were calculated using the program *CONTACT* (Winn *et al.*, 2011) applying distance cutoffs of 3.3 and 4.0 Å, respectively. Structural superpositions were performed with *SUPERPOSE* (Winn *et al.*, 2011) and schematic representations of all three crystal structures were prepared with *MolSoft* (Raush *et al.*, 2009). The programs *LIGPLOT* (Wallace *et al.*, 1995) and *MolScript* (Kraulis, 1991) were used to depict the interactions (Figs. 7 and 8), while *BobScript* (Esnouf, 1997) was employed for schematic representation of the MCU electron-density map (Fig. 3c). *Raster3D* (Merritt & Bacon, 1997) was also used to render the images. Structural classification of *St*GE2 folding was performed using the *CATH* database of domain structures server. The topology of each structure was extracted using
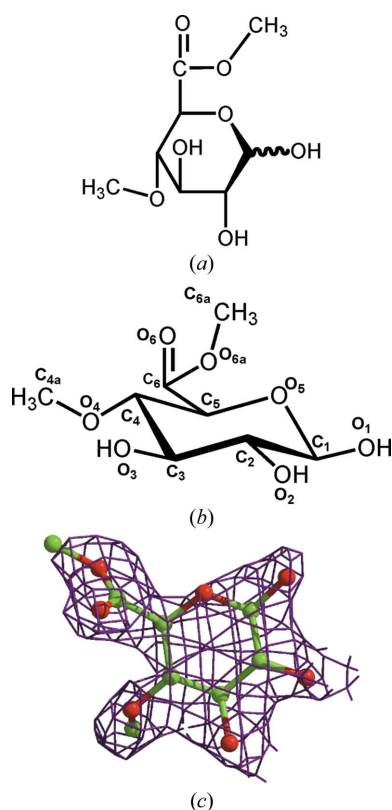
Figure 3
(a) Chemical structure of methyl 4-O-methyl-D-glucopyranuronate. (b) The anomer bound to the S213A mutant, also indicating the numbering system used. (c) Schematic representation of the $2F_{\text{o}} - F_{\text{c}}$ electron-density map contoured at the 1.0σ level of the refined MCU bound in the catalytic cavity of *St*GE2.

PDBsum (http://www.ebi.ac.uk/). Solvent-accessible areas were calculated using the PDBe *PISA* platform (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; Krissinel & Henrick, 2007). Structural homologues of *St*GE2 were sought with the aid of the *DALI* server (Holm & Rosenström, 2010; Figs. 2 and 5).

2.6. PDB accession code

The atomic coordinates and structure factors for the crystal structures of *St*GE2, its S213A mutant and the S213A mutant in complex with methyl 4-O-methyl-β-D-glucopyranuronate have been deposited in the Protein Data Bank (http://www.pdb.org) under accession codes 4g4g, 4g4i and 4g4j, respectively.

3. Results and discussion

3.1. Sequence analysis

Analysis of the wild-type *St*GE2 sequence (UniProt code G2QJR6) was performed with *BLASTP* (Altschul *et al.*, 1997), resulting in a long list of putative conserved domains that all belonged to hypothetical proteins, with the exception of Cip2_GE (56% identity and 72% homology with 92% coverage). According to the *BLASTP* results, residues 206–227, with amino-acid sequence RLGVTGCSRNGKGAFIT-GALVD and containing the G*XXXX*GK motif known as the P-loop motif (Walker *et al.*, 1982), were suggested to have a DUF463 domain architecture (Punta *et al.*, 2012). However, the *e*-value analysis of the results did not provide sufficient evidence to support the existence of the suggested motif. In addition, to the best of our knowledge, esterases involved in
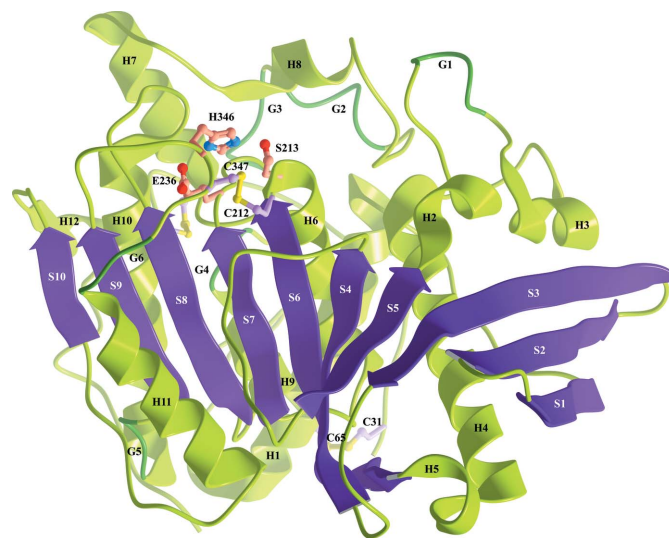
Figure 4
The overall structure of *St*GE2 determined at 1.55 Å resolution. The twisted β-sheet (shown in violet and labelled S) is sandwiched between two layers of α-helices (shown in light green and labelled H) and 3_{10}-helices (shown in dark green and labelled G). The catalytic triad residues are indicated in ball-and-stick representation and the disulfide bonds formed are also highlighted. This figure was prepared with *MolSoft* (Raush *et al.*, 2009).

the degradation of plant biomass do not require phosphates for activity; therefore, it is unlikely that the presence of this motif has a functional role.

