Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the catalytic core module of the *Chaetomium thermophilum* family GH6 cellobiohydrolase Cel6A

Andrew J. Thompson,[a] Tia Heu,[b] Tarana Shaghasi,[b] Romil Benyamino,[b] Aubrey Jones,[b] Esben P. Friis,[c] Keith S. Wilson[a]* and Gideon J. Davies[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England, [b]Novozymes Inc., 1445 Drew Avenue, Davis, California, USA, and [c]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

Correspondence e-mail: keith.wilson@york.ac.uk, gideon.davies@york.ac.uk

Received 24 November 2011
Accepted 16 April 2012

PDB Reference: *Ct*Cel6A, 4a05.

Cellulases, including cellobiohydrolases and endoglucanases, are important enzymes involved in the breakdown of the polysaccharide cellulose. These catalysts have found widescale industrial applications, particularly in the paper and textile industries, and are now finding use in 'second-generation' conversion of biomass to biofuels. Despite this considerable biotechnological application, and undoubted future potential, uncertainty remains as to the exact reaction mechanism of the inverting cellulases found in the GH6 family of carbohydrate-active enzymes. In order to gain additional understanding as to how these societally beneficial biocatalysts function, the crystal structure of a GH6 cellobiohydrolase from *Chaetomium thermophilum*, *Ct*Cel6A, has been solved. This structure reveals a distorted α/β-barrel fold comprising a buried tunnel-like active site quite typical of Cel6A enzymes. Analysis of an enzyme–product complex (cellobiose in the −3 and −2 subsites and cellotetraose in subsites +1 to +4) supports the hypothesis that this group of enzymes act *via* an atypical single-displacement mechanism. Of particular note in this analysis is an active-centre metal ion, Li^+, the position of which matches the position of the positively charged anomeric carbon of the oxocarbenium-ion-like transition state.

1. Introduction

The crystalline polysaccharide cellulose represents one of the most abundant biopolymers on earth. Historical analyses have suggested that cellulose accounts for approximately half of all known biomass (Hess *et al.*, 1928; Brown & Montezinos, 1976). More recent work has suggested an annual production of plant biomass on the >10 gigatonne scale, suggesting a global terrestrial pool of greater than 2000 gigatonnes (Falkowski *et al.*, 2000; Zhao & Running, 2010). Plant biomass therefore has considerable potential for exploitation as a renewable energy source. Cellulose consists of β-1,4-linked D-glucopyranosides which form long chains; the chains assemble to create a crystalline rod-like polymer and are linked together *via* hydrogen-bonding and van der Waals interactions. Unlike similar biopolymers such as amylose or amylopectin, cellulose presents as an entirely linear-chain molecule with no branches or coiling (the structure of cellulose has recently been reviewed in Nishiyama, 2009). This fibril-like architecture confers a relatively large tensile strength and allows the molecule to serve as a major structural component of plant cell walls. Cellulose is highly stable, with an estimated half-life of approximately four million years; consequently, its biodegradation is quite challenging. Nature has evolved a large variety of cellulases (cellulose-specific glycoside hydrolases; GHs), which are found in many of the CAZy sequence-based

GH families (for the CAZy classification, see Cantarel *et al.*, 2009). These enzymes possess a range of distinct activities in order to maximize degradation of this immense biopolymer (for reviews, see Hildén & Johansson, 2004; Sandgren *et al.*, 2005; Koivula *et al.*, 1998; Xu *et al.*, 2009; Gilbert *et al.*, 2008; Himmel *et al.*, 2007).

Cellulases are, in general, modular enzymes, often consisting of a catalytic domain and one or more carbohydrate-binding modules [CBMs; previously termed carbohydrate-binding or cellulose-binding domains (CBDs); Van Tilbeurgh *et al.*, 1986; Linder & Teeri, 1997; Boraston *et al.*, 2004; Shoseyov *et al.*, 2006]. From early studies, cellulases have traditionally been grouped into two broad categories: cellobiohydrolases (CBHs; sometimes, but perhaps controversially, termed exo-glucanases) and endo-glucanases. This classification reflects their respective catalytic functions: cellobiohydrolases act processively, liberating cellobiose from the ends of exposed cellulose polysaccharides (Fig. 1), whilst endo-glucanases cleave glycosidic bonds within these chains, creating additional terminal sugars for their more 'exo-acting' counterparts (Karlsson *et al.*, 1999; Väljamäe *et al.*, 1999).

Cellulases are widely used in the paper and textile industries, in biological washing products and in the environmentally friendly processing of plant-based waste, and increasingly for the production of 'second-generation' (cellulosic) biofuels (Carroll & Somerville, 2009; Xu *et al.*, 2009; Gilbert *et al.*, 2008; Himmel *et al.*, 2007; Viikari *et al.*, 2007). For many years, the cellular machinery of thermophilic and extremophilic organisms has been of central interest within many industrial fields (reviewed in Haki & Rakshit, 2003; Viikari *et al.*, 2007; Heinzelman *et al.*, 2009). Characteristic properties such as high thermostability and broad pH tolerance make the enzymes employed by these microbes highly desirable through their potential to enhance both the efficacy and the profitability of a wide range of commercial processes (Haki & Rakshit, 2003). The CAZy classification shows that the thermophilic soil-dwelling fungus *Chaetomium thermophilum* possesses at least seven β-glucosidase/cellulase enzymes from GH families 3, 6 and 7 (Cantarel *et al.*, 2009). The cellobiohydrolases from the GH6 family of enzymes are processive enzymes that remove the disaccharide cellobiose in a processive manner, acting from the nonreducing end of cellulosic polysaccharides (Rouvinen *et al.*, 1990; Koivula *et al.*, 1996). This disaccharide-liberating activity is believed to reflect the structural properties of the cellulose molecule. Each β-1,4-glucose moiety is rotated 180° relative to the next along the axis of the chain; thus, only every second glycosidic linkage is presented to the catalytic machinery in the correct orientation to allow bond hydrolysis. The catalytic core module of the Cel6A cellobiohydrolases has been shown to form a modified TIM-barrel structure (Rouvinen *et al.*, 1990) in which a buried cleft spanning the C-terminal region of the domain and capable of receiving β-1,4-glucose polysaccharides of various lengths is thought to contain the catalytic active site (Koivula *et al.*, 1996; Varrot, Hastrup *et al.*, 1999). It has been demonstrated that Cel6A CBHs perform catalysis with inversion of anomeric configuration (Knowles *et al.*, 1988), suggesting a single-displacement mechanism (discussed further below in light of the three-dimensional structure presented here).

