STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Structures of exoglucanase from *Clostridium cellulovorans*: cellotetraose binding and cleavageLi-Chu Tsai,^{a*} Imamaddin Amiraslanov,^b Hung-Ren Chen,^a Yun-Wen Chen,^b
Hsiao-Lin Lee,^c Po-Huang Liang^c and Yen-Chywan Liaw^{b*}

Received 1 June 2015

Accepted 25 August 2015

Edited by F. T. Tsai, Baylor College of Medicine,
Houston, USA**Keywords:** oligosaccharide binding and cleavage; GH48 exoglucanase; glycoside hydrolase; β -1,4-glycosidic bonds; reducing end of cellulose; *Clostridium cellulovorans*; (α/α)₆ fold.**PDB references:** ExgS, complex with PEG fragments, 4xwl; complex with cellobiose, 4xwm; complex with cellotetraose, 4xwn**Supporting information:** this article has supporting information at journals.iucr.org/f^aMolecular Science and Engineering, National Taipei University of Technology, 1, Section 3, Chung-Hsiao E. Road, Taipei 10608, Taiwan, ^bInstitute of Biological Chemistry, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan, and ^cInstitute of Molecular Biology, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan.
*Correspondence e-mail: lichu@ntut.edu.tw, mbycliaw@gate.sinica.edu.tw

Exoglucanase/cellobiohydrolase (EC 3.2.1.176) hydrolyzes a β -1,4-glycosidic bond from the reducing end of cellulose and releases cellobiose as the major product. Three complex crystal structures of the glycosyl hydrolase 48 (GH48) cellobiohydrolase S (ExgS) from *Clostridium cellulovorans* with cellobiose, cellotetraose and triethylene glycol molecules were solved. The product cellobiose occupies subsites +1 and +2 in the open active-site cleft of the enzyme–cellotetraose complex structure, indicating an enzymatic hydrolysis function. Moreover, three triethylene glycol molecules and one pentaethylene glycol molecule are located at active-site subsites –2 to –6 in the structure of the ExgS–triethylene glycol complex shown here. Modelling of glucose into subsite –1 in the active site of the ExgS–cellobiose structure revealed that Glu50 acts as a proton donor and Asp222 plays a nucleophilic role.

1. Introduction

The degradation of polysaccharides by hydrolysis is a biological process with potential industrial applications. Four cellulases possess the primary functions necessary for the breakdown of cellulose: endocellulase (EC 3.2.1.4) randomly cleaves internal β -1,4-glycosidic bonds, exocellulase (EC 3.2.1.91) cleaves the bond from the nonreducing end, cellobiohydrolase (EC 3.2.1.176) acts on the reducing end of cellulose, and β -glucosidase (EC 3.2.1.21) cleaves cellobiose into glucose units (Doi & Kosugi, 2004; Gilbert, 2007). These specialized enzymes may be secreted independently or, in some organisms, are involved in the formation of the cellulosome, a protein complex comprising an array of hydrolytic enzymes that act in concert to efficiently degrade cellulose (Fontes & Gilbert, 2010).

Cellobiohydrolases (EC 3.2.1.176) have been found in glycoside hydrolase families 7 (GH7) and 48 (GH48). GH7 cellobiohydrolase enzymes exhibit a jelly-roll β -sandwich fold structure and utilize two catalytic glutamates in a retaining hydrolytic mechanism (Parkkinen *et al.*, 2008; Vasella *et al.*, 2002). The GH48 family are predominately classified as (α/α)₆-fold enzymes that may utilize aspartate and glutamate as base and acid donors, respectively, for substrate cleavage and release of cellobiose as the main product through an inverting mechanism. Exoglucanase S (ExgS) from *Clostridium cellulovorans* (Tamaru *et al.*, 2000; Doi *et al.*, 1994) belongs to the GH48 family and consists of a catalytic domain (629 residues) followed by a dockerin domain (61 residues) at

© 2015 International Union of Crystallography

Table 1

Macromolecule-production information.

Source organism	<i>C. cellulovorans</i>
Forward primer	5'-AATTCGCGGCGCCGGTGG-3' 5'-CGCGGCGCCGGTGG-3'
Reverse primer	5'-GTTAATCGCCTTTCAGACCCGG-3' 5'-TCGAGTTAATCGCCTTTCAGACCCGG-3'
Cloning vector	pHTPP13
Expression vector	pHTPP13
Expression host	<i>E. coli</i>
Complete amino-acid sequence of the construct produced	632 amino acids

the C-terminus with 54–62% sequence identity to other clostridial enzymes. Bacterial cellulosomes from mesophilic clostridia, including *C. cellulolyticum*, *C. cellulovorans*, *C. josui* and *C. acetobutylicum*, containing a single scaffoldin may interact with the target proteins *via* the cohesin and dockerin domains (Bayer *et al.*, 2004). Dockerin, which is appended to a carbohydrate-binding module (CBM) or catalytic domain of an enzyme, has been proposed to bind the cohesin domains of scaffoldin in the cellulosome (Doi & Kosugi, 2004; Nagy *et al.*, 2007).

