Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution

Annabelle VARROT*, Sven HASTRUP†, Martin SCHÜLEIN† and Gideon J. DAVIES*[1]

*Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York Y01 5DD, U.K., and †Novo Nordisk A/S, Novo Allé, DK-2880 Bagsvaerd, Denmark

The three-dimensional structure of the catalytic core of the family 6 cellobiohydrolase II, Cel6A (CBH II), from *Humicola insolens* has been determined by X-ray crystallography at a resolution of 1.92 Å. The structure was solved by molecular replacement using the homologous *Trichoderma reesei* CBH II as a search model. The *H. insolens* enzyme displays a high degree of structural similarity with its *T. reesei* equivalent. The structure features both O- (α-linked mannose) and N-linked glycosylation and a hexa-co-ordinate Mg^{2+} ion. The active-site residues are located within the enclosed tunnel that is typical for cellobiohydrolase enzymes and which may permit a processive hydrolysis of the cellulose substrate. The close structural similarity between the two enzymes implies that kinetics and chain-end specificity experiments performed on the *H. insolens* enzyme are likely to be applicable to the homologous *T. reesei* enzyme. These cast doubt on the description of cellobiohydrolases as exo-enzymes since they demonstrated that Cel6A (CBH II) shows no requirement for non-reducing chain-ends, as had been presumed. There is no crystallographic evidence in the present structure to support a mechanism involving loop opening, yet preliminary modelling experiments suggest that the active-site tunnel of Cel6A (CBH II) is too narrow to permit entry of a fluorescenyl-derivatized substrate, known to be a viable substrate for this enzyme.

INTRODUCTION

The enzymic hydrolysis of cellulose, by cellulases, plays a central role in the natural environment, where it is involved in the recycling of plant biomass. Cellulases have attracted much industrial interest for selective cotton-fibre modification and their potential in enzymic de-inking and waste biomass conversion [1]. They are glycoside hydrolases, which use general acid/base chemistry to hydrolyse the β-1,4 linkages in cellulose. Glycoside hydrolases have been classified into over 67 families on the basis of amino-acid sequence similarities [2–5] and cellulases are found in 12 of these (5–9, 12, 26, 44, 45, 48, 60, 61). Three-dimensional structures of cellulases have been obtained for representatives of families 5–9, 12, 45 and 48 (recently reviewed in [6]). Fungal cellulases normally function as modular entities in which a catalytic core domain is linked to one or more additional domains, such as those involved in cellulose binding [7]. In anaerobic organisms, the organization may be more complex, with many catalytic and additional domains linked together in a huge supramolecular complex, termed the cellulosome [8].

Cellulases have been traditionally classified into endoglucanases and cellobiohydrolases (CBHs). This reflects their respective catalytic activities on intact crystalline cellulose: this substrate is hydrolysed rapidly by the CBHs but only poorly by the endoglucanases. *In vitro*, the differentiation is often made on the basis of their respective catalytic activities on an artificial substrate, carboxymethyl-substituted cellulose (CMC), which is hydrolysed more rapidly by the endoglucanases. These differences, coupled with the modest degree of synergy displayed by endoglucanases and CBHs [9], have led to a model for the enzymic degradation of cellulose in which endoglucanase action provides new chain ends for the exo-CBHs (reviewed in [10,11]).

Recently, the proposal that CBHs are true exo-enzymes has come under increasing scrutiny. This stems from the fact that many of the properties displayed by these enzymes are intermediate between the endo- and exo-types. In order to resolve some of these apparent inconsistencies, Henrissat and colleagues synthesized a variety of chain-end-modified substrates [12]. The rationale was that substrates selectively modified at the chain ends would allow discrimination between exo-CBHs and endoglucanases. Somewhat surprisingly, these substrates are hydrolysed with high efficiency by the so-called exo-CBHs, despite their complete lack of a true polysaccharide chain end. Indeed, even bulky fluorescein-modified derivatives are hydrolysed by CBHs (Figure 1; [13]). These and other results cast doubt on the description of CBHs as exo-enzymes and led some to the development of a model based on a processive attack. In this, following an initial attack (exo or endo), the enzyme proceeds along the substrate, making multiple attacks without diffusing away from the substrate [14]. This concept has been widely accepted in polysaccharide-hydrolysing enzymes for many years [15] and was first proposed for CBH II as a result of the X-ray structure determination [16].

This chain-end recognition work has, on the whole, been performed on the *Humicola insolens* enzyme, casting doubt on its applicability to other related systems such as the *Trichoderma reesei* enzyme, whose structure was known [16]. In this paper, we present the three-dimensional structure of the catalytic core domain of the family 6 CBH II [hereafter Cel6A (CBH II)] from *H. insolens*, determined by X-ray crystallography at a resolution of 1.92 Å. The structure is similar to the CBH II from *T. reesei*, a feature which was utilized in the molecular replacement structure determination. This strongly suggests that kinetic results for modified substrates obtained with the *H. insolens* enzyme, which cast doubt on the ability of family 6 CBHs to recognize non-reducing chain ends, are applicable to the *T. reesei* enzyme and presumably to other CBHs from family 6.

Abbreviations used: CBH, cellobiohydrolase; Cel6A (CBH II), family 6 cellobiohydrolase II; CMC, carboxymethyl-substituted cellulose.

[1] To whom correspondence should be addressed (e-mail davies@yorvic.york.ac.uk).

Co-ordinates for the structure described in this paper have been deposited with the Brookhaven Protein Data Bank with accession code 1BVW.PDB.

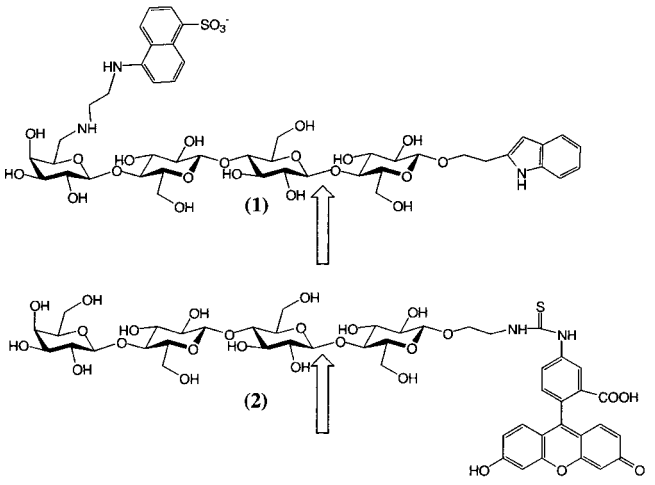


Figure 1 Modified oligosaccharide substrates that are hydrolysed by CBHs but which lack saccharide chain-ends [12,13]

The fluoresceinyl-derivatized substrates, such as compound 2, have bulky substituents whose dimensions are considerably larger than the diameter of the active-site tunnel in CBHs. The positions of hydrolysis of these compounds by the *H. insolens* Cel6A (CBH II) are indicated by vertical arrows.