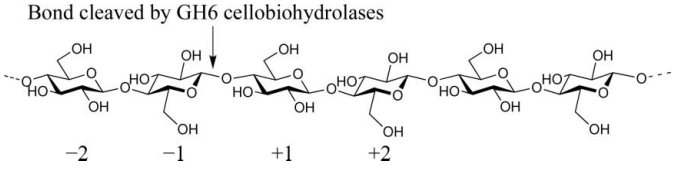

Figure 1
Action of GH6 cellobiohydrolases: the hydrolysis of cellulose strands in a processive manner moving along from the nonreducing end to liberate predominantly the disaccharide cellobiose.

Elucidation of both the physical and the kinetic properties of novel enzymes is often beneficial prior to their successful application in a widescale industrial setting. Recombinant enzymes need to be available in large quantities, to have broad operational ranges in terms of reaction conditions and to maintain high catalytic efficiency across this range. Thus, in recent years much emphasis has been placed on the need for the improvement of existing commercial biocatalysis systems through implementation of new or evolved enzymes that are able to more adequately satisfy these requirements (Szijártó *et al.*, 2008). Here, we present the crystal structure of the catalytic core domain of *C. thermophilum* Cel6A (*Ct*Cel6A) at a resolution of 1.9 Å in complex with molecules of both cellobiose and cellotetraose. We also show that this enzyme is similar to existing thermophile-derived Cel6A structures, thus supporting previous observations of an indirect single-displacement mechanism for catalysis and confirming the conserved structural features responsible for the increased thermostability observed within this group of proteins. Of particular note in this analysis is that an active-centre ion, in all likelihood Li^{+}, mimics the position adopted by the positively charged anomeric carbon of the oxocarbenium-ion-like transition state during catalysis.

2. Materials and methods

2.1. Cloning, expression and purification

A deletion mutation was performed on a vector (pAll017) encoding the cDNA of wild-type *C. thermophilum* CBHII using a QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, California, USA) to loop out the ∼285 bp N-terminal CBM-linker region. Oligonucleotides 066224 (5′-CTCTTGCGGCCACTTCGCTGGCTACGGCCAGCTACAACGGCAACCCG-3′) and 066225 (5′-CGGGTTGCCGTTGTAGCTGGCCGTAGCCAGCGAAGTGGCCGCAAGAG-3′) were designed to perform the deletion. One plasmid designated pTH154 comprising the *C. thermophilum* CBHII signal sequence and core without the CBM linker was verified by sequencing using oligonucleotides priming outside the coding region.

Aspergillus oryzae JaL250 (WO 99/61651) protoplasts were prepared according to the method of Christensen *et al.* (1988) and transformed with pTH154. 12 transformants of pTH154 were isolated to individual PDA plates. Once the cultures

from each spore-purified transformant were confluent and had sporulated, spore stocks were made by applying 5 ml sterile-filtered 0.01% Tween-80 (diluted with glass-distilled water) onto the centre of each PDA plate and using a sterile spreader to scrape the spores into solution. 8 µl spore stock was used to inoculate 1 ml MDU-2BP medium in 24-well culture plates, which were incubated at 307 K for 5 d. Broth samples were harvested on day 5. SDS–PAGE was carried out using Criterion Tris–HCl (5% resolving) gels with The Criterion System (Bio-Rad Laboratories Inc., Hercules, California, USA). The transformant showing the highest expression of *C. thermophilum* CBHII core based on the protein gel was designated *A. oryzae* TH125. *A. oryzae* TH125 was cultivated in 500 ml MDU-2BP at 307 K at 220 rev min^{-1} for 5 d.

Approximately 250 ml cleared supernatant from *A. oryzae* expressing the catalytic core module of *Ct*Cel6A was desalted into 20 m*M* sodium acetate buffer pH 5 using a 400 ml desalting column packed with Sephadex G-25. 350 ml desalted sample was then concentrated to 50 ml using an Amicon ultrafiltration apparatus with 10 kDa molecular-weight cutoff membrane. 50 m*M* Tris buffer pH 8 was added to the concentrated material, a sample of which was then loaded onto a 70 ml Q Sepharose High Performance column equilibrated with 20 m*M* Tris buffer pH 8. A ten column-volume salt gradient of 0–50% 20 m*M* Tris buffer pH 8 plus 1 *M* sodium chloride was used to elute bound material. Collected fractions were analysed by SDS–PAGE and pure fractions were pooled.

Purified *Ct*Cel6A was deglycosylated by the addition of EndoH (endoglycosidase H; NE BioLabs). 3000 units of EndoH were added to 5–10 mg *Ct*Cel6A in a 5 ml reaction containing 50 m*M* sodium acetate buffer pH 5. The reaction mixture was sterile-filtered and incubated at 310 K for 3 d. Samples without EndoH were included as a control. The reaction mixture was purified into 20 m*M* Tris buffer pH 8 using anion-exchange chromatography (MonoQ 5/5) with a sodium chloride gradient for elution. Collected fractions were analysed by SDS–PAGE and pure fractions were pooled.