Five crystal structures of the catalytic domain of GH48 cellobiohydrolases (Parsiegla *et al.*, 1998, 2000, 2008; Guérin *et al.*, 2002; Guimarães *et al.*, 2002; Sukharnikov *et al.*, 2012; Brunecky *et al.*, 2013; Kostylev *et al.*, 2014) with and without oligosaccharides revealed the active-site topology of GH48 exoglucanases, featuring a tunnel and cleft in which substrate subsites −7 to −1 and +1 to +4 (Davies *et al.*, 1997) are located, respectively. In CelF and CelS (Parsiegla *et al.*, 2000; Guimarães *et al.*, 2002), the authors proposed that substrate enters the tunnel and slides along to the open cleft (in the direction from subsite positions −7 to +2). In addition, the key residues that are involved in the enzymatic functions of CelF and CelS have been investigated (Parsiegla *et al.*, 2008; Saharay *et al.*, 2010). Site-directed mutagenesis and hydrolysis assays of Cel48A showed that Asp225 (equivalent to Asp222 in ExgS) is the catalytic base (Kostylev & Wilson, 2011). Here, we report three crystal structures of the catalytic domain of *C. cellulovorans* exoglucanase S (ExgS) with cellobiose, cellotetraose and 1PE-PGE. The enzymatic hydrolysis function and the detailed interaction of the active-site residues and substrates/1PE-PGE fragments will be discussed. To further understand the interaction between ExgS and dockerin-cohesin, we docked dockerin-cohesin to ExgS to identify the putative interfacial residues that might be required for optimal complex formation interactions.

2. Materials and methods

2.1. Cloning

Synthetic cDNA encoding full-length ExgS optimized for *Escherichia coli* expression was purchased from GeneArt, Germany. The quick and optimal protein expression system developed by Wang *et al.* (2009) was used for cloning and expression. The coding sequence of the catalytic domain of ExgS was amplified by PCR with two primer sets (Table 1).

Table 2

Crystallization.

Method	Vapour diffusion
Plate type	Hanging drop, 24-well
Temperature (K)	295
Protein concentration (mg ml ^{−1})	10.0
Buffer composition of protein solution	10 mM Tris–HCl, 100 mM NaCl pH 7.0
Composition of reservoir solution	2 M (NH ₄) ₂ SO ₄ , 2% PEG 400 in 0.1 M HEPES pH 7.5

The PCR conditions consisted of activation at 368 K for 5 min followed by 35 cycles of 368 K for 30 s, 328 K for 30 s and 345 K for 2 min; final extension took place at 345 K for 10 min. These PCR-amplified DNAs were mixed and phosphorylated at 310 K for 2 h, heated at 368 K for 10 min to denature the double-stranded DNAs, renatured at 338 K for 5 min and cooled to room temperature to generate a sticky-end DNA with an EcoRI site at the 5'-terminus and an XhoI site at the 3'-terminus. This facilitated insertion into the expression vector pHTPP13 that we had previously constructed (Wang *et al.*, 2009), resulting in a His₆ tag at the N-terminus of the expressed protein (Table 1). The plasmid was transformed into *E. coli* strain JM109 and a colony with the correct sequence of the cloned gene was selected for protein expression.

2.2. Protein expression and purification

8 l fresh Terrific Broth (TB) medium containing 100 µg ml^{−1} ampicillin was inoculated with an overnight culture of a single transformant (200 ml). It was then incubated at 310 K with shaking until the medium reached an absorbance at 600 nm of 0.6–1.0. The temperature of the medium was then reduced to 292 K. The inducer isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. After further growth for 26 h, the bacterial cells were harvested by centrifugation at 7000g and immediately resuspended in lysis buffer consisting of 20 mM Tris–HCl, 300 mM NaCl, 10 mM imidazole pH 7.5. The cells were then disrupted using a Cell Disruption System (Constant Systems) and the cell debris was removed by centrifugation at 17 000g. The cell-free extract was loaded onto an Ni²⁺-nitrilotriacetic acid column (Sigma) which had been equilibrated with 10 mM Tris–HCl, 400 mM NaCl, 10 mM imidazole pH 7.0. The column was washed using the same buffer, and the His₆-tagged proteins were subsequently eluted with 400 mM imidazole. The protein solution was changed to protein buffer (10 mM Tris–HCl, 50 mM NaCl pH 7.0) by dialysis. The protein solution was then further purified by size-exclusion chromatography filtration in protein buffer. The purity was checked by 10% SDS–PAGE. The purified proteins were concentrated using 30 kDa cutoff membranes in Amicon Ultra-15 centrifugal filtration units (Millipore). The protein concentration was determined using the Bradford assay (Bio-Rad, Hercules, California, USA). Prior to crystallization, the purified protein was concentrated to 10 mg ml^{−1} in protein buffer.

