The three-dimensional structure of the cellobiohydrolase Cel7A from *Aspergillus fumigatus* at 1.5 Å resolution

Olga V. Moroz,^a Michelle Maranta,^b Tarana Shaghasi,^b Paul V. Harris,^b Keith S. Wilson^{a*} and Gideon J. Davies^{a*}

^aDepartment of Chemistry, University of York, York Structural Biology Laboratory, York YO10 5DD, England, and ^bNovozymes Inc., 1443 Drew Avenue, Davis, CA 95618, USA

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

Received 9 October 2014
Accepted 15 December 2014

PDB references: Cel7A, apo, 4v1z; product complex, 4v20

The enzymatic degradation of plant cell-wall cellulose is central to many industrial processes, including second-generation biofuel production. Key players in this deconstruction are the fungal cellobiohydrolases (CBHs), notably those from family GH7 of the carbohydrate-active enzymes (CAZY) database, which are generally known as CBHI enzymes. Here, three-dimensional structures are reported of the *Aspergillus fumigatus* CBHI Cel7A solved in uncomplexed and disaccharide-bound forms at resolutions of 1.8 and 1.5 Å, respectively. The product complex with a disaccharide in the +1 and +2 subsites adds to the growing three-dimensional insight into this family of industrially relevant biocatalysts.

1. Introduction

The enzymatic deconstruction of plant cell-wall cellulose is central to the carbon cycles and is key to a plethora of industrial processes, notably those that seek to digest cellulose for societal benefit. Such biotechnological procedures include the digestion/modification of cellulose fibres in the pulp and paper industry, in detergents and in the production of second-generation biofuels (reviewed in, for example, Ragauskas *et al.*, 2006; Himmel *et al.*, 2007; Gilbert *et al.*, 2008). Cellulose is a fiendishly difficult substrate for industrial-level breakdown. Its β -1,4-glycosidic bonds are chemically highly resistant to attack (the half-life for uncatalysed cellulose hydrolysis is estimated to be 4.5 million years; Wolfenden & Snider, 2001), whilst its crystalline and highly ordered structure renders it recalcitrant and resistant to enzyme action.

Nature has evolved a complex set of solutions to the problems of cellulose hydrolysis. Fungi and bacteria secrete a battery of different enzymes, often acting in partial synergy, to degrade crystalline cellulose to the disaccharide cellobiose and ultimately the monosaccharide glucose. The 'classical' model for this action involves the synergistic action of endoglucanases, endo-active enzymes that create chain breaks upon which processive (exo-)cellobiohydrolases (processive enzymes perform multiple attacks whilst maintaining the polysaccharide chain) can act (reviewed in Kostylev & Wilson, 2012). It is now known that for many organisms this model needs to be supplanted by one featuring the initial oxygenating attack of the newly discovered and characterized lytic polysaccharide monooxygenases (Phillips *et al.*, 2011; Quinlan *et al.*, 2011; Beeson *et al.*, 2012; Horn *et al.*, 2012; Li *et al.*, 2012; Hemsworth *et al.*, 2013; Wu *et al.*, 2013; Forsberg *et al.*, 2014; Vu *et al.*, 2014; Kjaergaard *et al.*, 2015), but the concept of chain breakage for processive degradation by cellobiohydrolases remains.

Industrial enzyme suppliers use a limited diversity of organisms for the production of the cellulases, hemicellulases and other enzymes required for the efficient degradation of lignocellulosic plant residues. For many years the primary production organisms have been mutant strains of the mesophilic fungus *Hypocrea jecorina* (previously known as *Trichoderma reesei*), largely by virtue of their ability to secrete proteins at levels of 100 g l⁻¹ or more (discussed, for example, in Cherry & Fidantsef, 2003), their relatively potent cellulolytic system and their historical usage in various industries such as food and feed processing, laundry detergents and textile processing (Kitpreetchavanich *et al.*, 1986). However, the natively produced cellulases and