2.2. Crystallization, data collection and structure solution

Purified *Ct*Cel6A was screened for crystallization at a concentration of approximately 10 mg ml^{-1} following incubation with 10 m*M* cellobiose. Preliminary hits were obtained in several conditions from the PACT premier screen (Molecular Dimensions Ltd, Newmarket, England). Crystals of *Ct*Cel6A were subsequently grown by hanging-drop vapour diffusion at 292 K in both equal and 2:1 ratios of protein:reservoir solution [20%(*w*/*v*) polyethylene glycol (PEG) 6000, 0.1 *M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.0, 0.14 *M* lithium chloride]. All crystals were cryoprotected by incubating them for approximately 10 s in reservoir solution supplemented with 10%(*w*/*v*) glycerol. *Ct*Cel6A diffraction data were collected on beamline I04-1 of Diamond Light Source, Didcot, England and all images were processed using *iMOSFLM* and *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011). The crystal structure of *Ct*Cel6A was solved by molecular replacement to 1.9 Å resolution using the *CCP*4 implementation of the *MOLREP* program (Vagin & Teplyakov, 2010) with the coordinates of *Humicola insolens* Cel6A (*Hi*Cel6A; PDB entry 1oc6; Varrot, Macdonald *et al.*, 2003) as a search model using default input parameters. Model building was undertaken using *Buccaneer* (Cowtan, 2006), with maximum-likelihood refinement of the *Ct*Cel6A model conducted through numerous cycles of *REFMAC* (Murshudov *et al.*, 2011) and manual correction using *Coot* (Emsley & Cowtan, 2004).

3. Results and discussion

The crystals of *Ct*Cel6A belonged to the orthorhombic space group $P2_12_12_1$ (unit-cell parameters a = 50, b = 76, c = 107 Å, $\alpha = \beta = \gamma = 90^\circ$) with one protein molecule present in each asymmetric unit. The final model consisted of a single

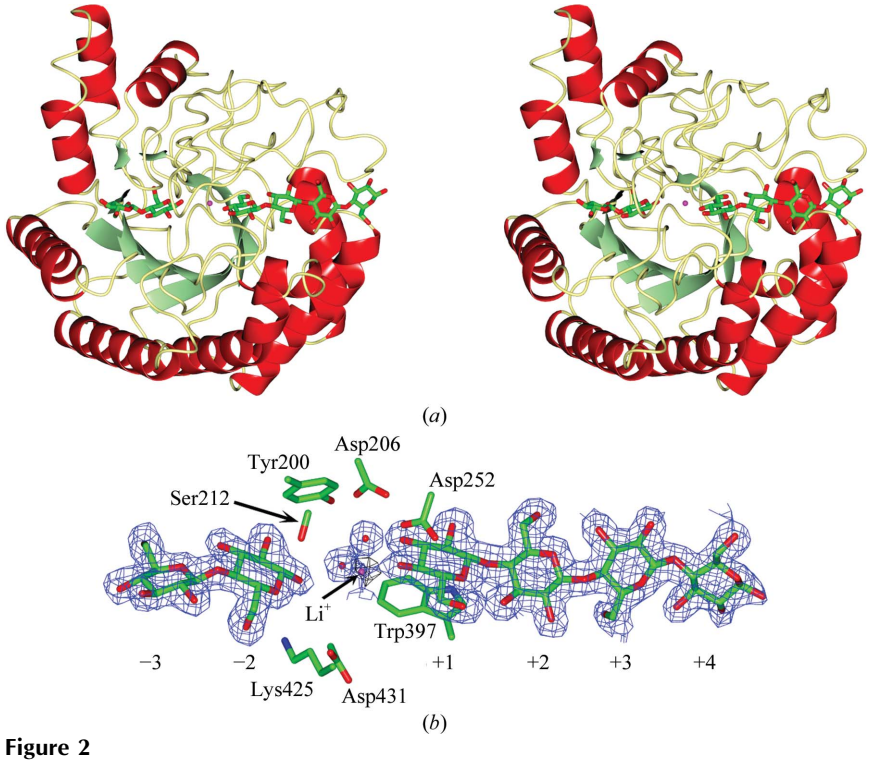

Figure 2
Structure of *Ct*Cel6A. (*a*) Divergent ('wall-eyed') stereo cartoon of *Ct*Cel6A with the ligand in green/red liquorice. (*b*) Observed electron density (maximum-likelihood weighted $2F_{obs} - F_{calc}$ contoured at 0.36 e Å^{-3}) for the ligands in the active centre of *Ct*Cel6A. The disaccharide cellobiose is bound in the −3 and −2 subsites, whilst cellotetraose is observed in the +1 to +4 sites. Residual $F_{obs} - F_{calc}$ difference density in the −1 site is interpreted as an active-centre Li^+ ion (Fig. 3). This figure was drawn using *CCP*4*mg* (McNicholas *et al.*, 2011).

Table 1
*Ct*Cel6A data-processing and refinement statistics.

Values in parentheses are for the outer resolution shell.