MATERIALS AND METHODS

Protein purification

H. insolens Cel6A (CBH II) had been cloned previously by Sven Hastrup (unpublished work and [17]). Recombinant Cel6A (CBH II) was produced by sub-cloning the gene into an *Aspergillus oryzae* expression system under the control of a fungal amylase promoter and an amyloglucosidase terminator from *A. niger* [18]. The transformed *Aspergillus* was fermented, in a 10 litre fermenter, and the extracellular fluid passed through a DEAE-Sepharose column (50 mM Tris/HCl, pH 7.5). The unbound material was freeze-dried and contained a mixture of full-length (65 kDa) and core (50 kDa) Cel6A (CBH II). This was treated with endoglycosidase F: 50 mg of the Cel6A (CBH II) mixture was solubilized in 100 mM potassium phosphate buffer, pH 7.5, containing 50 mM EDTA and 150 μl (9 units) of endoglycosidase F (Boehringer Mannheim, Mannheim, Germany) and incubated for 60 h at 37 °C. Full-length Cel6A (CBH II) was separated by treatment with 25 mg of Avicel and removal of the bound material. The unbound core domain was purified using high performance Q (HPQ)–Sepharose in 50 mM Tris/HCl buffer, pH 7.5, and eluted using an NaCl gradient. The fractions containing the core enzyme were concentrated using an Amicon cell with a 10 kDa molecular-mass cut-off membrane. N-terminal sequencing (results not shown) revealed that the catalytic core domain started at Tyr^{89} with the following sequence: Y^{89}NGNPFEGVQL (numbering starts with the first residue of the mature protein).

Crystallization, data collection and processing

Prior to crystallization, the enzyme was washed on Filtron Microsep™ 10 kDa cut-off membranes with distilled water and concentrated to 20 mg·ml^{-1}. The protein was crystallized by the hanging-drop vapour diffusion method using 22 % poly(ethylene glycol) 8000 and 200 mM magnesium acetate as precipitants and 100 mM triethanolamine, pH 7.0, as buffer. Crystals were mounted in a rayon-fibre loop and placed in a boiling nitrogen stream at 100 K. A cryoprotectant solution was made by inclusion of glycerol, to 20 % (v/v), to the harvesting solution. X-ray diffraction data for the native enzyme were collected from a single crystal, to a resolution of 1.92 Å, at the Daresbury Synchrotron Radiation Source, beamline PX9.6 ($\lambda = 0.872$) using a MAR Research 345 image-plate detector. The crystal was mounted with the $b(b^*)$ axis along the spindle axis so that 180° of data were collected with an oscillation angle of 1° per image. Data were processed and reduced using the HKL suite of programs [19]. All further computing involved the CCP4 suite [20], unless otherwise stated.

Model building and refinement

The structure was solved by molecular replacement using the deposited Y169F mutant of the *T. reesei* CBH II (Protein Data Bank code 1CB2) as a search model [21]. The program AMoRE [22] was used in conjunction with data in the resolution range 20–4 Å and an outer radius of Patterson integration of 30 Å. Structure solution revealed just one significant solution. For refinement, 5 % of the observations were immediately set aside for cross-validation analysis [23] and were used to monitor various refinement strategies, such as geometric and temperature-factor restraint values, the insertion of solvent water and as the basis for the maximum-likelihood refinement using REFMAC [24]. Manual corrections of the model using the X-FIT routines of the program QUANTA (Molecular Simulations Inc., San Diego, CA, U.S.A.) were interspersed with cycles of least-squares refinement. Water molecules were added in an automated manner using the ARP program [25] and verified manually prior to co-ordinate deposition. Co-ordinates and observed structure-factor amplitudes for the protein structure described in this paper have been deposited with the Brookhaven Protein Data Bank [26] with accession code 1BVW.PDB.

RESULTS AND DISCUSSION

Quality of the final model structure

Crystals of Cel6A (CBH II) grow over a period of 5–10 days to a maximum size of $0.5 \times 0.1 \times 0.05$ mm. They are in space group $P2_1$, with cell dimensions $a = 47.5$ Å, $b = 68.1$ Å, $c = 53.7$ Å, $\alpha = \gamma = 90$ °, $\beta = 110.93$ °. A packing density of 2.01 Å^3·Da^{-1}, corresponding to a solvent content of 38 % [27], is obtained, assuming there is one molecule of Cel6A (CBH II) core in the

Table 1 Data collection and quality statistics for the native *H. insolens* Cel6A (CBH II) catalytic core-domain structure

$R_{merge} = \Sigma_{hkl}\ \Sigma_i |I_{hkli} - \langle I_{hkl}\rangle| / \Sigma_{hkl}\ \Sigma_i \langle I_{hkl}\rangle$

Lower limit (Å)	Upper limit (Å)	R_{merge}	$I/\sigma I$	Completeness	Multiplicity
20.00	4.13	0.028	41.08	99.4	3.75
4.13	3.28	0.035	30.10	99.6	3.79
3.28	2.87	0.061	17.17	99.8	3.74
2.87	2.60	0.077	15.62	99.7	3.78
2.60	2.42	0.082	15.82	99.8	3.80
2.42	2.28	0.093	14.96	99.8	3.80
2.28	2.16	0.103	13.48	99.4	3.80
2.16	2.07	0.124	11.07	99.6	3.78
2.07	1.99	0.153	9.08	99.3	3.79
1.99	1.92	0.187	7.34	99.1	3.73
	Overall ...	0.063	19.7	99.5	3.78

Table 2 Refinement and structure-quality statistics for the *H. insolens* CBH II

r.m.s., root mean square.