3.2. Analysis of the *St*GE2 crystal structures

Diffracting crystals of *St*GE2 were grown in the orthorhombic lattice, space group $P2_12_12_1$, with one molecule per asymmetric unit and the three-dimensional structure was refined to 1.55 Å resolution. A total of 367 amino acids were well defined in the $2F_o - F_c$ electron-density map and were incorporated in the model. However, there was no density for 14 amino acids at the N-terminus prior to Cys31; therefore, they were excluded from the structure. Similarly, the S213A mutant was prepared by employing previously established protocols (Topakas *et al.*, 2010) to shed light on the structure–function relationships of the enzyme. Its three-dimensional structure was determined at 1.9 Å resolution from well diffracting crystals grown in the orthorhombic lattice although lacking one twofold axis (space group $P2_122_1$) compared with *St*GE2. The *St*GE2 structure was used as a starting model and two further N-terminal residues were modelled (369 amino acids in total) as suggested by the electron-density maps. However, there was no density for 12 amino acids at the N-terminus prior to Asp29; therefore, they were excluded from the structure. Structural studies of the S213A mutant followed. The S213A mutant complex with methyl 4-*O*-methyl-D-glucopyranuronate was formed by soaking preformed crystals and the structure was determined at 2.35 Å resolution. In general, all residues that were included in the final structures lay in allowed regions of the Ramachandran plot. Molecules of ethylene glycol and glycerol, which were used either in the crystallization medium or as a cryoprotectant, were also observed in all three structures, mainly trapped at the interface of the packed monomers in the unit cell and reinforcing the crystal lattice without posing any further modifications (Figs. 4 and 6). However, careful inspection of their binding sites facilitated structure interpretation.

3.2.1. The overall structure of *St*GE2.
*St*GE2 belongs to the α/β-hydrolase superfamily and is a serine-type hydrolase; its overall architecture follows the three-layer $\alpha\beta\alpha$-sandwich hydro-

lase fold with a Rossmann-fold topology, as revealed by its three-dimensional structure (Fig. 4). However, deviations from the canonical α/β-hydrolase fold are observed; the canonical fold features a β-sheet of eight strands in the core of the structure intercalated between two clusters of α-helices (four and two α-helices bilaterally; Ollis *et al.*, 1992). The β-sheet is expanded by the insertion of two antiparallel β-strands at the N-terminus of *St*GE2, resulting in a twisted β-sheet. The latter is sandwiched between two layers containing a total of 18 helices. Eight α-helices and two 3_{10}-helices are observed on one side of the β-sheet instead of two in the canonical fold, while on the opposite side an octet of α-helices and 3_{10}-helices that are equally divided, rather than only four, are recorded. The positions of two of the putative catalytic triad residues, namely Ser213 and Glu236, comply with those suggested by the canonical fold. Detailed analysis of the secondary structure according to *PROMOTIF* (Hutchinson & Thornton, 1996) as implemented in *PDBsum* (Laskowski *et al.*, 1997) is presented in Supplementary Table S1.