Data processing	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 49.7, b = 76.1, c = 107.2
Wavelength (Å)	0.917
Molecules in asymmetric unit	1
Resolution range (Å)	49.71–1.90 (2.00–1.90)
R_{merge}†	0.086 (0.190)
$\langle I/\sigma(I)\rangle$	14.6 (7.4)
Completeness (%)	99.9 (99.8)
Multiplicity	7.1 (7.1)
Refinement	
No. of reflections	31106
R_{cryst}†	0.17
R_{free}†	0.19
Mean B values (Å^2)	
Protein	20
Ligands/ions	25
Solvent	26
R.m.s.d. bonds (Å)	0.009
R.m.s.d. angles (°)	1.22
Ramachandran statistics (%)	
Preferred regions	95.20
Allowed regions	3.95
Outliers	0.85
PDB code	4a05

† Formulae for R_{merge} and R (R_{cryst} and R_{free}) as applied within *SCALA* (Winn *et al.*, 2011) and *REFMAC* (Murshudov *et al.*, 2011) are shown below: $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, $R = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl} |F_{obs}|$.

continuous chain spanning residues Tyr116–Pro475. The enzyme is at least partially N-glycosylated, with electron density visible for two β-1,4-linked *N*-acetylglucosamines covalently bound to the side-chain N atom of Asn167. Typical of Cel6A catalytic core domains, *Ct*Cel6A has a single domain with a distorted α/β-barrel architecture (Fig. 2*a*). A central sheet motif composed of six β-strands forms an incomplete circular structure, with the remaining space occupied by an unstructured N-terminal loop region. Surrounding the central β-sheet, eight major and two minor α-helices form the outer surface of the catalytic core module. A buried cleft containing the catalytic active site runs laterally above the inner β-barrel and is formed by two large flexible loops and two minor α-helices. The cleft contains one cellobiose molecule and one cellotetraose molecule (Fig. 2*b*), which occupy subsites −2 and −3 and +1 to +4, respectively (for a review of subsite nomenclature, see Davies *et al.*, 1997). The cellotetraose molecule is likely to arise from the presence of longer oligosaccharides as impurities within the cellobiose sample, although the possibility of a reverse hydrolysis reaction cannot be excluded given the excess concentration of cellobiose used. As part of the historical debate (see, for example, Ståhlberg *et al.*, 1993; Armand *et al.*, 1997; Varrot, Schülein *et al.*, 1999; Boisset *et al.*, 2000) over whether cellobiohydrolases are classical exo enzymes whose structure demands that they excise two sugar units from the nonreducing end, it is important here to note the existence of a −3 subsite, confirming that there is no steric barrier to sugar binding at this position of the chain within this particular Cel6A enzyme. Indeed, the existence of such a subsite has been hinted at previously, with earlier work on Cel6A (previously CBHII) enzymes showing cleavage of MeUmb(Glc)$_5$ to yield MeUmb(Glc)$_2$ and Glc$_3$ in addition to the expected products MeUmb(Glc)$_3$ and Glc$_2$ (Van Tilbeurgh *et al.*, 1986; Claeyssens *et al.*, 1989). However, it still has yet to be proved whether this subsite (and additional associated activity) might play a functional role in nature or is simply an artefact that is only observable through the use of small-molecule substrates and ligands under *in vitro* conditions. It is unclear why cellobiose occupies subites −3/−2 as opposed to −2/−1; this may reflect the impaired binding of a nondistorted and noncharged sugar moiety coupled with the binding of Li$^+$ in the active centre (discussed below).

A metal ion, most likely Li$^+$, observed in data sets collected from two separate crystals and derived from the crystallization conditions is observed in the −1 (catalytic) subsite, where it is coordinated by three active-site water molecules and the O4 hydroxyl of cellotetraose (Figs. 2 and 3). Several lines of evidence indicate that the metal is Li$^+$, as opposed to any other metal. Firstly, LiCl is present in the crystallization conditions at 0.14 *M*. The second line of evidence is the electron density:

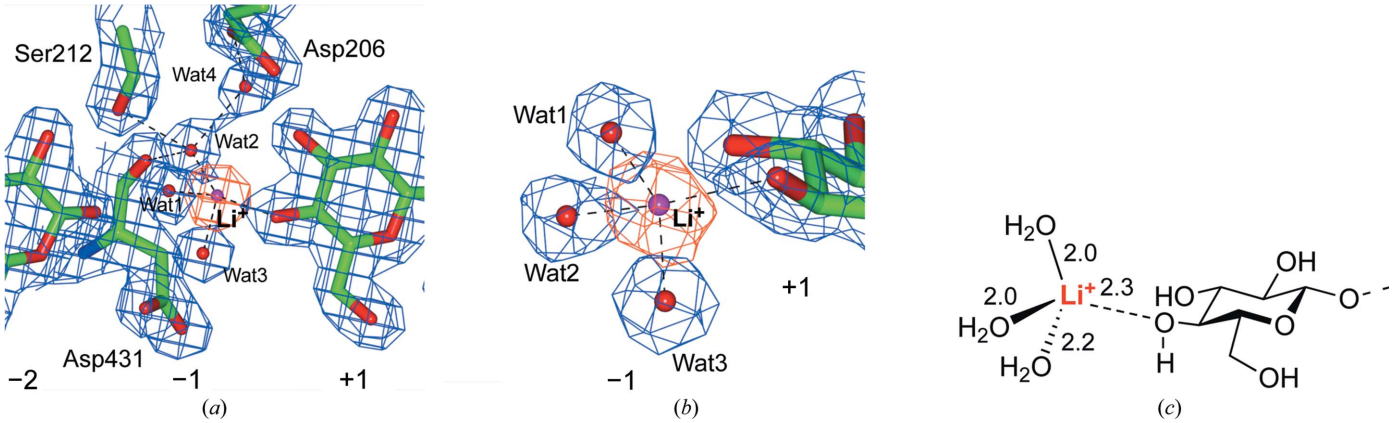

Figure 3
Geometry of an active-centre Li$^+$ ion. (*a*) Electron density (*REFMAC*-weighted $2F_{obs} - F_{calc}$ in blue and $F_{obs} - F_{calc}$ 'difference' density in red, contoured at 0.46 and 0.18 e Å^{-3}, respectively) showing the −2 to +1 subsites and the key catalytic residues discussed in the text. (*b*) Close up of the Li$^+$ site. (*c*) Schematic diagram of the coordination geometry of the Li$^+$ ion. (*a*) and (*b*) were drawn with *CCP*4*mg* (McNicholas *et al.*, 2011).

