Structural Changes of the Active Site Tunnel of *Humicola insolens* Cellobiohydrolase, Cel6A, upon Oligosaccharide Binding[†,‡]

Annabelle Varrot,[§] Martin Schülein,[||] and Gideon J. Davies*[,§]

Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York Y010 5DD, U.K., and Novo-Nordisk A/S, Novo allé, Bagsvaerd, DK-2880, Denmark

Received February 18, 1999; Revised Manuscript Received April 29, 1999

ABSTRACT: The mechanisms of crystalline cellulose degradation by cellulases are of paramount importance for the exploitation of these enzymes in applied processes, such as biomass conversion. Cellulases have traditionally been classified into cellobiohydrolases, which are effective in the degradation of crystalline materials, and endoglucanases, which appear to act on "soluble" regions of the substrate. *Humicola insolens* Cel6A (CBH II) is a cellobiohydrolase from glycoside hydrolase family 6 whose native structure has been determined at 1.9 Å resolution [Varrot, A., Hastrup, S., Schülein, M., and Davies, G. J. (1999) *Biochem. J. 337*, 297−304]. Here we present the structure of the catalytic core domain of *Humicola insolens* cellobiohydrolase II Cel6A in complex with glucose/cellotetraose at 1.7 Å resolution. Crystals of Cel6A, grown in the presence of cellobiose, reveal six binding subsites, with a single glucose moiety bound in the −2 subsite and cellotetraose in the +1 to +4 subsites. The complex structure is strongly supportive of the assignment of Asp 226 as the catalytic acid and consistent with proposals that Asp 405 acts as the catalytic base. The structure undergoes several conformational changes upon substrate binding, the most significant of which is a closing of the two active site loops (residues 174−196 and 397−435) with main-chain movements of up to 4.5 Å observed. This complex not only defines the polysaccharide−enzyme interactions but also provides the first three-dimensional demonstration of conformational change in this class of enzymes.

Cellulose is the most abundant biopolymer on Earth where, as a result of its presence in the plant cell wall, it represents about half of the known biomass. It is constituted of β-1,4-linked glucosyl units, which polymerize as linear chains. Through hydrogen bonding and van der Waals forces, these form a crystalline polymer. Naturally occurring cellulose crystals, termed cellulose I, are composed of parallel chains (reviewed in ref *1*). Cellulose is used in the manufacture of paper and textiles and as a raw material for processed food, and hence its degradation is of great commercial interest. The stability of crystalline cellulose renders its degradation difficult, and in nature this task is performed by specific glycoside hydrolases named cellulases. These enzymes find widespread use in the detergent and paper industry (*2*), and their application in the environmentally friendly processing of municipal waste has been under investigation since the 1970s. These processes require enzymes with both high activity in applied conditions and high specificity for different forms of cellulose. While our understanding of cellulases and their sequences, structures, and mechanisms has increased enormously over the past decade, it is still not sufficient to permit their large-scale application in many processes.

Cellulases are frequently modular, with a catalytic domain linked to one or more noncatalytic domains (*3*). Some anaerobic bacteria produce a supramolecular complex, termed the cellulosome, which is formed by the association of numerous catalytic and noncatalytic domains (*4*). Cellulases are traditionally classified into two distinct types reflecting their activities on crystalline cellulose. This substrate is hydrolyzed by the cellobiohydrolases (EC 3.2.1.91) but is resistant to attack by the endoglucanases (EC 3.2.1.4) (*5*−*7*). Cellulases are found in 12 out of the 71 glycoside hydrolases (families: 5, 6, 7, 8, 9, 12, 26, 44, 45, 48, 60, 61) (*8*−*11*). The fungus *Humicola insolens* produces at least seven cellulases: five endoglucanases belonging to families 5, 6, 7, 12, and 45 and two cellobiohydrolases, Cel7A (formally CBH I; enzyme nomenclature is discussed in ref *12*) and Cel6A (CBH II),[1] belonging to families 7 and 6, respectively (*6*, *13*). Three-dimensional X-ray structures for enzymes from *H. insolens* are known for families 6, 7, and 45 (Cel6A, Cel7B, and Cel45) (*14*−*16*). With the exception of family 45, all the enzymes from *H. insolens* display high levels of sequence similarity with their extensively studied counterparts from *Trichoderma reesei*.

[†] This work was funded, in part, by the Biotechnology and Biological Sciences Research Council, The European Union (EC-BIOTECH 003R04292), Novo-Nordisk A/S, and the University of York. G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates and structure factors for the structures described in this paper have been deposited with the Protein Data Bank (accession numbers 1BVW, 2BVW, and 2BVF-SF).

* Corresponding author. Telephone: 44-1904-432596. Fax: 44-1904-410519. E-mail: davies@yorvic.york.ac.uk.

[§] University of York.

[||] Novo-Nordisk A/S.

[1] Abbreviations: CBH, cellobiohydrolase; CBH II, cellobiohydrolase from glycoside hydrolase family 6; CCD, charge-coupled device; CMC, carboxymethyl-substituted cellulose; HPLC, high-pressure liquid chromatography; NCS, noncrystallographic symmetry; rms, root mean square.

10.1021/bi9903998 CCC: $18.00
Published on Web 06/23/1999

One feature of cellulase systems is the synergistic action on crystalline materials (*17*). This, together with the observed "pointed-tip" morphology resulting from the action of CBH II on cellulose crystals (*18*), gave rise to the widely quoted "endo−exo" model for cellulase action (reviewed in ref *19*). In this model, endoglucanases are believed to cut cellulose chains in its "amorphous regions", thereby creating exposed polysaccharide chain ends upon which the "exo"-cellobiohydrolases begin their action. The different macroscopic action of endoglucanases and cellobiohydrolases can be easily rationalized by their respective active site topographies, as was first predicted as a result of the *T. reesei* cellobiohydrolase II structure (*20*): endoglucanases displaying an open active site in a cleft whereas cellobiohydrolases enclose their active site within a tunnel. In family 6, this tunnel is formed predominantly by two loops, hereafter named the "N-terminal and C-terminal loops" (residues 174−196 and 397−435, respectively), which presumably trap a single polysaccharide chain within the tunnel and prevent it from readhering to the crystalline matrix.

