Structural Basis for the Exocellulase Activity of the Cellobiohydrolase CbhA from *Clostridium thermocellum*†

Florian D. Schubot,‡,§ Irina A. Kataeva,‡ Jessie Chang,‡ Ashit K. Shah,‡ Lars G. Ljungdahl,‡ John P. Rose,‡ and Bi-Cheng Wang*,‡

Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, Georgia 30602

Received September 5, 2003; Revised Manuscript Received November 25, 2003

ABSTRACT: Numerous bacterial and fungal organisms have evolved elaborate sets of modular glycoside hydrolases and similar enzymes aimed at the degradation of polymeric carbohydrates. Presently, on the basis of sequence similarity catalytic modules of these enzymes have been classified into 90 families. Representatives of a particular family display similar fold and catalytic mechanisms. However, within families distinctions occur with regard to enzymatic properties and type of activity against carbohydrate chains. Cellobiohydrolase CbhA from *Clostridium thermocellum* is a large seven-modular enzyme with a catalytic module belonging to family 9. In contrast to other representatives of that family possessing only endo- and, in few cases, endo/exo-cellulase activities, CbhA is exclusively an exocellulase. The crystal structures of the combination of the immunoglobulin-like module and the catalytic module of CbhA (Ig-GH9_CbhA) and that of an inactive mutant Ig-GH9_CbhA(E795Q) in complex with cellotetraose (CTT) are reported here. The detailed analysis of these structures reveals that, while key catalytic residues and overall fold are conserved in this enzyme and those of other family 9 glycoside hydrolases, the active site of GH9_CbhA is blocked off after the −2 subsite. This feature which is created by an extension and altered conformation of a single loop region explains the inability of the active site of CbhA to accommodate a long cellulose chain and to cut it internally. This altered loop region is responsible for the exocellulolytic activity of the enzyme.

Due to its relative abundance as a key constituent of plant cell walls, cellulose has enormous potential as a resource for renewable energy. The development of economically feasible means of converting this raw material into a useful fuel is the subject of considerable research effort (*1*). One promising approach is the utilization of a wide array of naturally occurring hydrolytic enzymes that target the various polysaccharides present in plant cell walls. These enzymes have evolved in numerous bacterial and fungal organisms that aim to exploit plant materials as a nutritional resource. Glycoside hydrolases constitute an extensive group, and codons have been identified for more than 1800 proteins so far. They commonly possess modular architectures, being composed of one or more catalytic modules (*2*) and a variety of accessory modules. The accessory modules serve numerous functions. Carbohydrate binding modules (CBM)[1] enhance the association of the catalytic modules to insoluble carbohydrates (*3*, *4*), surface-layer-homology modules bind enzymes to cell walls (*5*), while others modify substrate surfaces (*6*), and finally dockerins mediate the integration of the enzymes into multiprotein complexes, termed cellulosomes (*7*, *8*). Functions of some "X" modules are yet to be established. The catalytic modules of glycoside hydrolases have been classified into 90 families based on sequence similarity. Representatives of each family usually possess common fold and catalytic mechanism, and have highly conserved catalytic residues. Similarly, the carbohydrate-binding modules have been classified into 32 families. The classification system of glycoside hydrolases and related enzymes as well as their binding modules is presently available at the carbohydrate-active enzymes (CAZy) server at URL: http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html. Over 170 members are included in glycoside hydrolases of family 9. Representatives of this family feature an $(\alpha/\alpha)_6$-barrel fold and are characterized by an inverting mechanism of catalysis with an aspartic acid acting as a catalytic nucleophile and a glutamic acid residue serving as a catalytic proton donor. Most family members with established enzymatic functions display endoglucanase activity. The processive endo/exocellulase Cel9A (former E4) from *Thermobifida fusca* (*9*), and a semi-processive endo/exoglucanase Cel9B (former CenC) from *Cellulomonas fimi* (*10*) have a mixed

† Support for this research was provided in part by funds to B.-C. Wang from the University of Georgia Research Foundation, the Georgia Research Alliance, and by the U.S. Department of Energy Grant No. DE-FG02-93ER20127/A009 to L.G.L.