at approximately 0.45 e Å^{-3} there is no density at the metal centre in the $2F_{obs} - F_{calc}$ synthesis (Figs. 3*a* and 3*b*), yet there is clear spherical difference density in the $F_{obs} - F_{calc}$ synthesis (Figs. 3*a* and 3*b*). Although collected to Bragg spacings of 1.9 Å, the data are of exceptionally high quality, with a mean outer resolution shell $\langle I/\sigma(I)\rangle$ of 7.5 and an R_{merge} of 0.19 (Table 1). Beyond the active-centre ion, there are neither difference density peaks with coordination suggesting Li^+ ions elsewhere in the structure nor do any peaks assigned as water molecules show the short interaction distances indicative of a metal ion. Furthermore, the coordination is entirely consistent with that known for Li^+ in both the distance to ligands and the coordination geometry. The most frequently observed 'small-molecule' geometry for Li^+ is a tetrahedral coordination (Li^+ coordination chemistry is summarized in the classical review by Olsher *et al.*, 1991). This geometry is exactly what is observed in Cel6A. The four ligands of the metal are three water molecules and the O4 hydroxyl group of the sugar in the +1 subsite (Figs. 3*b* and 3*c*). Importantly for correct assignment of the metal, the metal–ligand distances refine to 2.0, 2.0, 2.1 and 2.2 Å, which are exactly as observed for Li^+ bonding to ligands in tetrahedral geometry (mean distance approximately 1.9–2.1 Å; Olsher *et al.*, 1991). Such distances are not consistent with the mean distance of 2.4 Å observed for Na^+ in a similar geometry (Harding, 2004).

Binding of the Li^+ ion almost certainly reflects the chemistry of catalysis and thus provides catalytic insight. Glycosidase catalysis occurs *via* oxocarbenium-ion-like transition states in which partial positive charge develops on the anomeric carbon of the substrate (Fig. 4*a*). Indeed, mimicry of the positive charge of the transition state is a common strategy for glycosidase inhibition (for reviews, see Bols, 1998; Lillelund *et al.*, 2002; Gloster & Davies, 2010). This favouring of binding of positively charged species is also the basis for the serendipitous inhibition of glycosidases by Tris buffers (Roberts & Davies, 2012), which has even been exploited in glycosidase inhibitors (Taylor *et al.*, 2007). It is therefore insightful that structural overlays with the *Hi*Cel6A D416A variant in complex with cellobio-derived isofagomine (PDB entry 1ocn; Varrot, Macdonald *et al.*, 2003) show the *Ct*Cel6A Li^+ ion to be located at a position equivalent (Fig. 4*b*) to the positively charged ring N atom of isofagomine, indicating a conservation of positive charge analogous to that present in the oxocarbenium-ion-like transition state. Furthermore, overlap with the known (seven in total) Michaelis complexes of family GH6 confirms that the Li^+ ion binds in the same location as the anomeric C atom of the Michaelis complex (Fig. 4*c*), consistent with the partial positive charge in a distorted sugar revealed through computational work (Davies *et al.*, 2012).

Li^+ binding may also be favoured through the location of its interacting water ligands. Five published GH6 complexes in which the −1 site is unoccupied show the positions of three water molecules (Fig. 3*b*) coordinating lithium to be 'conserved'. Two of these solvent positions reflect the binding positions for the O5 endocyclic O atom and the exocyclic O2 in the active-site-spanning complexes, with the third above the plane of the ring, approximately 2 Å from C1. The observation of Li^+ in the active centre of Cel6A reported here may suggest a future potential strategy for glycosidase inhibition through the incorporation of metal-binding scaffolds.

The *Ct*Cel6A catalytic module shows strong similarity to previously solved structures for this class of enzyme. Indeed, as expected for enzymes in the same CAZy classification, the *Ct*Cel6A structure shows both high structural and sequence similarity (Fig. 5) to previously solved enzymes within this class, including *Hi*Cel6A (Varrot, Hastrup *et al.*, 1999) and *Hypocrea jecorina* Cel6A (*Hj*Cel6A; Koivula *et al.*, 1996). These proteins show r.m.s.d.s of 0.71 and 0.77 Å mapped over 353 and 351 matching C^α positions, respectively (calculated using *SUPERPOSE* from the *CCP*4 program suite; Winn *et al.*, 2011). *Ct*Cel6A shows 77 and 63% sequence identity to *Hi*Cel6A and *Hj*Cel6A, respectively, with strong areas of sequence conservation observed around the putative catalytic residues. Superposition of the three Cel6A structures shows the location and formation of the catalytic site to be highly conserved. The conserved tryptophan residues 163, 397, 300