Cellobiohydrolases display numerous properties, which appear to be inconsistent with an exo mode of action (reviewed in refs *11*, *14*, *21*, and *22*). Many of the contradictory properties of cellobiohydrolases could, perhaps, more easily be rationalized were the active site tunnel to open, occasionally, to permit an endo attack, as must indeed be the case for their action on chain-end-modified substrates and xyloglucans, for example (*23*−*25*). In native enzyme structures in family 6, those of the cellobiohydrolases from *T. reesei* (*20*) and *H. insolens* (*14*), the active site enclosing loops appear relatively static, and no structural evidence for their opening has been presented. Here we present the structure of an oligosaccharide complex of the cellobiohydrolase II, Cel6A, from *H. insolens*. The structure reveals binding of a single glucose in the −2 subsite and of cellotetraose/cellotriose in the +1 to +4 subsites. In addition to the description of the enzyme−polysaccharide interactions through the substrate-binding tunnel, this structure also reveals, for the first time, significant movement of the two active site loops in order to produce a more enclosed conformation of the tunnel upon substrate binding.

EXPERIMENTAL PROCEDURES

Crystallization, Data Collection, and Processing. The catalytic core domain of Cel6A from *H. insolens* was prepared as described previously for the native enzyme (*14*). Initial experiments in which ligands were soaked into crystals resulted in crystal cracking, and so cocrystallization with cellobiose was performed. These crystals were prepared by the hanging drop vapor diffusion method using 2 μL drops containing a 50:50 (v/v) mix of protein (19 mg mL^{-1} in aqueous solution) and of reservoir solution with the addition of 20 mM β-D-cellobiose (Sigma Chemical Co., Poole, U.K.). A sparse-matrix screen revealed a single successful condition which was optimized to give final crystallization conditions of 200 mM ammonium acetate and 100 mM sodium acetate, pH 4.6, as buffer and 30% PEG 4000 as precipitant. Microseeding using crushed crystals as seeds was used in order to both improve crystal quality and initiate rapid nucleation. Cocrystals grow over a period of 1 month, display a platelike morphology, and reach a maximum size of 0.2 mm × 0.15 mm × 0.02 mm. Crystals were harvested in the solution described above, with the addition of glycerol, to a final concentration of 25% (v/v), as cryoprotectant. The crystals were loop-mounted and flash-cooled to 120 K using an Oxford Cryosystems cryostream. X-ray diffraction data were collected from two crystals at the Daresbury Synchrotron Radiation Source, beamline PX 9.6, using an ADSC Quad-4 CCD detector, up to resolutions of 2.1 and 1.7 Å, respectively. Two crystals were used due to a machine error at the synchrotron resulting in incomplete data from the first crystal. Data were processed using MOSFLM (A. Leslie, MRC-LMB, Cambridge, U.K.). All further computing used the CCP4 suite (*26*), unless otherwise stated.

Structure Determination. The structure of Cel6A in complex with cellobiose was solved by molecular replacement using the program AMoRE (*27, 28*). Data in the resolution range 20−4 Å were used together with an outer radius of Patterson integration of 33 Å. The native Cel6A structure (PDB code 1BVW), without solvent, was used as a search model (*14*). Structure solution revealed two significant solutions corresponding to the two molecules expected in the asymmetric unit. The structure was refined using REFMAC (*29*). Five percent of the observations were set aside for cross-validation analysis (*30*). They were used to monitor various refinement strategies such as geometry and temperature factor restraint values and the insertion of solvent water and as the basis for maximum-likelihood refinement (*29*). Tight noncrystallographic symmetry (NCS) restraints were applied initially, and these were slowly released as judged appropriate by the behavior of the cross-validation reflections. Initial model building utilized maximum-likelihood weighted $2F_o - F_c$ maps, which were averaged according to the NCS using DM (*31*). Stereochemical dictionaries for refinement of carbohydrate moiety were calculated from model coordinates generated with QUANTA (Molecular Simulation Inc., San Diego, CA). Manual corrections to the coordinates were performed using X-FIT routines of QUANTA and were interspersed with cycles of maximum-likelihood refinement. Water molecules were introduced automatically using ARP (*32*) and verified manually prior to coordinate deposition. Coordinates have been deposited with the Protein Data Bank (*33*). Details of data and model quality are given in Table 1.

RESULTS AND DISCUSSION

Structure Solution. Crystals of the glucose/cellotetraose complex of Cel6A are in space group $P2_1$ with cell parameters of $a = 48.56$ Å, $b = 154.43$ Å, $c = 51.04$ Å, $\alpha = \gamma = 90°$, and $\beta = 118.31°$. There are two molecules in the asymmetric unit with a packing density of 2.01 Å^3 Da^{-1} and a solvent content of approximately 38% (*34*). The crystals present strong $C222_1$ pseudo-symmetry; indeed, the position of reciprocal lattice points is entirely consistent with space group $C222_1$. Data reduction in this space group gives poor merging for symmetry-equivalent reflections (R_{merge} values in excess of 38%). Data were, therefore, processed in the space group $P2_1$ and the data from the two crystals reduced to give a final data set consisting of 277 256 observations of 72 987 unique reflections, with a completeness of 89.6% to 1.7 Å (95% complete to 2 Å resolution). The final data have an overall R_{merge} of 0.078, $I/\sigma I$ of 7.4, and a mean multiplicity of 4.4 observations per reflection (Table 1). Following structure solution with AMoRe, the

