Crystal structures of *Melanocarpus albomyces* cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding

TARJA PARKKINEN,¹ ANU KOIVULA,² JARI VEHEMAANPERÄ,³ AND JUHA ROUVINEN¹

¹Department of Chemistry, University of Joensuu, 80101 Joensuu, Finland

²VTT Technical Research Centre of Finland, 02044 VTT, Finland

³ROAL Oy, 05201 Rajamäki, Finland

(RECEIVED February 12, 2008; FINAL REVISION May 14, 2008; ACCEPTED May 14, 2008)

Abstract

Cellobiohydrolase from *Melanocarpus albomyces* (Cel7B) is a thermostable, single-module, cellulose-degrading enzyme. It has relatively low catalytic activity under normal temperatures, which allows structural studies of the binding of unmodified substrates to the native enzyme. In this study, we have determined the crystal structure of native *Ma* Cel7B free and in complex with three different cello-oligomers: cellobiose (Glc₂), cellotriose (Glc₃), and cellotetraose (Glc₄), at high resolution (1.6–2.1 Å). In each case, four molecules were found in the asymmetric unit, which provided 12 different complex structures. The overall fold of the enzyme is characteristic of a glycoside hydrolase family 7 cellobiohydrolase, where the loops extending from the core β-sandwich structure form a long tunnel composed of multiple subsites for the binding of the glycosyl units of a cellulose chain. The catalytic residues at the reducing end of the tunnel are conserved, and the mechanism is expected to be retaining similarly to the other family 7 members. The oligosaccharides in different complex structures occupied different subsite sets, which partly overlapped and ranged from –5 to +2. In four cellotriose and one cellotetraose complex structures, the cello-oligosaccharide also spanned over the cleavage site (–1/+1). There were surprisingly large variations in the amino acid side chain conformations and in the positions of glycosyl units in the different cello-oligomer complexes, particularly at subsites near the catalytic site. However, in each complex structure, all glycosyl residues were in the chair (⁴C₁) conformation. Implications in relation to the complex structures with respect to the reaction mechanism are discussed.

Keywords: cellulase; cellobiohydrolase; substrate complex; crystal structure; reaction mechanism; *Melanocarpus albomyces*; thermophilic

Reprint requests to: Juha Rouvinen, Department of Chemistry, University of Joensuu, PO Box 111, 80101 Joensuu, Finland; e-mail: juha.rouvinen@joensuu.fi; fax: 358-13-2513390.

Abbreviations: CBH, cellobiohydrolase; CBM, carbohydrate-binding module; GH, glycoside hydrolase; *Tr* Cel7A, *Trichoderma reesei* (*Hypocrea jecorina*) cellobiohydrolase Cel7A; *Ma* Cel7B, *Melanocarpus albomyces* cellobiohydrolase Cel7B; *Pc* Cel7D, *Phanerochaete chrysosporium* Cel7D; *Te* Cel7A, *T. emersonii* Cel7A; Glc₂, cellobiose; Glc₃, cellotriose; Glc₄, cellotetraose.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.034488.108>.

Cellulose, the main component of plant cell walls, is the most abundant natural polymer on Earth. It is composed of glucosyl units linked by β-1,4-glycosidic bonds that have alternating orientations, and the repeating structural unit is cellobiose, a disaccharide. The linear cellulose chains form a crystalline polymer through hydrogen bonding and van der Waals forces. In nature, cellulose can exist in ordered, crystalline, and less-ordered, amorphous forms. The complete hydrolysis of cellulose

requires the concerted action of many enzymes, produced either as free cellulase systems or cellulosomes by microorganisms. Cellulases are enzymes that work on a solid-liquid interface and are relevant to many industrial applications, ranging from textile processing to biomass hydrolysis for bioethanol production (for review, see Bhat and Bhat 1997). Cellulases can be broadly divided into two types of enzymes, endoglucanases (EG; EC 3.2.1.4) and cellobiohydrolases (CBH; EC 3.2.1.91), which show synergism in their action on crystalline cellulose (Barr et al. 1996; Bayer et al. 1998; Boisset et al. 2000; Lynd et al. 2002). EGs hydrolyze internal bonds at random positions on more amorphous regions and produce new free chain ends for CBHs, which cleave off (mainly) cellobiose units in a processive manner. CBHs are important enzymes for crystalline cellulose degradation and are predominantly produced by various fungal species. Many fungal cellulases have a modular structure, where a small carbohydrate-binding module (CBM) is connected to the catalytic module with a presumably flexible linker peptide. The binding of CBM to the surface of a crystalline substrate improves the activity on insoluble substrates but does not affect the activity on small soluble substrates (Bayer et al. 1998). Cellulases and other glycoside hydrolases (GH) can be classified into different families, based on sequence homology and the structure of their catalytic modules (CAZy database at <http://www.cazy.org/>) (Davies and Henrissat 1995; Henrissat and Davies 1997). Cellulases are currently found in over 10 different GH families. Cellulases and other GHs utilize general acid hydrolysis mechanisms in order to cleave the O-glycosidic bond, leading either to retention or inversion of the configuration of the anomeric (C1) center. The enzymes within the same GH family also have the similar (retaining or inverting) catalysis mechanism (Davies and Henrissat 1995; Yip and Withers 2004).

The GH-7 family consists of endoglucanases and cellobiohydrolases found exclusively in eukaryotic organisms. GH-7 cellulases hydrolyze glycosidic bonds using a double-displacement mechanism (S_N2), which leads to a net retention of the configuration at the anomeric carbon of the substrate. CBHs belonging to family 7 have been shown to hydrolyze the cellulose from the reducing end of a chain (Divne et al. 1998; Boisset et al. 2000). The three-dimensional structures of GH-7 CBH catalytic modules from mesophilic fungi *Trichoderma reesei* (also known as *Hypocrea jecorina*) (Divne et al. 1994) and *Phanerochaete chrysosporium* (Muñoz et al. 2001), and from a thermophilic fungus *Talaromyces emersonii* (Grassick et al. 2004) have been solved. The overall fold is, in each case, a β -sandwich, where loops extending from the β -sandwich form an enclosed cellulose-binding tunnel. The complex structures of the catalytically inactivated *T. reesei* Cel7A with different cello-oligomers have pro-