Resolution of data (Å)	20–1.92
No. of protein atoms/molecule (residues 89–450)	2829
No. of solvent waters	531
No. of O-glycosylation atoms (2 × mannose)	22
No. of N-glycosylation atoms (1 × *N*-acetylglucosamine)	14
No. of magnesium atoms	1
No. of glycerol atoms (1 × glycerol)	6
Resolution used in refinement (Å)	15–1.92
R_{cryst}	0.14
R_{free}	0.21
r.m.s deviation 1–2 bonds (Å)	0.008
r.m.s deviation 1–3 bonds (Å)	0.023
r.m.s deviation chiral volumes (Å^3)	0.099
Average main-chain B (Å^2)	12.9
Average side-chain B (Å^2)	15.5
Average solvent B (Å^2)	31.1
Main chain ΔB, bonded atoms (Å^2)	1.4

asymmetric unit. Data collected at the Daresbury Synchrotron Radiation Source consist of 90 589 observations of 23 966 unique reflections. During the data-reduction procedure, 1308 observations were rejected. The final data are 99.5% complete to 1.92 Å resolution, with an overall R_{merge} ($\Sigma_{hkl} \Sigma_i |I_{hkli} - \langle I_{hkl} \rangle| / \Sigma_{hkl} \Sigma_i \langle I_{hkl} \rangle$) of 0.063, a mean $I/\sigma(I)$ of 19.7 and a mean multiplicity of observations of 3.8 observations/reflection. A summary of data quality and completeness is given in Table 1.

The open reading frame for the *H. insolens Cel6A* (*CBH II*) gene encodes a mature protein of 450 amino acids. In common with most cellulases [7], Cel6A (CBH II) is a modular protein. It has an N-terminal cellulose-binding domain (equivalent to the *T. reesei* cellulose-binding domain elucidated by NMR [28]) linked to the catalytic core domain via a short linker region. The full catalytic core consists of residues 89–450, but in the three-dimensional X-ray structure only residues 91–450 are clearly visible in the electron-density maps. The final model has a crystallographic R value of 0.14, with a corresponding R_{free} of 0.21 for all observed data between 15 and 1.92 Å resolution. This model contains 2829 protein atoms, 531 solvent water molecules, a single Mg^{2+} ion, 1 glycerol molecule and 36 atoms of carbohydrate resulting from both O- and N-glycosylation. The overall structure is well ordered with an average main-chain temperature factor of just 12.9 Å^2. This model has deviations from stereochemical target values of 0.008 and 0.023 Å (corresponding to approximately 1.1°) for 1–2 and 1–3 bonds respectively. Final refinement statistics for the 1.92 Å native structure are given in Table 2. All the non-glycine residues have conformational angles (ϕ, ψ) in permitted regions of the Ramachandran Plot [29], with none of these in the 'generously allowed' or 'disallowed' regions as defined by PROCHECK [30].

Description of the structure

The structure of the catalytic core of Cel6A (Figure 2) presents the distorted α/β barrel topology, first described for a family 6 glycoside hydrolase by the *T. reesei* CBH II structure determination [16]. It is a single domain whose barrel displays an elliptical cross-section and whose central β-sheet core is composed of seven, not eight, parallel β-strands. The topology thus presents seven parallel β-strands, with eight major and two minor helices and an additional β-strand. Additionally, β-strands VI and VII are not connected by an α-helix but by an extended loop. The absent strand in the barrel results in the formation of a substrate-binding crevice between strands I and VII. Barrel closure is permitted by a single hydrogen bond between the main-chain amide of Phe^{135} on strand I and the carbonyl oxygen of Trp^{397} from strand VI. The additional β-strand lies outside the barrel, where it hydrogen-bonds to barrel strand II.

Cel6A displays both O- and N-linked glycosylation. Asn^{141} is found in the classic N-glycosylation motif Asn-Xaa-Ser/Thr. Following treatment with endoglycosidase F prior to crystallization, only the Asn-linked *N*-acetylglucosamine remains (Figure 3A). Two O-glycosylation sites have been found. Both Thr^{118} and Ser^{127} are O-glycosylated, both displaying a single α-linkage to the C-1 atom of mannose (Figure 3B). No density is visible for any additional sugars linked to the mannose, despite the fact that both mannose units are involved in crystal contacts. The electron density for Ser^{127} reveals two conformations, only one of which is glycosylated, hinting at glycosylation heterogeneity in this sample. A single magnesium ion is found, presumably as a result of the magnesium present in the crystallization conditions (Figure 3C). It is also involved in the crystal contacts. The magnesium ion displays the tetragonal bipyramidal co-ordination, typical of that normally observed for

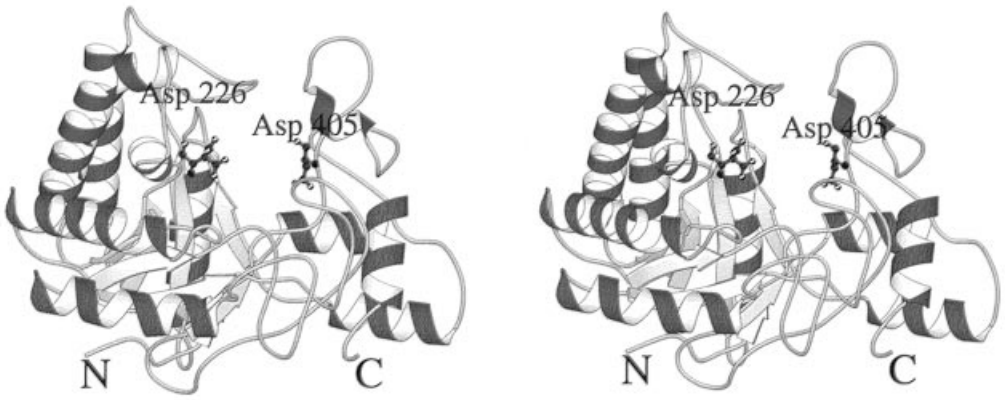


Figure 2 Schematic representation of the Cel6A (CBH II) from *H. insolens*, in divergent stereo

The catalytic residues Asp^{226} (acid) and Asp^{405} (potential base) are shown in ball-and-stick representation. This figure was drawn with the MOLSCRIPT program [54].

