The Crystal Structure of a 2-Fluorocellotriosyl Complex of the *Streptomyces lividans* Endoglucanase CelB2 at 1.2 Å Resolution[†,‡]

Gerlind Sulzenbacher,[§,||] Lloyd F. Mackenzie,[⊥] Keith S. Wilson,[§] Stephen G. Withers,[⊥] Claude Dupont,[#] and Gideon J. Davies*[,§]

Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, Y010 5DD, U.K., European Molecular Biology Laboratory (EMBL), DESY, Notkestrasse 85, 22603 Hamburg, Germany, Department of Chemistry, University of British Columbia, Vancouver, V6T 1Z1, Canada, and Centre de Microbiologie et Biotechnologie, INRS−Institut Armand-Frappier, Laval-des-Rapides, C.P. 100, Québec, Canada, H7N4Z3

Received November 6, 1998; Revised Manuscript Received January 19, 1999

ABSTRACT: Glycoside hydrolases have been classified into over 66 families on the basis of amino acid sequence. Recently a number of these families have been grouped into "clans" which share a common fold and catalytic mechanism [Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696]. Glycoside hydrolase Clan GH-C groups family 11 xylanases and family 12 cellulases, which share the same jellyroll topology, with two predominantly antiparallel β-sheets forming a long substrate-binding cleft, and act with net retention of anomeric configuration. Here we present the three-dimensional structure of a family 12 endoglucanase, *Streptomyces lividans* CelB2, in complex with a 2-deoxy-2-fluorocellotrioside. Atomic resolution (1.2 Å) data allow clear identification of two distinct species in the crystal. One is the glycosyl-enzyme intermediate, with the mechanism-based inhibitor covalently linked to the nucleophile Glu 120, and the other a complex with the reaction product, 2-deoxy-2-fluoro-β-D-cellotriose. The active site architecture of the complex provides insight into the double-displacement mechanism of retaining glycoside hydrolases and also sheds light on the basis of the differences in specificity between family 12 cellulases and family 11 xylanases.

Cellulose is the most abundant polymer of the biosphere, and the principles which govern its enzymatic breakdown are consequently of great fundamental and industrial interest. Since the realization that our planet does not possess infinite resources, the idea that cellulose may be utilized both as a source of renewable energy and as a foodstuff has become attractive for governmental bodies (*1*). Enzymes produced by cellulolytic organisms are obvious candidates for these roles. Superficially, cellulose itself is a very simple polymer of glucose units joined by β-1,4-glycosidic linkages. Extensive intra- and intermolecular hydrogen bonds confer a high degree of structural order to this plant cell wall polymer, rendering it extremely resistant to enzymatic attack. Xylan, the major component of hemicellulose, is structurally less ordered than cellulose. The absence of the C-5 hydroxymethyl group in xylose substantially decreases the number of stabilizing intra- and intermolecular hydrogen bonds and confers a more hydrophobic character on xylan compared to cellulose. Furthermore, random substitutions of acetyl, arabinofuranosyl, and glucuronyl moieties at the O-2 and O-3 positions render the xylan polymer structurally and chemically more heterogeneous than cellulose (*2*−*4*).

Glycoside hydrolases have, thus far, been classified into over 66 sequence-based families (*5*). At least 10 of these families contain cellulases (5−9, 12, 26, 44, 45, 48, and 61), and two families, 10 and 11, contain xylanases. Some glycoside hydrolase families share the same overall fold, active site architecture, and catalytic mechanism and have been grouped into clans which may contain enzymes hydrolyzing a variety of different substrates (*5*−*7*). Only two enzyme families have been assigned to clan C, to date: family 11 xylanases and family 12 cellulases. Whereas there is a large body of information on the 3-D structures and mechanisms of family 11 xylanases (for example, references *8*−*11*), little is known about the nature of substrate binding to family 12 enzymes.

GH-C clan enzymes perform catalysis with net retention of anomeric configuration. The mechanism is widely believed to be a double-displacement in which a covalent glycosyl-enzyme intermediate is formed and subsequently hydrolyzed, with acid/base assistance, via oxocarbenium ion transition states (*12*, *13*). A reliable method for trapping the covalent glycosyl-enzyme intermediate is through the use of 2-deoxy-2-fluoro-substituted sugars (*14*). The fluorine at C-2 destabilizes the oxocarbenium ion-like transition states both by the loss of crucial stabilizing interactions and through its inductive effect. The rates of both formation and breakdown

[†] This work was funded, in part, by the Biotechnology and Biological Sciences Research Council, The European Union (Contracts BIO4-CT97-2303 and BIO4-CT97-5035), The Protein Engineering Centre of Excellence of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Universities of York and British Columbia. G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates and observed structure factor amplitudes for the structure described in this paper have been deposited with the Protein Data Bank (accession numbers 2NLR and 2NLR-SF).

* To whom correspondence should be addressed. Telephone: 44-1904-432596. Fax: 44-1904 410519. E-mail: davies@yorvic.york.ac.uk.

[§] University of York.

[||] EMBL Hamburg.

[⊥] University of British Columbia.

[#] Institut Armand-Frappier.

10.1021/bi982648i CCC: $18.00
Published on Web 03/25/1999

of the intermediate are thus reduced. The retardation of the rate of formation of the intermediate can, however, be overcome by the use of a good leaving group such as 2,4-dinitrophenol which renders the intermediate kinetically accessible. Such reagents have found widespread use in the structural analysis of retaining glycoside hydrolases (*15*−*18*).

We recently determined the native structure of the catalytic core domain of the family 12 endoglucanase CelB2[1] from *Streptomyces lividans* (*19*). The overall fold of this cellulase and its active site topology were remarkably similar to those observed for the xylanases from family 11, as had been predicted by hydrophobic cluster analysis (*20*). 2-Fluoro-substituted sugars have recently been used to trap the covalent glycosyl-enzyme intermediate for CelB and confirmed the assignment of Glu 120 as the catalytic nucleophile (*21*). Here we describe the crystal structure of the catalytic domain of CelB2 in complex with a 2-deoxy-2-fluorocellotriosyl species. Two distinct species are seen in the crystal, and atomic resolution X-ray data permit identification and analysis of both species. One is the trapped covalent glycosyl-enzyme intermediate, with the proximal sugar in a $^4C_1$ conformation through an α-linkage to the nucleophile, Glu 120. The other species is the hydrolysis product of the glycosyl-enzyme intermediate, 2-deoxy-2-fluorocellotriose. These structures, in addition to confirming the nucleophile as Glu 120, allow a mapping of the noncovalent interactions in the −3 to −1 subsites and give some insight into the specificity of clan GH-C enzymes.