Table 1: Refinement and Structure Quality Statistics for the *H. Insolens* Cel6A Complex

data	
resolution of data (outer shell) (Å)	20−1.70 (1.79−1.70)
R_{merge}[a] (outer shell)	0.078 (0.359)
mean $I/\sigma I$ (outer shell)	7.4 (1.9)
completeness (outer shell) (%)	89.6 (89.6)
multiplicity (outer shell)	4.4 (4.6)
refinement	
no. of protein atoms	5684
no. of N-glycosylation atoms	28 (1 GlcNAc per protein molecule)
no. of ligand atoms	103
no. of glycerol atoms	42
no. of solvent waters	617
resolution used in refinement (Å)	20−1.70
R_{cryst}	0.175
R_{free}	0.225
rms deviation 1−2 bonds (Å)	0.008
rms deviation 1−3 bonds (Å)	0.024
rms deviation chiral volume (Å^3)	0.105
av chain A B (Å^2)	16.82
av chain B B (Å^2)	17.84
av ligand B (Å^2) cellotetraose (A molecule only)	22.36
av ligand B (Å^2) cellotriose (B molecule only)	21.33
av ligands B (Å^2) glucose	12.84
av glycerol B (Å^2)	31.62
av solvent B (Å^2)	28.61
main-chain ΔB, bonded atoms (Å^2)	1.4

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

molecular packing revealed that the two molecules were related by an angle of 177°, consistent with the self-rotation function output, enough to cause breakdown of the $C222_1$ symmetry. The final model has a crystallographic *R* factor of 17.5 and an R_{free} of 22.5 against all reflections in the resolution range from 20 to 1.70 Å. The overall root-mean-square deviations from ideality are 0.008 Å in bond lengths, 1.2° in bond angles, and 1.5 Å^2 for the ΔB between bonded atoms. This model (PDB code 2BVW) is composed of all atoms for residues 90−450 in chain A and for residues 91−450 in chain B, 7 glycerol molecules, 617 water molecules, and 1 N-linked *N*-acetylglucosamine per protein molecule. Additionally, the two active sites both contain a single glucose unit in the −2 subsite. The A molecule substrate-binding tunnel also contains a single tetrasaccharide species (subsites +1 to +4) which has been modeled as cellotetraose. In the B molecule, this oligosaccharide was best modeled as cellotriose (+1 to +3). Refinement statistics are summarized in Table 1. The Ramachandran plot (*35*) as defined by PROCHECK (*36*) shows that 88.3% of the non-glycine residues are in the most favored region and 11.4% in the additionally allowed regions and a single residue, Asp 226, is in the "generously allowed" region. Importantly, Asp 226 is the catalytic acid; it has ϕ, ψ angles of 76.2°, −21.8° and is in good electron density with an overall *B* factor of 15 Å^2.

Description of the Structure. The two molecules of Cel6A in the asymmetric unit overlap with an rms difference of 0.24 Å for the 359 equivalent Cα atoms [calculated using LSQMAN (*37*)]. Differences between the A and B molecules occur only in surface loops and are small in magnitude. The overall temperature factor for the A molecule is 16.8 Å^2 and for molecule B 17.8 Å^2. Strangely, water molecules appear to be better defined in molecule A where 358 discrete solvent molecules were modeled, compared to 259 in the B molecule. For these reasons, all further comparisons will use the A molecule, unless otherwise stated. An overlap of the native *H. insolens* Cel6A structure [PDB code 1BVW (*14*)] with its oligosaccharide complex reveals significant differences (Figure 1). The rms deviation between the native and complex A-chain coordinates is 0.97 Å for 359 equivalent Cα atoms (*37*). If the different elements of the secondary structure are taken in account (Figure 2), the rms deviation is approximately 0.35 Å for the central β-barrel and about 0.45 Å for the peripheral α-helices α_4, α_6, α_7, and α_8 and their interconnecting loops. There are, however, more significant differences (from 1.0 to 4.5 Å) for the remaining α-helices and their associated loops (Figure 3). These regions include helices α_1 and α_2 (residues 106−126), helix α_3 (residues 159−162), and helix α_5 (residues 238−246). The most important movements concern the two active site enclosing loops. Elements of both the N-terminal active site loop and the C-terminal active site loop have moved substantially (up to 4.5 Å) in response to substrate binding, discussed further, below.

Conformational Changes between Native and Complexed Structures. The conformational changes observed in the native and oligosaccharide-complexed forms of Cel6A involve residues involved both in crystal packing and in substrate binding. Helices α_1 and α_2 and their interconnecting loop (which was mobile in the native structure with *B* values in excess of 45 Å^2 (*14*)) have made a "rigid-body" movement of about 2.3 Å. This is due to crystal lattice contacts with both the $\alpha_3-\beta_{II}$ loop of one symmetry-related molecule and the α_4 helix from another, and this is also reflected in slight movements for all these structural elements. As would be expected, the A and B molecules show similar structural changes compared to the native, as would be expected since their lattice contacts are essentially the same due to the $C222_1$ pseudo-symmetry.

The Cel6A complex with glucose/cellotetraose reveals key conformational changes in the two active site enclosing loops (residues 174−196 and 397−435) which appear to relate to oligosaccharide binding and not crystal contacts. An overlap of the complex and native structures reveals a rigid-body movement of the C-terminal active site loop (residues 397−435). This movement is on an average of 1.1 Å up to 1.9 Å for Ala 413 Cα and displaces the loop toward the center of the substrate-binding tunnel. It does not, however, involve substantial main- or side-chain conformational torsion angle changes. This movement is connected to a similar movement of the $\beta_{VI}-\beta_{VII}$ loop (Thr 363 to Ile 375) by virtue of the disulfide bridge between Cys 372 and Cys 419. As a consequence, Trp 371 in the $\beta_{VI}-\beta_{VII}$ loop moves approximately 1 Å in order to optimize the stacking interaction with the glucose moiety in the +1 subsite of the enzyme. Also, residues 431−432 form the base for binding of the β-face of the pyranosyl ring in the −2 subsite. The largest movement is observed for residues 412−416, which form the lid of the active site tunnel. This results in solvent-mediated interaction between Arg 414 and the O-3 hydroxyl of the glucose in the −2 subsite. The most significant of the loop movements involves the more N-terminal of the two loops, in particular residues Ala 183 to Gly 188, (Figure 4). In addition to rigid-body movement of this loop into the active site tunnel with an average displacement of ap-