vided detailed information about the cellulose binding (Ståhlberg et al. 1996; Divne et al. 1998). The active site tunnel of *T. reesei* Cel7A is 50 Å long and contains 10 subsites (−7 to +3) for the glycosyl units of a cellulose chain. The cellulose chain is likely to enter the active site from one end of the tunnel (from −7 subsite) and thread through the whole tunnel in such a manner that the cellulose chain is twisted almost upside down (Divne et al. 1998). The tunnel-shaped active site of cellobiohydrolases provides numerous protein interactions and is supposed to be responsible for their processive action on crystalline cellulose. Alternate loop conformations have been detected with some GH-6 family cellobiohydrolases (Boisset et al. 2000). These could facilitate cellulose chain sliding in the tunnel and possibly also occasional opening of the loops to allow endo-type cleavages. Three conserved acidic residues, Glu212, Asp214, and Glu217, located between the subsites −1 and +1 at the end of the tunnel of *Tr* Cel7A, have been shown to be directly involved in the cleavage of the cellulose chain. Glu212 and Glu217, which are located at opposite sides of the catalytic site, act as the nucleophile and the acid/base in the retaining mechanism, respectively (Ståhlberg et al. 1996). The role of Asp214 is thought to control the correct position and protonation state of Glu212. The cleavage of the glycosidic bond can only occur when the cellulose chain in the tunnel is positioned so that the glycosidic bond at −1/+1 is pointing close to the acid/base residue Glu217.

A thermophilic ascomycete fungus *Melanocarpus albomyces* (formerly known also as *Myriococcum albomyces* or *Thielavia albomyces*) produces xylanases and cellulases with pronounced thermal stability and activity in the neutral to alkaline pH range (Maheshwari et al. 2000). These *Melanocarpus* glycoside hydrolases have potential in various industrial applications, and three neutral cellulases have also been expressed at high levels in an industrial production host *T. reesei* (Haakana et al. 2004; Miettinen-Oinonen et al. 2004). Two of these cellulases, *Ma* Cel7A and *Ma* Cel45A, are endoglucanases that belong to the GH-7 and GH-45 families, respectively, while the third cellulase, Cel7B, is a GH-7 cellobiohydrolase. Interestingly, none of the three cellulases has a cellulose-binding module (CBM) (Haakana et al. 2004). *Ma* Cel7B is composed of 430 amino acids with a theoretical molecular weight of 47.5 kDa. The main soluble product from microcrystalline cellulose hydrolysis by *Ma* Cel7B is cellobiose (S.P. Voutilainen, H. Boer, T. Puranen, J. Vehmaanperä, and A. Koivula, in prep.). When compared to *T. reesei* Cel7A cellobiohydrolase, *Ma* Cel7B has about a 3°C improved unfolding temperature, while the catalytic rate on soluble substrates (at room temperature) is threefold lower. In addition, *Ma* Cel7B shows high end-product cellobiose inhibition

($K_d \sim 6 \mu\text{M}$), which is about threefold stronger than that for *Tr* Cel7A (S.P. Voutilainen, H. Boer, T. Puranen, J. Vehmaanperä, and A. Koivula, in prep.). Random mutagenesis has recently been used to further improve the thermostability of *Ma* Cel7B (Voutilainen et al. 2007). In addition, we have previously reported the crystallization of the *Ma* Cel7B (Parkkinen et al. 2007).

Here, we report the three-dimensional structure of the native *M. albomyces* cellobiohydrolase Cel7B at 1.6 Å resolution and three other Cel7B structures that are the results of soaks with cellobiose (Glc₂), cellotriose (Glc₃), and cellotetraose (Glc₄). Since four *Ma* Cel7B molecules were in the asymmetric unit, we have 16 different structures of Cel7B, examples of the extensively conformational heterogeneity of the enzyme that offer new insights into the function and mechanism of retaining cellobiohydrolases.

Results

Overall structures

Statistics for the crystallographic data, refinement, and final models are shown in Table 1. The *Ma* Cel7B crystals were pseudo-merohedrally twinned. As a consequence,

the SHELX program (Sheldrick and Schneider 1997) was used for the refinement. This program, and high-resolution diffraction data (1.6–2.1 Å), produced good quality electron density maps that allowed us to unambiguously detect glycosyl units in the active site. Four molecules were present in the asymmetric unit, and each molecule contained a complete mature protein (residues 1–430). The first amino acid at the N terminus is a pyroglutamate residue, also observed in the other enzymes of this family (Divne et al. 1994; Kleywegt et al. 1997; Sulzenbacher et al. 1997; Mackenzie et al. 1998; Muñoz et al. 2001; Grassick et al. 2004). The root-mean-square deviations (RMSD) between the molecules in the asymmetric unit suggest quite large structural differences between the molecules: native, 0.5–0.6 Å, Glc₂, 0.5–0.6 Å, Glc₃, 0.5–0.7 Å, and Glc₄, 0.7–0.9 Å. These RMSD values are based on 430 C^α atoms.

M. albomyces Cel7B has a characteristic overall fold of GH-7 cellobiohydrolases. This structure consists of two antiparallel β-sheets packed face-to-face to form a β-sandwich. Both β-sheets contain six strands, and they are highly curved and form concave and convex surfaces of the sandwich. In Figure 1A the loops that extend from the concave face of the β-sandwich define the shape of the substrate-binding tunnel. The loops are stabilized by nine

Table 1. X-ray data collection and model refinement statistics

	Native	Cellobiose	Cellotriose	Cellotetraose
A. Data collection				
Beamline	X12	X12	X11	X11
Space group	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁
Unit cell dimensions				
<i>a</i> (Å)	50.9	50.9	50.8	51.0
<i>b</i> (Å)	94.5	94.7	94.4	94.8
<i>c</i> (Å)	189.8	190.0	189.9	190.4
β (°)	90.0	90.0	90.2	90.0
Resolution (Å) ^a	25–1.6 (1.7–1.6)	25–1.7 (1.8–1.7)	25–1.8 (1.9–1.8)	25–2.1 (2.2–2.1)
No. of observations	882,338	740,535	617,959	396,409
No. of unique reflections	235,645	197,103	161,992	105,275
Redundancy	3.7	3.8	3.8	3.8
<i>R</i> _{sym} (%) ^a	7.2 (30.2)	7.4 (29.9)	9.9 (35.0)	17.7 (45.3)
<i>I</i> / <i>σI</i> ^a	13.6 (4.8)	14.9 (4.8)	10.3 (4.1)	7.6 (3.1)
Completeness (%) ^a	99.6 (99.4)	99.7 (99.9)	97.7 (97.2)	99.5 (97.9)
B. Refinement				
<i>R</i> -factor (%)	19.9	18.4	18.5	20.9
<i>R</i> _{free} (%)	24.3	23.8	23.7	28.1
Protein atoms	13,332	13,332	13,332	13,332
Ligand atoms		92	204	160
Water molecules	689	1137	863	464
Average B value (Å²)				
Protein	21.6	21.4	26.2	33.3
Ligand		22.4	31.1	31.2
Water	21.6	21.7	23.3	21.5
RMSD from ideal geometry				
Bonds (Å)	0.009	0.007	0.006	0.005
Angles (Å)	0.026	0.026	0.024	0.021

^a Values in parentheses are for highest resolution shell.