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

	ExgS–1PE-PGE	ExgS–cellobiose	ExgS–cellotetraose
Diffraction source	BL13B1, NSRRC	BL13B1, NSRRC	BL13B1, NSRRC
Wavelength (Å)	1.0	1.0	1.0
Temperature (K)	100	100	100
Detector	ADSC Quantum 315 CCD	ADSC Quantum 315 CCD	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	231.72	229.76	369.67
Rotation range per image (°)	0.5	0.25	0.5
Total rotation range (°)	107	90	90
Exposure time per image (s)	10	5	20
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
a, b, c (Å)	108.24, 108.24, 182.99	109.14, 109.14, 182.15	108.85, 108.85, 182.38
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Mosaicity (°)	0.476	0.318	1.030
Resolution range (Å)	30–2.05 (2.09–2.05)	30–1.70 (1.76–1.70)	30–2.88 (2.98–2.88)
Total No. of reflections	1640500	2924083	1287379
No. of unique reflections	68867	120479	25466
Completeness (%)	100.0 (100.0)	99.5 (99.7)	99.7 (100)
Multiplicity	8.6 (8.6)	6.6 (6.6)	6.8 (7.1)
$\langle I/\sigma(I) \rangle$	41.1 (5.7)	34.3 (7.4)	25.3 (8.7)
R_{merge} (%)	7.5 (40.1)	5.1 (26.5)	8.1 (25.3)
Overall B factor from Wilson plot (Å ²)	31.2	21.0	41.4

Table 4

Structure solution and refinement.

Values in parentheses are for the outer shell.

	ExgS–1PE-PGE	ExgS–cellobiose	ExgS–cellotetraose
No. of reflections, working set	68475	119728	25330
No. of reflections, test set	1999	2000	1998
Final R_{cryst} (%)	14.2	14.6	16.4
Final R_{free} (%)	17.3	16.3	19.3
No. of non-H atoms			
Protein	5107	5152	5012
Ligand + ion	649	75	81
Water	472	601	27
R.m.s. deviations			
Bonds (Å)	0.008	0.008	0.006
Angles (°)	1.055	1.186	1.107
Average B factors (Å ²)			
Protein	24.6	13.5	34.0
Ligand	50.3	30.7	45.5
Water	35.2	25.4	27.6
Ramachandran plot			
Most favoured (%)	98.06	97.90	97.94
Allowed (%)	1.84	1.80	1.74
Outliers (%)	0.00	0.3	0.32

2.3. Crystallization, data collection and structure determination

The initial crystallization conditions were screened using a screening kit from Hampton Research, mixing 0.5 μ l protein solution with an equal volume of reservoir solution. Several combinations of crystallization conditions produced crystals. Drops (2 μ l) of the protein solution were mixed with 2 μ l reservoir solution (Table 2). Crystals appeared after 4 d and grew slowly to maximum dimensions of 0.5 $\times$ 0.5 $\times$ 0.3 mm after three weeks at 22°C. The ExgS protein crystals were soaked with a saturated solution of cellobiose or cellotetraose in each different reservoir solution and were incubated for 2 h. The crystals were used to carry out diffraction experiments to

screen for the best diffraction conditions. The crystals of the protein–cellobiose complex that gave the best diffraction data were obtained using 12.5% PEG 3350, 0.2 M disodium tartrate, whereas those of the protein–cellotetraose complex were obtained using 25% PEG 6000 in HEPES buffer pH 7.5. The native crystals were soaked with 5% and then 10% glycerol in reservoir solution before cooling, and the complex crystals were directly used for cryo data collection. Data were collected using synchrotron X-ray radiation on beamline BL13B1 at NSRRC with an ADSC Quantum 315 CCD detector to best diffraction data resolutions of 1.70, 2.88 and 2.05 Å for the ExgS–cellobiose, ExgS–cellotetraose and ExgS–1PE-PGE complexes, respectively. Diffraction statistics are listed in Table 3.

The structure of the catalytic domain of free ExgS was solved by molecular replacement, using a previously determined structure of cellulase Cel48F from *C. cellulolyticum* (PDB entry 1f9o; Parsiegla *et al.*, 2000) as the search model, which has a sequence identity of 64% to ExgS. Cross-rotation, translation-function and packing-function analysis by *Phaser* (McCoy *et al.*, 2007) within the *PHENIX* package (Adams *et al.*, 2010) clearly identified a single solution. Following analysis with *AutoBuild*, 634 residues with side chains plus 472 water molecules were built automatically and refined to an R factor of 17.3%. Manual rebuilding of the model and the addition of ligand molecules were performed using *Coot* (Emsley *et al.*, 2010). Further translation/libration/screw (TLS) refinement was carried out using the *phenix.refine* algorithm within the *PHENIX* package (Afonine *et al.*, 2012) for all data to 2.05 Å resolution. The structures of the cellobiose and cellotetraose complexes were solved using the structure of the free enzyme as a template, implementing a similar algorithm. The program randomly selected 5.03, 5.11 and 2.9% of the observed reflections for the calculation of R_{free} during the refinement process. The final refinement statistics for the ExgS complexes