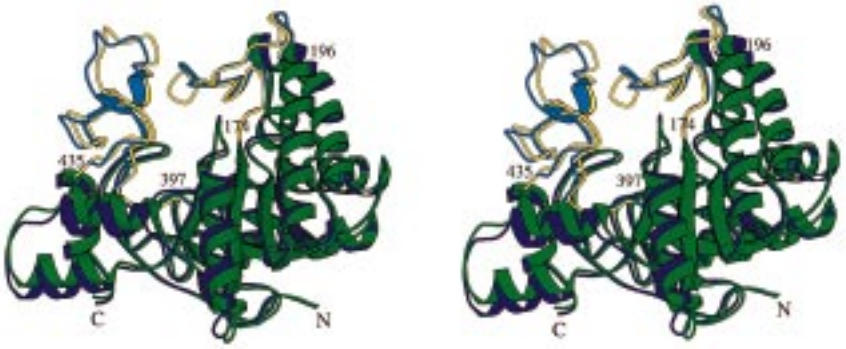

FIGURE 1: Stereographic representation of the overlay of the native and the cellobiose complex structures from *H. insolens* Cel6A, in divergent stereo. Native Cel6A is shown in blue and the oligosaccharide complex in green. The two active site loops that move in response to ligand binding are shown in light blue and yellow for the native and the complex, respectively. This figure was drawn with MOLSCRIPT (*56*).

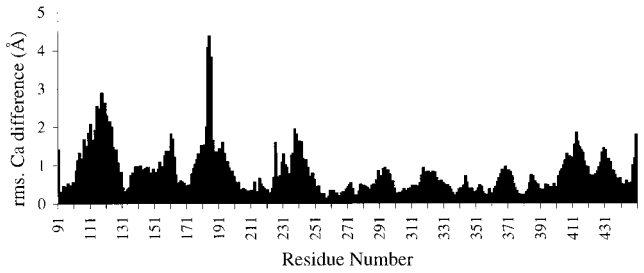

FIGURE 2: Plot of the residue-by-residue difference in Cα position between the native and oligosaccharide-complexed Cel6A.

proximately 1.3 Å, residues Asn 187, Ser 186, and Ala 185 also undergo main-chain conformational torsion angle changes. As before, the presence of a disulfide bond, this time Cys 181−Cys 240, results in a concerted movement of a second region, residues Asn 230 to Ser 245 from helix $\alpha_{5'}$ and the $\alpha_{5'}-\alpha_5$ loop, on an average of 1.5 Å. Together, these structural rearrangements both increase the contacts between the two active site loops and change the environment within the substrate-binding tunnel. The main-chain nitrogen of Ala 184 from the N-terminal loop now hydrogen-bonds to the side-chain OD2 of Asp 416. Asn 187 was inside the active site tunnel in the native structure where its OD1 was hydrogen-bonded to the NH1 of the B conformer of Arg 414 from the C-terminal loop. In the present complex, the side chain of Asn 187 has moved over 7 Å to the surface of the protein, where it is now exposed to solvent and interacts with the NH1 of Arg 414 through its main-chain carbonyl. Ser 186 was outside the tunnel in the native structure and hydrogen-bonded to Asp 180. Now, following main-chain conformational changes, it is involved in substrate binding in the −2 subsite and may also interact with the −1 subsite in a true substrate complex. The OG of this residue has moved almost 6 Å from outside to inside the tunnel. Structural changes in the main chain of Ala 185 place its main-chain carbonyl oxygen inside the tunnel, where it forms a weak hydrogen bond (distance 3.4 Å) with the NH1 of Arg 414. Taken together, these movements lead to a more enclosed conformation of the tunnel and optimization of the contacts between the protein and ligand. It is intriguing that, by virtue of the two disulfide bonds, the protein−ligand interactions in the −2 subsite are influenced by ligand-binding and resultant loop-closure events in the +1 and +2 subsites. Such a communication process, in which the binding energy of one subsite is influenced by ligand binding in another, distant, subsite may provide a partial explanation for the processivity of these enzymes.

Carbohydrate Binding. In the first "difference" electron density map, density for carbohydrate was evident throughout the substrate-binding site. A single glucose molecule is found in the −2 subsite (nomenclature according to ref *38*), (Figure 5). It refines with a temperature factor of approximately 12 Å^2 in both A and B molecules and certainly does not appear to be a disordered disaccharide. Glucose may result either from contamination in the commercial cellobiose preparation or by degradation of cellobiose. HPLC analysis of commercial cellobiose samples reveals significant contamination not only with glucose but also with cellotriose and cellotetraose (data not shown). The −2 subsite corresponds to the "A subsite" in the *T. reesei* CBH II, which is known to have a high-affinity site for glucose (*20*). Difference electron density was also clear for cellotriose in the +1, +2, and +3 subsites. During the refinement, it became clear that an extra unit could be modeled in the +4 subsite of the A molecule only. These oligosaccharides have, therefore, been modeled as cellotetraose and cellotriose in the A and B molecules, respectively. Again, it is not clear whether these longer species originate by condensation of cellobiose or, more likely, from the contaminating longer oligosaccharides revealed by HPLC analysis. In the B molecule, density for the fourth glucose unit is present but is too disordered to model appropriately. All glucose units present a $^4C_1$ chair conformation. They refine with average isotropic temperature factors for each glucose unit of 12.7, 15.9, 25.7, and 31 Å^2 for the +1, +2, +3, and +4 subsites in the A molecule and 15.1, 18.2, and 29.5 Å^2 for the +1, +2, and +3 subsites in the B molecule. Interactions with the protein appear most numerous in the +1 and +2 subsites (C and D subsite in the *Trichoderma* CBH II) but less numerous in the +3 and +4 subsite. Only the −2 to +2 subsites are enclosed by the two active site loops in the tunnel. These six subsites are consistent with kinetic measurements on this enzyme (*6, 39*). There is little apparent structural barrier to a −3 subsite, known to exist from kinetic studies (*23, 40, 41*), but perhaps small rearrangements of Arg 140 would be required to accommodate a sugar in this site.