hemicellulases of *H. jecorina* are not necessarily optimal with respect to specific activity or thermostability, and considerable effort has been expended in attempts to find or engineer superior enzymes (Kellermann *et al.*, 2014; Chandel *et al.*, 2012). Thermostability and thermoactivity in particular have been targeted (Blumer-Schuetz *et al.*, 2014) since there are potential benefits of performing lignocellulose hydrolysis at elevated temperature, such as reduced substrate viscosity, reduced cooling requirements following hydrothermal pretreatment, a reduced potential for bacterial contamination during saccharification and of course increased specific activity of the enzymes if they remain stable at elevated temperature. *Aspergillus fumigatus* is a ubiquitous thermotolerant saprophytic fungus that is known to produce a very active cellulolytic enzyme complex when grown on lignocellulose (Sharma *et al.*, 2011; Soni & Kango, 2013). In addition, several glycoside hydrolases secreted by *A. fumigatus* are quite heat-stable, for example β -glucosidase (Kim *et al.*, 2007), endoglucanase (Liu *et al.*, 2011), β -xylosidase (Kitpreechavanich *et al.*, 1986) and endo- β -1,4-xylanase (Souza *et al.*, 2012).

Cellobiohydrolase I (CBHI) is arguably the most important cellulase from these fungal species. Although the exact 'exo' (chain-end) action of cellobiohydrolases remains controversial (primarily because loop opening and closing can afford some degree of endo activity; discussed, for example, in Varrot *et al.*, 2003; Davies *et al.*, 2000; Varrot, Schülein *et al.*, 1999; Varrot, Hastrup *et al.*, 1999; Armand *et al.*, 1997), they are widely agreed to be processive enzymes. Using the CAZY sequence-based classification, key fungal cellobiohydrolases are found in CAZY glycoside hydrolase families GH6 and GH7 (Henrissat & Davies, 1997; Cantarel *et al.*, 2009), with CBHII enzymes (family GH6) liberating cellobiose at the nonreducing end (EC 3.2.1.91) and CBHI enzymes (family GH7) releasing cellobiose at the reducing end (Chanzy & Henrissat, 1985; Fig. 1). The first three-dimensional structure of a CBHI enzyme was solved by the Jones group in 1994 (Divne *et al.*, 1994) and this *T. reesei* (*H. jecorina*)

Cel7A (CBHI) enzyme has remained one of the best characterized structurally (discussed further below).

Cellobiohydrolases, by virtue of their ability to degrade cellulose substantially, are key industrial enzymes of the 21st century. There is considerable drive to uncover new enzymes, determine their three-dimensional structures and assess them for cellulose deconstruction. In this light, here we report the three-dimensional structure of a CAZY family GH7 fungal cellobiohydrolase, AfCel7A, from *A. fumigatus* in its uncomplexed (hereafter referred to as apo) form and in complex with a cellobiose product mimic at resolutions of 1.8 and 1.5 Å, respectively. This enzyme is of particular industrial importance as inferred from the more than 50 patents and patent applications that reference it.

2. Materials and methods

2.1. Gene expression and protein purification

The catalytic core domain, residues 1–440, of AfCel7A was provided as a kind gift from Novozymes A/S. Briefly, the gene encoding the Cel7A catalytic core was amplified from plasmid pEJG93 (Vlasenko *et al.*, 2012) using the following forward and reverse primers, with bases in *italics* having identity to the expression vector used for subsequent subcloning and a stop codon added (**bold**): forward primer, *ACTGGATTACCATGCTGGCCTCCACCTTCTCCTACC*; reverse primer, *TCACCTCTAGTTAATTAACTAACCGCTGTTGAAGGTCGAGCCG*. The amplification reaction was composed of 283 ng DNA, 200 nM dNTPs, 1 pmol of each primer, 1× 'Expand 2' buffer and 1 µl 'Expand 1' DNA polymerase (the compositions of these buffers are as provided by the manufacturer; Roche Diagnostics Corporation, Indianapolis, Indiana, USA) in a total volume of 50 µl. The amplification conditions were one cycle at 95°C for 60 s, 30 cycles of 95°C for 30 s, 60°C for 30 s and

Figure 1

Action of cellobiohydrolases. (a) The processive action of a cellobiohydrolase I in liberating cellobiose at the reducing end of the polysaccharide chain. (b) The double-displacement reaction mechanism in which a covalent glycosyl-enzyme intermediate is formed and subsequently hydrolysed.

Table 1

X-ray data and structure-refinement statistics.

Values in parentheses are for the highest resolution shell.