The three-dimensional structure of *St*GE2 with one molecule per asymmetric unit follows the monomeric form that *St*GE2 adopts in solution, as suggested by HPLC gel filtration (data not shown). The latter was confirmed by analysis of the
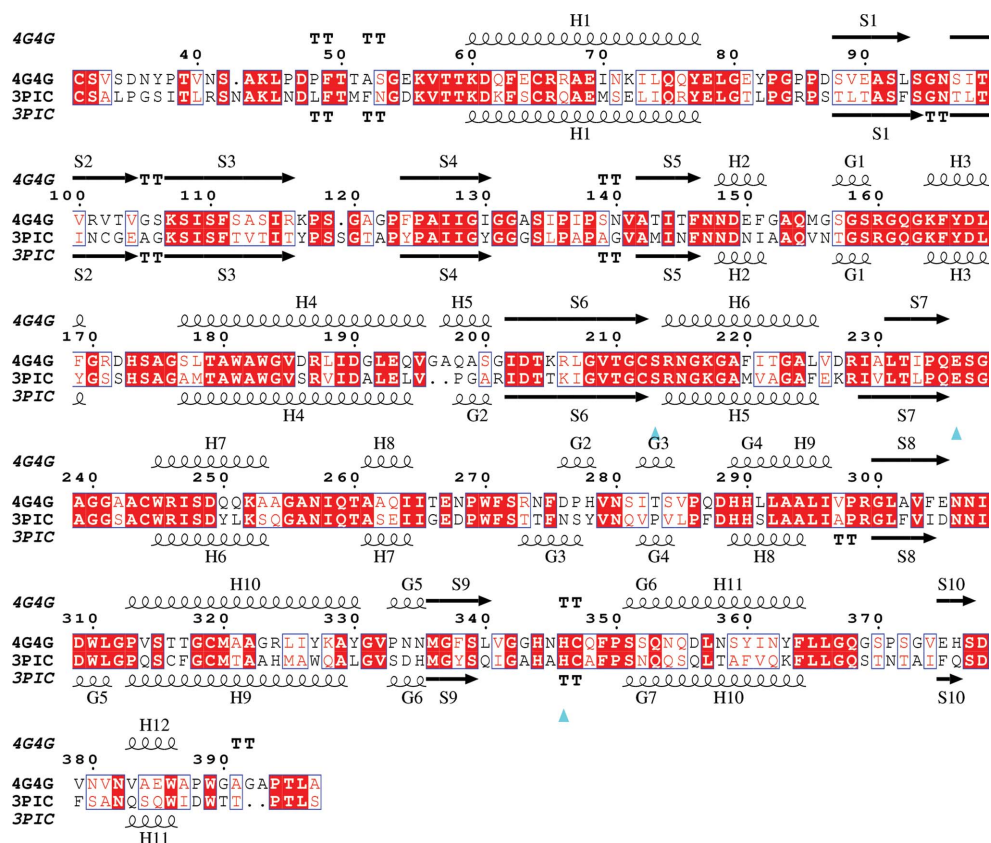
Figure 5
Sequence alignment of *St*GE2 with Cip2_GE. The secondary-structure elements are highlighted in the consensus line. Identical and similar residues are shown in white on a red background and in red on a white background, respectively. The residues of interest are indicated by cyan triangles in the same line. The secondary-structure elements α-helices, 3_{10}-helices, β-strands and strict β-turns are denoted H, G, S and TT, respectively.

protein interfaces using the *PISA* server (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html; Krissinel & Henrick, 2007), revealing the absence of specific interactions that could result in the formation of stable quaternary structures. The overall structure is stabilized by three disulfide bonds, namely Cys31–Cys65, Cys212–Cys347 and Cys244–Cys319.
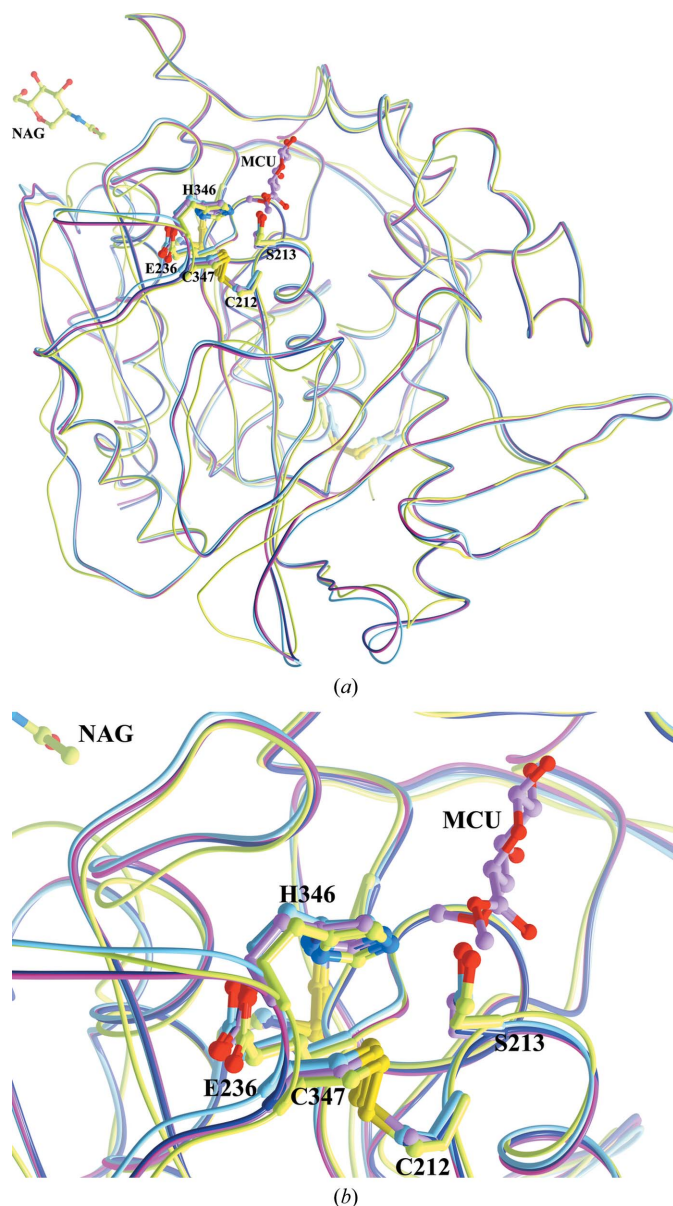
Residues Ser213, Glu236 and His346 of *St*GE2 could be proposed as putative catalytic triad residues since their architecture complies with that proposed for the catalytic