Protein−Ligand Interactions. The Cel6A complex with glucose/cellotetraose permits description of the interactions from the −2 to +4 subsites (Figure 6). In the −2 subsite, the glucose ring binds by stacking with Trp 137. Direct

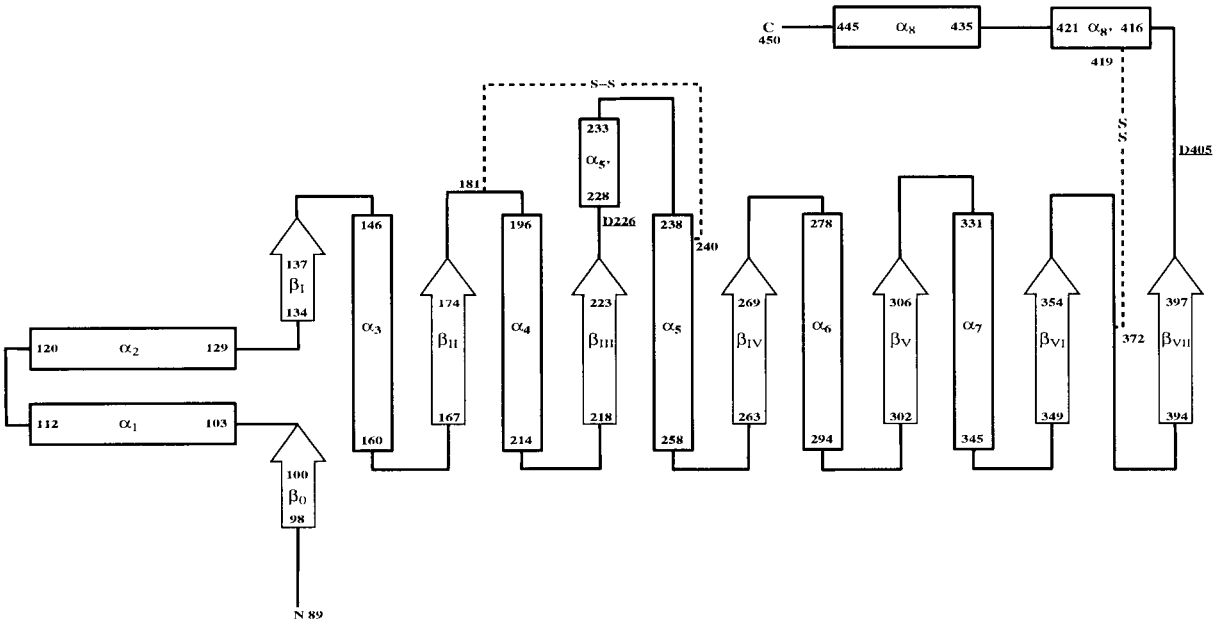

FIGURE 3: Topology diagram for Cel6A from *H. insolens*.

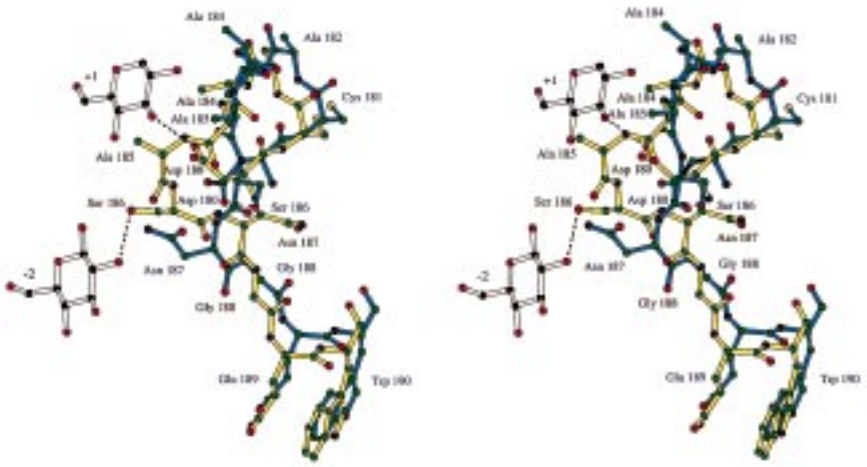

FIGURE 4: Ball and stick representation of the overlay of the N-terminal loop in the native (light blue) and cellobiose complex (yellow) structure of *H. insolens* Cel6A. Hydrogen bonds are defined with dash lines in the corresponding color. This figure was drawn with MOLSCRIPT (*56*).

hydrogen bonds are observed with the protein: O-6 with Lys 399 NZ and Glu 403 OE1, O-3 with Asp 139 OD2, and O-2 and O-1 with Ser 186 OG. The O-4 hydroxyl is hydrogen-bonded through two water molecules to Trp 137 NH1 and Tyr 104 OH. The O-2 and O-3 hydroxyls also make solvent-mediated interactions with the protein. In the +1 subsite stacking interactions are provided by Trp 371. The O-3, O-4, and O-6 hydroxyls H-bond directly to Asp 180 OD1 (2.8 Å), Asp 226 OD2 (2.6 Å), and Asn 310 ND2 (2.9 Å), respectively. O-3 is, additionally, hydrogen-bonded to Asp 180 OD2 through a solvent-mediated interaction. Contacts between the protein and O-2 are mediated by solvent. The hydrogen bond between Asp 226 OD2 and the O-4 hydroxyl is important—since this mimics the interaction between the catalytic acid and the glycosidic oxygen during the reaction.