are listed in Table 4. *MolProbity*, a component of the *PHENIX* package, was used to calculate the Ramachandran statistics (Chen *et al.*, 2010). The atomic coordinates and structural factors have been deposited in the PDB as entries 4xwl (ExgS–1PE–PGE), 4xwm (ExgS–cellobiose) and 4xwn (ExgS–cellotetraose). The side chains of the five lysine residues 102, 194, 216, 478 and 549 were truncated to the C γ atom because no continuous electron-density map was obtained in the 4xwm structure. In addition, many side-chain residues had two conformations in the 4xwm structure.

2.4. Dockerin–cohesin docking

We used the online docking program *ZDOCK* (Pierce *et al.*, 2011), which uses a fast Fourier transform to perform a three-dimensional search of the spatial degrees of freedom between two molecules, with default parameters. Two dockerin–cohesin structures (PDB entries 2vn5 and 2vn6; Pinheiro *et al.*, 2008) were used to dock to the catalytic domain of ExgS to obtain structural insight into the enzyme–dockerin conformation.

3. Results and discussion

3.1. Overall structure of ExgS

The crystal structures of ExgS (632 residues) with the substrate cellotetraose, with the product cellobiose and with 1PE–PGE were determined at 2.88, 1.70 and 2.05 Å resolution, respectively. The overall structure of ExgS consists of 12 helices arranged in an (α/α) $_6$ -barrel and eight β -strands, as shown in Fig. 1(a). In the ExgS–cellotetraose complex, one

cellobiose was found in the open cleft at subsites +1 and +2, and one cellopentaose at subsites –2 to –6 in the tunnel. In the ExgS–cellobiose complex, three cellobiose molecules were identified in the active site, with one cellobiose located in the open cleft at subsites +1 and +2 and the other two located in the tunnel at subsites –2 and –3 and at subsites –5' and –6' (shifted from –5 and –6) (Fig. 1b). The enzyme without substrate (referred to here as ExgS–1PE–PGE) contains 21 glycol fragments (EDO, PEG, PGE, PG4, 1PE and P33) of polyethylene glycol 400, including four fragments at the active site, one PGE at subsite –2, a 1PE at subsites –3 to –5 and two PGEs at subsite –6. The remaining PGE fragments were found on the surface of the protein owing to the fact that the ExgS enzyme was crystallized in a PEG 400 buffer solution. Ethylene glycol fragments found co-crystallized with a protein have also been reported in other enzymes (Abbott *et al.*, 2009). The high *B* factor for the PGE molecules and cellobiose in the ExgS–cellobiose complex may reflect a structure with a high degree of flexibility. The cellobiose may slide along the substrate channel, where three cellobiose molecules are present, and hence result in a high *B* factor.

Based on the location of sugars in the active site, we surmised that cellobiose occupied subsites +1 and +2, subsites –2 and –3 and subsites –5' and –6' in ExgS–cellobiose, while cellobiose occupied subsites +1 and +2 and cellopentaose was positioned at subsites –2 to –6 in the ExgS–cellotetraose complex. In the ExgS–cellotetraose structure, one cellobiose found at subsites +1 and +2 revealed the enzymatic function, while the cellopentaose located at subsites –2 to –6 in the tunnel appears to result from two overlapping cellotetraoses.

Figure 1

(a) The overall structures of three ExgS complexes superimposed in grey (PDB entry 4xwm) are shown as a cartoon drawing of the C α atoms. The molecules in the active site, cellotetraose (cyan and red), cellobiose (green and red) and three PGEs and one 1PE (magenta and red), are shown. (b) The OMIT ($2F_o - F_c$) electron-density map is contoured at the 1.0σ level around the molecules in the active site of ExgS. Three PGE molecules and one 1PE molecule (top), three cellobiose molecules (middle) and the substrate cellotetraose (bottom) are shown in the active site. The structure is oriented so that the reducing end of the sugar is at the top right of the protein molecule.

One of the possible explanations might be that cellotetraoses individually occupy multiple subsites of the enzyme (for example -2 to -5 and -1 to -6) in the crystal, leading to continuous electron density between subsites -2 and -6 . The higher heterogeneity of the substrate-bound enzyme might also explain the lower resolution of the ExgS–cellotetraose crystal. These subsite-binding profiles are similar to those observed previously for the CelS–cellohexaose (Guimarães *et al.*, 2002), CelF–cellotetraose (Parsiegla *et al.*, 2000) and CelF–cellohexaose (Parsiegla *et al.*, 2000) complex structures. CelF–cellohexaose has one cellotriose molecule located at subsites $+1$ to $+3$, indicating the enzymatic hydrolysis function, and five sugar subunits at subsites -2 to -6 . Moreover, one cello-tetraose and one thio-oligosaccharide inhibitor (PIPs-IG3) were found at subsites -1 to $+3$ and $+1$ to $+4$ in the CelF complex, respectively (Parsiegla *et al.*, 2000).