Figure 6
(*a*) Superposition of the crystal structures of *St*GE2 (shown in cyan), the S213A mutant (shown in dark blue), the S213A mutant in complex with MCU (shown in magenta) and the glycosylated (*N*-acetylglucosamine; NAG) Cip2_GE (shown in lime green; Pokkuluri *et al.*, 2011). The catalytic triad residues are also indicated (the colour scheme follows the relevant structure). The three disulfide bonds that are formed in all structures are highlighted in yellow. (*b*) Expanded view of the active site indicating the ester bond hydrolyzed by *St*GE2. This figure was prepared with *MolSoft* (Raush *et al.*, 2009).

mechanism of α/β hydrolases. Specifically, Ser213 is a nucleophilic serine located between strand S6 and helix H6 forming the so-called 'nucleophilic elbow', with Ramachandran angles lying in the generously allowed regions (Ollis *et al.*, 1992; Nardini & Dijkstra, 1999; Quevillon-Cheruel *et al.*, 2005). His346 (which acts as an acid/base complementing the nature of the substrate upon binding) bridges strand S9 and helix G6, while Glu236 (which is suggested to adjust the basic character of His346) is situated between strand S7 and helix H7 (Correia *et al.*, 2008). The exact locus and organization of the catalytic triad residues is stabilized by the tight hydrogen-bond interactions that are formed between them. The hydroxyl group of Ser213 interacts with His346 NE2, while Glu236 OE1 is hydrogen-bonded to His346 ND1 (at distances of 2.7 and 2.8 Å, respectively).

A total of 14 ethylene glycol molecules (present in the crystallization medium) and two glycerol molecules (used as cryoprotectant) were also identified in the structure of *St*GE2. They are predominantly positioned at the interface of the packed monomers, but without providing evidence for specific binding.

3.2.2. Comparison of *St*GE2 with Cip2_GE from *H. jecorina*. Comparison of the *St*GE2 structure with that of Cip2_GE, the only known member of the CE15 family (Pokkuluri *et al.*, 2011) for which a structure has been determined to date, showed that the overall structure of the enzymes is homologous and that they share 56% sequence identity (over the residues observed in the crystal structure, with 97% coverage) and a *Z*-score of 59 with an r.m.s.d. of 1.1 Å (as determined by the *DALI* server for monomer *B*, the closest structural homologue).

However, superposition of the two crystal structures on C$^{\alpha}$ atoms (Winn *et al.*, 2011) gave an r.m.s.d. of 3.5 Å, revealing changes in selected secondary-structure elements and in the N- and C-termini. Detailed comparison of the two structures showed that the differences detected at sequence level involve either the insertion/deletion of residues leading to alterations in the secondary structure, the presence/absence of a glycosylation site or the packing of the monomers in the crystal. Specifically, the insertion of two residues led to the formation of an α-helix (H5, amino acids 197–200) instead of the 3$_{10}$-helix observed in the Cip2_GE structure (Figs. 4, 5 and 6). The resulting structure is stabilized through a water-mediated interaction formed between Gln188 and Glu81′ from a symmetry-related molecule. Similar changes were recorded for residues 276–278, which form a 3$_{10}$-helix (G2) that is shorter in length compared with the Cip2_GE structure. Also, residues 289–292 form a 3$_{10}$-helix (G4) and an α-helix (H9) rather than only an α-helix as in Cip2_GE, and residues 309–311 form a loop instead of a 3$_{10}$-helix as in Cip2_GE. All of the aforementioned changes are observed at the interface of the monomers with symmetry-related molecules in the unit cell. Of these, it is the latter modification involving residues 309–311 that is considered to be most significant since it is situated in the vicinity of the catalytic site of *St*GE2. This is accompanied by a difference in the number of residues forming strand S9 (residues 336–340), which contains one more residue

compared with the corresponding strand in the Cip2_GE structure (Figs. 5 and 6) and adopts a slightly more extended conformation. The succeeding residues belong to a tight turn bearing the catalytic triad residue His346 (see §3.2.4).

Additional profound changes were observed; however, these could be attributed to shifts imposed by intermolecular interactions of Cip2_GE monomers (with three molecules in the asymmetric unit) as well as interactions with symmetry-related molecules dictated by crystal packing as in the case of the N- and C-termini (Pokkuluri *et al.*, 2011). Moreover, the residues lining the H12 α-helix and the preceding loop region are distorted compared with those in the Cip2_GE structure since the N-linked glycosylation sequence motif (Asn-*X*-Ser) and the *N*-acetylglucosamine molecule bound at Asn447 observed in the latter are missing in *St*GE2 (Asn-*X*-Ala is the corresponding motif in *St*GE2; Fig. 6).