	Uncomplexed	'Cellobiose' complex
Data collection		
Beamline	Diamond I03	Diamond I02
Wavelength (Å)	0.98	0.98
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters		
a, b, c (Å)	79.2, 131.4, 45.9	79.6, 130.8, 46.1
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	65.69–1.78 (1.88–1.78)	43.49–1.50 (1.54–1.50)
$R_{\text{merge}}^{\dagger}$ (%)	9.9 (67.6)	5.7 (62.8)
$\langle I/\sigma(I) \rangle$	10.7 (2.9)	19.8 (2.9)
Completeness (%)	99.3 (98.2)	99.9 (99.7)
Multiplicity	5.9 (6.1)	7.1 (7.1)
$CC_{1/2}^{\ddagger}$	0.996 (0.727)	1.000 (0.798)
Model refinement		
No. of reflections used	44036	73170
$R_{\text{cryst}}/R_{\text{free}}$ (%)	15.4/19.4	13.9/16.3
No. of protein atoms	3425	3644
No. of ligands	2	33
No. of waters	472	740
Mean B factor, protein (Å ²)	20	21
ESU §	0.10	0.06
R.m.s.d., bond lengths (Å)	0.018	0.015
R.m.s.d., bond angles (°)	1.55	1.54
Ramachandran plot ¶ (%)		
Favoured	96.8	96.9
Allowed	2.87	2.82
Outliers	0.26	0.31
PDB code	4v1z	4v20

† R_{merge} is defined as $100 \times \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl . ‡ $CC_{1/2}$ values for I_{mean} are calculated by splitting the data randomly in half. § Estimated standard uncertainty, based upon R_{free} , calculated using *REFMAC*. ¶ Calculated using the validation options in *Coot*.

72°C for 60 s, followed by one cycle at 72°C for 7 min. The reaction was run on a 1% agarose gel and the predominant band was excised and purified using the QIAquick Gel Extraction Kit (catalogue No. 28704, Qiagen, Valencia, California, USA) as described by the manufacturer. The product was subcloned into expression vector pAllLo2 (Brown *et al.*, 2005) using the InFusion Advantage system as described by the manufacturer. The vector was previously cut with NcoI, PacI and SpeI. The expression plasmid was purified from *Escherichia coli* transformants carrying the correct construct as verified by DNA sequencing. This DNA was used to transform *A. oryzae* strain JaL250 as described previously (Wu *et al.*, 2009). Six transformants were picked and grown in 1 ml MDU2BP medium in stationary 24-well plates at 34°C for 4 d. All six transformants produced protein bands of approximately the correct size (~50 kDa) as judged by SDS–PAGE. A spore stock from the best producer was inoculated into two 2.8 l flasks containing 500 ml each of MDU2BP medium and incubated for at 34°C and 250 rev min^{−1} for 4 d. MDU2BP medium was composed of 45 g maltose, 1 g MgSO₄·7H₂O, 1 g NaCl, 2 g K₂HPO₄, 12 g KH₂PO₄, 2 g urea, 500 µl AMG trace-metals solution and deionized water to 1 l; the pH was adjusted to 5.0 and the solution was then filter sterilized with a 0.22 µm filtering unit. AMG trace-metals solution was composed of 14.3 g ZnSO₄·7H₂O, 2.5 g CuSO₄·5H₂O, 0.5 g NiCl₂·6H₂O, 13.8 g FeSO₄·H₂O, 8.5 g MnSO₄·7H₂O, 3 g citric acid and deionized water to 1 l.

The fermentation broth was concentrated and desalted by tangential flow ultrafiltration against 20 mM Tris–HCl pH 8.0 (buffer A). This was applied onto a 20 ml Mono Q column equilibrated with buffer A and unbound proteins were eluted with ten column volumes of buffer A at a flow rate of 4 ml min^{−1}. Bound proteins were eluted with a 20 column-volume linear gradient from 0 to 600 mM NaCl in buffer A. AfCel7A eluted at a conductivity of 7–14 mS cm^{−1}. Frac-

tions eluting at 11–14 mS cm^{−1} were pooled and 1 µl Endo H (New England Biolabs) was added per 10 mg of protein as determined by the BCA assay (Thermo Scientific) using BSA as standard. This was incubated for 24 h at 37°C. Deglycosylated AfCel7A was applied at a flow rate of 1 ml min^{−1} to a 5 ml Mono Q column and eluted as described above. Fractions showing a band of approximately 50 kDa by SDS–PAGE were pooled (12 ml total) and applied at a flow rate of 1 ml min^{−1} to a 75 ml Superdex 75 column equilibrated with buffer A plus 150 mM NaCl. Fractions showing a band of approximately 50 kDa were pooled (10 ml total at 2.75 mg ml^{−1} protein).