The ExgS structure shares a similar topology with other GH48 enzymes from *C. cellulolyticum* (CelF; Parsiegla *et al.*, 2008), *C. thermocellum* (CelS; Guimarães *et al.*, 2002), *Hahella chejuensis* KCTC 2396 (Sukharnikov *et al.*, 2012), *Caldicellulosiruptor bescii* DSM6725 (Brunecky *et al.*, 2013) and *Thermobifida fusca* (Cel48A; Kostylev *et al.*, 2014). A comparison of ExgS with the five other enzymes gave an overall r.m.s.d. of 0.58 – 0.77 Å for 589 – 607 corresponding C^α atoms. The core structures are almost identical and superimposed well except for the interconnecting loops and the N- and C-terminal regions. In the previously solved CelF–cellobiose (Parsiegla *et al.*, 2000; PDB entry 1fae) and CelS–cellobiose (Guimarães *et al.*, 2002; PDB entry 111y) structures, one cellobiose was found at subsites $+1$ and $+2$, but cellobiose was not found in the tunnel segment of the enzyme. A superposition of the hemi-thiocelooligosaccharides in the CelF E55Q complex (Parsiegla *et al.*, 2008; PDB entry 2qno), *C. bescii* DSM6725 (Brunecky *et al.*, 2013; PDB entry 4txt) and Cel48A from *T. fusca* (PDB entry 4jjj; Kostylev *et al.*, 2014) with the ExgS–cellotetraose complex is displayed in Fig. 2. The substrates from four structures all superimposed well, with the exception of subunit -7 .

al., 2002; PDB entry 111y) structures, one cellobiose was found at subsites $+1$ and $+2$, but cellobiose was not found in the tunnel segment of the enzyme. A superposition of the hemi-thiocelooligosaccharides in the CelF E55Q complex (Parsiegla *et al.*, 2008; PDB entry 2qno), *C. bescii* DSM6725 (Brunecky *et al.*, 2013; PDB entry 4txt) and Cel48A from *T. fusca* (PDB entry 4jjj; Kostylev *et al.*, 2014) with the ExgS–cellotetraose complex is displayed in Fig. 2. The substrates from four structures all superimposed well, with the exception of subunit -7 .

3.2. The active site

The active site, located between the core $(\alpha/\alpha)_6$ barrel and mixed loops and strands, with substrate subsites -7 to $+4$ is approximately 50 Å in length. The tunnel region accommodates substrate subsites -7 to -2 , while the open cleft contains subsites $+1$ to $+4$, with subsite -1 between them. There are 22 residues involved in substrate binding via hydrogen bonds (<3.5 Å) or van der Waals stacking interactions of aromatic residues. Five aromatic residues, Trp302, Trp304, Tyr291, Trp149 and Trp403, are involved in stacking interactions with substrate subunits -7 , -5 , -3 , -2 and $+2$, respectively. The side chains of 18 residues, Glu39, Glu50, Thr105, Asn171, Gln174, Asp208, Gln210, Gln214, Lys216, Thr218, Asp222, Lys266, Tyr267, Tyr315, Tyr395, Asp485, Glu543 and Trp611, make hydrogen-bond interactions with the substrate, while Phe173, Tyr482, Gly483 and Asn219 use

Figure 2

(a) Superimposition of the C^α backbones of ExgS–cellotetraose (red; PDB entry 4xwn), the CelF E55Q complex (green; PDB entry 2qno), CelA (pink; PDB entry 4jjj) and Cel48A (blue; PDB entry 4txt). The four structures superimposed well, except for a few loops on the molecule surface. The r.m.s.d. values for the backbone atoms are 0.58 Å (PDB entries 4xwn and 2qno, 607 atoms), 0.79 Å (PDB entries 4xwn and 4jjj, 591 atoms) and 0.64 Å (PDB entries 4xwn and 4txt, 599 atoms). (b) The substrate site is highlighted. The cellobiose in subsites $+1$ and $+2$ and the pentose in subsites -2 to -6 in ExgS–cellotetraose are shown in red. The cellobiose molecule located at subsites $+1$ and $+2$ and substrate subsites -2 to -7 are shown in CelA (pink; PDB entry 4jjj). Subsites $+2$ to -7 in the CelF E55Q complex (green) and subsites $+1$ to $+3$ in Cel48A are shown in blue. (c) Sequence logos of the four enzymes. The conserved active-site residues are Glu39, Glu50, Thr105, Trp149, Asn171, Phe173, Gln174, Asp208, Gln210, Asp222, Gln214, Lys216, Thr218, Asn219, Lys266, Tyr267, Tyr291, Trp302, Trp304, Tyr315, Tyr395, Trp403, Tyr482, Gly483, Asp485, Glu543 and Trp611 in the ExgS enzyme.