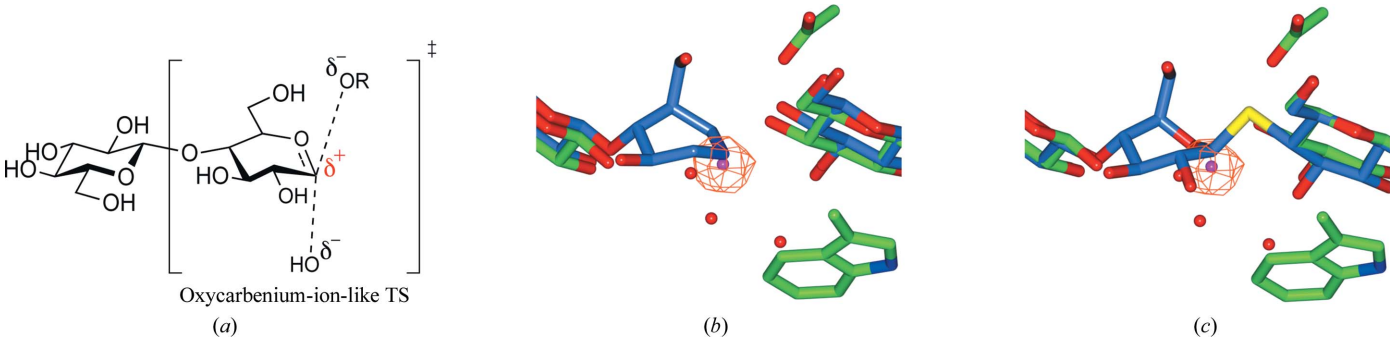

Figure 4
Catalytic relevance of the active-centre Li^+. (*a*) Oxocarbenium-ion-like transition state for catalysis with inversion of anomeric configuration by Cel6A. Note the partial positive charge on the anomeric C atom of the transition state. (*b*) Overlay (−1 to +1 subsites only) of the ligand complex observed here (pale green) with the cellobiose-derived isofagomine complex (blue) of *Hi*Cel6A (PDB entry 1ocn; Varrot, Macdonald *et al.*, 2003). (*c*) Overlay as in (*b*) but with the 'cellobiose-S-cellobiose' Michaelis complex of *Hi*Cel6A (PDB entry 1gz1) shown in blue. The active-centre lithium ion observed in *Ct*Cel6A, shown here as $F_{obs} - F_{calc}$ density at 0.18 e Å^{-3}, is observed to bind in essentially the same location as the positively charged nitrogen centre of the isofagomine/the anomeric C atom of the Michaelis complex.

and 303 in *Ct*Cel6A make stacking interactions with the glucose sugar rings in the −2, +1, +2 and +4 subsites, respectively.

Whereas the retaining glycosidase mechanism occurs *via* a conventional double-displacement reaction that is now widely understood (for a recent review, see Vocadlo & Davies, 2008), catalysis by GH6 enzymes occurs with inversion of anomeric configuration. Furthermore, whilst many inverting enzymes display a classical acid and base catalytic dyad, the exact mechanism of GH6 enzymes is far less clear. Previous structural and biochemical work has identified the catalytic acid residue as Asp221 in *Trichoderma reesei* (now known as *H. jecorina*) Cel6A (Rouvinen *et al.*, 1990), equating to Asp226 in the *H. insolens* enzyme (Varrot, Hastrup *et al.*, 1999; Varrot, Schülein *et al.*, 1999); this assignment is also consistent with elegant solution work on the related CenA endoglucanase (Damude *et al.*, 1995; for its positional equivalent within the various Cel6A homologues, see http://www.cazypedia.org/index.php/GH6). In *Ct*Cel6A, the equivalent Asp252 is positioned immediately above the O4 hydroxyl of cellotetraose, poised perfectly to act as the Brønsted acid. Although Cel6 enzymes perform catalysis with inversion of anomeric configuration, assignment of a 'classical' catalytic base residue has remained elusive; site-directed mutagenesis has suggested a base (Damude *et al.*, 1995), but none of the currently available structures show any proton acceptor within an acceptable hydrogen-bonding distance of the putative nucleophilic water molecule. Whilst kinetic evidence has shown a significant reduction in catalysis upon the introduction of a D405N (or equivalent) mutation in, for example, *Hi*Cel6A (Varrot, Frandsen *et al.*, 2003), the apparent lack of a coordinated water molecule makes it unclear as to how the Cel6 enzyme would successfully complete a classical single-displacement reaction. In order to account for these apparent inconsistencies, various structural and kinetics studies, notably on *T. reesei* (*Tr*Cel6A; Koivula *et al.*, 1996, 2002), led to the proposal of an unusual mechanism for base-assisted catalysis in which an aspartate (Asp175 in *Tr*Cel6A, although with a possible role for Asp401) acts as an 'indirect' proton acceptor *via* a chain of coordinated water molecules in a so-called Grotthus-type mechanism (von Grotthus, 1806; Agmon, 1995; Koivula *et al.*, 2002). This proposal is consistent with structural data on inhibitor complexes of Cel6 enzymes (for example, Varrot, Macdonald *et al.*, 2003), in which the equivalent residue Asp206 forms a hydrogen bond to the O3 hydroxyl of the +1 subsite glucosyl moiety of cellotetraose whilst both groups simultaneously coordinate a water molecule (Wat 4 in Fig. 4*b*). An additional water molecule (Wat 2) is positioned within hydrogen-bonding distance of the first water, the side-chain hydroxyl group of Ser212, the backbone hydroxyl group of Asp431 and the active-site lithium ion. These key residues are all conserved in the GH6 enzymes (as shown in the partial sequence alignments in Fig. 5). Wat 2 is appropriately positioned (relative to the Li^+ ion) to act as a catalytic nucleophile at the −1 subsite, allowing proton transfer *via* an intermediate water to the Asp206 acceptor and thus completing the single-displacement reaction.

```
CtCel6A                                 QSCSSVWGQCGGINYNGPTCCQSGSVCAYLNDWYSQ 60
HiCel6A                                 QNCAPTWGQCGGIGFNGPTCCQSGSTCVKQNDWYSQ 36
TrCel6A                                 QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQ 60
                                        * *   ******    ***** *** *   ** ***

CtCel6A         CIPGQAQPGTTSTTARTTSTSTTSTSSVRPTTSNTPVTTAPP-TTTIPGGASSTASYNGN 119
HiCel6A         CLPG-SQVTTTSTTSTSSSSTTSRATSTTRTGGVTSITTAPTRTVTIPGGATTTASYNGN 95
TrCel6A         CLPG----AASSSSSTRAASTTSRVSPTTSRSSSATPPPGST-TTRVPPVGSGTATYSGN 115
                * **       *         *                      *   *     ** * **

CtCel6A         PFSGVQLWANTYYSSEVHTLAIPSLS-PELAAKAAKVAEVPSFQWLDRNVTVDTLFSGTL 178
HiCel6A         PFEGVQLWANNYYRSEVHTLAIPQITDPALRAAASAVAEVPSFQWLDRNVTVDTLLVETL 155
TrCel6A         PFVGVTPWANAYYASEVSSLAIPSLT-GAMATAAAAVAKVPSFMWLDT-LDKTPLMEQTL 173
                ** **  *** ** ***  ****          *  ** **** ***       *   **

CtCel6A         AEIRAANQRGANPPYAGIFVVYDLPDRDCAAAASNGEWSIANNGANNLQRYIDRIRELLI 238
HiCel6A         SEIRAANQAGANPPYAAQIVVYDLPDRDCAAAASNGEWAIANNGANNYKGYINRIREILI 215
TrCel6A         ADIRTANKNGGN--YAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVV 231
                 ** **  * *  **   ************ *****  **  *      **  **

CtCel6A         QYSDIRTILVIEPDSLANMVTNMNVQKCSNAASTYKELTVYALKQLNLPHVAMYMDAGHA 298
HiCel6A         SFSDVRTILVIEPDSLANMVTNMNVAKCSGAASTYRELTIYALKQLDLPHVAMYMDAGHA 275
TrCel6A         EYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHA 291
                  ** ** ********** ***    **  * * * *   **  ** ** **** *****

CtCel6A         GWLGWPANIQPAAELFAQIYRDAGRPAAVRGLATNVANYNAWSIASPPSYTSPNPNYDEK 358
HiCel6A         GWLGWPANIQPAAELFAKIYEDAGKPRAVRGLATNVANYNAWSISSPPPYTSPNPNYDEK 335
TrCel6A         GWLGWPANQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEK 351
                ********  *** ***  *  *  * * ********** * * *** **  *  * **

CtCel6A         HYIEAFAPLLRNQGFD-AKFIVDTGRNGKQPTGQLEWGHWCNVKGTGFGVRPTANTGHEL 417
HiCel6A         HYIEAFRPLLEARGFP-AQFIVDQGRSGKQPTGQKEWGHWCNAIGTGFGMRPTANTGHQY 394
TrCel6A         LYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSL 411
                 ** *  ***   *   * ** * ** *******  ** ***  ***** ** ****

CtCel6A         VDAFVWVKPGGESDGTSDTSAARYDYHCGLSDALTPAPEAGQWFQAYFEQLLINANPPF- 476
HiCel6A         VDAFVWVKPGGECDGTSDTTAARYDYHCGLEDALKPAPEAGQWFQAYFEQLLRNANPPF- 453
TrCel6A         LDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL 471
                 * ********* *****  * * * ** * *** *** ** ****** *** **** *
```