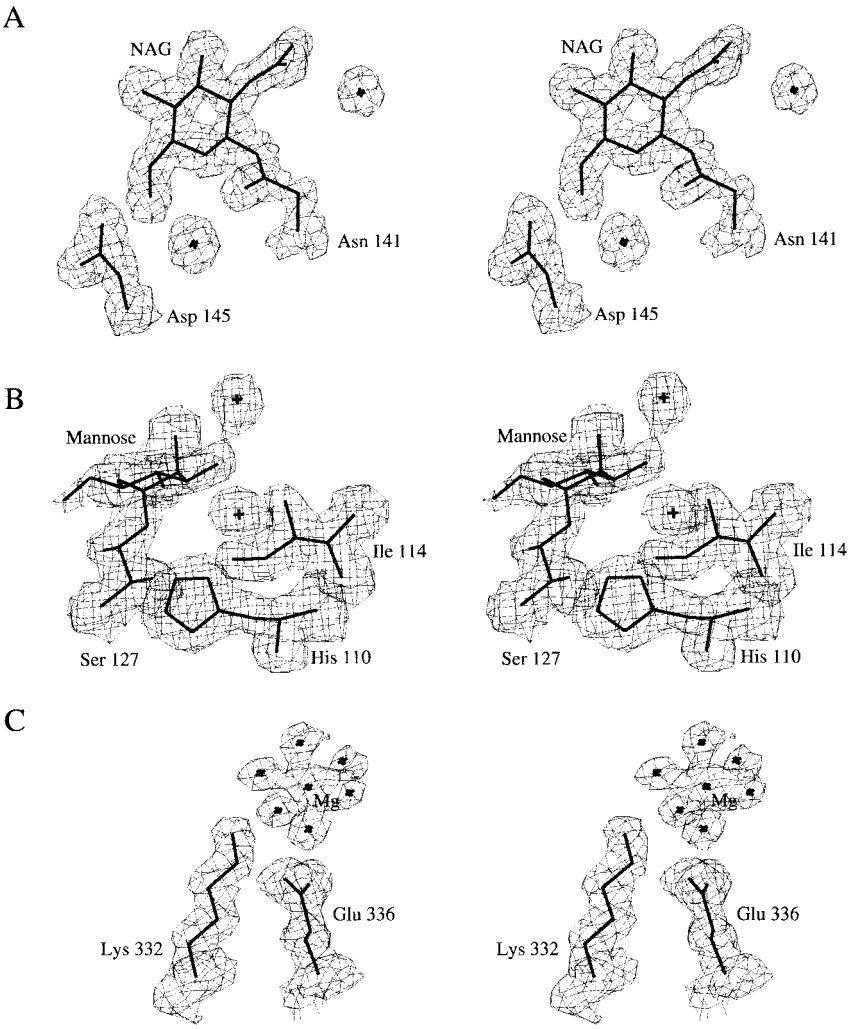


Figure 3 Electron density, maximum-likelihood-weighted $2F_{obs} - F_{calc}$ syntheses

Contour level was approximately 0.4 electrons Å^{-3} for (**A**) the N-linked *N*-acetylglucosamine (NAG) on residue Asn^{141}, (**B**) O-linked α-mannose on Ser^{127}, and (**C**) a hexa-co-ordinate Mg^{2+} ion near Lys^{332} and Glu^{336}.

Mg^{2+} ions, with six water ligands (distances from 1.85 to 2.39 Å). These water ligands in turn interact with the protein via Glu^{336} and Lys^{332}. A single glycerol molecule is found in the substrate-binding tunnel. Unusually, Cel6A has five *cis*-proline residues: Pro^{165}, Pro^{325}, Pro^{362}, Pro^{427} and Pro^{448}. One of these, Pro^{427}, lies in one of the two active-site loops, and thus *cis–trans* proline isomerization might, conceivably, play a role in loop flexibility as has been proposed for other glycoside hydrolases [31].

Similarity to other family 6 enzymes

The Cel6A (CBH II) structure is, as expected, extremely similar to that described for the *T. reesei* CBH II. The two proteins present 64 % sequence identity. The secondary structures are identical and the structures superpose extremely well, with a root-mean-square deviation of 0.69 Å for 356 out of 359 equivalent Cα atoms (calculated using LSQMAN [32]). While these two CBHs are extremely alike, the similarity with the endoglucanase E2 from *Thermomanospora fusca* is considerably less. Endoglucanase E2 and Cel6A (CBH II) have just 25 % sequence identity and overlap reveals just 233 equivalent Cα atoms, which overlap with a root mean square of 1.49 Å. The comparison between the two CBHs and the endoglucanase reveals that the β-barrel core is better conserved than position of the peripheral α helices (Figure 4). The significant difference between the endoglucanases and CBHs is that two loops which enclose the active sites in the CBHs are absent, to reveal an open substrate-binding groove, in the endoglucanase. This feature was predicted based on the structure of the CBH II alone [16] before being confirmed by the endoglucanase E2 structure determination [33]. The active site of Cel6A (CBH II), as for the *T. reesei* CBH II, is enclosed by two extended surface loops (residues 174–196 and 407–435). This results in the formation of a long substrate-binding tunnel, over 20 Å long. Both of these loops are stabilized by disulphide bridges (Cys^{181}–Cys^{240}, Cys^{372}–Cys^{419}). One of the two active-site loops (407–435) present in Cel6A (CBH II) is absent from E2. The other surface loop (174–196) is 'pulled back' in the E2 structure in order to expose the active site. This results in an active-site tunnel for Cel6A (CBH II), as opposed to a cleft in E2 (Figure 5). Cel6A (CBH II) contains at least four subsites, -2 to $+2$, with the catalytic centre enclosed by the tunnel's loops. Whereas the processive nature of hydrolysis may be easily explained, one