Previous studies of the effect of temperature on *St*GE2 stability showed that *St*GE2 remains active at 323 K for at least 24 h, maintains more than 70% of its activity at 333–343 K and exhibits half-lives of 22 h at 328 K and 0.5 h at 333 K (Topakas *et al.*, 2010), while Cip2_GE is only quite stable at 313 K, with half-lives of 10 and 2 min at 323 and 333 K, respectively (Li *et al.*, 2007). A comparative analysis of the structures of the thermophilic *St*GE2 and its mesophilic homologue Cip2_GE was performed to monitor the differences in the structural determinants that could affect their thermostabilities. The typical indicators employed to assess protein thermostability are rather diverse. At the sequence level the protein composition of prolines has been examined, taking into account their limited allowed configurations and the restrictions that they impose on the preceding residue (Prajapati *et al.*, 2007). The primary structure of *St*GE2 contains 22 proline residues compared with 17 in Cip2_GE. Moreover, the number of thermolabile residues such as Met, Cys, Asn and Gln was considered. *St*GE2 contains 48 thermolabile residues compared with 54 in Cip2_GE, as is observed in the majority of thermophilic proteins when compared with their mesophile homologues (Kumar *et al.*, 2000). The ratio of Arg/(Arg+Lys) has also been used as a indicator for this purpose. *St*GE2 exhibits an increased Arg/(Arg+Lys) ratio compared with Cip2_GE (0.56 and 0.5, respectively; Mrabet *et al.*, 1992). Similarly, electrostatic interactions are among the structural features which are known to be shared by thermostable proteins. The structures of *St*GE2 and Cip2_GE present the same number of salt bridges and cation–π interactions, as calculated using the *ESBRI* server (Costantini *et al.*, 2008) and *CaPTURE* (Gallivan & Dougherty, 1999).

3.2.3. The overall structures of the unliganded and bound forms of the S213A mutant. The S213A mutant was prepared following a previously developed protocol (Topakas *et al.*, 2010) to shed light on the structure–function relationships of the enzyme and the putative role of the nucleophilic Ser in the substrate-recognition/catalytic mechanism. Determination of its three-dimensional structure at 1.9 Å resolution revealed that overall it remained unchanged with reference to the *St*GE2 crystal structure except for some areas involved in crystallographic symmetry interactions. Superposition of the S213A mutant atomic coordinates with those of *St*GE2 on C$^\alpha$ atoms showed that the two structures differed only slightly by an r.m.s.d. of 0.4 Å. Similarly, the crystal structure of the S213A mutant in complex with MCU indicated that the substrate analogue bound at the catalytic site of the enzyme without affecting the overall structure (r.m.s.d. on C$^\alpha$ atoms of 0.40 and 0.12 Å with reference to the *St*GE2 and the S213A mutant structures, respectively; Fig. 6). Ethylene glycol and glycerol molecules were also bound in both structures as in the case of *St*GE2. Of these, two ethylene glycol molecules were located in the active-site region of the S213A mutant structure (Table 1, Fig. 6; see §3.2.4).

3.2.4. *St*GE2 catalytic site. Overall, the catalytic site in the monomeric three-dimensional structure of *St*GE2 lies on the surface of the molecule; it is mostly exposed to the solvent and is not affected by symmetry-related packing interactions. All residues lining the putative catalytic triad face the solvent,
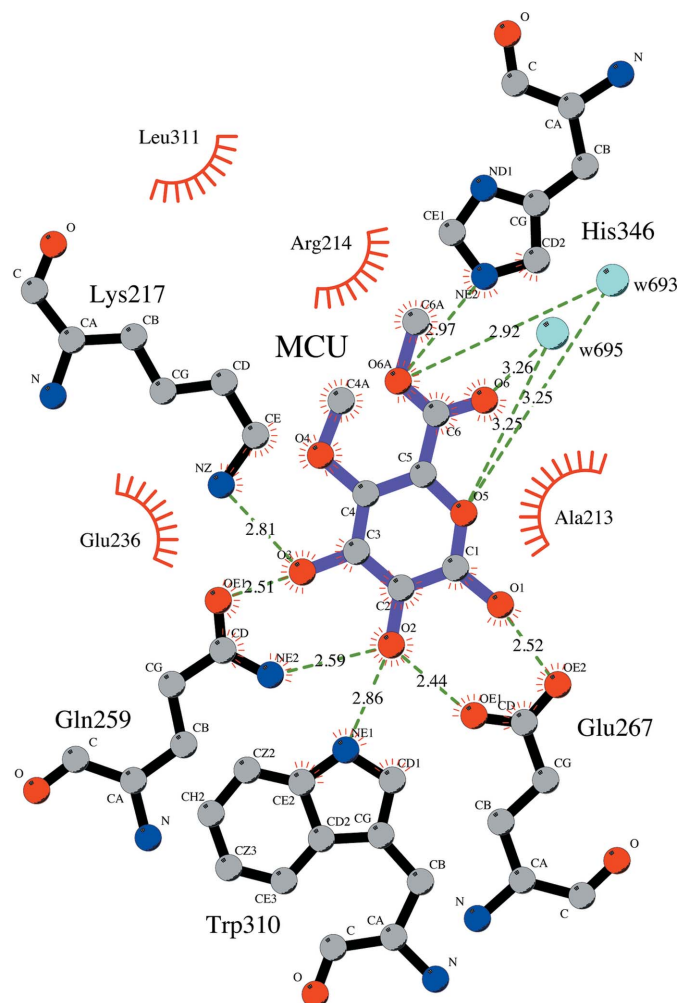
Figure 7
LIGPLOT diagram of MCU interacting with the active-site residue lying in the vicinity. The substrate-analogue bonds are shown in purple, while bonds of the residues lining the site are shown in black. Hydrogen bonds are shown as black dashed lines with distances indicated in Å. Additional residues forming van der Waals interactions with MCU are represented by red semicircles with radiating spokes.