their main-chain carbonyl group or nitrogen to hydrogen-bond to the substrate. The aromatic Trp149, Tyr291, Trp302, Trp304 and Trp403 exhibit hydrophobic stacking interactions thought to reduce the driving force of the sugar ring as it slides

along the pathway. The amino-acid residues involved in hydrogen-bonding interactions or van der Waals stacking with the substrate in ExgS–cellotetraose are shown in Fig. 3. These active-site residues involved in substrate binding are

Figure 3

Stereoview of the active site in ExgS. Subsites –7 to –1 are located in the tunnel, while subsites +1 to +4 are located in the cleft. The aromatic residues Trp149, Trp302, Trp304 and Trp403 are stacked with subsites –2, –7, –5 and +2, respectively. The residues Tyr291 and Trp611 make hydrogen bonds to subsites –4 and +1, respectively. The remaining residues (grey) form hydrogen bonds to the substrate. The key residues Glu50 and Asp222 are shown in blue, while cellobiose and cellopentaose are shown in yellow and red. Subsites –7, –1, +3 and +4 (red and cyan) are modelled in the structure.

Figure 4

(a) Superimposition of the PGE and 1PE molecules (magenta) and tetraose (cyan) in the active sites of ExgS–1PE–PGE and ExgS–cellotetraose. 1PE occupied subsites –3 and –4, while PGE1 occupied the position of subsite –2. Part of the 1PE molecule protrudes towards the residue Tyr395 (magenta) and causes its side chain to rotate 120° away from the active channel. (b) Superimposition of the substrates in ExgS–cellotetraose (cyan) and the CelF E44Q complex (grey). The substrate in CelF E44Q has a different orientation of subsites –1 to –3. Tyr395 (magenta) in ExgS–1PE–PGE and Tyr403 in the CelF E44Q complex have the same rotation and stack with subsite –1 of the hemithioglucosaccharide.

conserved in known GH48 structures and have been found in processive cellulases or other sugar-transportation channels (Varrot *et al.*, 2003; Guimarães *et al.*, 2002; Parsiegla *et al.*, 2008).

In the ExgS–1PE–PGE complex structure, one PGE molecule is located in subsite –2, while a 1PE molecule occupies the region around subsites –3 to –5. In contrast to ExgS–cellotetraose, a significant difference in the ExgS–1PE–PGE structure is that the side chain of Tyr395 rotates approximately 120° away from the active site (Fig. 4*a*) owing to part of the 1PE protruding towards Tyr395. In contrast, the rotation of Tyr395 was equivalent to that of Tyr403 as previously observed in the CelF E44Q–hemithiocollooligosaccharide complex (PDB entry 1g9j; Parsiegla *et al.*, 2008). Tyr403 stacked with subsite –1 (glucose ring) was found in CelF E44Q–hemithiocollooligosaccharide, as shown in Fig. 4(*b*). The superimposition of the C α atoms of CelF E44Q–hemithiocollooligosaccharide with those of the ExgS–1PE–PGE and ExgS–cellotetraose complexes revealed that Tyr395/Tyr403 had similar positions and stacked with subsite –1 (Fig. 4*b*). Before rotation, Tyr395 forms three hydrogen bonds to the O3 atom of subsite –2 as well as the O5 and O6 atoms of subsite –3 found in ExgS–cellotetraose. Tyr395, located at the hinge point between the tunnel and open cleft, was proposed to provide a gate to the tunnel and open-cleft regions, while the substrate slides into the active site and stacks with the glucose ring (Fig. 4*b*). In addition, Glu50 shifts away from subsite +1 and moves closer to Asp222 (with a distance of 2.6 Å), forming a hydrogen bond, in the ExgS–1PE–PGE structure. Coincidentally, Glu55 (equivalent to Glu50 in ExgS) formed a hydrogen bond to Asp230 (equivalent to Asp222 in ExgS) and was also found in the structures of the CelF enzyme and its E44Q mutant (Parsiegla *et al.*, 2008). These results confirm the conformational changes of Glu50 when substrate subsite +1 is not occupied by substrate (Parsiegla *et al.*, 2008).

3.3. The cellobiohydrolase function of ExgS

In the ExgS–cellotetraose complex structure, five sugar subunits were found in the tunnel segment. Additionally, one cellobiose is located in the open cleft at subsites +1 and +2 (see

Fig. 1*b*). As a product of hydrolysis, this may imply that the ExgS enzyme catalyzed a glycosidic bond cleavage during the crystallization procedure. The key residues Glu50 and Asp222, one water molecule (W1) in the ExgS–cellobiose structure and a model subsite –1 are shown in Fig. 5. The water molecule (W1) bridges the C1 atom of subsite –1 and the O $^{\delta 1}$ atom of Asp222 at a distance of 3.4 Å, contributing to the mechanism of hydrolysis. Asp222 acts as general base to remove an H atom from water W1, allowing the resulting hydroxyl group to attack the anomeric C atom at subsite –1. In support of this finding, Saharay *et al.* (2010) identified that Glu87 (equivalent to Glu50 in ExgS) in the GH48 CelS exoglucanase acts as a general acid and Asp255 (equivalent to Asp222 in ExgS) acts as a general base in the catalytic function.