In the +2 subsite, Trp 277 stacks the glucose ring. The O-6 hydroxyl is hydrogen-bonded to Asn 234 ND2 and also, via solvent, to Thr 233 OG and Asn 234 OD1. The O-3 hydroxyl makes a direct hydrogen bond to His 271 NE2 and makes water-mediated interactions with Asn 310 OD1 and the main-chain carbonyl of Trp 368. Very few contacts are observed in the +3 and +4 subsites. In the +3 subsite, O-2 is hydrogen-bonded to Asn 280 ND2 (3 Å) and O-3 is bonded through a water to Thr 233 OG and Asn 234 OD1. In subsite +4 there is only one solvent-mediated interaction with the protein from O-6 to Asn 280 OD1, and there is an aromatic stacking interaction with Trp 277, a residue implicated in the hydrolysis of crystalline cellulose (*42*). The −1 subsite is not occupied with sugar in this structure. It is surrounded by residues Tyr 174, Asp 180, Ser 186, Thr 309, Lys 399, and Asp 405. No stacking interactions appear possible in this subsite.

Active Site and Catalytic Mechanism. Cel6A catalyzes the hydrolysis of the β-1,4-glycosidic bonds in cellulose by what is widely assumed to be a single displacement mechanism, leading to the inversion of anomeric configuration [as described originally by Koshland (*43*)]. Cel6A is an e → a enzyme in the Sinnott nomenclature (*44*, *45*). It is generally believed that two catalytic carboxylate residues are involved

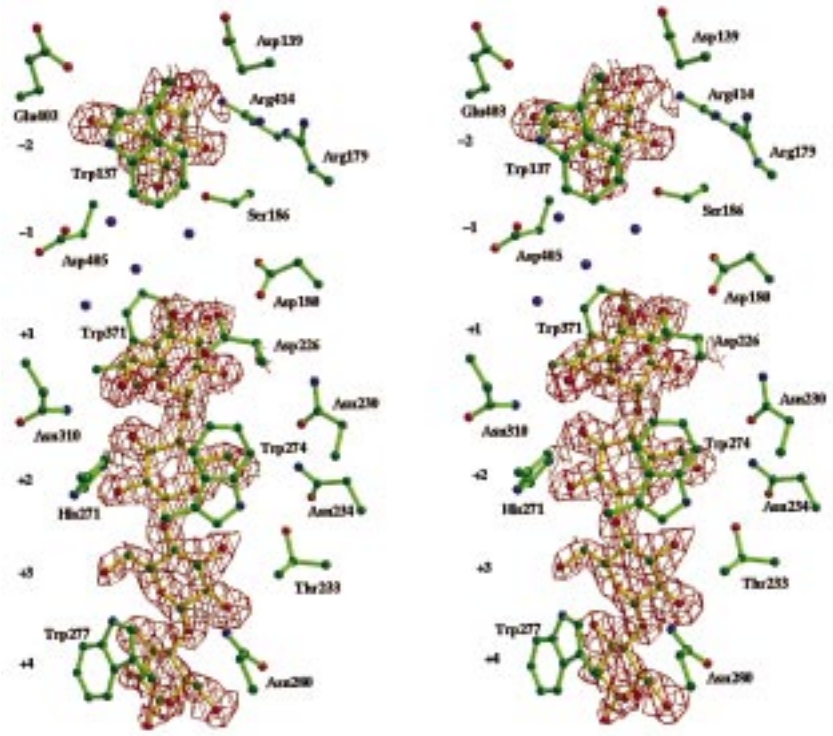

FIGURE 5: Divergent stereographic representation of the maximum-likelihood $2F_o - F_c$ density map, contoured at 0.4 e/Å^3 around the glucose and cellotetraose moiety observed in the Cel6A active site. Subsites and residues involved in the interactions with the substrate moieties are labeled.

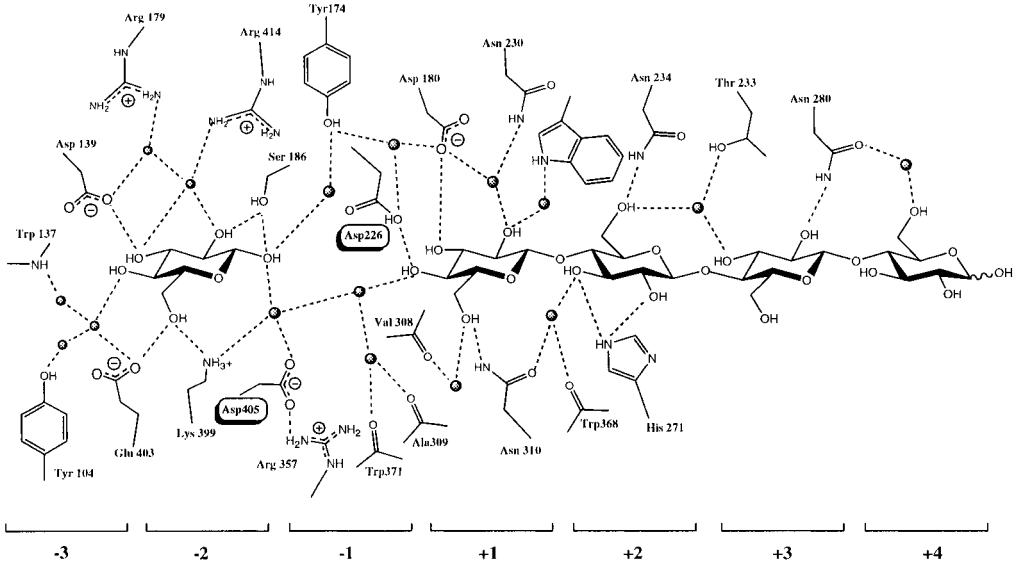

FIGURE 6: Schematic diagram of the enzyme−oligosaccharide hydrogen bond network in the Cel6A−cellobiose complex. Water molecules are shown as shaded spheres.

in such a single-displacement mechanism. One acts as an acid, to protonate the glycosidic bond and promote the departure of the leaving group, while the other functions as a base and deprotonates an incoming water molecule for nucleophilic attack at the anomeric carbon (Figure 7). In the case of CBH II, much controversy has surrounded the identification, indeed the existence, of the potential base function (reviewed briefly in ref *14*).