Table 2
Direct hydrogen bonds and water-mediated hydrogen-bond interactions of MCU in the catalytic pocket of S213A *St*GE2.

There are a total of seven direct hydrogen-bond interactions.

Atom	Protein atom	Distance (Å)	Angle (°)
O1	Glu267 OE2	2.5	119.9
O2	Gln259 NE2	2.6	127.1
	Glu267 OE1	2.4	119.2
	Trp310 NE1	2.9	169.8
O3	Lys217 NZ	2.8	0.0
	Gln259 OE1	2.5	127.6
O5	Wat693 O	3.3	—
	Wat695 O	3.3	—
O6	Wat695 O	3.3	—
O6a	His346 NE2	3.0	146.0
	Wat693 O	2.9	—
	Arg214 NH1	3.2	—

Table 3
van der Waals interactions of MCU in the catalytic pocket of S213A *St*GE2.

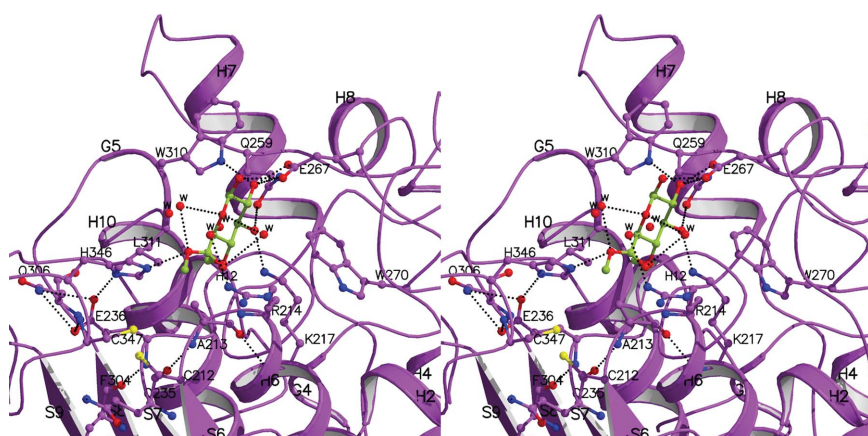There are a total of 59 van der Waals interactions.

Substrate atom	Protein atoms	No. of contacts
C1	Glu267 OE1, OE2	2
C2	Gln259 OE1, NE2; Glu267 OE1, OE2, CD; Trp310 NE1	6
C3	Lys217 NZ; Gln259 OE1	2
C4	Lys217 NZ	1
C4a	Ala213 CA, CB; Lys217 NZ; Glu236 O; Leu311 CD2; Wat670; Wat712	7
C5	Wat693; Wat695; Wat712	3
C6	Ala213 CB; His346 NE2; Wat693; Wat695	4
C6a	Ala213 CB; Arg214 NH1; His346 NE2, CD2; Wat694; Wat693	6
O1	Glu267 CD, OE1; Wat695	3
O2	Gln259 OE1, CD; Glu267 CD, OE2; Trp310 CD1, CE2	6
O3	Lys217 CE; Gln259 CD, NE2; Wat670	4
O4	Ala213 CA, CB, C; Arg214 N; Lys217 CE, NZ	6
O6	Ala213 CB; Arg214 CB, CG, CD, NH1, N	6
O6a	Ala213 CB; His346 CD2; Wat694	3

Figure 8
Stereo diagram of the interactions between the S213A mutant and MCU bound at the catalytic pocket as well as the hydrogen bonds formed by the catalytic triad residues. The side chains and backbone atoms of protein residues involved in ligand binding are shown in ball-and-stick representation. Water molecules (w) are depicted as spheres and hydrogen-bond interactions are shown as dotted lines. This figure was prepared with *MolScript* (Kraulis, 1991) and was rendered with *Raster*3D (Merritt & Bacon, 1997).

with accessible surface areas for Ser213, Glu236 and His346 of 27.1, 2.7 and 75.7 $Å^2$, respectively. The catalytic residues Ser213 and Glu236 are not occluded by symmetry-related polypeptide chains, except for those in the vicinity of His346. The *St*GE2 structure suggests that strands S9 and S7 shape the active site, coordinating His346 and Glu236 towards forming the catalytic triad with Ser213 (Fig. 4). The catalytic triad residues are involved in a network of hydrogen bonds that contribute to the rigid architecture of the catalytic pocket. Focusing on the nucleophilic Ser213, its hydroxyl group interacts directly with His346 NE2, which in turn forms a hydrogen bond between its ND1 atom and Glu236 OE1. The setting is complemented by additional direct and water-mediated polar interactions involving residues Arg214, Gly216, Lys217, Gln235, Phe304 and Asn306 (Supplementary Table S2). This mode of interaction provides structural evidence for the nucleophilic role of Ser213, while His346 is expected to act as an acid–base complementing the nature of the substrate upon binding and Glu236 is expected to play a role as a regulator of the basic character of histidine (Correia *et al.*, 2008).