2.2. Thermal stability of *Aspergillus fumigatus* Cel7A

The comparison of *A. fumigatus* Cel7A and *H. jecorina* Cel7A (with their linkers and family 1 carbohydrate-binding 'CBM1' domains) was performed as previously described (examples 34 and 49 in McBrayer *et al.*, 2013). Briefly, the enzyme mixture consisted of 35% (w/w) Cel7A, 25% Cel6A CBHII from *Myceliophthora thermophila*, 15% Cel5A endoglucanase from *Thermoascus aurantiacus*, 15% GH61A from *T. aurantiacus*, 5% GH3 β-D-glucosidase from *A. fumigatus* and 5% GH10 xylanase 3 from *A. fumigatus* (UniProt Q4WLG5). Total protein loading was 3 mg per gram of cellulose and the substrate was acid-pretreated corn stover from the US Department of Energy National Renewable Energy Laboratory. Incubations were for 72 h in 50 mM sodium acetate pH 5.0 with 1 mM manganese sulfate. Reactions were carried out in duplicate and the conversion of cellulose to glucose was quantified by HPLC as described previously (example 34 in McBrayer *et al.*, 2013).

2.3. Crystallization and data collection

Initial crystallization screening for the apoenzyme was carried out using sitting-drop vapour diffusion with drops set up using a Mosquito Crystal liquid-handling robot (TTP Labtech, UK) with 150 nl protein solution plus 150 nl reservoir solution in 96-well format plates (MRC 2-well crystallization microplates, Swissci, Switzerland) equilibrated against 54 µl reservoir solution. Experiments were carried out at room temperature with a number of commercial screens. Initial crystals were obtained in PACT screen conditions containing PEG in the presence of zinc, with the best crystals being obtained in B12 [20% (w/v) PEG 6K, 10 mM ZnCl₂, MES pH 6.0]. These were manually optimized in a 24-well Linbro tray (hanging drops) with a grid of PEG concentrations starting from 10%.

2',4'-Dinitrophenyl 2-deoxy-2-fluoro-β-D-cellobioside was used for complex formation in order to try to trap a covalent glycosyl-enzyme intermediate (although a noncovalent product complex was subsequently observed; see below). Crystals of the complex were obtained in a manual screen with conditions successful for the apoprotein using micro-seeding. The initial seeding stock was prepared by crushing the apo crystals, adding 100 µl of the mother liquor and vortexing the mixture for 1 min using a seed bead (Hampton Research). The complex was formed by adding a nonspecified amount of ligand to the apoprotein using the following procedure: a needle was first dipped into an Eppendorf tube containing ligand powder and then into a small aliquot (5–15 µl) of mother liquor. The solution was briefly mixed and added to an equal volume of protein solution; the resulting mixture was distributed on cover slips as 2 µl crystallization drops to which 0.5 µl of 1:10 diluted seeding stock was added. These were used in a hanging-drop setup in a 24-well Linbro plate.

2.4. Structure solution and refinement

All computation was carried out using programs from the *CCP4* suite (Winn *et al.*, 2011) unless otherwise stated. For the apoenzyme structure, data were collected on beamline I03 at Diamond Light Source to 1.78 Å resolution. The space group was $P2_12_12$, with unit-cell parameters $a = 79.2$, $b = 131.4$, $c = 45.9$ Å. The images were integrated with *MOSFLM* (Battye *et al.*, 2011) and then scaled with *SCALA*. The structure was solved using *MOLREP* (Vagin & Teplyakov, 2010) with Cel7A from *Talaromyces emersonii* (PDB entry 1q9h; Grassick *et al.*, 2004) as a search model. For the ligand complex, data were collected on beamline I02 at Diamond Light Source to 1.5 Å resolution and were processed using *xia2* (Winter, 2010). The space group was $P2_12_12$, with unit-cell parameters $a = 79.5$, $b = 130.8$, $c = 46.1$ Å. The structure was solved by *MOLREP* using apo AfCel7A as the search model.