MATERIALS AND METHODS

Crystallization and Data Collection. The preparation of the catalytic core of CelB2 has been described previously (*19*). To obtain a complex with 2F-cellotriose, a 10 mM protein solution was incubated for 5 min with a solution containing a stoichiometric quantity of DNP-2F-cellotrioside (*22*), buffered at pH 7.0 with 10 mM TRIS-HCl. Crystals were grown under the same conditions as used for the native enzyme at pH 4.5. X-ray data to 1.2 Å resolution, Table 1, were collected on the EMBL X-11 beam line ($\lambda = 0.9107$ Å) at the DORIS storage ring in Hamburg from a single crystal in a stream of N_2 gas at 100 K. A 30 cm Mar-Research Imaging Plate Scanner was used as detector. Data processing and reduction were carried out with the HKL program suite (*23*, *24*), and all further computing used the CCP4 suite of programs (*25*) unless otherwise stated.

Structure Solution and Refinement. Since it transpired that the complex of CelB2 was not isomorphous with the native crystals, molecular replacement was carried out with the program AMoRe (*26*, *27*) using the native structure of CelB2 as a search model. Of the reflection data, 5% were set aside for cross-validation purposes (*28*), and the same test set of free reflections was used during all subsequent stages of refinement. Simulated annealing was performed with XPLOR (*29*, *30*) with a starting temperature of 3000 K and a step size of 25 K, followed by grouped *B*-factor refinement. Further refinement was carried out with the maximum likelihood method as implemented in the program REFMAC (*31*) making use of low-resolution bulk solvent corrections and anisotropic F_o *vs* F_c scaling (*32*). Maximum likelihood refinement was coupled with cycles of manual model rebuilding with the program O (33). The solvent model was automatically updated with the program ARP (*34*) and carefully inspected by eye.

Once isotropic refinement with REFMAC had converged, anisotropic treatment of the individual atomic displacement parameters was introduced with the program SHELX-97 (*35*). The free-*R* factor was monitored to calibrate restraints and weights and to establish whether this approach was justified. "Rigid-bond" restraint and "similarity restraints" were applied to the atomic displacement parameters of protein atoms, and an approximately isotropic behavior was imposed on solvent atoms. The high-quality electron density maps resulting from restrained anisotropic refinement allowed building of discrete disorder and the location of missing atoms. The occupation factors of equivalent components of disordered groups were made complementary by the use of free variables. Further waters were added to the model based on strong peaks in the difference Fourier, and diffuse solvent was modeled according to Babinet's principle as implemented in SHELX-97 (*35*). "Antibumping restraints", applied to all possible symmetry equivalents, prevented water molecules from diffusing into protein regions.

At this point the trisaccharide was included into the model and treated isotropically for a few cycles, prior to anisotropic refinement. Bonding distances for the ligand were restrained to target values derived from the crystal structure of methyl-β-D-cellotrioside (*36*), but no restraint was imposed on the glycosidic linkage between enzyme and saccharide. The atomic displacement parameters of the equivalent components of the disordered sugar in the −1 subsite were restrained to be similar. Hydrogen atom positions were generated automatically by adopting geometric criteria and refined using a "riding" model. The isotropic atomic displacement parameters for hydrogen atoms were made dependent on the equivalent isotropic displacement parameter of the atoms to which they were attached. A number of solvent molecules were refined with occupancy of 0.5, or with their occupancy coupled in complementary fashion to disordered parts of the protein. A last conjugate gradient least-squares refinement cycle was carried out against all data, including the cross-validation data set. To obtain the estimated standard deviations in the atomic parameters, a final cycle of overlapped block-matrix (BM) least-squares refinement was carried out, in which the restraints were turned off.

RESULTS

Endoglucanase CelB from *S. lividans* consists of a compact catalytic domain which is connected via a flexible linker to a family IIa cellulose-binding domain (CBD). A catalytically competent form missing the linker and the CBD was constructed and overexpressed in a *S. lividans* expression system and the native structure described (*19*). The overall fold of endoglucanase CelB2, typical for clan GH-C enzymes, is a half-opened "right-hand", with two predominantly antiparallel β-sheets twisted around a deep depression. An embedded single α-helix and an α-helical turn make up part of the "palm". A long loop crossing the substrate-binding groove and terminating it at the reducing end, known as the "cord", is a feature of all family 11 and family 12 structures.

[1] Abbreviations: CelB2, catalytic core domain of the family 12 endoglucanase CelB from *Streptomyces lividans*; DNP-2F-cellotrioside, 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-cellotrioside; CBD, cellulose-binding domain.

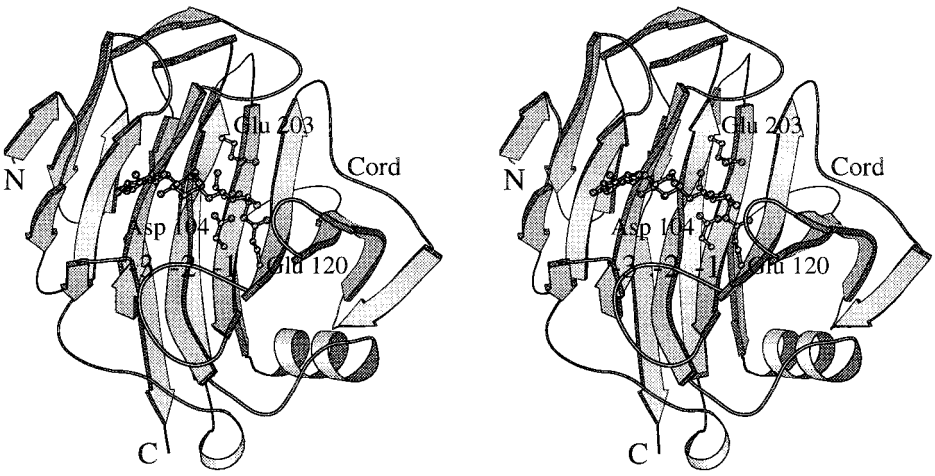

FIGURE 1: Schematic diagram, in divergent stereo, of the catalytic core domain of *Streptomyces lividans* CelB2 in complex with a 2-F-cellotrioside. The catalytic acid/base, the nucleophile, and the covalent cellotriosyl-enzyme intermediate are shown in ball-and-stick representation. This figure was drawn with MOLSCRIPT (*52*).