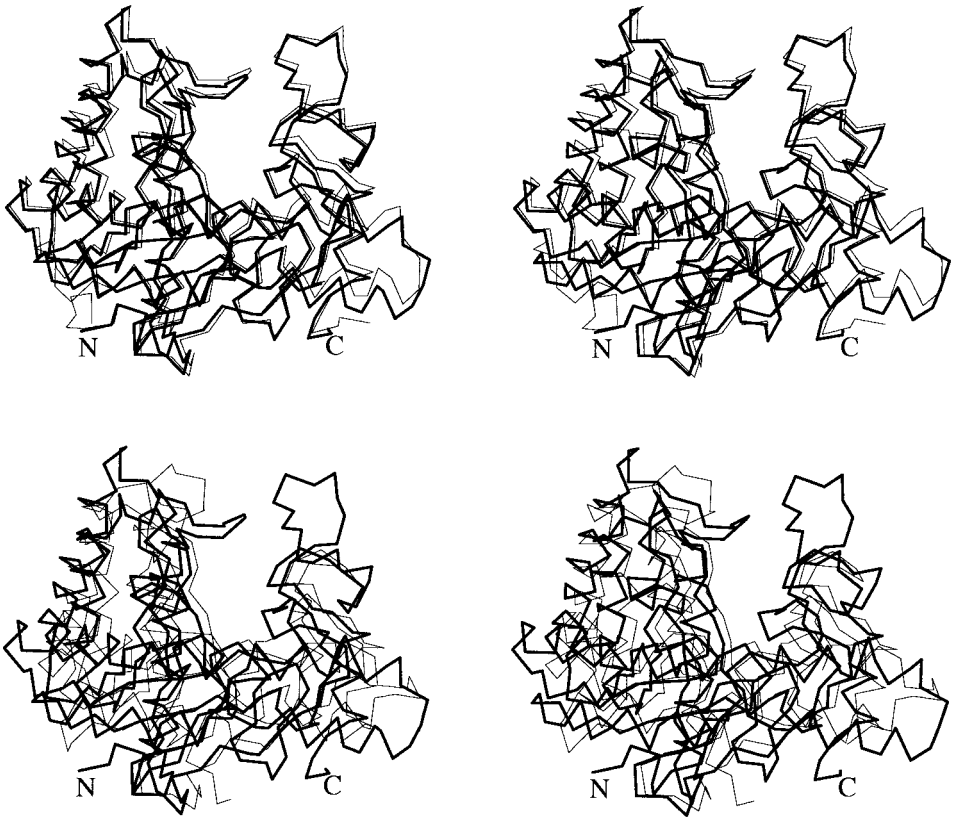


Figure 4 Stereographic representation of the overlap of family 6 enzymes of known structure

(Top) *H. insolens* Cel6A (CBH II) (bold), *T. reesei* CBH II (faint); (bottom) *H. insolens* Cel6A (CBH II) (bold) *Therm. fusca* endoglucanase E2 (faint). This figure was drawn with the MOLSCRIPT program [54].

of the implications of an active-site tunnel for CBHs is that some of the, apparently contradictory, 'endo-like' properties of the enzyme are more difficult to rationalize. The possibility that the active-site loops might occasionally open to permit an endo-attack has been proposed [14] and recently discussed in more detail [34]. Analysis of the temperature-factor distributions for these two loops, however, gives no hint that they are excessively mobile (Figure 6). The two tunnel loops display *B* values below 20 Å^2, lower than that observed in the previous *T. reesei* CBH II structure determination.

The two extended surface loops that enclose the active site present just a single direct contact with each other: Asn^{187} hydrogen-bonding to Arg^{414}. All the other contacts between the two loops are mediated through water molecules. Arg^{414} adopts two discrete conformations. In the 'B' conformation, the NH1 is 2.9 Å from the OD1 of Asn^{187}, with a much weaker interaction existing for the 'A' conformation, with a corresponding distance of 3.25 Å. The NH2 of the 'A' conformation results in an additional hydrogen bond with the OE1 of Glu^{430} (2.9 Å). Since Arg^{414} makes the only direct link between the two loops, the equilibrium between the two conformations of Arg^{414} need only be perturbed slightly for loop movement to occur. The 'A' conformation neither encloses the active site fully nor makes contacts with the other tunnel-enclosing loop. In the N-terminal loop, the lid of the tunnel includes residues Ala^{182}–Gly^{188}. In order for this loop to open as much as is seen in the endoglucanase E2 structure a large conformational change of about 10 Å for the main chain of the Ala^{183}–Asn^{187} would be required.

Catalytic mechanism

CBH II catalyses glycosidic-bond hydrolysis with inversion of anomeric configuration: it is an e → a enzyme using Sinnott and co-worker's nomenclature [35]. In the classical single-displacement mechanism outlined by Koshland [36], this requires the presence of two catalytic carboxylate groups: a proton donor to protonate the glycosidic bond and promote leaving-group departure, and a catalytic base to activate the hydrolytic water molecule for nucleophilic attack at the anomeric centre. For inverting enzymes, these groups are typically separated by approximately 9–10 Å [37] (Scheme 1).

The initial structure determination of the *T. reesei* CBH II [16] clearly revealed a candidate for the proton donor, i.e. Asp^{221} [equivalent to Asp^{226} in the *H. insolens* Cel6A (CBH II)], an assignment later confirmed by site-directed mutagenesis of a related enzyme [38]. The identification, indeed the presence, of the catalytic base in the CBH II reaction is less clear. Initial examination of the CBH II structure failed to find an invariant residue in a suitable position to act as a catalytic base. Asp^{401} seemed appropriately placed from a geometrical view, but was not invariant amongst the family 6 sequences and made close salt-links with both lysine and arginine residues; an unusual environment for a proton acceptor. The structure determination both in native conformation and in complex with cellobiose for the family 6 endoglucanase, E2, from *Therm. fusca* [33], however, led to the tentative proposal that the equivalent residue, Asp^{265} in E2 numbering, was indeed the catalytic base. At this time, a

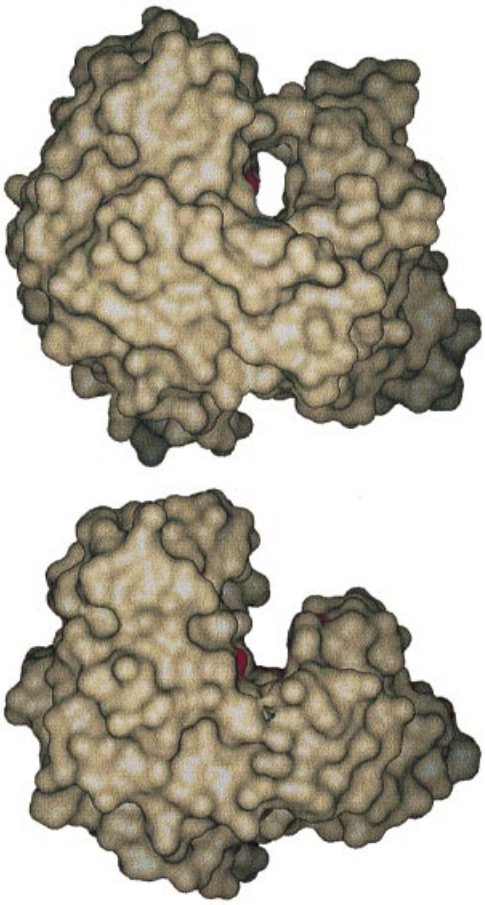
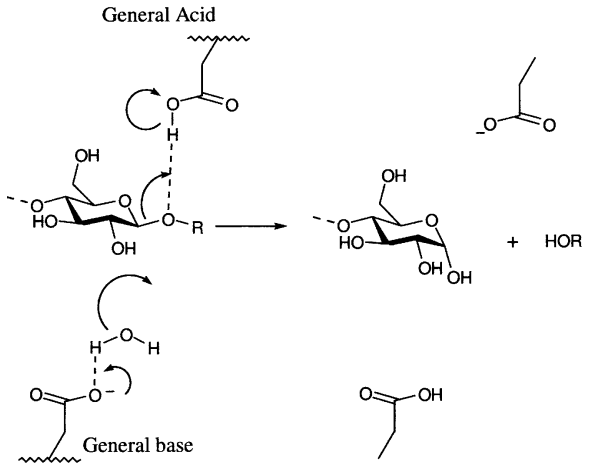