Structures were refined using *REFMAC* (Murshudov *et al.*, 2011) iterated with manual model building/correction in *Coot* (Emsley *et al.*, 2010). Refinement dictionaries were created using *JLigand* (Lebedev *et al.*, 2012). Processing and refinement statistics for both structures are given in Table 1. Structure figures were prepared using *CCP4mg* (McNicholas *et al.*, 2011).

3. Results and discussion

3.1. Activity on pretreated corn stover

In order to assess the high-temperature performance of *A. fumigatus* Cel7A in comparison to *H. jecorina* Cel7A, we employed an enzyme mixture selected to perform well over a period of 3 d at 50–65°C using an industrially relevant lignocellulosic biomass substrate: acid-pretreated corn stover. The enzymes chosen were all from thermophilic or thermotolerant fungi and included cellobiohydrolase II, endoglucanase, β -D-glucosidase, lytic polysaccharide mono-

Figure 2
Conversion of the cellulose in pretreated corn stover to glucose by AfCel7A and *H. jecorina* Cel7A. A mixture of partially purified enzymes was used for hydrolysis at the indicated temperature for 72 h as described in §2. The enzyme mixture either lacked Cel7A CBHI (diamonds) or contained CBHI from *H. jecorina* (triangles) or *A. fumigatus* (circles) at 35% (w/w) of the total protein.

oxygenase and xylanase. As shown in Fig. 2, the enzyme mixture without CBHI produced a relatively low level of cellulose conversion (~36%) with a fairly flat temperature profile between 50 and 65°C. Addition of CBHI produces a dramatic increase in conversion that is strongly temperature-dependent, reflecting the thermoactivity and thermostability of the respective CBHI enzymes. At 50°C, the two CBHI enzymes perform similarly; however, at higher temperatures the superiority of the *A. fumigatus* CBHI is evident and allows the use of the full cellulase mixture at temperatures in the critical 55–60°C range where bacterial contamination becomes much less of an issue.

3.2. Three-dimensional structure of AfCel7A

The structure of AfCel7A was solved by molecular replacement using the known three-dimensional structure of *T. emersonii* Cel7A as a search model. The apoenzyme structure and that of a complex with a cellobiosyl moiety in the +1 and +2 subsites (subsite nomenclature is given in Davies *et al.*, 1997) were refined at 1.78 and 1.5 Å resolution, respectively (Table 1). The three-dimensional structure starts, as is often the case for secreted fungal enzymes, with a post-translationally produced pyroglutamate entity. Thereafter, the AfCel7A structure can be traced up to residue 440 with short breaks at residues 100–102 and 196–200. A single *N*-glycosylation site, modelled with a single GlcNAc (after deglycosylation with EndoH), is found at Asn270. There are two *cis*-peptides: Tyr388–Pro389 and Ser439–Gly440. Two metal ions modelled as Zn^{2+} are present and are presumed to have come from the crystallization conditions, with one involved in crystal contacts bridging Asp46 from one molecule in the crystal and His418 in an adjacent molecule. The metal ions were described as zinc based on the fact that only crystallization conditions where zinc was present produced any hits (the best hit was PACT B12 with 10 mM zinc chloride) and further optimizations only worked when zinc was added. The ligands (Asp, His and waters) and bond distances are also in the range expected for zinc: close to 2.0 Å, with one longer distance of 2.4 Å (Harding, 2006).

The three-dimensional structure of AfCel7A forms the classical β -sandwich structure first reported for this enzyme class by Divne and coworkers for the structure of *T. reesei* Cel7A (CBHI), the study of which (Divne *et al.*, 1994, 1998; Ståhlberg *et al.*, 1996) has recently been augmented by a series of ligand complexes in tandem with computational approaches (Knott *et al.*, 2014). AfGH7A similarly displays two large antiparallel β -sheets that stack to form a sandwich, the concave face of which, enclosed by extended loop regions, forms the substrate-binding tunnel (Fig. 3a). At the time of writing, there are 11 known structures in family GH7 (http://www.cazy.org/GH7_structure.html). As expected, AfCel7A shows the greatest structural similarity (using the *PDBFold* server; Krissinel & Henrick, 2004; Table 2) to other fungal cellobiohydrolases (from both ascomycetes and basidiomycetes) in family GH7 (Table 2). The closest classical endocellulase to AfCel7A is *Fusarium oxysporum* Cel7A (EG1; described, for example, in Sulzenbacher *et al.*, 1997; PDB entry 2ovw; *Q* score 0.62; 363 C α atoms overlapping with an r.m.s.d. of 1.38 Å).