A number of residues located in this loop, amongst them Pro 133, are conserved throughout clan C members. The function of this loop remains unknown, but speculations pointing toward a possible loop movement upon substrate binding have been put forward (*10*). There is a 35 Å long substrate-binding cleft. It is lined with aromatic and hydrophilic residues, as commonly observed in sugar-binding proteins. Approximately 20 Å from the nonreducing end of the cleft the two invariant catalytic residues Glu 120 and Glu 203 point into the cleft from opposite sides (Figure 1).

Native CelB2 crystallized in such a way that the "cord" region of one molecule filled the substrate-binding cleft of a symmetry-related molecule, thus precluding conventional soaking methods. Additionally, by analogy with other systems (*18*, *37*, *38*), it was considered unlikely that the covalent intermediate could be formed at low pH where the nucleophile would become protonated. To obtain a complex of CelB2, the protein was first reacted in solution, at higher pH, with the mechanism-based inhibitor DNP-2F-cellotrioside. The reaction was monitored by the appearance of the distinctive yellow 2,4-dinitrophenolate ion, and, upon completion, crystallization was performed as for the native protein. Crystals of the CelB2 complex, displaying a different morphology to the native crystals, appeared after 3 months and reached a maximum size of 0.2 × 0.2 × 0.2 mm. This turned out to be a new crystal form belonging to space group $P3_121$ with cell dimensions of $a = b = 65.96$ Å, $c = 88.74$ Å, $\alpha = \beta = 90°$, and $\gamma = 120°$. There is one molecule of CelB2 in the asymmetric unit with a solvent content of approximately 47% (*39*). As for the native enzyme, there are several tight intermolecular contacts between side-chain carboxylate groups of symmetry-related molecules, explaining the requirement of low pH levels for crystallization. A total of 415 327 observations were merged to give 69 645 unique reflections, leading to an overall mean $\langle I/\sigma I\rangle$, multiplicity of observations, and R_{merge} for symmetry-equivalent reflections of 14.9, 5.9, and 0.054, respectively. The overall completeness is 99.1%, with 100% for the 25.0−2.58 Å low-resolution shell, indicating that all the strong terms have been measured for that resolution range. The overall temperature factor as estimated from Wilson statistics is 10.9 Å^2. An overview of data quality and refinement statistics is given in Table 1.

Table 1: Data Quality, and Structure Refinement Statistics for the 2F-Cellotriose Complex of *Streptomyces lividans* CelB2

resolution of data (outer shell) (Å)	25.0−1.20 (1.22−1.20)
R_{merge}[a] (outer shell)	0.054 (0.408)
$\langle I/\sigma I\rangle$ (outer shell)	14.93 (3.74)
completeness (outer shell) (%)	99.1 (97.2)
multiplicity (outer shell)	5.96 (4.49)
resolution used in refinement (Å)	25.0−1.20
no. of protein atoms	1663
no. of water molecules (B < 50 Å^2)	270
discretely disordered residues	32
no. of ligand atoms	34
R_{cryst}[b]	11.12 (19.1)
R_{free}	14.23 (22.0)
R_{cryst} for all data	11.19 (19.0)
mean protein atoms B (Å^2)	14.27
mean main-chain atoms B (Å^2)	12.74
mean side chains B (Å^2)	16.43
mean solvent atoms B (Å^2)	28.75
mean ligand atoms B (Å^2)	13.19 (subsite −3, 16.25; −2, 14.40; −1, 11.14)
B-factor from Wilson statistics (Å^2)	10.96
rmds 1−2 bonds (Å)	0.017
rmds 1−3 bonds (Å)	0.035
rmds chiral volumes (Å^3)	0.082

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkli} - \langle I_{hkli}\rangle|/\sum_{hkl}\sum_i\langle I_{hkli}\rangle$. [b] $R_{cryst} = \sum||F_o| - |F_c||/\sum|F_o|$.

The structure was solved in a straightforward manner by molecular replacement using the native CelB2 structure as search model. Refinement with XPLOR and later with REFMAC proceeded smoothly. σ_A-weighted $F_o - F_c$ and $2F_o - F_c$ electron density maps were of high quality and allowed unambiguous identification of most protein atoms. Some difficulties were encountered in the location of residues Asp 95 and Gly 96, which are situated close to a 2-fold axis and exhibit complementary double conformations. Clear difference electron density could be observed in the active site region, and at a contour level of 0.53 electron/Å^3, most of the ligand atoms were visible and well resolved, indicating that atomic resolution had been achieved. Anisotropic F_o vs F_c scaling in REFMAC was statistically justified since R_{free} as well as R_{cryst} fell by 2% (*32*). After convergence of maximum likelihood refinement and the addition of 218 water molecules to the model, the R_{cryst} and R_{free} for all data between 25 and 1.2 Å were 19.6% and 21.3%, respectively. As these values are rather high and because of the occurrence of appreciable $F_o - F_c$ electron density peaks on almost every