Figure 5
Partial sequence alignment of GH6 cellobiohydrolases of known three-dimensional structure (illustrated by *Hi*Cel6A and *Tr*Cel6A in addition to the *Ct*Cel6A enzyme reported here). Key catalytic residus discussed in the text and shown in Fig. 3(*a*) are highlighted in bold.

The structure of the *Ct*Cel6A catalytic core module thus conforms with those of other Cel6As, supporting the current biological model of the GH6 family. The synergistic action of cellobiohydrolases and endoglucanases allows the efficient breakup and biocatalysis of large cellulose polymers with a net inversion of anomeric configuration, an activity that has been widely demonstrated to be of great use in many industrial processes. More recently, the requirement for cellulases with high thermostability and broad pH tolerances has been highlighted in additional ventures

such as the hydrolysis of lignocellulose to yield fermentable sugars for use in the production of biofuels (Viikari *et al.*, 2007). The characterization of both biological and engineered enzymes has most often sought to identify targets with generally increased tolerance to both temperature and pH. Features such as these are highly desirable in commercial biocatalysts, as reactions carried out at higher temperatures and more extreme pH are known to reduce substrate viscosity (particularly for large biopolymers such as cellulose), increase conversion rates with lower net energy consumption and reduce the risk of biological contamination. However, candidate biological enzymes falling into this category still often have limited commercial potential owing to very low natural production levels by their host organisms. With this in mind, studies such as that of Heinzelman *et al.* (2009) have attempted to engineer chimaeric enzymes with both desirable physical tolerances and increased catalytic activity under these extreme reaction conditions using recombination techniques based on known structural, kinetic and sequence data. The structure of *Ct*Cel6A presented here will augment the relatively limited structural data within this interesting class of enzymes and will aid in the future design of commercially viable cellobiohydrolases that are able to enhance many of the increasingly important biocatalytic challenges facing society.

Research on plant cell-wall degrading enzymes in the York laboratory is funded by the Biotechnology and Biological Sciences Research Council (BBSRC) through grant BB/I014802. AJT is a BBSRC-funded student. GJD is a Royal Society/Wolfson Research Merit award recipient. KSW thanks Novozymes A/S for funding.