3.3. Ligand binding to AfCel7A

The catalytic centre of Cel7 enzymes comprises two key residues, an acid/base and a nucleophile (Fig. 1b), to allow a mechanism with net retention of anomeric configuration *via* the formation and subsequent breakdown of a covalent glycosyl-enzyme intermediate. A crystal of AfCel7A was obtained in the presence of 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-cellobioside with the aim of trapping the covalent glycosyl-enzyme inhibitor in the –1 and –2 subsites (as

described, for example, in Sulzenbacher *et al.*, 1999). In fact, what was observed was a disaccharide moiety lying in the +1 and +2 subsites, suggesting that the DNP species has both reacted to form an intermediate and subsequently been hydrolysed and has migrated to the +1 and +2 subsites known to be most favoured for binding the disaccharide moiety. Electron density cannot distinguish between cellobiose and 2-fluorocellobiose, but as only the latter was present the ligand has been modelled as 2-fluorocellobiose in refinement. The electron density is clear and unambiguous (Fig. 3*b*), allowing a description of all of the interactions with the enzyme in these subsites (Fig. 3*c*).

The overall structure of the ligand-bound complex is essentially the same as for the apoenzyme (including the Zn^{2+} sites). The Q -score is 0.995 for 440 C^α atoms overlapping with an r.m.s.d. of 0.2 Å. One difference from the apo structure is that the main chain in the region Thr246-Tyr247-Ser248 has an alternative conformation, partially moving away upon ligand binding. This region is a part of the 'exo-

Table 2

Closest structural matches to AfCel7A calculated using *PDBFold* (<http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver>).

The top five matches based upon *PDBFold* Q -score (excluding duplicates) are shown.

Enzyme	PDB code	Q score	Residues aligning	Sequence identity (%)	C^α r.m.s.d. (Å)
Cel7A, <i>Talaromyces emersonii</i>	3pl3	0.96	436/436	75	0.55
Cel7A, <i>Heterobasidion annosum</i>	2xsp	0.91	431/440	66	0.72
Cel7A, <i>Humicola grisea</i>	4csi	0.90	429/437	64	0.74
Cel7D, <i>Phanerochaete chrysosporium</i>	1z3t	0.89	422/431	65	0.70
Cel7A, <i>Trichoderma reesei</i>	1dy4	0.88	423/434	66	0.76

loop' B3 (residues 244–252 in AfCel7A) and has been suggested to be involved in moving the substrate along the binding tunnel (von Ossowski *et al.*, 2003). In a recent structure of the *H. grisea* enzyme the corresponding loop adapted a conformation different from the

Figure 3

The three-dimensional structure of AfCel7A. (a) The three-dimensional structure of AfCel7A coloured according to secondary structure, with the cellobiose ligand (in subsites +1 and +2) shown as coloured shaded spheres and the two Zn^{2+} ions from the crystallization conditions as grey spheres. The single GlcNAc *N*-glycosylation site is indicated. (b) Observed electron density for the 2-fluorocellobiosyl moiety in the +1 and +2 subsites; the map is shown as an unbiased maximum-likelihood weighted $F_o - F_c$ synthesis at 3σ with phases calculated prior to the incorporation of any ligand atoms in refinement. (c) Interactions of the 2-fluorocellobiosyl moiety in the +1 and +2 subsites. All panels are shown in divergent ('wall-eyed') stereo and were drawn using *CCP4mg* (McNicholas *et al.*, 2011).

Figure 4

Superposition of the exo-loop regions from apo AfCel7A (purple), its ligand complex (green), *H. grisea* Cel7A (PDB entry 4csi; blue; Momeni *et al.*, 2014) and *H. jecorina* Cel7A (PDB entry 4c4c; ice blue; Knott *et al.*, 2014). The alternate conformations of the region around Tyr247 further confirm the flexibility of the exo-loop.

other known Cel7A structures, and the authors suggest that this is an indication of flexibility of loop B3 (Momeni *et al.*, 2014; Fig. 4). Our results are consistent with this observation. The cellobiosyl moiety lies in the +1 and +2 subsites, with the +1 sugar unit making a stacking interaction with Trp383 and with the remainder of the interactions primarily being hydrogen-bonding in character (Fig. 3c), all of which are essentially invariant in family GH7 cellobiohydrolases. Interestingly, the O4 atom of the −1 sugar does not make a direct interaction with the acid/base Glu212; indeed, it is some 7 Å distant from this residue, suggesting that the complex may represent not an initial formed product complex in the +1/+2 subsites but rather one in the process of departing from the active centre. The equivalent complex for other cellobiohydrolases (for example, PDB entry 4hap; not shown; Kern *et al.*, 2013) show the +1 sugar approximately 1.5 Å further back in the active tunnel, closer to the acid/base.