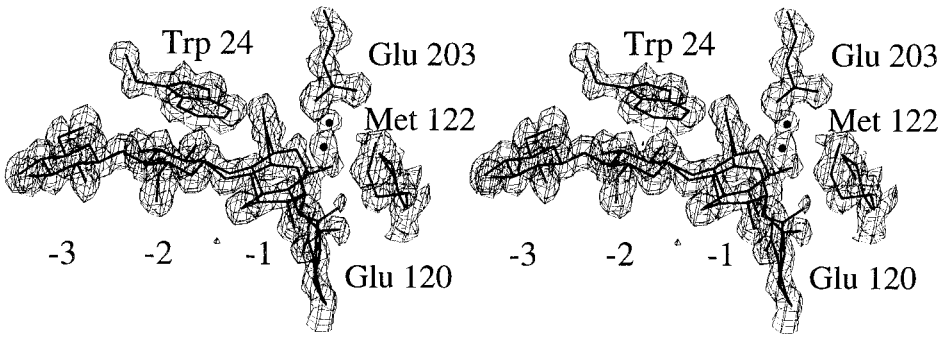

FIGURE 2: Active-site region of CelB2 in complex with 2-F-cellotriose, shown in divergent stereo. The map shown is a $2F_o - F_c$ synthesis, contoured at 0.5 electron/Å^3. Residues mentioned in the text are indicated. This figure was prepared using O (*33*).

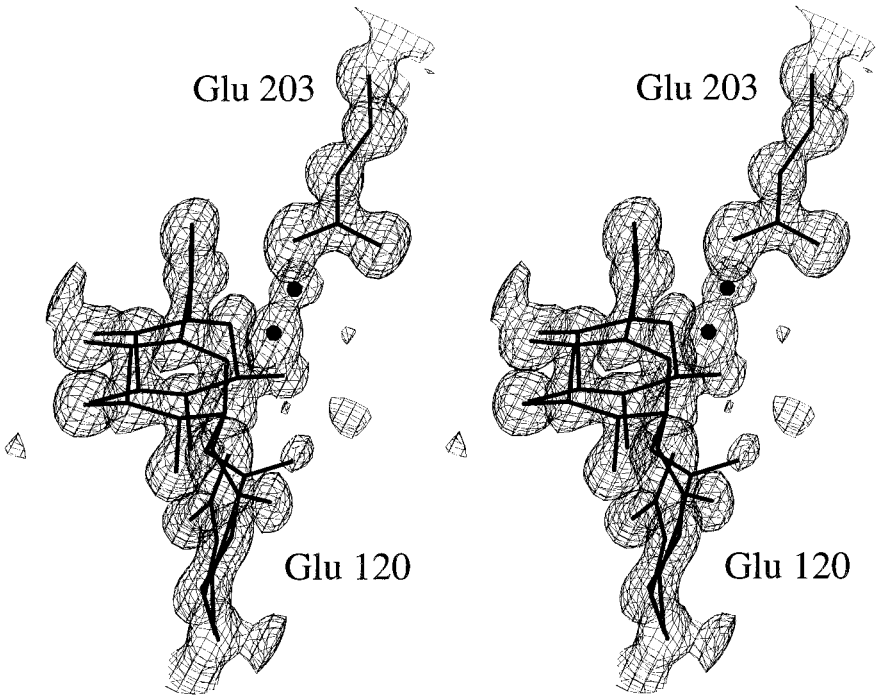

FIGURE 3: Divergent stereo closeup of the −1 subsite disorder. The two pyranoside species are shown, together with the multiple conformations observed for Glu 120, Met 122, and a single water molecule. The contour level is 0.5 electron/Å^3. The figure is a $2F_o - F_c$ synthesis and was prepared using O (*33*).

atom, an anisotropic description of the atomic displacements was deemed necessary. The drop in both R_{cryst} and R_{free} by 5% and 4%, respectively, unambiguously demonstrated that the anisotropic refinement was meaningful. The better phase estimates resulting from the restrained anisotropic refinement improved the map quality substantially, and modeling of disorder and the location of missing atoms became straightforward. Introduction of the ligand, postponed up to this point in order to avoid model bias, was carried out very cautiously, and extreme care was taken of the static disorder observed in the −1 subsite. Although a covalently bound glycosyl-enzyme intermediate was expected, as observed in similar studies on other glycoside hydrolases (*15*−*18*), we were careful to consider the possibility of other stable enzyme−oligosaccharide species.

Identification of the Bound Species; Quality of the Final Model Structure. The difference electron density was excellent for the sugar moieties in the −3 and −2 subsites and was well-defined for subsite −1, so that all three sugar rings could be unambiguously positioned (Figure 2). Static disorder manifested itself near the catalytic center, where the electron density at the O-5 position was elongated in a vertical fashion with respect to the pyranose plane. The electron density corresponding to the anomeric carbon was split, with a consistent portion being continuous with the electron density of the nucleophile Glu 120, which in turn exhibited discrete disorder. The other part of the electron density for the anomeric carbon extended beyond the scissile point in a vertical fashion and abruptly terminated at about 3 Å with a few well-resolved electron density peaks at the aglycon (+1) subsite. The sum of these observations led us to conclude that two species were present in the crystal (Figure 3). One species was the covalently bound glycosyl-enzyme intermediate, with the pyranose ring in the −1 subsite in a standard $^4C_1$ chair conformation. The other species was most likely either the uncleaved inhibitor [as observed in a related study on a family 5 endoglucanase from *Bacillus agaradhaerens* (*18*)] or the product, 2F-cellotriose. The former possibility was carefully examined but finally discarded since the atomic resolution electron density in the −1 subsite could only be modeled by pyranoside in a $^4C_1$ conformation. An unhydrolyzed substrate complex would have required modeling of the sugar in a more distorted conformation in order to avoid steric clashes of the aglycon with the protein. In addition,

the well-resolved peaks in the +1 subsite could be modeled neither as a planar dinitrophenyl group nor as a pyranoside and clearly corresponded to solvent water molecules. The second species in the −3, −2, and −1 subsites was easily modeled as 2F-cellotriose, the reaction product.