3.4. The Ca $^{2+}$ -binding site

A calcium ion is located on the surface of the protein molecule, bound with distorted pentahedral bipyramidal geometry to four O atoms (from Gln178, Glu183 and Asp397), the backbone carbonyl O atom of Gln178 and two water molecules (Supplementary Fig. S1). The distances between the calcium ion and the ligand atoms are listed in Supplementary Table S1. The carbonyl O atom of Gln178 and a water molecule are bound at the apical positions, while the other atoms are arranged in a distorted pentagonal plane. The calcium ion is located below the active site, distant from the proposed catalytic residues Glu50 and Asp222 (~21 Å). A calcium ion was also identified at a similar location in CelF from *C. cellulolyticum* (Supplementary Table S1), but not in CelS from *C. thermocellum* or the GH48 enzyme from *H. chejuensis* KCTC 2396 (Sukharnikov *et al.*, 2012). Although the calcium ion is not required for enzymatic function, it also appears at similar locations in other GH48 enzymes. Calcium ions found in proteins are commonly bound with a pentagonal bipyramidal geometry. The calcium ion found in the CelF enzyme is also bound with a pentagonal bipyramidal geometry to four O atoms (from Glu185, Asp408 and Glu560), the backbone carbonyl O atom of Arg180 and two water molecules (see Supplementary Table S1). The distances between the calcium ion and its ligands are listed in Supplementary Table S1.

Figure 5

The catalytic residues in the active site of ExgS. Asp222 acts as a base to accept a hydrogen ion from a water molecule (W1). The remaining hydroxyl group is then free to attack the anomeric C atom at subsite –1. Glu50 acts as an acid to provide an H atom to the leaving O4 atom at subsite +1. The sugar –1 subunit is modelled. The dashed line between two atoms represents a hydrogen bond and the distance is shown in Å.

3.5. Docking ExgS with dockerin–cohesin

In a review article (Fontes & Gilbert, 2010), the authors reported that dockerin bridges the cohesin and the enzyme. To further understand the interaction between the enzyme and the dockerin in the cellulosome assembly, we modelled a dockerin–cohesin structure to dock ExgS in order to gain insight into the conformation of the enzyme. Previous reports proposed that the cohesin structure serves as a bridge between the dockerin and the enzyme (Fontes & Gilbert, 2010). The crystal structure of dockerin–cohesin (PDB entries 2vn5 and 2vn6; Pinheiro *et al.*, 2008) showed that the dockerin contains dual binding sites to interact with the cohesin. The structural information obtained from the above-mentioned structures was employed to predict the docking site for dockerin–cohesin on ExgS in this study. The ExgS dockerin sequence exhibits approximately 50% identity to that of the dockerin from *C. cellulolyticum* (Pinheiro *et al.*, 2008). After ZDOCK analysis, we chose one model (Supplementary Fig. S2) from the top five that predicted the dockerin N-terminus to be positioned near the C-terminus of ExgS. However, the model indicated that the interactions between cohesin and ExgS were not as have previously been reported (Fontes & Gilbert, 2010; Gilbert, 2007). Ten residues of the dockerin, Tyr5, Asp7, Asp10, Gly12, Val14, Asn40, Leu41, Lys55, Tyr59 and Met60, and seven residues of the cohesin, Tyr52, Asn81, Thr83, Ser85, Thr123, Lys125 and Asp126, form hydrogen bonds to ExgS residues in the model structure. Most interactions are between the main chains of residues or between main chain and side chain. There is no aromatic stacking present in either the ExgS–cohesin or ExgS–dockerin models. A detailed list of the interacting residues between dockerin–cohesin (PDB entry 2vn5) and ExgS in the complex model is provided in Supplementary Table S2.

4. Conclusions

We have determined X-ray crystal structures of the *C. cellulovorans* exoglucanase S (ExgS) with cellobiose, cellotetraose and 1PE-PGE. The crystal structures of ExgS in complex with cellobiose and cellotetraose, corresponding to enzyme–substrate and enzyme–product complexes, provide an opportunity to elucidate the cellotetraose-cleavage mechanism catalyzed by ExgS. The cellobiose bound at subsites +1 and +2 of ExgS–cellotetraose revealed that this enzyme has the cellobiohydrolase function of cellotetraose binding and cleavage. In the enzyme–product complex, the cellobiose molecule could slide inside the substrate channel, where three cellobiose molecules are present. Structural comparisons reveal a high similarity of the active-site tunnel to other known GH48 enzymes. The PGE1 at subunit –2 and the rotation of the side chain of Tyr395 found in the ExgS–1PE-PGE complex reveal an induced-fit enzyme–substrate mechanism. The structure of ExgS–cellobiose with a modelled subunit –1 shows that Glu50 and Asp222 together with one water molecule are likely to provide the general acid and base required for catalytic function. Glu50 has the freedom to shift towards Asp222 when

subsites +1 and +2 are not occupied by product. In addition, our structural model suggests that ExgS may interact with dockerin and cohesin. Further studies of the functional and structural roles of the key catalytic residues and the residues in the interface between the enzyme and dockerin–cohesin are in progress.