4. Summary

Fungal GH7 cellobiohydrolases are one of the major enzyme classes in the deconstruction of lignocellulosic biomass. The key role of fungal CBHI in cellulose degradation, through its processive action, has long been established. The structure of AfCel7A, an enzyme that still maintains high activity after incubation at 65°C, adds to the growing insight into this phenomenon and will inspire and direct protein-engineering strategies for improved, bespoke, cellulases for biomass conversion.

The staff of the Diamond Light Source are thanked for the provision of data-collection facilities.

References

- Armand, S., Drouillard, S., Schülein, M., Henrissat, B. & Driguez, H. (1997). *J. Biol. Chem.* **272**, 2709–2713.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst. D* **67**, 271–281.
- Beeson, W. T., Phillips, C. M., Cate, J. H. & Marletta, M. A. (2012). *J. Am. Chem. Soc.* **134**, 890–892.
- Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., Conway, J. M., Adams, M. W. W. & Kelly, R. M. (2014). *FEMS Microbiol. Rev.* **38**, 393–448.
- Brown, K., Harris, P., Zaretsky, E., Re, E., Vlasenko, E., McFarland, K. & Lopez, D. L. A. (2005). Patent WO 2005074647 A3.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
- Chandel, A. K., Chandrasekhar, G., Silva, M. B. & Silvério da Silva, S. (2012). *Crit. Rev. Biotechnol.* **32**, 187–202.
- Chanzy, H. & Henrissat, B. (1985). *FEBS Lett.* **184**, 285–288.
- Cherry, J. R. & Fidantsef, A. L. (2003). *Curr. Opin. Biotechnol.* **14**, 438–443.
- Davies, G. J., Brzozowski, A. M., Dauter, M., Varrot, A. & Schülein, M. (2000). *Biochem. J.* **348**, 201–207.
- Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochem. J.* **321**, 557–559.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T. & Jones, T. A. (1994). *Science*, **265**, 524–528.
- Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. (1998). *J. Mol. Biol.* **275**, 309–325.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Forsberg, Z., Mackenzie, A. K., Sorlie, M., Rohr, A. K., Helland, R., Arvai, A. S., Vaaje-Kolstad, G. & Eijsink, V. G. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 8446–8451.
- Gilbert, H. J., Ståhlbrand, H. & Brumer, H. (2008). *Curr. Opin. Plant Biol.* **11**, 338–348.
- Grassick, A., Murray, P. G., Thompson, R., Collins, C. M., Byrnes, L., Birrane, G., Higgins, T. M. & Tuohy, M. G. (2004). *Eur. J. Biochem.* **271**, 4495–4506.
- Harding, M. M. (2006). *Acta Cryst. D* **62**, 678–682.
- Hemsworth, G. R., Davies, G. J. & Walton, P. H. (2013). *Curr. Opin. Struct. Biol.* **23**, 660–668.
- Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
- 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.
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. (2012). *Biotechnol. Biofuels*, **5**, 45.
- Kellermann, M., Huang, R. A., Forbes, A. B. & Rehbein, S. (2014). *Res. Vet. Sci.* **97**, 199–203.
- Kern, M., McGeehan, J. E., Streeter, S. D., Martin, R. N., Besser, K., Elias, L., Eborall, W., Malyon, G. P., Payne, C. M., Himmel, M. E., Schnorr, K., Beckham, G. T., Cragg, S. M., Bruce, N. C. & McQueen-Mason, S. J. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 10189–10194.
- Kim, K.-H., Brown, K. M., Harris, P. V., Langston, J. A. & Cherry, J. R. (2007). *J. Proteome Res.* **6**, 4749–4757.
- Kitpreechanavich, V., Hayashi, M. & Nagai, S. (1986). *Agric. Biol. Chem.* **50**, 1703–1711.
- Kjaergaard, C. H., Qayyum, M. F., Wong, S. D., Xu, F., Hemsworth, G. R., Davies, G. J., Walton, P. H., Johansen, K. S., Hedman, B., Hodgson, K. O. & Solomon, E. I. (2015). *Proc. Natl Acad. Sci. USA*, **111**, 8797–8802.
- Knott, B. C., Momeni, M. H., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J. & Beckham, G. T. (2014). *J. Am. Chem. Soc.* **136**, 321–329.
- Kostylev, M. & Wilson, D. (2012). *Biofuels*, **3**, 61–70.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst. D* **60**, 2256–2268.
- Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. & Murshudov, G. N. (2012). *Acta Cryst. D* **68**, 431–440.
- Li, X., Beeson, W. T., Phillips, C. M., Marletta, M. A. & Cate, J. H. (2012). *Structure*, **20**, 1051–1061.