Figure 1. (A) The overall structure of *Ma* Cel7B complexed with cellobiose. The cellobiose is bound at glycosyl subsites +1 and +2. (B) The superimposition of Cα backbone of all 16 *Ma* Cel7B structures. Native structures are in black, cellobiose complex structures are in orange, cellotriose complex structures are in red, and cellotetraose complex structures are in purple. (C) The final electron density ($2|F_o| - |F_c|$) maps for Glc₄, after the SHELX refinement (all model atoms were used for phase calculation, the level of contouring 1.0 σ), resolution 2.1 Å.

disulphide bridges, which are located between residues 19–25, 49–69, 59–65, 138–395, 172–210, 176–209, 230–256, 238–243, and 261–329. The substrate-binding tunnel is ~50 Å long and contains nine glycosyl binding sites, –7 to +2, four of them aligned by tryptophan residues (subsites –7, –4, –2, and +1) (see below for a more detailed discussion). The three conserved carboxylic acid residues between subsites –1 and +1, Glu212, Asp214, and Glu217, are in a similar position as in *Tr* Cel7A, and it has been suggested they act as the nucleophile, acid/base, and the assisting amino acid in controlling the correct position and protonation state of the nucleophile.

Ma Cel7B complex structures

Complex structures were obtained by soaking crystals in three different solutions that contain cello-oligomers. The complex structures all have the same space group (*P*2₁)

and similar cell parameters as the native *Ma* Cel7B crystals. Crystal twinning was also present. All structures have four molecules in the asymmetric unit (A, B, C, D). The RMSD between the native and complex structures are Glc₂, 0.4 Å, Glc₃, 0.5 Å, and Glc₄, 0.7 Å. The electron densities of bound ligands are well defined for most of the glycosyl units (except for the Glc₃ structure, molecule C, has poor density for glycosyl units at subsites –5 and –4), and they allow us to determine the placement and orientation of cello-oligosaccharides (Fig. 1C). The bound ligands all have the same directionality, with their reducing end to the right in Figure 1C. This confirms that the hydrolysis proceeds from the reducing end of the cellulose chain in a similar manner as reported for other GH-7 family cellobiohydrolases. In addition, the glycosyl units in all the different complex structures are in phase and correctly oriented at the –1/+1 site for catalysis (i.e., in the so-called productive binding mode). The subsite occupancy for each cello-oligomer in each of the four molecules in the complex structures is shown in Table 2. No major differences could be detected in the protein main chain conformations between the various models. Smaller conformational changes are seen in two loops (97–102 and 381–390) located at both ends of the active site tunnel upon cello-oligosaccharide binding. Clear changes can also be seen in many side chain conformations in the active site (further discussed below).

Cellobiose

The *Ma* Cel7B Glc₂ complex structure was solved at 1.7 Å resolution. In all four molecules of the asymmetric unit, the Glc₂ molecules are bound in the product site of the tunnel (subsites +1 and +2) (Fig. 2A). The glycosyl in subsite +1 packs against Trp375, but there are ~0.5 Å differences in the binding of cellobiose within the plane defined by Trp375 in different molecules (A, B, C, and D). The hydroxyl group O4 of glucose (corresponding to

Table 2. Occupied glycoside binding sites for each *Ma* Cel7B complex

a.u.	Soak	–5	–4	–3	–2	–1	+1	+2
A	Glc ₂						x	x
B	Glc ₂						x	x
C	Glc ₂						x	x
D	Glc ₂						x	x
A	Glc ₃	x	x	x		x	x	x
B	Glc ₃					x	x	x
C	Glc ₃	x	x	x		x	x	x
D	Glc ₃					x	x	x
A	Glc ₄		x	x			x	x
B	Glc ₄				x	x	x	x
C	Glc ₄						x	x
D	Glc ₄			x	x		x	x

Figure 2. (A–C) Active site superposition of the four molecules in the asymmetric unit in stereo. (A) Cellobiose, (B) cellotriose, and (C) cellotetraose soaked structures. Observed cello-oligomers are in green and red. The side chains of amino acid residues in the cellulose-binding site are in gray. (D) Superposition of all 12 active sites with cellulose chain modeled in *Tr* Cel7A structure (blue) in stereoview.

the glycosidic oxygen) in the subsite +1 in molecules B, C, and D is close to the carboxylic oxygen of the acid/base catalyst Glu217 (2.4–2.7 Å), while in molecule A the distance is longer, 4.6 Å.

Cellotriose

The *Ma* Cel7B Glc₃ complex structure was solved at 1.8 Å resolution. The molecules showed two kinds of binding modes (Fig. 2B). Molecules A and C both have two Glc₃ molecules bound in the active site tunnel, one occupying subsites –5 to –3 and the other –1 to +2. Molecules B and D have only one Glc₃, bound in subsites –1 to +2. The electron densities for Glc₃ molecules in sub-

sites –1 to +2 are well defined and continuous over the catalytic site, as shown in Figure 2B. The glycosyl in subsite –1 pack partially against Trp366 and in subsite +1 against Trp375, as in cellobiose complexes. The differences in the binding of Glc₃ in different molecules are larger than in Glc₂ complexes, as large as 2 Å. The distance from the glycosidic oxygen to the carboxylic oxygens of acid/base catalyst Glu217 is over 4.3 Å, which explains the inability of *Ma* Cel7B crystals to hydrolyze Glc₃. In all four cellotriose complex structures the side chain of Glu127 has turned slightly, thus making hydrogen bonds with O5 or O6 of the glycosyl residue of subsite –1. In addition, the two carboxylic oxygens of the

other catalytic residue, the nucleophilic Glu212, are also over 4.4 Å away from C1 of the glycosyl residue at subsite -1.