Figure 5 Van der Waals surface representations of the family 6 CBH Cel6A from *H. insolens* (top), and the endoglucanase E2 from *Therm. fusca* (bottom)



Scheme 1 Canonical single-displacement reaction mechanism for an 'inverting' (e → a) glycoside hydrolase

re-evaluation of the known family 6 sequences resulted in identification and correction of some rogue sequences, the revised alignments revealing that the proposed general-base residue was truly invariant. Further evidence as to the existence and identity of the general base came from elegant studies on a homologous family 6 enzyme, CenA from *Cellulomonas fimi*. Bell-shaped pH profiles indicated both general-acid and general-base contributions to catalysis [39]. Site-directed mutagenesis clearly identified these residues and demonstrated quantitatively the role of a third aspartate in the elevation of the pK_a of the Brønsted acid [38]. These assignments have yet to attract widespread support. Indeed, the idea that deprotonation of the catalytic water occurs via a 'Grotthus' mechanism in which deprotonation of the nucleophile occurs via a string of water molecules has also been considered (reviewed in [35]). Much of the published kinetics on these systems is further confused since experiments are often performed in acetate buffer or with CMC as substrate, which can give rise to buffer- or substrate-assisted catalysis, respectively, even in the absence of enzymic acid/base functions.

The *H. insolens* Cel6A (CBH II) structure is consistent with the previously proposed catalytic machinery in which the Brønsted acid is Asp^{226}, and the base is Asp^{405}. The respective identification of these residues in family 6 enzymes of known structure or function is shown in Table 3. These residues lie either side of the active-site tunnel with their carboxylate groups separated by

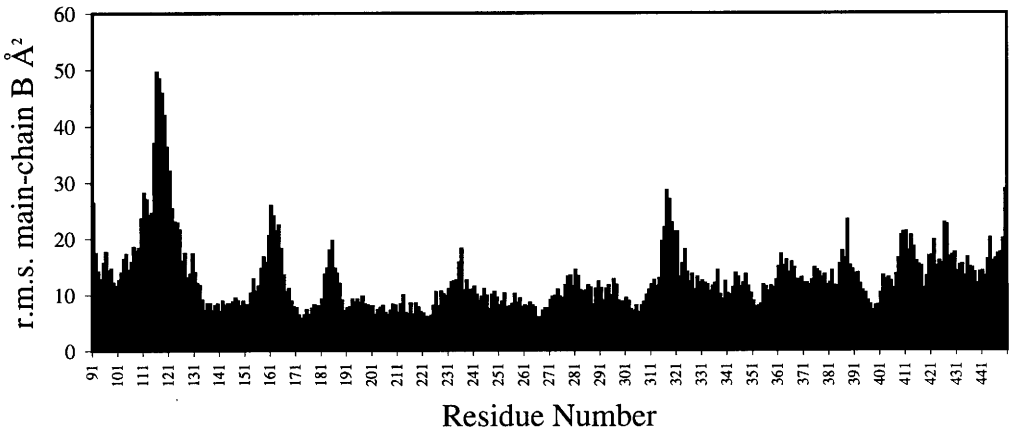


Figure 6 Plot of main-chain root mean square (r.m.s.) *B* values against residue number for the *H. insolens* Cel6A (CBH II)

The two loops enclosing the active site, residues 174–196 and 407–435, display higher mobility than the molecular core, but are still quite static, with average temperature factors of around 20 Å^2.

Table 3 Catalytic residue equivalencies in family 6

Enzyme	Catalytic acid	Postulated catalytic base	pK_a modulator
CBH II (*T. reesei*) [16]	221	401	263
E2 (*Therm. fusca*) [33]	117	265	156
CenA (*C. fimi*) [38]	252	392	287
Cel6A (*H. insolens*)	**226**	**405**	**268**

approximately 9.5 Å. The catalytic acid, Asp^{226}, is present in a double conformation, these two conformations being equivalent to the two different orientations found for this residue in the CBH II and E2 structure determinations [16,33]. In common with other similar systems [31,40], these two conformations may represent different protonation states of this residue at pH 7.0. The presence of two distinct conformations for the catalytic acid is important. In one of the conformations (corresponding to that observed in the E2 structure), the distance from the Asp^{226} OD1 to the OD1 atom of the proposed base is 9.5 Å, consistent with the distance observed in an analysis of many inverting enzymes [37]. In the alternative conformation (corresponding to that previously observed in *T. reesei* CBH II) the carboxylate oxygens lie 11.3–13.2 Å distant. Asp^{405} cannot therefore be excluded as a basic catalyst on distance grounds and thus the *H. insolens* Cel6A (CBH II) structure is entirely consistent with a standard Koshland [36] single-displacement reaction with both acid and base catalytic functions, as proposed by Damude and co-workers [38,39].

Damude and colleagues identified a third catalytic aspartate, equivalent to Asp^{268} in the *H. insolens* Cel6A (CBH II), which functions to elevate the pK_a of the Brønsted acid. In the Cel6A (CBH II) structure it lies approximately 4.8 Å away from Asp^{226}, linked by a solvent-mediated hydrogen bond, and is thus appropriately placed to assist in pK_a elevation. A fourth catalytic residue has been implicated in pK_a elevation in Cel6A (CBH II). This residue, Asp^{180} in the *H. insolens* enzyme, lies in a position equivalent to Asp^{175} in the *T. reesei* protein. It has been noted, however, that although apparently conserved in sequence, this residue is not conserved in terms of structure, since the equivalent residue in the corresponding *Therm. fusca* endoglucanase lies 4–5 Å distant from its position in the CBHs and hence 11 Å away from the proton donor. Consistent with its position, mutations at this residue have only a small effect on catalysis and suggest a less important role [16,38].