References

Agmon, N. (1995). *Chem. Phys. Lett.* **244**, 456–462.
Armand, S., Drouillard, S., Schülein, M., Henrissat, B. & Driguez, H. (1997). *J. Biol. Chem.* **272**, 2709–2713.
Boisset, C., Fraschini, C., Schülein, M., Henrissat, B. & Chanzy, H. (2000). *Appl. Environ. Microbiol.* **66**, 1444–1452.
Bols, M. (1998). *Acc. Chem. Res.* **31**, 1–8.
Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
Brown, R. M. & Montezinos, D. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 143–147.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Carroll, A. & Somerville, C. (2009). *Annu. Rev. Plant Biol.* **60**, 165–182.
Christensen, T., Woeldike, H., Boel, E., Mortensen, S. B., Hjortshoej, K., Thim, L. & Hansen, M. T. (1988). *Nature Biotechnol.* **6**, 1419–1422.
Claeyssens, M., Van Tilbeurgh, H., Tomme, P., Wood, T. M. & McRae, S. I. (1989). *Biochem. J.* **261**, 819–825.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, R. C. & Warren, R. A. (1995). *Biochemistry*, **34**, 2220–2224.
Davies, G. J., Planas, A. & Rovira, C. (2012). *Acc. Chem. Res.* **45**, 308–316.
Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Falkowski, P. *et al.* (2000). *Science*, **290**, 291–296.
Gilbert, H. J., Stålbrand, H. & Brumer, H. (2008). *Curr. Opin. Plant Biol.* **11**, 338–348.
Gloster, T. M. & Davies, G. J. (2010). *Org. Biomol. Chem.* **8**, 305–320.
Grotthus, C. J. T. von (1806). *Ann. Chem. Phys. (Paris)*, **58**, 54–73.
Haki, G. D. & Rakshit, S. K. (2003). *Bioresour. Technol.* **89**, 17–34.
Harding, M. M. (2004). *Acta Cryst.* D**60**, 849–859.
Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J. & Arnold, F. H. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5610–5615.
Hess, K., Haller, R. & Katz, J. R. (1928). *Die Chemie der Zellulose und ihrer Begleiter: Micellartheorie und Quellung der Zellulose.* Leipzig: Akademische Verlagsgesellschaft.
Hildén, L. & Johansson, G. (2004). *Biotechnol. Lett.* **26**, 1683–1693.
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.
Karlsson, J., Medve, J. & Tjerneld, F. (1999). *Appl. Biochem. Biotechnol.* **82**, 243–258.
Knowles, J. K. C., Lentovaara, P., Murray, M. & Sinnott, M. L. (1988). *J. Chem. Soc. Chem. Commun.*, pp. 1401–1402.
Koivula, A., Kinnari, T., Harjunpää, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T. A. & Teeri, T. T. (1998). *FEBS Lett.* **429**, 341–346.
Koivula, A., Reinikainen, T., Ruohonen, L., Valkeajärvi, A., Claeyssens, M., Teleman, O., Kleywegt, G. J., Szardenings, M., Rouvinen, J., Jones, T. A. & Teeri, T. T. (1996). *Protein Eng.* **9**, 691–699.
Koivula, A. *et al.* (2002). *J. Am. Chem. Soc.* **124**, 10015–10024.
Lillelund, V. H., Jensen, H. H., Liang, X. & Bols, M. (2002). *Chem. Rev.* **102**, 515–553.
Linder, M. & Teeri, T. T. (1997). *J. Biotechnol.* **57**, 15–28.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
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.
Nishiyama, Y. (2009). *J. Wood Sci.* **55**, 241–249.
Olsher, U., Izatt, R. M., Bradshaw, J. S. & Dalley, N. K. (1991). *Chem. Rev.* **91**, 137–164.
Roberts, S. M. & Davies, G. J. (2012). *Methods Enzymol.* **510**, 141–168.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. & Jones, T. A. (1990). *Science*, **249**, 380–386.
Sandgren, M., Ståhlberg, J. & Mitchinson, C. (2005). *Prog. Biophys. Mol. Biol.* **89**, 246–291.
Shoseyov, O., Shani, Z. & Levy, I. (2006). *Microbiol. Mol. Biol. Rev.* **70**, 283–295.
Ståhlberg, J., Johansson, G. & Pettersson, G. (1993). *Biochem. Biophys. Acta*, **1157**, 107–113.
Szijártó, N., Siika-Aho, M., Tenkanen, M., Alapuranen, M., Vehmaanperä, J., Réczey, K. & Viikari, L. (2008). *J. Biotechnol.* **136**, 140–147.
Taylor, E. A., Clinch, K., Kelly, P. M., Li, L., Evans, G. B., Tyler, P. C. & Schramm, V. L. (2007). *J. Am. Chem. Soc.* **129**, 6984–6985.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Väljamäe, P., Sild, V., Nutt, A., Pettersson, G. & Johansson, G. (1999). *Eur. J. Biochem.* **266**, 327–334.
Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. & Pettersson, G. (1986). *FEBS Lett.* **204**, 223–227.
Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schülein, M. & Davies, G. J. (2003). *Structure*, **11**, 855–864.
Varrot, A., Hastrup, S., Schülein, M. & Davies, G. J. (1999). *Biochem. J.* **337**, 297–304.
Varrot, A., Macdonald, J., Stick, R. V., Pell, G., Gilbert, H. J. & Davies, G. J. (2003). *Chem. Commun.*, pp. 946–947.
Varrot, A., Schülein, M. & Davies, G. J. (1999). *Biochemistry*, **38**, 8884–8891.

Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J. & Siika-aho, M. (2007). *Adv. Biochem. Eng. Biotechnol.* **108**, 121–145.
Vocadlo, D. J. & Davies, G. J. (2008). *Curr. Opin. Chem. Biol.* **12**, 539–555.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, Q., Singh, A. & Himmel, M. E. (2009). *Curr. Opin. Biotechnol.* **20**, 364–371.
Zhao, M. & Running, S. W. (2010). *Science*, **329**, 940–943.