* Corresponding author: Dr. Bi-Cheng Wang, Room B204A Life-Sciences Building, Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, GA 30602. E-mail: wang@bcl1.bmb.uga.edu. Phone: (706) 542-1747. Fax: (706) 542-3077.

‡ Department of Biochemistry & Molecular Biology, The University of Georgia, Athens, GA 30602.

§ Current Address: Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702, USA.

[1] Abbreviations: GH9, glycoside hydrolase of family 9; Ig, immunoglobulin; CBM, carbohydrate-binding module; CTT, cellotetraose; X, module of unknown function; PCR, polymerase-chain reaction; PNP, *p*-nitrophenyl; MIRAS, multiple isomorphous replacement including anomalous scattering.

10.1021/bi030202i CCC: $27.50
Published on Web 01/15/2004

type of activity. The two highly homologous cellobiohydrolases from *Clostridium thermocellum* cellulosome, CelK and CbhA, which display over 90% sequence identity between their catalytic modules and originated as a result of a gene duplication event (*11*), are the only known exocellulases of family 9 (*4*, *6*). Endo- and exocellulases differ in the topology of their active sites (*12*, *13*) resulting in the internal (endo-) or external (exo-) modes of carbohydrate chain cleavage.

Recently, representatives of family 9 were subdivided into themes A, B, C, and D on the basis of their modular arrangement (*8*). Theme A incorporates plant enzymes consisting of only the catalytic module. Theme B combines enzymes containing a catalytic module and a C-terminal family 3c CBM (CBM3c). Glycoside hydrolases of theme C contain an N-terminal Ig-like module attached to a catalytic module. Enzymes of theme D contain an N-terminal family 4 CBM (CBM4) in addition to the Ig-like module and a catalytic module. Some enzymes of subgroups B, C, and D possess other accessory modules, but order of the above modules is characteristic for each theme. Crystal structures are currently available for four family 9 members: (a) the combination of the catalytic module and the C-terminal CBM3c in the structures of *T. fusca* Cel9A (*15*) and *Clostridium cellulolyticum* Cel9G (*16*) corresponding to theme B; (b) the combination of an N-terminal Ig-like module and a catalytic module in the nonprocessive endoglucanase Cel9A (former CelD) from *C. thermocellum* (*17*) representing theme C; and (c) the nonprocessive endoglucanase Cel9M from *C. cellulolyticum* composed of only the catalytic module and in this regard representative of theme A (*18*). Comparison of the intimate topologies of endo-, endo/exo-, and exocellulases belonging to the same family might give clues toward the understanding of evolutionary modifications leading to the conversion of one type activity to another. Cellobiohydrolase CbhA, a representative of theme D, is a multimodular enzyme comprising an N-terminal CBM4 (residues 40−200), an Ig-like module (residues 210−309), a catalytic module (residues 310−815), $X1_1$ and $X1_2$ modules (residues 820−912 and 914−1000, respectively), CBM3 (residues 1014−1151), and a dockerin module (residues 1164−1230) (*19*). In the present paper, we report the crystal structure of a module pair, the Ig-like module and the family 9 catalytic module (GH9_CbhA), of CbhA (Ig-GH9_CbhA) from *C. thermocellum*. On the basis of the structure of the inactive Ig-GH9_CbhA(E795Q) mutant in complex with cellotetraose (CTT), a discussion is given of the key structural elements, that define its exoactivity.