Cellotetraose

The *Ma* Cel7B Glc₄ complex structure was solved at 2.1 Å resolution. All four molecules in the asymmetric unit show different glycosyl binding modes (Fig. 2C). The complete Glc₄ molecule is only seen in molecule B at subsites -2 to +2. In the rest of the molecules, the enzyme has cleaved the Glc₄ during soaking, and either one or two Glc₂ molecules can be detected bound in the active site of the *Ma* Cel7B. Molecule A has two Glc₂ molecules bound at subsites -4 to -3 and +1 to +2. Molecule C has only one Glc₂ bound in the product site (+1 to +2), and molecule D has two Glc₂ molecules bound at subsites -3 to -2 and +1 to +2. The cellobiose bound at subsites +1 to +2 in molecules C and D is superimposable upon the bound ligand in the cellobiose complex structures. On the other hand, the binding of the glycosyl residue at subsite +1 in molecule A is clearly different compared to other structures. Here, the glycosyl ring packs more intensively against Trp375, and its position has shifted 3 Å compared to the corresponding glycosyl unit in other structures. This also extends the distance of the carboxylic oxygen of Glu217 (acid/base) to the O4 of the glycosyl unit (which corresponds to the glycosidic oxygen of the cleaved linkage) to 5.4 Å. In molecule B, a Glc₄ molecule is detected bound to the subsites -2 to +2, thus extending well over the catalytic cleavage site in both directions. The binding mode of Glc₄ resembles the binding of Glc₃ molecules, especially in one important context: The distances from the glycosidic oxygen to the two carboxylic oxygens of Glu217 are

4.4 and 4.7 Å, which suggests that in this binding mode the acid/base catalyst Glu217 is not able to protonate the glycosidic oxygen.

Protein-carbohydrate interactions at different subsites

The occupied glycosyl binding sites in the determined *Ma* Cel7B complex structures are summarized in Table 2, and the protein-carbohydrate-hydrogen bonding interactions are shown in Figure 3. The interactions at different subsites are described in more detail below.

The putative binding sites -7 and -6

The protein mediated interactions in subsites -7 and -6 cannot be determined on the basis of the solved *Ma* Cel7B complex structures. The subsites are in the complex structures occupied by neighboring *Ma* Cel7B molecule in such a way that the amino acids Glu99 and Tyr100 penetrate into the entrance of the cellulose-binding tunnel. In *Tr* Cel7A, Trp40 packs against a glycosyl residue occupying the subsite -7. The same residue (Trp40) also exists in *Ma* Cel7B, which suggests the existence of subsites -7 (and -6) in this enzyme. The Trp40 at subsite -7 has the same side chain conformation as in *Tr* Cel7A except in the molecule A in the native *Ma* Cel7B structure, and in the Glc₃ and Glc₄ complex structures in which the tryptophan side chain has flipped and no longer stacks against the glucose ring at subsite -7.

Binding site -5

The binding of glycosyl at site -5 is observed in two *Ma* Cel7B molecules soaked with Glc₃. The observed protein-carbohydrate interactions include hydrogen bonds between the glycosyl O2 hydroxyl and the amide

Figure 3. Protein-carbohydrate-hydrogen-bonding interactions in the active site tunnel of *Ma* Cel7B. The interactions are collected from all 12 complex structures. Hydrogen bonds are indicated with broken lines.

group of Asn103, and the O6 hydroxyl and the main chain oxygen of Arg39. Only the first mentioned hydrogen bond is detected in both molecules. In the other molecule, O6 hydroxyl makes a hydrogen bond with Glu99 from the neighboring molecule. O4 hydroxyl is also hydrogen bonded to the same Glu99.

Binding site -4

Occupation of binding site -4 is observed in three molecules, in two Glc₃ and one Glc₄ complex structures. Two asparagine residues, Asn103 and Asn37, are in close contact with a glycosyl ring. The hydrogen bond between the amide group of Asn37 and glycosyl O2 hydroxyl is shown in all three molecules. The other asparagine, Asn 103, is hydrogen bonded to O6 hydroxyl (Glc₃ soaked) or O4 hydroxyl (Glc₄ soaked). In addition to hydrogen bonds, the protein-carbohydrate interactions include hydrophobic stacking with an indole ring of Trp38 residue.

Binding site -3

Binding site -3 is occupied in two Glc₃ soaked and in two Glc₄ soaked molecules. The glycosyl units are bound in a slightly different orientation in all four molecules. The amino acids, which are in close contact with the ligand, include two aspartic acids and an arginine. Glycosyl O6 hydroxyl is hydrogen bonded to Arg107 and Asp179, and O2 hydroxyl forms a hydrogen bond with the carbonyl oxygen of Asp367.

Binding site -2

Binding site -2 is occupied in two Glc₄ soaked molecules, B and D. In molecule B, this site is occupied by Glc₄, bound at subsites -2 to +2, and in molecule D the hydrolysis product Glc₂ is bound to subsites -3 to -2. It has been observed in *Tr* Cel7A complex structures that the cellulose chain is twisted twice in the tunnel, between subsites -4 and -3, and -3 and -2, leading to an overall twist of about 130° (Fig. 2D; Divne et al. 1998). These twists are also detected in *Ma* Cel7B complex structures (Fig. 2A-C). The glycosyl unit at subsite -2 in the Glc₄ complex is not twisted as much as the corresponding glycosyl in the Glc₂ complex; therefore it does not form any hydrogen bonds with the *Ma* Cel7B protein. In contrast, the glycosyl at subsite -2 in the Glc₂ structure forms several hydrogen bonds with the protein. Due to this difference between Glc₄ and Glc₂ complexes, the corresponding glycosyl units at subsite -2 are not exactly in the same position. O2 hydroxyl is hydrogen bonded to hydroxyl groups of Tyr145 and Ser364. O3 hydroxyl bonds to Arg107 and O2 hydroxyl to Tyr247. The fifth hydrogen bond is between the O4 hydroxyl and hydroxyl group of Tyr171. However, this bond cannot

be formed when O4 is a part of a glycosidic linkage. The protein-carbohydrate interactions in site -2 also include hydrophobic stacking with Trp366. The conformation of this tryptophan side chain is turned in the Glc₄ complex structure molecule C, thus preventing its interaction with the cellulose chain in the active site tunnel (Fig. 2C).