Simply in terms of bond cleavage, at least, CBH II is not an efficient catalyst with a k_{cat} for reduced cellohexaose of just 4 s^{-1} (over 20 times lower than the corresponding endoglucanase from family 6) [41–43]. These low catalytic constants are a feature of CBH II from all species and led Sinnott and co-workers to propose that binding energy is utilized not merely for catalysis, but also to disrupt the cellulose crystallites [44] in what Jencks described as a 'coupled vectorial process' [45]. This proposal is given more force by the observation that reversal of CBH II action may lead to the formation of crystalline cellulose [46]. CBH II is also unusual in that it appears to hydrolyse the 'wrong' α-cellobiosyl fluoride according to Michaelis–Menten kinetics [44] and not, as would be expected, via the Hehre resynthesis mechanism [47]. The reason for this is unclear since the corresponding endoglucanases from this family do, indeed, hydrolyse α-cellobiosyl fluoride via a synthesis–resynthesis mechanism [39].

Macroscopic considerations: exo versus endo

Cellulases are broadly classified into two subgroups, endoglucanases (EC 3.2.1.4) and CBHs (EC 3.2.1.91). This differentiation reflects the higher activity of the CBHs on intact crystalline substrates. Although this difference is undoubtedly genuine, the differentiation between CBHs and endoglucanases is typically made upon the basis of their respective hydrolyses of an artificial substrate, CMC. This is a good substrate for the endoglucanases, but a relatively poor one for CBHs. The justification for this is that the carboxymethyl substituents on CMC may not enter the enclosed active site of an exo-enzyme, whereas they could enter the more open active-site cleft of an endoglucanase.

CBHs show a small degree of synergy with the endoglucanases [9]. This has widely been interpreted as evidence that the CBHs are exo-enzymes, requiring polymer-chain ends for catalytic activity, and that synergy arises because the endoglucanases make internal cuts, generating free-chain ends upon which the CBHs act (for review, see [11]). This, together with the observed pointed-tip morphology that results from CBH II action [48,49], gave rise to the belief that CBH II acts at the non-reducing end of the substrate, consistent with its description as an exo-enzyme. What is strange, however, is that there appear to be no reports of genuine synergy on soluble forms of cellulose, such as phosphoric acid-swollen cellulose, which would certainly have been expected were a genuine exo/endo synergy operating.

Recently, proposals that CBH II is a true exo-enzyme have come under increasing scrutiny. CBH II and CBH I also display a small degree of synergy [50], wholly inconsistent with their proposed exo-action at opposite chain-ends. A plot of specific fluidity versus reducing sugar for CBH II-catalysed hydrolysis of CMC reveals a small increase in fluidity, indicative of a significant amount of internal (endo) cleavage [51]. Others have also shown, on the basis of the reducing-end production, that the description of CBH II as a true exo-enzyme is unjustified [52]. More recently, striking proof that CBHs can make an initial endo-attack within a polysaccharide chain comes from an analysis of the *A. niger* and *T. reesei* CBH systems. Both CBH I and CBH II make internal cuts in mixed-linkage and substituted-polysaccharide substrates such as β-glucan, lichenan and xyloglucan. Indeed, analysis of xyloglucan-oligosaccharide hydrolysis patterns reveals that undecorated regions of this polymer are hydrolysed by CBH II regardless of their position within the polymer. Hence, a polymer with a degree of polymerization of 14 is cleaved in true endo-fashion to yield two molecules with degrees of polymerization of 7 [53].

In order to differentiate between chain-end-recognizing cellulases and endoglucanases, Armand and co-workers [12] synthesized terminally modified cellodextrins in which fluorescent groups were located at both chain ends (Figure 1, compound 1). These reagents were hydrolysed rapidly by CBHs and demonstrate that CBHs do not recognize polymer chain ends [12]. Additionally, the fluorescent groups were too large to enter the active-site tunnel of the CBHs, as calculated on the basis of the published X-ray structures. Further work with even more bulky fluoresceinyl-derivatized substrates, such as compound 2 in Figure 1, clearly renders the concept of CBHs as exo-enzymes (in the normal meaning of the word, at least) as untenable. These experiments were performed on the *H. insolens* enzyme, leading to speculation that conclusions might not be applicable to other CBHs, such as the *T. reesei* enzyme, whose structure was known at that time. It was possible that the *H. insolens* Cel6A (CBH II) might have displayed a different active-site topology, making hydrolysis of these compounds more likely. The 1.92 Å resolution *H. insolens* Cel6A structure, presented here, reveals an extremely

close structural similarity with the *T. reesei* family 6 CBH. This strongly suggests that proposals concerning chain-end specificity made for one enzyme are likely to hold good for the other. The active-site tunnel of the *H. insolens* Cel6A (CBH II) constricts to a minimum dimension of approximately 3.5 Å. A fluorescein group, such as that found on the reducing end of compound 2 (Figure 1), has dimensions of approximately $16 \times 13 \times 8$ Å.

It is, however, possible to model the fluoresceinyl-derivatized substrates into the active site of Cel6A (CBH II) into such a position that the active-site residues interact with the known position of cleavage (shown in Figure 1) on the substrate. The fluoresceinyl group would lie outside the tunnel and does not clash with the protein. CBH II, however, has been proposed to allow threading of the cellulose chain through the active-site tunnel with a corresponding 'processive' hydrolysis [16]. So, whereas the three-dimensional structures of the *T. reesei* CBH II and *H. insolens* Cel6A (CBH II) are fully consistent with a catalytically competent state of the enzyme, hydrolysis of the fluoresceinyl-derivatized substrates cannot possibly occur without a conformational change of the active-site loops to permit their initial entry into the active site.

In order for substrate entry to occur, the N-terminal loop would have to be displaced approximately 10 Å in a position similar to that found in the E2 structure, and Arg^{414} from the second tunnel-enclosing loop would have to move further towards Glu^{430}. These movements would seem sufficient to permit hydrolysis of the modified substrates. The possibility that the loops move in response to substrate binding is not without precedent. Movements of up to 8 Å have been observed on two completely unrelated cellulases, the endoglucanase V from *H. insolens* [40] and the Cel5A from *Bacillus agaradhaerens* (A. Varrot, M. Schülein and G. J. Davies, unpublished work). We see no evidence, however, for such a conformational change in the native *H. insolens* Cel6A (CBH II) structure. Yet these enzymes are able to hydrolyse both polymeric and end-substituted substrates in an endo-fashion. Clearly further experiments are required before the action of CBHs on polymeric substrates is understood.

The authors would like to thank the staff of the CCLRC Daresbury Laboratory for provision of data-collection facilities. This work was funded, in part, by the European Union (under contract BIO4-CT97–2303) and Novo-Nordisk A/S. The Structural Biology Laboratory at York is supported by the Biotechnology and Biological Sciences Research Council. G.J.D. is a Royal Society University Research Fellow.