MATERIALS AND METHODS

Site-Directed Mutagenesis. Glutamic acid in position 795 of Ig-GH9_CbhA was selectively mutated to glutamine residue using the following oligonucleotide primers: 5′-CCAGTTAACAGTTAT**CTG**GTTGGTTGACCATGAGTC-3′ and 5′-GACTCATGGTCAACCAAC**CAG**ATAACTG-TTAACTGG-3′. Plasmid pET-21b(+) containing the DNA fragment encoding Ig-GH9_CbhA served as a template. PCR with mutagenesis primers was carried out using the QuickChange Site-Directed Mutagenesis kit (Stratagene Cloning Systems, La Jolla, CA). The PCR product was used to transform BL21(DE3)pLys competent cells (Stratagene). Plasmid DNA was isolated and sequenced to confirm the mutation.

Purification. Native Ig-GH9_CbhA and mutant E795Q were purified as described previously (*5*) by affinity chromatography using a Ni-nitrilotriacetic acid agarose column (Qiagen Inc., Valencia, CA), and gel-exclusion chromatography with a prepacked TSK-3000SW column (Toso-Haas, Montgomeryville, PA). The proteins were homogeneous as ascertained by sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

Activity Assays. The activity of CbhA was assayed in 50 mM sodium citrate buffer (pH 6.0), at 60 °C with *p*-nitrophenyl(PNP)-cellobioside as a substrate. The activity was monitored by measuring the release of PNP and expressed as micromoles of PNP released per minute per mg of enzyme. Concentrations of proteins were determined with a Coomassie Protein Assay reagent (Pierce, Rockford, IL).

Crystallization. Crystals of the native Ig-GH9_CbhA domains were obtained via modified microbatch crystallization under oil at 273 K using 1 μL drops (*20*). A setup contained 0.5 μL of protein solution at 20 mg/mL in 25 mM Tris-HCl buffer (pH 8.5) and 0.2 M NaCl, and 0.5 μL reservoir solution containing 20% PEG4000, 0.2 M $(NH_4)_2SO_4$, 0.1 M Tris, pH 7.0.

Crystals of the inactive E795Q mutant in complex with cellotetraose were obtained from the same setup with the addition that the reservoir solution contained 1 mM cellotetraose and 0.1 M $SrCl_2$. These slightly modified conditions resulted in crystals with hexagonal $P6_1$ space group symmetry as opposed to the orthorhombic $P2_12_12_1$ of the native crystal.

Data Collection. Data sets for the native crystals, and the K_2PtCl_4 derivative were collected on a Rigaku R-Axis IV image plate detector on loop-mounted (*21*) and flash-frozen (93 K) crystals using 5kW Cu Kα radiation generated by a Rigaku FRD system and focused with Rigaku Max Screen optics. The data were indexed, integrated, and scaled using HKL 1.9.1 (*22*). The data for the mercury derivative of the native protein and the cellotetraose complex were collected on a Bruker SMART 6000 CCD detector using the same generator and Rigaku Hi Res optics. In this case, the data were processed with the Proteum software package. Details of data collection and processing for all data sets are shown in Table 1.

Structure Solution and Refinement. For phasing Platinum and Mercury salt derivatives of the native enzyme crystals were generated. The derivative crystals were obtained by soaking native protein crystals for 3 days in reservoir solutions containing 10 mM K_2PtCl_4 and $HgCl_2$, respectively. The heavy atom derivatives were analyzed using SOLVE (*23*). The obtained MIRAS phases were directly channeled into RESOLVE (*24*). The initial 3.2 Å map and partial backbone trace created by RESOLVE were of excellent quality, exhibiting clear protein/solvent boundaries and recognizable features of secondary structure. The density modified map allowed fitting of almost the entire backbone and most of the side chains in the molecular modeling programs O (*25*) and XTALVIEW (*26*). After several rounds of phase combination using SIGMAA (*27*) and model building, the remaining residues and side chains could be fit into the electron density map. The model was refined using the simulated annealing (SA) protocol of CNS (*28*) followed by manual adjustment of the model using SA omit maps. The structure of the complex was obtained via Molecular