Binding site -1

Binding site -1 is occupied in all four Glc₃ soaked molecules and in one Glc₄ soaked molecule, which contains an intact Glc₄ molecule. Similarly to the *Tr* Cel7A structure, the nucleophile, Glu212, and acid/base, Glu217, are positioned on opposite sides of the glycosidic linkage between sites -1 and +1. The distance between their carboxylate groups is ~6Å. Asp214 adapts different side chain conformations, making hydrogen bonds with Glu214 or Glu217. It is difficult to deduce its role exactly; however, the corresponding mutation Asp214Asn in *Tr* Cel7A has reduced activity (Ståhlberg et al. 1996). The glycosyl unit at subsite -1 makes two hydrogen bonds with the *Ma* Cel7B. The O6 hydroxyl is hydrogen bonded to a carboxylate group of Glu217 in all four complex structures. The other hydrogen bond, between O2 hydroxyl and an amide group of Gln175, has only been detected in the Glc₄ complex structure. In addition to these two bonds, the free O4 hydroxyl is also hydrogen bonded to Tyr171 and Asp173.

Binding sites +1 and +2

The product binding sites +1 and +2 are occupied in all of the structures. Because of the large amount of structural information, there is some variation in hydrogen bonding between the molecules. The number of protein-carbohydrate interactions in product sites is higher than in other sites; almost every hydroxyl group is hydrogen bonded with *Ma* Cel7B. The site +1 is the fourth subsite with hydrophobic stacking interactions. Trp375 is packed against the β -face of the glycosyl ring. O2 hydroxyl makes a hydrogen bond with carbonyl oxygen of Asn259; the bond is observed in eight molecules. O3 and O6 hydroxyl groups have several amino acids in close contact. It has been observed that the O3 hydroxyl makes hydrogen bonds with Asp214, Glu217, and His228 and O6 with Gln175, Thr246, and Arg251. In addition, in the Glc₂ soaked complexes, O4 hydroxyl is hydrogen bonded to Glu217.

In most of the complexes, Arg392 hydrogen bonds bidentally to the glycosyl O1 and O6 hydroxyl groups in site +2. In some complexes, O1 is also hydrogen bonded to Asp338 and O6 to Asp262 and Arg267. O3 hydroxyl makes a hydrogen bond with Arg251. Altogether, bonds have been observed in 10 complex structures.

Discussion

Side chain conformation changes in the active site

In the active site of *Ma* Cel7B, many amino acid side chains have a large variation of conformations in the different complex structures, particularly at subsites -1 , $+1$, and $+2$. A clear conformation change was seen in two tryptophan residues, Trp40 and Trp366, as mentioned above. The amino acids at the cleavage site ($-1/+1$), Glu212, Asp214, and Glu217, have several conformations (Fig. 2). Furthermore, conformational changes were observed in various other side chains in the active site, including the arginines 251, 267, and 392 at subsites $+1$ and $+2$. These side chain changes indicate that the active site of *Ma* Cel7B is flexible. However, it was not possible to find a correlation between the bound ligands and individual side chain conformations.

Overall comparison of M. albomyces Cel7B with other GH-7 cellobiohydrolases

The three previously solved structures of the GH-7 cellobiohydrolase catalytic modules from *T. reesei* Cel7A (PDB 1CEL), *P. chrysosporium* Cel7D (PDB 1GPI), and *T. emersonii* Cel7A (PDB 1Q9H) have sequence identities of 50%, 48%, and 51% with *Ma* Cel7B, respectively. All four catalytic modules have a similar overall fold, and the superimposing of the C $^{\alpha}$ -traces of *Tr* Cel7A, *Pc* Cel7D, and *Te* Cel7A on *Ma* Cel7B structure gave RMSD values between 1.0 and 1.2. The structural differences are mostly found in the loop structures. The nine disulphide bridges that stabilize the structure of *Ma* Cel7B are conserved in all four enzymes, and the *Tr* Cel7A has an additional 10th disulphide bridge between residues 4 and 72. These cysteines are replaced by glycine and methionine in *Ma* Cel7B and by glycine and alanine in *Pc* Cel7D and *Te* Cel7A. Most residues involved in the substrate binding are highly conserved between the four enzymes, including both the stacking of the four Trp side chains against the glycosyl units and direct hydrogen bonds from the protein. Only three residues directly involved in substrate binding in *Ma* Cel7B are not totally conserved. Thr246 is missing in the *Pc* Cel7D structure at subsite $+1$. Asn259 is an aspartic acid in other enzymes; however, the carboxyl group of Asn259 interacts with the substrate, and this change has no effect in binding. The third difference is that Asp338 is replaced by glycine in *Tr* Cel7A at subsite $+2$.

Although the structures of the substrate binding tunnels are similar, there are differences in loops covering the binding site cleft near subsites -7 , -1 , and $+2$ that presumably affect the enzyme activity. The biggest differences can be seen in a loop situated at the entrance of the

Ma Cel7B tunnel (residues 97–102). This loop has one residue insertion compared to *Tr* Cel7A and a four-residue insertion compared to *Pc* Cel7D and *Te* Cel7A structures. Due to this insertion, the entrance around subsite -7 of *Ma* Cel7B is more closed. In the turn of this loop there is a tyrosine (Tyr100) residue, which may have a guiding effect up on the substrate-binding tunnel as suggested for Tyr47 in *Pc* Cel7D (Muñoz et al. 2001). On the other hand, in *Tr* Cel7A, Gln101 makes hydrogen bonds with the glycopyranoses at sites -5 and -6 . This glutamine is replaced by glycine in *Ma* Cel7B, and, as in the *Pc* Cel7D and *Te* Cel7A structures, the hydrogen bonds cannot be formed. Consequently, there are more protein–carbohydrate interactions at the entrance of the tunnel in *Tr* Cel7A than in the other three enzymes.