After modeling the disordered parts of the ligand, the protein and the solvent, and location of hydrogen atoms, refinement converged to an *R*-factor of 11.12% and an R_{free} of 14.23%. Final refinement against all 69 645 reflections, including those previously set aside for cross-validation purposes, resulted in an *R*-factor of 11.19%, and the largest peak and hole in the difference Fourier synthesis are 0.35 and −0.30 e Å^{-3}, respectively. The final model for the complex of the catalytic domain of endoglucanase CelB includes 1663 protein atoms, 34 ligand atoms, 216 water atoms with full occupancy, and 54 water atoms with half-occupancy, or with the occupation factor coupled to disordered protein. The average conventional *B*-factors are 12.7 Å^2 for main-chain atoms, 16.4 Å^2 for side-chain atoms, 13.2 Å^2 for ligand atoms (16.3 Å^2 for subsite −3, 14.4 Å^2 for subsite −2, 11.1 Å^2, for subsite −1), and 28.8 Å^2 for solvent atoms. Out of 222 residues, 32 have alternate conformations, and most are involved either in ligand binding or in intermolecular crystal contacts. Met 122, located near the catalytic center, was present in oxidized form in the native crystal structure. This was not the case for the complex structure, indicating that the presence of the ligand protected this group from oxidation. Among the non-glycine and non-proline residues, 91.2% have (φ/ψ) values in the "most favored regions" of the Ramachandran plot, 7.7% in the "additionally-allowed" regions, and none in the "disallowed regions" as defined by PROCHECK (*40*). Two residues are located in the "generously allowed region", and these are the mobile Val 98 and the well-ordered Thr 67. The overall root-mean-square deviations from ideality are 0.017 and 0.035 Å for 1−2 and 1−3 bonds, respectively. Coordinates and structure factor amplitudes have been deposited with the Protein Data Bank (*41*).

CelB2−Trisaccharide Interactions: −3 and −2 Subsites. Atomic resolution data allowed dissection of two distinct species in the crystal: the glycosyl-enzyme intermediate and the reaction product. For both moieties, a trisaccharide occupies the −3, −2, and −1 subsites with the positions and interactions of the −2 and −3 subsite sugars identical for both (Figure 4). The differences in the −1 subsite are discussed below. The overall structure of CelB2 undergoes small conformational changes upon binding of 2-fluorocellotriose, reflected in an rms difference of 0.42 Å for all 222 C_α atoms (Figure 5). In addition to a number of changes associated with the different crystal-packing environments, there are systematic movements of the loops flanking the nonreducing end of the active site crevice. Shifts of approximately 0.7 Å for the main-chain atoms in loops involved in ligand binding have the effect of narrowing the rim, whereas no appreciable main-chain structural changes could be detected for the residues making up the floor of the cleft. There are movements of up to 2 Å for some aromatic side chains involved in stacking interactions with the pyranose rings. The biggest main-chain shift involving active site residues is for the loop connecting β-strands β7 and β8, which was partially disordered in the native structure (temperature factors ∼ 40−50 Å^2). This loop, adjacent to the catalytic center, is well-defined in the complex and carries Asn 155 and Asn 158, two residues in intimate contact with the fluorine at the C-2 position and the nucleophile Glu 120, respectively. From the native structure it was not clear which oxygen atoms of the Brønsted acid/base Glu 203 and the nucleophile Glu 120 participated directly in catalysis since they were separated by an average of 7 Å, somewhat longer than that typically observed (*42*). A conformational change upon substrate binding was therefore expected to take place. It is clear from the complex structure that a rearrangement of the nucleophile, Glu 120, has indeed occurred, placing it just 5.8 Å from the acid/base in the complex structure.

The oligosaccharide maintains its linear conformation. The intramolecular hydrogen bonds between O-3(i) and O-5(i+1), as observed in the crystal structure of methyl β-cellotrioside (*36*), are preserved, and all the glucopyranosyl rings are found in the full $^4C_1$ chair conformation. The sugar ring of the −3 subsite is sandwiched between the aromatic rings of Phe 8 and Tyr 66, both of which have undergone conformational changes of over 2 Å upon oligosaccharide binding. The O-6 hydroxyl group hydrogen bonds to OD1 of the invariant Asn 22. All the other sugar hydroxyls in subsite −3 establish water-mediated hydrogen bonds with polar residues either in the binding cleft or originating from symmetry-related molecules. The number of direct hydrogen-bonding interactions is greater for subsite −2. The O-2 hydroxyl group interacts with Asn 22 ND2 and a well-ordered water molecule (*B*-factor 10.5 Å^2) which seems crucial for the mediation of substrate−enzyme interactions. It is conserved with respect to the native structure and sequestered between the polar side chains of the invariant residues Asp 104, Trp 106, and Gln 199. The O-3 hydroxyl of subsite −2 accepts a hydrogen bond from His 65, and the O-6 hydroxyl is part of an extensive hydrogen bonding network involving several water molecules as well as the side chains of Tyr 66 and Ser 157. An additional stabilization of the pyranosyl group in −2 derives from a stacking interaction with the aromatic ring of the invariant Trp 24.

−1 Subsite Interactions in Product and Intermediate Complexes. The two trisaccharide species in the crystal manifest themselves in two different conformations for the sugar in the −1 subsite (Figure 3). This is paralleled by multiple conformations for the nucleophile and for two residues in intimate contact with it, Asp 104 and Met 122. The $^4C_1$ chair planes (C-2, C-3, C-5, O-5) for the two species are related by an angle of approximately 20° which results in a distance of 1.3 Å between the two C-1 positions. The hydrogen bonding interactions in the −1 subsite are essentially the same for both species. The 2-fluorine is at a distance of 2.8 Å from Asn 155 ND2. It is significant that the C-2 substituent in β-retaining glycoside hydrolases frequently interacts with an amide group as observed in a number of unrelated glycoside hydrolase structures (*15*, *18*, *43*). The O-3 atom makes water-mediated contacts with Asn 158, which may be important for the stabilization of Glu 120. The O-6 hydroxyl group H-bonds to the NE1 of Trp 24 and to the OE2 of the acid/base Glu 203. It is likely that this C6−OH accepts a hydrogen bond from Trp 24 and donates a hydrogen to the general acid/base Glu 203. These are from the *anti* position, thereby permitting the OE2 atom to transfer its proton to the glycosidic oxygen of the scissile point *syn* to its carboxyl group.