It is generally accepted that Asp 226 (or its equivalent) functions as the catalytic acid (*20*, *46*, *47*). This role is confirmed by the presence of the H-bond from this residue to the O-4 atom of the leaving group subsite sugar, shown here. The identification of the catalytic base is less certain. On the basis of the native enzyme structure, we supported proposals (*47*, *48*) that the catalytic base would be Asp 405. This role of this residue has been questioned in the *T. reesei* Cel6A structure on the basis of distance grounds (it was too distant from the catalytic acid) and because of its unusual environment (*20*). In the *H. insolens* Cel6A native structure, however, the double conformation of the catalytic acid, Asp 226, placed it at a distance of 9.5 Å from the Asp 405 OD1, consistent with the likely separation of acid and base in a inverting mechanism (*49*). Similarly, the position of the equivalent Asp 265 in the *Thermomonospora fusca* endo-

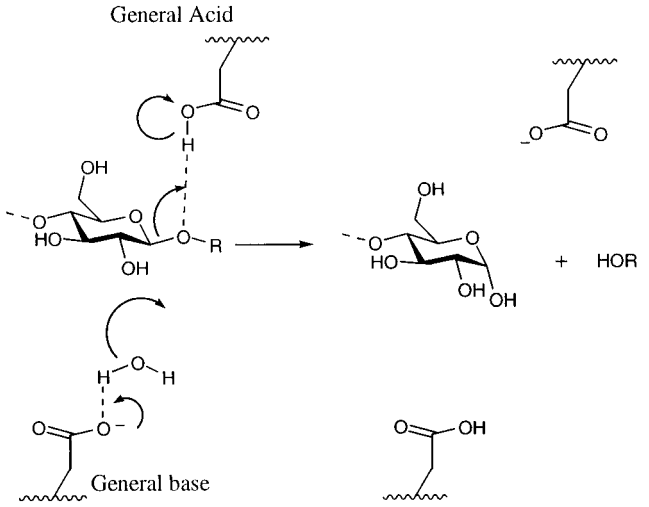

FIGURE 7: Canonical mechanism for an inverting (e → a) hydrolase.

glucanase E2 led to speculation of its role as the base (*46*). Nevertheless, it remains true that, in all these structures, this residue is found in an unusual environment where it makes a salt link with Arg 357.

In the present Cel6A complex structure the catalytic acid is found in just a single conformation, corresponding to that seen for the equivalent Asp 117 in the endoglucanase E2 structure. The OD2 carboxylate oxygen lies 8.8 and 10 Å distant from the OD1 and OD2 atoms, respectively, of Asp 405. Asp 226 has moved slightly in the complex structure in order to interact with the substrate in the +1 subsite. The main-chain oxygen and carbon have rotated about 65°, giving ψ and ϕ angles of −21.8° and 76.2°, respectively, instead of 45.6° and 55.5° in the native structure. This places Asp 226 in a generously allowed region of the Ramachandran plot. Both the conformation and environment of the proposed base, Asp 405, are unchanged between the native and complex structures. It remains salt-linked to Arg 357 and is just 0.5 Å closer to the acid in the complex structure (as a result of movement of the acid). Its stability had also been commented on in modeling studies on the related E2 structure (*50*). The absence of a sugar unit in the −1 subsite of Cel6A makes further speculation on the catalytic base difficult. Simple modeling of a cellooligosaccharide across the point of cleavage, however, indicates that Asp 405 both lies on the appropriate side of the subsite and is found at a suitable distance from C1, to activate a water for nucleophilic attack with inversion of configuration.

Further residues in the active site of family 6 have been implicated in catalysis. The role of Asp 268, or its equivalent, in the elevation of the pK_a of the catalytic acid has been elegantly demonstrated (*47, 48*). This buried residue undergoes no conformational change between native and complex structures. Its OD2 interacts with Asp 226 OD1 through a water molecule. A fourth "catalytic" acid, Asp 180, was postulated as also playing a role in pK_a elevation in the *T. reesei* Cel6A structure (*20*). This residue undergoes a large movement upon substrate binding. It moves some 110° around the Cα−Cβ bond in order to interact with the O-3 and O-2 hydroxyls of the glucose moiety in the +1 subsite and no longer interacts directly with the acid. This is consistent with mutagenesis results on related systems which have questioned its role in catalysis and pK_a elevation (*47*). Indeed, mutation of this residue to alanine in the related *T. reesei* CBH II resulted in only a 5-fold decrease in activity (*20*).

The structure of the *H. insolens* Cel6A oligosaccharide complex provides detailed mapping of the protein−ligand interactions through the −2 to +4 subsites, consistent with the six subsites known to contribute to catalysis (*6, 19, 39, 42*). The principal subsites, −2 to +2, are enclosed within the substrate-binding tunnel, while subsites +3 and +4 have been implicated in the threading of the cellulose chain into the active site and hence the catalytic activity on crystalline cellulose (*42*). The complex structure also reveals no structural barriers to a −3 or even a −4 subsite. This is consistent with subsite mapping and kinetic studies (*40, 41*); indeed, it is essential to explain the cleavage patterns of chain-end-modified substrates such as fluorescenyl-derived compounds (*23*). In the absence of a sugar unit in the −1 subsite, one can only speculate about the conformation of the sugar unit. No stacking interactions seem to be present, as is often the case for the −1 subsite and this may reflect the need for pyranose ring distortion during catalysis (as observed in refs *51*−*53*). Indeed, modeling of the *T. reesei* CBH II structure has indicated a role for ring distortion during catalysis (*19, 54*). Preliminary modeling appears to confirm that Asp 405 lies in an appropriate position, below the pyranoside ring, to function as a catalytic base. Flexibility in the active site enclosing loops is manifested in main-chain movements of up to 4.5 Å upon complexation with oligosaccharide. The dynamic nature of these loops has long been invoked in order to account for the many anomalous properties of these enzymes (*21, 23*−*25, 55*). The Cel6A complex throws little light on potential loop-opening events, but the capability for loop movement in response to substrate binding in this family of enzymes has been clearly demonstrated.