The second loop containing structural changes covers the catalytic site of *Ma* Cel7B and is formed by residues 243–254. In the *Pc* Cel7D structure, this loop is shorter (six-residue deletion); therefore, the active site is more exposed to solvent than in the other structures (Muñoz et al. 2001). In the other three cellobiohydrolase structures, the loop has an equal length and contains only minor conformational changes. Although the amino acid composition of the loop in the different cellulases varies, all the amino acids pointing inside the tunnel are conserved. However, Tyr247 together with Tyr370 makes the catalytic site enclosed in both *Ma* Cel7B and *Tr* Cel7A structures, while in the *Te* Cel7A structure the Tyr244 (an equivalent to Tyr247) is not shown and Tyr370 is replaced by alanine. Two conformations have been observed for the two tyrosine residues in free and complex *Tr* Cel7A structures (Divne et al. 1998; von Ossowski et al. 2003). In *Ma* Cel7B only one conformation was observed that corresponded to the cello-oligosaccharide bound *Tr* Cel7A conformation.

At the product site of the tunnel, a loop formed by residues 381–390 of *Ma* Cel7B contains one residue deletion in comparison to other enzymes. This makes the loop slightly more compact compared to other structures, but there is no effect in the hydrogen-bonding pattern with the substrate. In addition, several conformations in the *Ma* Cel7B structures imply that this loop is very flexible.

Comparison of the Ma Cel7B and Tr Cel7A complex structures

The *Tr* Cel7A-cello-oligosaccharide complexes, which are determined using catalytically deficient mutants, contain the following structures: 3CEL (E212Q, Glc₂), 5CEL (E212Q, Glc₄), 6CEL (E212Q, Glc₅), and 7CEL (E217Q, Glc₆) (Ståhlberg et al. 1996; Divne et al. 1998). Similar to these *Tr* Cel7A complex structures, in *Ma* Cel7B complex structures, no ring distortion at site -1

was detected. Divne et al. (1998) also modeled a cellulose chain in the tunnel of *Tr* Cel7A based on the determined complex structures. Here the conformation of modeled glycosyl in site -1 of *Tr* Cel7A was distorted and situated closer to the catalytic amino acids. Most of the glycosyl units in the *Ma* Cel7B complex structures are situated approximately in the same positions as in the modeled *Tr* Cel7A complex structure (Fig. 2D). Only the glycosyl unit at site -1 of *Ma* Cel7B is clearly bound differently, as no ring distortion was detected.

The superposition of the *Ma* Cel7B Glc₂ complex with the *Tr* Cel7A Glc₂ complex structure shows differences in Glc₂ binding. In *Ma* Cel7B, the Glc₂ is closer to the catalytic site and forms hydrogen bonds with Glu217 and His228 at subsite +1. In subsite +2 the hydrogen bonding is similar (only O6 conformation is different), though the position of the glycosyl is different. The bound Glc₂ molecules were also found in the *Ma* Cel7B Glc₄ soaked structure and in the *Tr* Cel7A Glc₆ structure. The Glc₂ binding in the Glc₄ soaked complex molecules C and D had only minor changes (small shifts in position and different orientation of O6 hydroxyl) compared to the Glc₂ soaked complex. Also, in the *Tr* Cel7A Glc₆ structure the Glc₂ is bound somewhat similarly as in the *Ma* Cel7B Glc₂ soaked structure. In the *Ma* Cel7B Glc₄ soaked molecule A, the glycosyl in site +1 is bound away from catalytic amino acids; thus, it does not form any hydrogen bonds with the enzyme. The glycosyl in site +2 occupies almost the same position as in other structures. These comparisons showed several different binding modes for Glc₂ in product site +1 to +2.

The unhydrolyzed Glc₄ molecules occupy the same subsites (-2 to +2) in *Ma* Cel7B Glc₄ and *Tr* Cel7A Glc₄. However, the positions of the glycosyl rings differ and the hydrogen bonding with the enzyme is also different. The distance between the ligand and the catalytic amino acids is shorter in the *Ma* Cel7B complex than in the *Tr* Cel7A Glc₄. In addition, the catalytic Glu217 has a different conformation in these two structures. In *Ma* Cel7B, Glu217 forms two hydrogen bonds with Glc₄ (O6 hydroxyl at site -1 and O3 hydroxyl at site +1), whereas in the *Tr* Cel7A complex the Glc₄ is not hydrogen bonded to Glu217. The differences in binding may be caused by the mutation Glu212Gln.

In *Tr* Cel7A Glc₅ complex, two Glc₅ molecules occupy subsites -6 to -2 and +1 to +4. When this structure is compared with the *Ma* Cel7B Glc₄ complex, the oligosaccharide molecules occupy different subsites. In addition, the overlapping glycosyl units have differences in orientation; however, the Glc₅ binding in the *Tr* Cel7A Glc₅ complex is similar to Glc₄ soaked molecule D, where two Glc₂ molecules are bound to subsites -3 to -2 and +1 to +2. The cellulose chain breakage clearly causes changes in glycosyl orientation at subsite +1. In structures

where the -1 site is unoccupied, the glycosyl in site +1 is closer to the two catalytic amino acids. In site -2, the glycosyl in the *Ma* Cel7B Glc₄ complex is not twisted as much as in the *Tr* Cel7A Glc₅ complex, since the -3 subsite is important for the twisting of the cellulose chain. In the product sites +1, +2, and +3 of *Tr* Cel7A, three arginine residues assist in the binding and positioning of the substrate and may play a role in the recognition of the reducing end of the cellulose chain. Arginine side chains (251, 267, and 392) are present at equivalent locations in *Ma* Cel7B. However, the Arg267, which makes a hydrogen bond with glycosyl at subsite +3 in *Tr* Cel7A, is hydrogen bonded to glycosyl at subsite +2 in some *Ma* Cel7B complexes. The additional subsites +3 and +4 shown in the *Tr* Cel7A Glc₅ complex are not possible for *Ma* Cel7B, since the Asp338 (glycine in *Tr* Cel7A) prevents the binding of glycosyl at the same position as the +3 subsite is in the *Tr* Cel7A structure.