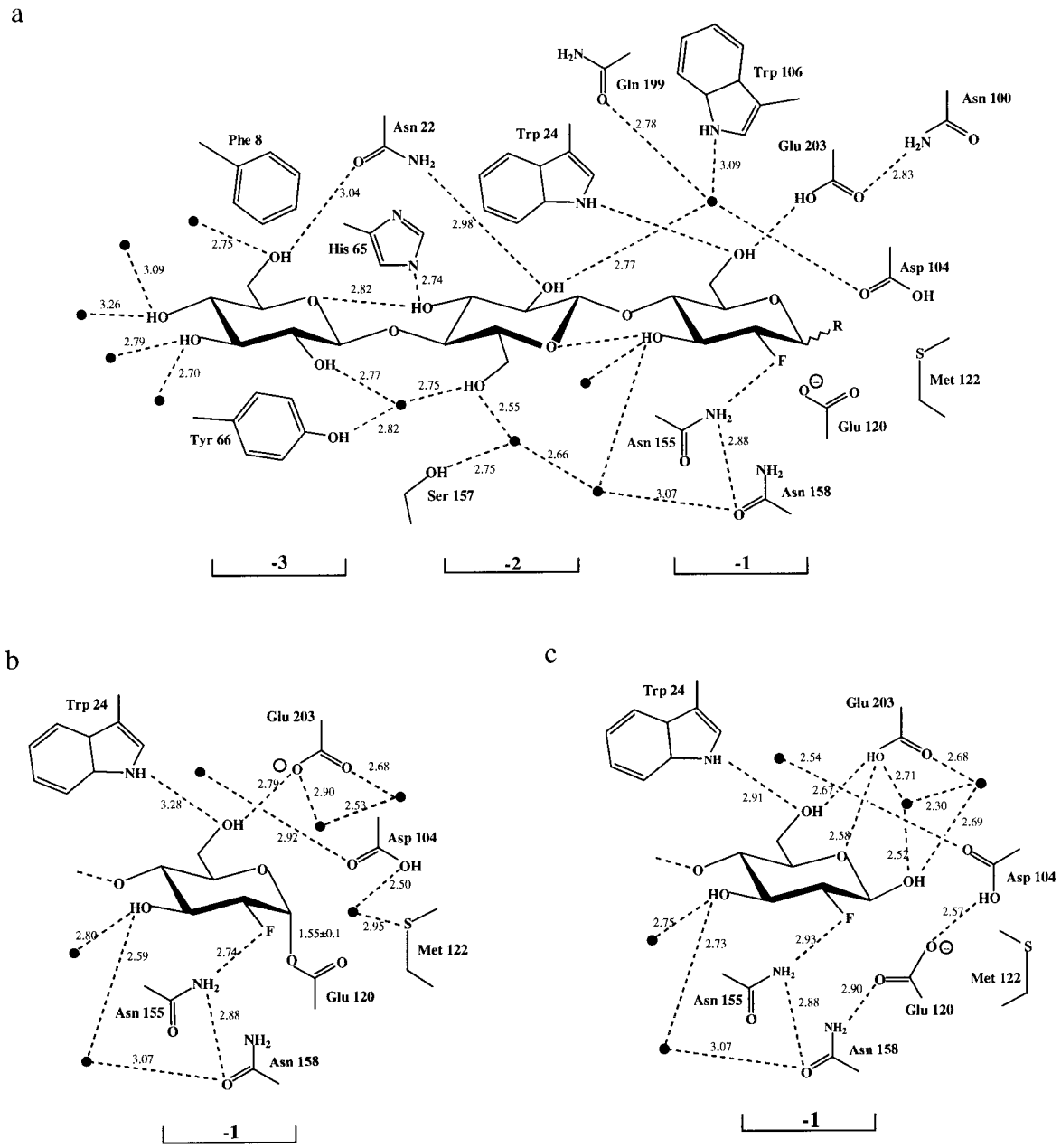

FIGURE 4: Schematic diagram of the interactions of CelB2 with 2-F-cellotriose. Closeups of the −1 subsite interactions for the intermediate and product species are shown in (b) and (c).

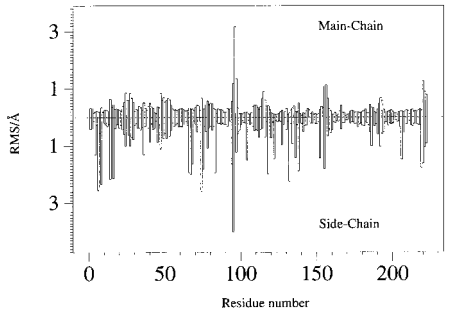

FIGURE 5: Root-mean-squared coordinate differences for the native and 2-F-cellotriose complexes of CelB2.

Covalent Glycosyl-Enzyme Intermediate. As was recently observed in a family 5 complex with a 2-F-oligosaccharide (*18*), the nucleophile appears to have reoriented to minimize the unfavorable interaction between the carbonyl group and the C-2 fluorine atom. This is somewhat different from what is seen in complexes of the *Cellulomonas fimi* CEX (*15*, *17*) and the *Sinapis alba* myrosinase (*16*), both of which show an interaction between the C2-F and the carbonyl group of the nucleophile. Indeed such an interaction in the true substrate is believed to contribute greatly to catalysis (*44*). In the second conformation, the nucleophile approaches the side chain of Met 122, and, correspondingly, this residue has a second conformation. In the glycosyl-enzyme intermediate, the C1 atom of the proximal saccharide has tetrahedral geometry and is bonded to the nucleophile Glu 120 by a covalent linkage in the α-anomeric configuration. The length of the linkage refined to a value of 1.55 Å, with an estimated standard uncertainty of 0.1 Å. This confirms the predictions made on the basis of the native enzyme structure (*19*) and places the identification of the nucleophile by mass spectrometry on a three-dimensional footing (*21*). The anomeric carbon of the attached saccharide is found *syn*

to the ester group of Glu 120, the preferred location for a nucleophilic attack by a carboxylic acid (*45*). A conserved water molecule is found 2.3 Å from the Brønsted acid/base Glu 203 and 3.4 Å from the C1 atom of the covalent intermediate, ideally poised for a nucleophilic attack.