- Liu, D., Zhang, R., Yang, X., Xu, Y., Tang, Z., Tian, W. & Shen, Q. (2011). *Protein Expr. Purif.* **79**, 176–186.
- McBrayer, B., Shaghasi, T. & Vlasenko, E. (2013). Patent US 8580536 B2.
- McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* **D67**, 386–394.
- Momeni, M. H., Goedegebuur, F., Hansson, H., Karkehabadi, S., Askarieh, G., Mitchinson, C., Larenas, E. A., Ståhlberg, J. & Sandgren, M. (2014). *Acta Cryst.* **D70**, 2356–2366.
- 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.* **D67**, 355–367.
- Ossowski, I. von, Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y., Driguez, H., Claeysens, M., Sinnott, M. L. & Teeri, T. T. (2003). *J. Mol. Biol.* **333**, 817–829.
- Phillips, C. M., Beeson, W. T., Cate, J. H. & Marletta, M. A. (2011). *ACS Chem. Biol.* **6**, 1399–1406.
- Quinlan, R. J. *et al.* (2011). *Proc. Natl Acad. Sci. USA*, **108**, 15079–15084.
- Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R. & Tschaplinski, T. (2006). *Science*, **311**, 484–489.
- Sharma, M., Soni, R., Nazir, A., Oberoi, H. S. & Chadha, B. S. (2011). *Appl. Biochem. Biotechnol.* **163**, 577–591.
- Soni, H. & Kango, N. (2013). *Recent Pat. Biotechnol.* **7**, 207–218.
- Souza, D. T., Bispo, A. S., Bon, E. P., Coelho, R. R. & Nascimento, R. P. (2012). *Appl. Biochem. Biotechnol.* **166**, 1575–1585.
- Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeysens, M., Teeri, T. T. & Jones, T. A. (1996). *J. Mol. Biol.* **264**, 337–349.
- Sulzenbacher, G., Mackenzie, L. F., Wilson, K. S., Withers, S. G., Dupont, C. & Davies, G. J. (1999). *Biochemistry*, **38**, 4826–4833.
- Sulzenbacher, G., Schüle, M. & Davies, G. J. (1997). *Biochemistry*, **36**, 5902–5911.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Varrot, A., Frandsen, T. P., von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schüle, M. & Davies, G. J. (2003). *Structure*, **11**, 855–864.
- Varrot, A., Hastrup, S., Schüle, M. & Davies, G. J. (1999). *Biochem. J.* **337**, 297–304.
- Varrot, A., Schüle, M. & Davies, G. J. (1999). *Biochemistry*, **38**, 8884–8891.
- Vlasenko, E., McBrayer, B., Stringer, M. & Schnorr, K. (2012). Patent US 20120272410 A1.
- Vu, V. V., Beeson, W. T., Phillips, C. M., Cate, J. H. & Marletta, M. A. (2014). *J. Am. Chem. Soc.* **136**, 562–565.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Wolfenden, R. & Snider, M. J. (2001). *Acc. Chem. Res.* **34**, 938–945.
- Wu, M., Beckham, G. T., Larsson, A. M., Ishida, T., Kim, S., Payne, C. M., Himmel, M. E., Crowley, M. F., Horn, S. J., Westereng, B., Igarashi, K., Samejima, M., Stahlberg, J., Eijssink, V. G. & Sandgren, M. (2013). *J. Biol. Chem.* **288**, 12828–12839.
- Wu, G., Teter, S. A., Harris, P. & Lai, W. (2009). Patent WO 2009103237 A1.