Implications for the function and reaction mechanism

The characteristic feature of *Ma* Cel7B is a strong product inhibition by cellobiose ($K_d \sim 6 \mu\text{M}$) (S.P. Voutilainen, H. Boer, T. Puranen, J. Vehmaanperä, and A. Koivula, in prep.), which may be an important factor for the low catalytic efficiency of the enzyme on cello-oligosaccharides under crystallization conditions. Glucose units were also detected at subsites +1 to +2 in all determined 12 complex structures. On the other hand, a Glc₂ complex structure has been reported only for the Glu212Gln mutant but not for the native *Tr* Cel7A. In this complex structure, the position of Glc₂ occupying subsites, +1 to +2, was shifted by $\sim 0.8 \text{ \AA}$ compared to the Glc₂ complex structures of *Ma* Cel7B. This may be enough to explain the difference in the cellobiose binding affinity of *Ma* Cel7B compared to that of *Tr* Cel7A. However, there are no clear structural reasons in enzymes for that because surrounding active site residues are identical. Therefore, the reason for slightly different binding is probably due to small structural changes throughout the structure, which result in minor changes in the positions of active site residues. It is also possible that small structural changes induced by crystal packing explain the observed differences in the Glc₃ and Glc₄ binding modes to different molecules in the asymmetric unit of *Ma* Cel7B structure.

Enzymatic hydrolysis of glycosidic linkage has been extensively studied during recent decades, including complex structure where a covalent intermediate is detected between the catalytic nucleophile (aspartic or glutamic acid) and glycosyl residue bound at the -1 subsite of a retaining hydrolase (Davies et al. 1998; Vocadlo et al. 2001). This has been the major reason that the previous suggestion regarding a carbocation intermediate in lysozyme has been rejected. There have also been detailed structural

studies of high-energy conformations of carbohydrate rings bound at the subsite -1 of retaining hydrolases (Sulzenbacher et al. 1996; Tews et al. 1996; Davies et al. 1998; Sidhu et al. 1999; van Aalten et al. 2001; Vocadlo et al. 2001). These studies are normally based on mutant enzymes or modified substrates (inhibitors) because it has been impossible to obtain actual substrate-enzyme complexes. The major advantage in this study has been that we have used native enzyme and oligomers of a natural substrate, cellulose. It is also advantageous to have 16 different snapshots from the same enzyme, which reveal a surprisingly large conformational heterogeneity both in the ligands and enzyme. This is especially important because we also see different conformations in the catalytic residues, in the acid/base catalyst Glu217, and in the nucleophilic Glu212.

The double-displacement reaction mechanism (Fig. 4) of retaining cellulases acting on β -bonds is based on earlier suggestions (Rye and Withers 2000; Zechel and

Withers 2000; Vasella et al. 2002) but we present it here in a slightly different way by splitting it into smaller steps. We suppose that the first step is the protonation of glycosidic oxygen by an acid/base catalyst. This is well in accordance with the fact that ether oxygens are weak Brønsted bases. After protonation, the first intermediate is formed. According to the measured structures of *Ma* Cel7B, the protonation of the substrate is easily obtained because we can see in Glc₂ complex structures that the corresponding glycosidic oxygen is very near the carboxylic oxygens of Glu217. It is probably that at higher temperatures structural movements would bring cello-triose and cellotetraose more close to catalytic residues, thus allowing their hydrolysis.

If the protonated substrate moves toward nucleophilic Glu212, a nucleophilic substitution of the S_N2 type may occur, leading to the formation of the second, covalent intermediate (route A). Another (but according to the

Figure 4. The discussed reaction mechanisms for *Ma* Cel7B (see text for further discussion).

current literature, less probable) route (B) would lead to the breaking of the C1–O bond (S_N1 reaction). The formed, short-lived carbocation would immediately react with the nucleophile, creating a similar covalent intermediate. The third alternative (route C) would be that the carbocation would react directly with a water molecule if available. The covalent intermediate would be released by a hydroxide ion (S_N2 reaction).

The measured structures in this study suggest that the protonation of glycosidic oxygen would be the first step in the reaction. The structures do not allow one to deduce which one of the three reaction routes would be the most probable for *Ma* Cel7B. One alternative is, of course, that one route is dominant and other routes are side routes. There were no signs of higher energy conformations (Nerinx et al. 2005; Stam et al. 2005) of the glycosyl ring at subsite –1 in the determined complex structures, but, of course, this does not exclude the possibility that some glucosyl units of the substrate would bind to the active site tunnel in a distorted conformation.

Materials and Methods

X-ray data collection

The production, purification, and crystallization of *Ma* Cel7B have been described previously (Parkkinen et al. 2007). For the substrate soaks, a crystal was transferred to a cryoprotectant solution containing cello-oligomer. The soaking time was about 30 s with Glc₂ and Glc₃ and 15 min with Glc₄. After soaking, the crystals were plunged into liquid nitrogen and stored for data collection. Data were collected on beamlines X11 and X12 at the DORIS storage ring, EMBL Hamburg Outstation, Germany, and processed using the XDS package (Kabsch 1993). Data collection and processing statistics are given in Table 1. The *Ma* Cel7B crystals were pseudo-merohedrally twinned and belonged to space group $P2_1$, with unit cell dimensions of $a = 50.9$ Å, $b = 94.5$ Å, $c = 189.8$ Å, $\beta = 90.0^\circ$.

Structure solution and refinement

The native *Ma* Cel7B structure was solved by molecular replacement using the MOLREP (Vagin and Teplyakov 1997) program from CCP4 (Collaborative Computational Project, Number 4 1994). The structure of *T. reesei* cellobiohydrolase Cel7A (PDB 1CEL) was used as a search model. The molecular replacement calculations yielded four monomers for an asymmetric unit. Following rigid body refinement the model building was performed with the use of O (Jones et al. 1991), and the refinement was performed with SHELXH (Sheldrick and Schneider 1997). Five percent of the reflections were set aside for *R*-free calculations and the twin operator $h, -k, -l$ was used during the refinement. Non-crystallographic symmetry (NCS) restraints were applied during the first cycles of refinement. During the final cycles of refinement, NCS restraints were released, and water molecules were added onto the structures with the use of SHELXWAT. The native structure was used as the starting model for refinement of the complex structures.

Statistics for the crystallographic refinement and the final models are given in Table 1. The quality of the models was checked with PROCHECK (Laskowski et al. 1993).

Data deposition

The coordinates for the four structures have been deposited with the Protein Data Bank with accession numbers 2RFW, 2RFY, 2RFZ, and 2RG0.

Acknowledgments

We thank Reetta Kallio-Ratilainen for excellent technical assistance. Funding from the MaBio-project by the European Social Fund and the State provincial office of Eastern Finland, Department of Education and Culture, is gratefully acknowledged. The support from the European Community Research Infrastructure Action under the FP6 “Structuring the European Research Area” program for the EMBL Hamburg Outstation, Contract number RII3-CT-2004-5060008, is also acknowledged. We also thank Dr. Manfred Weiss for providing support at the EMBL Hamburg.