Product Complex. Following the deglycosylation of the intermediate by a solvent water molecule, the reaction product 2-F-cellotriose is formed (Figure 4c). The newly formed anomeric carbon of the product, 2-F-cellotriose, has a β-configuration and lies 2.9 Å from the nucleophile. Presumably since the distance between Glu 120 and the 2-F has increased, the nucleophile has returned to its "relaxed" conformation, where it is stabilized by the ND2 of Asn 158. Asp 104, sitting below the proximal sugar, undergoes a conformational change, placing it just 2.5 Å from the nucleophile where it helps to maintain an appropriate charge on the nucleophile during the catalytic cycle (*46*). In the native and intermediate structures, Asp 104 is remote from Glu 120, suggesting that it plays a role in stabilizing the negative charge on the nucleophile in the presence of oligosaccharide.

DISCUSSION

The nucleophile of glycosyl hydrolase clan C enzymes was first identified in a family 11 xylanase by trapping of a covalent 2-deoxy-2-fluoroxylosyl-enzyme followed by tandem mass spectrometry (*47*). The identity of the Brønsted acid/base has been established for this *Bacillus circulans* xylanase by ^{13}C NMR titration of the ^{13}C-labeled enzyme (*46*). Such measurements have demonstrated an elegant pK_a cycling during the catalytic cycle which ensures maintenance of the correct ionization state of the residues involved in the hydrolytic event. Such pK_a cycling is not likely to be an exclusive preserve of family 11 xylanases and presumably applies to enzymes from family 12 as well. Several abortive epoxyxyloside complexes with *Trichoderma reesei* xylanase II (*48*) and a xylobiose complex with *Bacillus circulans* xylanase (*9*) have given information about enzyme−substrate interactions and directionality of substrate binding in family 11 xylanases. To date no information concerning the mode of substrate binding in family 12 cellulases was available.

The CelB2 complex structure gives insights into the mode of saccharide binding for family 12 cellulases and allows identification of some of the structural features responsible for substrate specificity in glycosyl hydrolase clan C enzymes. The major chemical difference between the two natural substrates xylan and cellulose is at the C-5 position, where xylose lacks the hydroxymethyl substituent of glucose. Furthermore, naturally occurring xylans carry substitutions of acetyl, arabinofuranosyl, and glucuronyl moieties at the C-2 and C-3 positions, which place specific constraints on the shape and polarity of the substrate-binding groove. Comparison of the structure of CelB2 with the xylobiose complex of the inactive mutant of xylanase from *Bacillus circulans* (*9*) and the native *B. pumilis* DSM family 11 xylanase (Sulzenbacher et al., in preparation) gives rms deviations of 2.1 Å for 158 Cα atoms and 2.0 Å for 174 Cα atoms, respectively, after pairwise least-squares-superposition with the program O (*33*). The major structural differences between CelB2 and family 11 xylanases are in the loop regions flanking the active site, where most of the residues responsible for substrate recognition are located. The superposition and comparison of catalytic and binding residues confirm that families 11 and 12 evolved from a common ancestor as anticipated from hydrophobic cluster analysis and the topologies of the native structures (*19*, *20*).

The catalytic machinery of glycosyl hydrolase clan C enzymes is highly conserved. The catalytic residues originate from equivalent β-strands: β-strand B4 for the Brønsted acid/base and β-strand B6 for the nucleophile. The exact orientation of the acid/base is different for the family 11 and 12 enzymes. In family 11, the position of the general acid/base is highly constrained by aliphatic stacking between Leu 46 and Tyr 80 (*B. pumilis* numbering) and by hydrogen bonding to Asn 44 and Tyr 95. A consequence of this is that the carboxylate of Glu 182 is perpendicular to the ring plane of the proximal sugar. In contrast in CelB2, Glu 203 is held in position by a single hydrogen bond to Asn 100, which places the carboxylate parallel to the −1 subsite sugar plane. It is ideally positioned to interact both with the glycosidic oxygen and with the O-6 hydroxyl of *gluco*-derived substrates. This suggests that ligand binding plays a role in the pK_a modulation of the acid−base. On the basis of the native enzyme structure (*19*), we suggested that a glucosyl moiety in the −1 subsite of CelB2 would hydrogen bond through the O-6 hydroxyl with Trp 24 NE1, located above subsite −2. This is indeed the case, and, from an overlap with the xylanase structures, we can see that an equivalent hydrogen bonding donor is absent in family 11. Trp 24 is conserved throughout glycosyl clan C enzymes where it is involved in stacking interactions with a saccharide in subsite −2. In family 11, Leu 46, or its equivalent, not only disfavors a hydroxymethyl substituent at C-5 but also causes displacement of the Trp to a position where the NE1 points away from subsite −1. Thus, in family 11, this invariant hydrophobic residue plays a multifunctional role: favoring binding of a more hydrophobic xylose-based polymer and preventing the binding of cellulose substrates through a combination of steric hindrance and the lack of an appropriately positioned hydrogen bonding partner for O-6.

The presence of the "thumb" in family 11 xylanases, a hairpin loop overlooking the catalytic center, renders the substrate-binding cleft very deep and narrow at subsites −1 and −2. This might explain why family 11 xylanases cannot tolerate large substituents at O-2 or O-3 in the −1 subsite (*49*, *50*). The invariant Pro 131 and the peptide planes of Ser 132 and Ile 133 expose a flat surface to the cleft, which is designed rather for a hydrophobic, unsubstituted, xylose-based substrate than a hydrophilic cellulose chain. In CelB2 this "thumb" is missing, and a wide cavity opens instead at subsites −1 and −2. The active site region of family 12 cellulases is much less restricted, and notable activity on xylan has been observed for other family 12 cellulases (*51*) but not in *S. lividans* CelB.