REFERENCES

1 Brennan, M. B. (1998) Chem. Eng. News **76**, 39–47
2 Henrissat, B. (1991) Biochem. J. **280**, 309–316
3 Henrissat, B. and Bairoch, A. (1993) Biochem. J. **293**, 781–788
4 Henrissat, B. and Bairoch, A. (1996) Biochem. J. **316**, 695–696
5 Henrissat, B. and Davies, G. J. (1997) Curr. Opin. Struct. Biol. **7**, 637–644
6 Davies, G. J. (1998) Biochem. Soc. Trans. **26**, 167–174
7 Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, Jr., R. C. and Warren, R. A. J. (1991) Microbiol. Rev. **55**, 303–315
8 Bayer, E. A., Morag, E. and Lamet, R. (1994) Trends Biotechnol. **12**, 379–386
9 Henrissat, B., Driguez, H., Viet, C. and Schülein, M. (1985) BioTechnology **3**, 722–726
10 Teeri, T. T. and Koivula, A. (1995) Carbohydr. Eur. **12**, 28–33
11 Teeri, T. T. (1997) Trends Biotechnol. **15**, 160–167
12 Armand, S., Drouillard, S., Schülein, M., Henrissat, B. and Driguez, H. (1997) J. Biol. Chem. **272**, 2709–2713
13 Boisset, C., Armand, S., Drouillard, H., Chanzy, H., Driguez, H. and Henrissat, B. (1998) in Carbohydrases from *Trichoderma reesei* and other Microorganisms (Claeyssens, M., Piens, K. and Nerinckx, W., eds.), pp. 124–132, Royal Society of Chemistry, London
14 Davies, G. and Henrissat, B. (1995) Structure **3**, 853–859
15 Robyt, J. F. and French, D. (1967) Arch. Biochem. Biophys. **122**, 8–16
16 Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. and Jones, T. A. (1990) Science **249**, 380–386
17 Wöldike, H. F., Hagen, F., Hjort, C. M. and Hastrup, S. (1991) World Patent WO 91 17 244
18 Christensen, T., Wöldike, H., Boel, E., Mortensen, S. B., Hjortshøj, K., Thim, L. and Hansen, M. T. (1988) BioTechnology **6**, 1419–1422
19 Otwinowski, Z. and Minor, W. (1997) Methods Enzymol. **276**, 307–326
20 Collaborative Computational Project Number 4 (1994) Acta Cryst. **D50**, 760–763
21 Koivula, A., Reinikainen, T., Ruohonen, L., Valkeajärvi, A., Claeyssens, M., Teleman, O., Kleywegt, G. J., Szardenings, M., Rouvinen, J., Jones, T. A. and Teeri, T. T. (1996) Protein Eng. **9**, 691–699
22 Navaza, J. (1994) Acta Cryst. **A50**, 157–163
23 Brünger, A. T. (1992) Nature (London) **355**, 472–475
24 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Acta Cryst. **D53**, 240–255
25 Lamzin, V. S. and Wilson, K. S. (1993) Acta Cryst. **D49**, 129–147
26 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr, E. T., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. **112**, 535–542
27 Matthews, B. W. (1968) J. Mol. Biol. **33**, 491–497
28 Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. and Gronenborn, A. M. (1989) Biochemistry **28**, 7241–7257
29 Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) J. Mol. Biol. **7**, 95–99
30 Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) J. Appl. Cryst. **26**, 282–291
31 Törrönen, A., Harkki, A. and Rouvinen, J. (1994) EMBO J. **13**, 2493–2501
32 Kleywegt, G. J. and Jones, T. A. (1994) ESF/CCP4 News **31**, 9–14
33 Spezio, M., Wilson, D. B. and Karplus, P. A. (1993) Biochemistry **32**, 9906–9916
34 Henrissat, B. (1998) Cellulose Commun. **5**, 84–90
35 Davies, G., Sinnott, M. L. and Withers, S. G. (1997) in Comprehensive Biological Catalysis, vol. 1 (Sinnott, M. L., ed.), pp. 119–209, Academic Press, London
36 Koshland, D. E. (1953) Biol. Rev. **28**, 416–436
37 McCarter, J. D. and Withers, S. G. (1994) Curr. Opin. Struct. Biol. **4**, 885–892
38 Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, Jr., R. C. and Warren, R. A. J. (1995) Biochemistry **34**, 2220–2224
39 Damude, H. G., Ferro, V., Withers, S. G. and Warren, R. A. J. (1996) Biochem. J. **315**, 467–472
40 Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C. and Schülein, M. (1995) Biochemistry **34**, 16210–16220
41 Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B. and Schülein, M. (1993) Eur. J. Biochem. **217**, 947–953
42 Schülein, M. (1997) J. Biotech. **57**, 71–81
43 Schülein, M., Tikhomirov, D. F. and Schou, C. (1993) in Proceedings of the second TRICEL symposium on *Trichoderma reesei* cellulases and other hydrolases, vol. 8 (Suominen, P. and Reinikainen, T., eds.), pp. 109–116, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland
44 Konstantinidis, A. K., Marsden, I. and Sinnott, M. L. (1993) Biochem. J. **291**, 883–888
45 Jencks, W. P. (1989) Methods Enzymol. **171**, 145–164
46 Sinnott, M. L. and Sweilem, N. S. (1998) in Carbohydrases from *Trichoderma reesei* and other Microorganisms (Claeyssens, M., Piens, K. and Nerinckx, W., eds.), pp. 13–20, Royal Society of Chemistry, London
47 Hehre, E. J., Brewer, C. F. and Genghof, D. S. (1979) J. Biol. Chem. **254**, 5942–5950
48 Chanzy, H. and Henrissat, B. (1985) FEBS Lett. **184**, 285–288
49 Chanzy, H., Henrissat, B., Vuong, R. and Schülein, M. (1983) FEBS Lett. **153**, 113–118
50 Fägerstam, L. G. and Pettersson, L. G. (1980) FEBS Lett. **119**, 97–100
51 Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, Jr., R. C., Warren, R. A. J. and Gilkes, N. R. (1995) J. Biol. Chem. **270**, 4383–4386
52 Ståhlberg, J., Johansson, G. and Pettersson, G. (1993) Biochim. Biophys. Acta **1157**, 107–113
53 Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E. and Kanda, T. (1996) J. Biochem. (Tokyo) **120**, 1123–1129
54 Kraulis, P. J. (1991) J. Appl. Cryst. **24**, 946–950

Received 20 July 1998/2 September 1998; accepted 9 October 1998