Mutation of Ser213 to Ala showed that the absence of the hydroxyl group did not affect the active-site architecture. The hydrogen bonds formed between the Ala213 backbone atoms (N and O) and nearby residues, as well as the contacts with His346 and Glu236, were maintained. The loop bearing His346 (residues 344–350) is subjected to shifts ranging from ∼0.5 to ∼1.0 Å in the S213A mutant owing to crystal-packing interactions. Despite these shifts, the position of His346 in the catalytic site is preserved (minor shifts of ∼0.5 Å of all atoms). The disulfide bond formed between Cys347 and Cys212 further enhances the rigidity of the catalytic site by bridging the two adjacent catalytic triad residues (His346 and Ser213, respectively). Additional electron density for two ethylene glycol molecules EDO406 and EDO408 was identified in the S213A mutant structure at the same site. The latter lay at the entrance of the cavity, forming five potential hydrogen bonds to the neighbouring residues Lys217 NZ, Gln259 OE1 and NE2, Glu267 OE2 and Trp310 NE1 (Supplementary Fig. S1). This finding reveals that the catalytic cavity is exposed to the solvent and has the potential to accommodate the substrate by forming favourable binding interactions (Supplementary Tables S3, S4 and S5).

With the aim of advancing our knowledge of the structure–function relationship of *St*GE2 and elucidating the underlying mechanism of enzyme action, structural studies of the S213A mutant in complex with the substrate analogue MCU were performed. The results revealed that the substrate analogue bound at the catalytic site, forming an extended network of interactions (seven direct and four indirect hydrogen bonds as well as 59 van der Waals

interactions) with the active-site residues (Tables 2 and 3, Figs. 7 and 8). The residues participating in the hydrogen-bond network were those observed in the S213A mutant structure with ethylene glycol together with His346 NE2. Two O atoms of the glucopyranose ring, O2 and O3, are aligned with the positions of the ethylene glycol hydroxyl groups shifted by ∼0.7 and ∼0.2 Å, respectively (Figs. 7 and 8, Supplementary Fig. S1).

Specifically, the substrate analogue rests in the cavity using the peripheral hydroxyl group O atoms of the glucopyranose ring as an anchor to promote its binding. Both the Gln259 and Glu267 side-chain atoms are hydrogen-bonded to O1, O2 and O3 of the sugar. The conformation of the Glu267 side chain is altered [rotation of its (χ_1, χ_3) dihedral angles by (∼74°, ∼127°)] compared with both the S213A mutant and the StGE2 structures, favouring tight binding of the substrate analogue. Trp310 NE1 is also in close contact with O2, while Lys217 NZ interacts with O3. An additional polar contact is formed between Lys217 NZ and O4 at a distance of 3.4 Å. O5 completes this hydrogen-bonding arrangement by taking part in water-mediated interactions with Arg214 NH1 (through water molecules Wat693 and Wat695). O6a, which pertains to the ester bond to be cleaved, is hydrogen-bonded to the NE2 atom of the imidazole ring of His346; this contact contributes to the suggestion that the position of the catalytic triad is rather concrete and the pocket is crafted to impose a 'ready for nucleophilic attack' orientation of the substrate regardless of the absence of the nucleophile. To this end, comparison of the StGE2 crystal structure with the complex structure demonstrates that the Ser213 OG atom is indeed facing the ester bond at a distance of ∼2.2 Å.

The van der Waals interactions of the substrate analogue mainly concern the same palette of residues as those involved in the aforementioned hydrogen bonds. Among these, attention is drawn to the backbone N atom of Arg214 and its counterparts, O4 of the methoxy moiety and O6 of the ester group (at a distance of ∼3.4 Å), which is also hydrogen-bonded to Ser213 OG, implying that Arg214 is positioned to be part of the so-called oxyanion hole (Figs. 7 and 8).

Superposition of the structures of the unliganded S213A mutant and the complex showed that MCU binds at the same position as previously occupied by one of the two ethylene glycol molecules bound in the catalytic pocket in the un-liganded structure. In particular, EDO408 mimics the positions of the C2, O2, C3 and O3 atoms of the glupyranose ring, while EDO406 lies at the vicinity of the site. The complex was prepared using preformed crystals of the S213A mutant; therefore, it appears that the substrate analogue displaced the ethylene glycol molecules from the catalytic pocket upon binding, explaining the quality of the electron density attributed to the ligand (Figs. 3c and 7 and Supplementary Fig. S1).