No information is available so far about substrate binding at the aglycon-binding site of glycosyl hydrolase clan C enzymes. From the relative position of the product in the −1 subsite of the CelB2 complex, it appears that a cellulose chain would have to undergo a slight bend within the enzyme's cleft to avoid steric clashes with Met 122. This residue is probably involved in stacking interactions with a glucosyl moiety in the +1 subsite, and further stabilizing

interactions might be provided from the side chain of Asn 100. Stacking interactions with a pyranose ring in the +2 subsite might conceivably be established with the main-chain peptide planes of Asn 51, Gly 52, and Ala 53, located in the extensive loop connecting β-strands A3 and B3. Interestingly, Gly 52 overlaps with the side chain of the conserved Tyr 184 in family 11 xylanases, a residue ideally positioned for stacking interactions with a pyranosyl ring. At the far end of the binding cleft, Gln 132, part of the "cord", seems well positioned for an interaction with the substrate. This residue is conserved in most family 12 enzymes and has a structural equivalent in an invariant tyrosine in family 11 xylanases. The position of the latter residue and Gln 132 in CelB2 is stabilized by a proline, Pro 133 in CelB2, located right in the middle of the "cord" and conserved in all glycosyl hydrolase clan C enzymes. It seems likely that, in contrast to previous proposals that the "cord" might undergo conformational changes upon substrate binding (*10*, *19*), the function of the "cord" region is to deliver residues for substrate binding, and that the invariant proline is vital to ensure the correct orientation of these residues. Further comparison of family 11 and 12 structures awaits the determination of a truly equivalent intermediate complex for family 11.

REFERENCES

1. Bayer, E. A., and Lamed, R. (1992) *Biodegradation 3*, 171−188.
2. Béguin, P., and Aubert, J.-P. (1994) *FEMS Microbiol. Rev. 13*, 25−58.
3. Gilbert, H. J., and Hazlewood, G. P. (1993) *J. Gen. Microbiol. 139*, 187−194.
4. Warren, R. A. J. (1996) *Annu. Rev. Microbiol. 50*, 183−212.
5. Henrissat, B., and Davies, G. J. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
6. Henrissat, B., Callebaud, I., Fabrega, S., Lehn, P., Mornon, J.-P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7090−7094.
7. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
8. Campbell, R. L., Rose, D. R., Wakarchuk, W. W., To, R., Sung, W., and Yaguchi, M. (1993) in *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 63−72, Foundation for Biotechnical and Industrial Fermentation Research, Espoo.
9. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) *Protein Sci. 3*, 467−475.
10. Törrönen, A., Harkki, A., and Rouvinen, J. (1994) *EMBO J. 13*, 2493−2501.
11. Törrönen, A., and Rouvinen, J. (1995) *Biochemistry 34*, 847−856.
12. Koshland, D. E. (1953) *Biol. Rev. 28*, 416−436.
13. Davies, G., Sinnott, M. L., and Withers, S. G. (1997) in *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.) pp 119−209, Academic Press, London.
14. Withers, S. G., Street, I. P., Bird, P., and Dolphin, D. H. (1987) *J. Am. Chem. Soc. 109*, 7530−7531.
15. White, A., Tull, D., Johns, K., Withers, S. G., and Rose, D. R. (1996) *Nat. Struct. Biol. 3*, 149−154.
16. Burmeister, W. P., Cottaz, S., Driguez, H., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663−675.
17. Notenboom, V., Birsan, C., Warren, R. A. J., Withers, S. G., and Rose, D. R. (1998) *Biochemistry 37*, 4751−4758.
18. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707−11713.
19. Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C., and Davies, G. J. (1997) *Biochemistry 36*, 16032−16039.
20. Törrönen, A., Kubicek, C. P., and Henrissat, B. (1993) *FEBS Lett. 321*, 135−139.
21. Zechel, D. L., He, S., Dupont, C., and Withers, S. G. (1998) *Biochem. J.* (in press).
22. Mackenzie, L. F., Wang, Q., Warren, R. A. J., and Withers, S. G. (1998) *J. Am. Chem. Soc. 120*, 5583−5584.
23. Otwinowski, Z. (1993) in *Data Collection and Processing: proceedings of the CCP4 study weekend* (Sawyer, L., Isaacs, N., and Bailey, S., Eds.) Science and Engineering Research Council, Daresbury, U.K.
24. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology: Macromolecular Crystallography, part A* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, London and New York.
25. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
26. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
27. Navaza, J., and Saludijan, P. (1997) *Methods Enzymol. 276*, 581−594.
28. Brünger, A. T. (1992) *Nature 355*, 472−475.
29. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
30. Brünger, A. T. (1988) *J. Mol. Biol. 203*, 803−816.
31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
32. Murshodov, G. N., Davies, G. J., Isupov, M., Krzywda, S., and Dodson, E. J. (1998) *CCP4 Newslett. 35*, 37−42.
33. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
34. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D49*, 129−147.
35. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol. 277*, 319−343.
36. Raymond, S., Heyraud, A., Qui, D. T., Kvick, A., and Chanzy, H. (1995) *Macromolecules 28*, 2096−2100.
37. Tull, D., and Withers, S. G. (1994) *Biochemistry 33*, 6363−6370.
38. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961−9969.
39. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
40. Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 282−291.
41. 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.
42. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
43. Sulzenbacher, G., Schülein, M., and Davies, G. J. (1997) *Biochemistry 36*, 5902−5911.
44. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) *Nat. Struct. Biol. 5*, 812−818.
45. Gandour, R. D. (1981) *Bioinorg. Chem. 10*, 169−176.
46. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958−9966.
47. Miao, S., Ziser, L., Aebersold, R., and Withers, S. G. (1994) *Biochemistry 33*, 7027−7032.
48. Havukainen, R., Törrönen, A., Laitinen, T., and Rouvinen, J. (1996) *Biochemistry 35*, 9617−9624.
49. Biely, P., Vršanská, M., and Bhat, M. K. (1997) in *Carbohydrases from Trichoderma reesei* and other microorganisms structures, biochemistry, genetics and applications (Claeyssens, M., Nerinckx, W., and Piens, K., Eds.) pp 94−112, Royal Society of Chemistry, London.
50. Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., and Kratky, C. (1998) *Biochemistry 37*, 13475−13485.
51. Saarilahti, H. T., Henrissat, B., and Pavla, E. T. (1990) *Gene 90*, 9−14.
52. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI982648I