References

- Barr, B.K., Hsieh, Y.-L., Ganem, B., and Wilson, D.B. 1996. Identification of two functionally different classes of exocellulases. *Biochemistry* **35**: 586–592.
- Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. 1998. Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**: 548–557.
- Bhat, M.K. and Bhat, S. 1997. Cellulose degrading enzymes and their potential industrial applications. *Biotechnol. Adv.* **15**: 583–620.
- Boisset, C., Fraschini, C., Schülein, M., Henrissat, B., and Chanzy, H. 2000. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Appl. Environ. Microbiol.* **66**: 1444–1452.
- Collaborative Computational Project, Number 4. 1994. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760–763.
- Davies, G. and Henrissat, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**: 853–859.
- Davies, G.J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A.M., Schülein, M., and Withers, S.G. 1998. Snapshots along an enzymatic reaction coordinate: Analysis of a retaining β -glycoside hydrolase. *Biochemistry* **37**: 11707–11713.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T., and Jones, T.A. 1994. The three-dimensional structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**: 524–528.
- Divne, C., Ståhlberg, J., Teeri, T.T., and Jones, A.T. 1998. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **275**: 309–325.
- Grassick, A., Murray, P.G., Thompson, R., Collins, C.M., Byrnes, L., Birrane, G., Higgins, T.M., and Tuohy, M.G. 2004. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *Eur. J. Biochem.* **271**: 4495–4506.
- Haakana, H., Miettinen-Oinonen, A., Joutsjoki, V., Mäntylä, A., Suominen, P., and Vehmaanperä, J. 2004. Cloning of cellulase genes from *Melanocarpus albomyces* and their efficient expression in *Trichoderma reesei*. *Enzyme Microb. Technol.* **34**: 159–167.
- Henrissat, B. and Davies, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**: 637–644.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for binding protein models on electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**: 110–119.

- Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**: 795–800.
- Kleywegt, G.J., Zou, J.-Y., Divne, C., Davies, G.J., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T.T., and Jones, T.A. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.* **272**: 383–397.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemistry quality of protein structures. *J. Appl. Crystallogr.* **26**: 282–291.
- Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**: 506–577.
- Mackenzie, L.F., Sulzenbacher, G., Divne, C., Jones, T.A., Wöldike, H.F., Schülein, M., Withers, S.G., and Davies, G.J. 1998. Crystal structure of the family 7 endoglucanase I (Cel7B) from *Hemicola isolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. *Biochem. J.* **335**: 409–416.
- Maheshwari, R., Bharadwaj, G., and Bhat, M.K. 2000. Thermophilic fungi: Their physiology and enzymes. *Microbiol. Mol. Biol. Rev.* **64**: 461–488.
- Miettinen-Oinonen, A., Londesborough, J., Joutsjoki, V., Lantto, R., and Vehmaanperä, J. 2004. Three cellulases from *Melanocarpus albomyces* for textile treatment at neutral pH. *Enzyme Microb. Technol.* **34**: 332–341.
- Muñoz, I.G., Ubhayasekera, W., Henriksson, H., Szabó, I., Pettersson, G., Johansson, G., Mowbray, S.L., and Ståhlberg, J. 2001. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. *J. Mol. Biol.* **314**: 1097–1111.
- Nerinckx, W., Desmet, T., Piens, K., and Claeysens, M. 2005. An elaboration on the *syn-anti* proton donor concept of glycoside hydrolases: Electrostatic stabilisation of the transition state as a general strategy. *FEBS Lett.* **579**: 302–312.
- Parkkinen, T., Koivula, A., Vehmaanperä, J., and Rouvinen, J. 2007. Preliminary X-ray analysis of cellobiohydrolase Cel7B from *Melanocarpus albomyces*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**: 754–757.
- Rye, C.S. and Withers, S.G. 2000. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**: 573–580.
- Sheldrick, G.M. and Schneider, T.R. 1997. SHELXL: High-resolution refinement. *Methods Enzymol.* **277**: 319–343.
- Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L., and Brayer, G.D. 1999. Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase. *Biochemistry* **38**: 5346–5354.
- Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeysens, M., Teeri, T.T., and Jones, T.A. 1996. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **264**: 337–349.
- Stam, M.R., Blanc, E., Coutinho, P.M., and Henrissat, B. 2005. Evolutionary and mechanistic relationships between glycosidases acting on α - and β -bonds. *Carbohydr. Res.* **340**: 2728–2734.
- Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G.J. 1996. Structure of the *Fusarium oxysporum* endoglucanase I with a non-hydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* **35**: 15280–15287.
- Sulzenbacher, G., Schülein, M., and Davies, G.J. 1997. Structure of the endoglucanase I from *Fusarium oxysporum*: Native, cellobiose, and 3,4-epoxybutyl β -D-cellobioside-inhibited forms, at 2.3 Å resolution. *Biochemistry* **36**: 5902–5911.
- Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K.S., and Vorgias, C.E. 1996. Bacterial chitinase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat. Struct. Biol.* **3**: 638–648.
- Vagin, A. and Teplyakov, A. 1997. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **30**: 1022–1025.
- van Aalten, D.M.F., Komander, D., Synstad, B., Gåseidnes, S., Peter, M.G., and Eijssink, V.G.H. 2001. Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc. Natl. Acad. Sci.* **98**: 8979–8984.
- Vasella, A., Davies, G.J., and Böhm, M. 2002. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **6**: 619–629.
- Vocaldo, D.J., Davies, G.J., Laine, R., and Withers, S.G. 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**: 835–838.
- von Ossowski, I., Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., et al. 2003. Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. *J. Mol. Biol.* **333**: 817–829.
- Voutilainen, S.P., Boer, H., Linder, M.B., Puranen, T., Rouvinen, J., Vehmaanperä, J., and Koivula, A. 2007. Heterologous expression of *Melanocarpus albomyces* cellobiohydrolase Cel7B, and random mutagenesis to improve its thermostability. *Enzyme Microb. Technol.* **41**: 234–243.
- Yip, L.Y. and Withers, S.G. 2004. Nature's many mechanisms for the degradation of oligosaccharides. *Org. Biomol. Chem.* **2**: 2707–2713.
- Zechel, D.L. and Withers, S.G. 2000. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**: 11–18.