ACKNOWLEDGMENT

The authors thank Dr. Simon Charnock for assistance with the Dionex HPLC analysis of oligosaccharide samples and Dr. Johan Turkenburg for assistance during data collection.

REFERENCES

1. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) *Curr. Opin. Struct. Biol. 8*, 548−557.
2. Schülein, M., Kauppinen, M., Lange, L., Lassen, S., Andersen, L., Klysner, S., and Nielsen, J. (1998) in *ACS Symposium Series 687: Enzyme Applications in Fiber Processing*, pp 66−74, American Chemical Society, Waschington, DC.
3. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1991) *Microbiol. Rev. 55*, 303−315.
4. Bayer, E. A., Morag, E., and Lamet, R. (1994) *Trends Biotechnol. 12*, 379−386.
5. Wood, T. M. (1992) *Biochem. Soc. Trans. 20*, 46−53.
6. Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B., and Schülein, M. (1993) *Eur. J. Biochem 217*, 947−953.
7. Teeri, T. T., and Koivula, A. (1995) *Carbohydr. Eur. 12*, 28−33.
8. Henrissat, B. (1991) *Biochem. J. 280*, 309−316.
9. Henrissat, B., and Bairoch, A. (1993) *Biochem. J. 293*, 781−788.
10. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
11. Henrissat, B., and Davies, G. J. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.

12. Henrissat, B., Teeri, T., and Warren, R. (1998) *FEBS Lett. 425*, 352−354.
13. Schülein, M. (1997) *J. Biotechnol. 57*, 71−81.
14. Varrot, A., Hastrup, S., Schülein, M., and Davies, G. J. (1999) *Biochem. J. 337*, 297−304.
15. Mackenzie, L. F., Sulzenbacher, G., Divne, C., Jones, T. A., Wöldike, H., Schülein, M., Withers, S. G., and Davies, G. J. (1998) *Biochem. J. 335*, 409−416.
16. Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G., and Schülein, M. (1993) *Nature 364*, 362−364.
17. Henrissat, B., Driguez, H., Viet, C., and Schülein, M. (1985) *Bio/Technology 3*, 722−726.
18. Chanzy, H., and Henrissat, B. (1985) *FEBS Lett. 184*, 285−288.
19. Teeri, T. T. (1997) *Trends Biotechnol. 15*, 160−167.
20. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K., and Jones, T. A. (1990) *Science 249*, 380−386.
21. Ståhlberg, J., Johansson, G., and Pettersson, G. (1993) *Biochim. Biophys. Acta 1157*, 107−113.
22. Henrissat, B. (1998) *Cellulose Commun. 5*, 84−90.
23. 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.
24. Armand, S., Drouillard, S., Schülein, M., Henrissat, B., and Driguez, H. (1997) *J. Biol. Chem. 272*, 2709−2713.
25. Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E., and Kanda, T. (1996) *J. Biochem. 120*, 1123−1129.
26. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
27. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
28. Navaza, J., and Saludijan, P. (1997) *Methods Enzymol. 276*, 581−594.
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
30. Brünger, A. T. (1992) *Nature 355*, 472−475.
31. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr D49*, 148−157.
32. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D49*, 129−147.
33. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. T., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
34. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
35. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95−99.
36. Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 282−291.
37. Kleywegt, G. J., and Jones, T. A. (1994) *ESF/CCP4 Newsl. 31*, 9−14.
38. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J. 321*, 557−559.
39. Schülein, M., Tikhomirov, D. F., and Schou, C. (1993) in *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 109−116, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.
40. Ruohonen, L., Koivula, A., Reinikainen, T., Valkejärvi, A., Teleman, A., Claeyssens, M., Szardenings, M., Jones, T. A., and Teeri, T. T. (1993) in *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 87−96, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.
41. Biely, P., Vrsanska, M., and Claeyssens, M. (1993) in *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 99−108, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.
42. Koivula, A., Kinnari, T., Harjunpaa, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T. A., and Teeri, T. T. (1998) *FEBS Lett. 429*, 341−346.
43. Koshland, D. E. (1953) *Biol. Rev. 28*, 416−436.
44. Sinnott, M. L. (1987) in *Enzyme Mechanisms* (Page, M. I., and Williams, A., Eds.) pp 259−297, Royal Society of Chemistry, London.
45. Davies, G., Sinnott, M. L., and Withers, S. G. (1997) in *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.) pp 119−209, Academic Press, London.
46. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) *Biochemistry 32*, 9906−9916.
47. Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1995) *Biochemistry 34*, 2220−2224.
48. Damude, H. G., Ferro, V., Withers, S. G., and Warren, R. A. J. (1996) *Biochem. J. 315*, 467−472.
49. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
50. Taylor, J., Teo, B., Wilson, D., and Brady, J. (1995) *Protein Eng. 8*, 1145−1152.
51. Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G. J. (1996) *Biochemistry 35*, 15280−15287.
52. Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996) *Nat. Struct. Biol. 3*, 638−648.
53. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707−11713.
54. Teeri, T. T., Koivula, A., Linder, M., Wohlfahrt, G., Divne, C., and Jones, T. A. (1998) *Biochem. Soc. Trans. 26*, 173−178.
55. Fägerstam, L. G., and Pettersson, L. G. (1980) *FEBS Lett. 119*, 97−100.
56. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
57. Raymond, S., Henrissat, B., Qui, D. T., Kvick, A., and Chanzy, H. (1995) *Carbohydr. Res. 277*, 209−229.
58. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. A. (1998) *J. Mol. Biol. 275*, 309−325.

BI9903998