4. Conclusions

The crystal structures of a recombinant novel thermophilic GE and of unliganded and bound forms of its S213A mutant were determined at 1.55, 1.9 and 2.35 Å resolution, respec-

tively. StGE2 is a member of the α/β-hydrolase superfamily and its overall structure follows the three-layer $\alpha\beta\alpha$-sandwich hydrolase fold with a Rossmann topology. Three disulfide bonds, one of which is situated at the entrance to the catalytic pocket, contribute to the rigidity of the structure and the active-site architecture. Mapping of the catalytic site using a methyl 4-O-methyl-β-D-glucopyranuronate substrate analogue revealed that the catalytic triad residues, namely Ser213, Glu236 and His346, participate in a concrete 'ready for nucleophilic attack' configuration. This is partially coordinated by strands S6, S7 and S9 of the β-sheet, which drive His346 and its triad counterparts into an orientation that makes substrate recognition possible. The complex structure also unveiled the inherent flexibility of residues shaping the pocket, such as Glu267, which alters its side-chain conformation to accommodate the substrate analogue. The setting is complemented by the Arg214 backbone N atom of the so-called oxyanion hole that lies opposite the ester and methoxy group and is proposed to stabilize the tetrahedral intermediate during catalysis. The methoxy group, comprising atoms O4 and C4a, might also play an additional role to the catalytic triad residues in substrate recognition by enhancing binding via an increased number of van der Waals interactions formed with the residues lining the site (Table 3).

StGE2 and Cip2_GE are the only characterized GEs for which three-dimensional structures have been determined to date. Direct comparison of the two enzymes brings to light the biochemical and structural determinants that promote StGE2 as a more suitable target that could be further explored for potential biotechnological applications in comparison to Cip2_GE. These comprise the thermophilicity of the former, the accessibility of the catalytic triad residues and the crafting of the active site towards substrate recognition, as well as its monomeric form in both the crystal and solution. We believe that the X-ray structure of StGE2 and the complex structure of its S213A mutant have uncovered the fingerprint of substrate binding to a GE for the first time, appraising its potential to be exploited as a model for developing tailor-made effective biocatalysts. The next step in this direction is a thorough investigation of StGE2 by site-directed mutagenesis to delineate its functional role and to develop a prototype for bioconversion.

References

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

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Correia, M. A., Prates, J. A., Brás, J., Fontes, C. M., Newman, J. A., Lewis, R. J., Gilbert, H. J. & Flint, J. E. (2008). *J. Mol. Biol.* **379**, 64–72.

Costantini, S., Colonna, G. & Facchiano, A. M. (2008). *Bioinformation*, **3**, 137–138.

Ďuranová, M., Hirsch, J., Kolenová, K. & Biely, P. (2009). *Biosci. Biotechnol. Biochem.* **73**, 2483–2487.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Gallivan, J. P. & Dougherty, D. A. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 9459–9464.

Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. & Foust, T. D. (2007). *Science*, **315**, 804–807.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kumar, S., Tsai, C.-J. & Nussinov, R. (2000). *Protein Eng.* **13**, 179–191.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L. & Thornton, J. M. (1997). *Trends Biochem. Sci.* **22**, 488–490.

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

Li, X.-L., Špániková, S., de Vries, R. P. & Biely, P. (2007). *FEBS Lett.* **581**, 4029–4035.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Mrabet, N. T. *et al.* (1992). *Biochemistry*, **31**, 2239–2253.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nardini, M. & Dijkstra, B. W. (1999). *Curr. Opin. Struct. Biol.* **9**, 732–737.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.

Pokkuluri, P. R., Duke, N. E. C., Wood, S. J., Cotta, M. A., Li, X.-L., Biely, P. & Schiffer, M. (2011). *Proteins*, **79**, 2588–2592.

Prajapati, R. S., Das, M., Sreeramulu, S., Sirajuddin, M., Srinivasan, S., Krishnamurthy, V., Ranjani, R., Ramakrishnan, C. & Varadarajan, R. (2007). *Proteins*, **66**, 480–491.

Punta, M. *et al.* (2012). *Nucleic Acids Res.* **40**, D290–D301.

Quevillon-Cheruel, S., Leulliot, N., Graille, M., Hervouet, N., Coste, F., Bénédetti, H., Zelwer, C., Janin, J. & Van Tilbeurgh, H. (2005). *Protein Sci.* **14**, 1350–1356.

Raush, E., Totrov, M., Marsden, B. D. & Abagyan, R. (2009). *PLoS One*, **4**, e7394.

Reddy, N. & Yang, Y. (2005). *Trends Biotechnol.* **23**, 22–27.

Schleberger, C., Sachelaru, P., Brandsch, R. & Schulz, G. E. (2007). *J. Mol. Biol.* **367**, 409–418.

Špániková, S. & Biely, P. (2006). *FEBS Lett.* **580**, 4597–4601.

Topakas, E., Moukouli, M., Dimarogona, M., Vafiadi, C. & Christakopoulos, P. (2010). *Appl. Microbiol. Biotechnol.* **87**, 1765–1772.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). *EMBO J.* **1**, 945–951.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

Weng, J.-K., Li, X., Bonawitz, N. D. & Chapple, C. (2008). *Curr. Opin. Biotechnol.* **19**, 166–172.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.