Acknowledgements

We are grateful for access to beamlines BL13B1 and BL13C1 at the National Synchrotron Radiation Research Center in Taiwan. This work was supported by Academia Sinica and the National Science Council, Taiwan, Republic of China (NSC96-3114-P-001-004, NSC97-3114-P-001-004 and NSC101-2311-B-027-001). We thank Mr Meng-Cheng Chiang for protein docking. We would like to express our gratitude for the comments on the manuscript and revision by Dr Jason L. J. Lin and the IMB English editing core.

References

- Abbott, D. W., Macauley, M. S., Vocadlo, D. J. & Boraston, A. B. (2009). *J. Biol. Chem.* **284**, 11676–11689.
- Adams, P. D. *et al.* (2010). *Acta Cryst. D* **66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst. D* **68**, 352–367.
- Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
- Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., Yang, S.-J., Resch, M. G., Adams, M. W. W., Lunin, V. V., Himmel, M. E. & Bomble, Y. J. (2013). *Science*, **342**, 1513–1516.
- 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.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
- Doi, R. H., Goldstein, M., Hashida, S., Park, J.-S. & Takagi, M. (1994). *Crit. Rev. Microbiol.* **20**, 87–93.
- Doi, R. H. & Kosugi, A. (2004). *Nature Rev. Microbiol.* **2**, 541–551.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Fontes, C. M. G. A. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
- Gilbert, H. J. (2007). *Mol. Microbiol.* **63**, 1568–1576.
- Guérin, D. M. A., Lascombe, M.-B., Costabel, M., Souchon, H., Lamzin, V., Béguin, P. & Alzari, P. M. (2002). *J. Mol. Biol.* **316**, 1061–1069.
- Guimarães, B. G., Souchon, H., Lytle, B. L., Wu, J. H. D. & Alzari, P. M. (2002). *J. Mol. Biol.* **320**, 587–596.
- Kostylev, M., Alahuhta, M., Chen, M., Brunecky, R., Himmel, M. E., Lunin, V. V., Brady, J. & Wilson, D. B. (2014). *Biotechnol. Bioeng.* **111**, 664–673.
- Kostylev, M. & Wilson, D. B. (2011). *Appl. Environ. Microbiol.* **77**, 6274–6276.
- 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.
- Nagy, T., Tunnicliffe, R. B., Higgins, L. D., Walters, C., Gilbert, H. J. & Williamson, M. P. (2007). *J. Mol. Biol.* **373**, 612–622.
- Parkkinen, T., Koivula, A., Vehmaanperä, J. & Rouvinen, J. (2008). *Protein Sci.* **17**, 1383–1394.
- Parsiegl, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J.-P., Driguez, H. & Haser, R. (1998). *EMBO J.* **17**, 5551–5562.

- Parsiegla, G., Reverbel, C., Tardif, C., Driguez, H. & Haser, R. (2008). *J. Mol. Biol.* **375**, 499–510.
- Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J.-P., Driguez, H. & Haser, R. (2000). *Biochemistry*, **39**, 11238–11246.
- Pierce, B. G., Hourai, Y. & Weng, Z. (2011). *PLoS One*, **6**, e24657.
- Pinheiro, B. A., Proctor, M. R., Martinez-Fleites, C., Prates, J. A. M., Money, V. A., Davies, G. J., Bayer, E. A., Fontes, C. M. G. A., Fierobe, H.-P. & Gilbert, H. J. (2008). *J. Biol. Chem.* **283**, 18422–18430.
- Saharay, M., Guo, H. & Smith, J. C. (2010). *PLoS One*, **5**, e12947.
- Sukharnikov, L. O., Alahuhta, M., Brunecky, R., Upadhyay, A., Himmel, M. E., Lunin, V. V. & Zhulin, I. B. (2012). *J. Biol. Chem.* **287**, 41068–41077.
- Tamaru, Y., Karita, S., Ibrahim, A., Chan, H. & Doi, R. H. (2000). *J. Bacteriol.* **182**, 5906–5910.
- Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schüle, M. & Davies, G. J. (2003). *Structure*, **11**, 855–864.
- Vasella, A., Davies, G. J. & Böhm, M. (2002). *Curr. Opin. Chem. Biol.* **6**, 619–629.
- Wang, H.-M., Shih, Y.-P., Hu, S.-M., Lo, W.-T., Lin, H.-M., Ding, S.-S., Liao, H.-C. & Liang, P.-H. (2009). *Biotechnol. Prog.* **25**, 1582–1586.