Table 1: Data Collection and Refinement Statistics

	CbhA_Gh9	CbhA_Gh9(E795Q):CTT	Pt derivative	Hg derivative
resolution [Å]	69−2.1	54−2.4	68−3.2	70−2.4
space group	$P2_12_12_1$	$P6_1$	$P2_12_12_1$	$P2_12_12_1$
cell [Å]	$a = 74.55$	$a = b = 108.567$	$a = 74.42$	$a = 74.73$
	$b = 75.76$	$c = 119.219$	$b = 75.67$	$b = 75.99$
	$c = 137.50$		$c = 137.67$	$c = 138.19$
completeness [%] (last shell)	91.6(2.18−2.1:65.9)	97.5(2.51−2.4:95.8)	93.6(3.36−3.2:93.1)	87.5(2.62−2.4:70.6)
redundancy	2.8(1.8)	3.8 (2.8)	4.8(4.6)	3.89(3.1)
I/σ_I	11.8(3.1)	8.53(1.58)	17.2(13.1)	10.2(4.8)
R_{sym} [%]	8.1(24.8)	6.4(23.6)	8.3(11.1)	9.3(16.3)
no. of water molecules	862	390		
R [%]	14.7	21.0		
R_{free} [%]	18.5	27.5		
rms bonds [Å]	0.008	0.009286		
rms angles [deg]	1.381	1.85779		

Replacement using CNS. The refinement statistics for the two models are given in Table 1. Water molecules were added to both structures using ARP/wARP (*28*) in combination with REFMAC (*26*).

In the native enzyme structure density for a bicyclic molecule, possibly a detergent molecule that might have been present in one of the used reagents, appeared on the product-binding side of the active site after addition of the water molecules. The identity of this molecule is still under investigation and the structure will be updated and submitted to the Protein Data Bank as soon as the issue has been resolved.

The model quality was assessed with PROCHECK (*29*). In both structures, all nonglycine residues resided either in the most favorable or in the allowed regions of the Ramachandran plot and their overall geometry was better than the average when compared to structures solved at the same resolution.

The structures of Ig-GH9_CbhA and Ig-GH9_CbhA-(E795Q):CTT were submitted to the Protein Data Bank (PDB) (*30*) (PDB codes 1UT9 and 1RQ5, respectively).

RESULTS AND DISCUSSION

The *C. thermocellum* cellobiohydrolase CbhA consists of all together seven modules. An N-terminal family 4 carbohydrate binding module (CBM) is followed by an immunoglobulin-like (Ig-like) module of unknown function, the large family 9 catalytic module (GH9_CbhA), $X1_1$ and $X1_2$ modules (previously considered as fibronectin type 3-like modules), a family 3 CBM, and a dockerin module (*19*). The $X1_1$ and $X1_2$ modules were recently shown to aid substrate hydrolysis by loosening the surface of insoluble cellulose (*6*). CbhA displays typical exocellulytic activity as it does not decrease viscosity of a carboxymethyl cellulose solution and produces almost exclusively cellobiose from acid-swollen cellulose, Avicel, and filter paper (*4*, *31*). The crystallized construct reported here consists of the catalytic module GH9 in conjunction with the N-terminal Ig-like module. Although the two modules have relatively independent folds, it has been demonstrated that the deletion of the entire (*32*) or even half (Kataeva, unpublished) of the Ig-like module leads to a complete loss of enzymatic activity.

The structure of the native module pair Ig-GH9_CbhA without bound carbohydrate substrate and the E795Q mutant in complex with cellotetraose were solved at a resolution of 2.1 and 2.4 Å, respectively. In the free enzyme structure electron density could be observed for the entire backbone structure ranging from residues 208 to 816, with the only exception being a gap between amino acids 600 and 602. In the complex structure this gap of disordered residues is more pronounced ranging from residues 599 to 604. This is likely the result of the lower quality of the data obtained for the complex crystals as apparent from the statistics in Table 1.

Crystal Structure of the Free Enzyme. The overall structure of the Ig-GH9_CbhA closely mirrors that of three enzyme structures that have been reported previously for family 9 glycoside hydrolases (Figure 1a). The catalytic module consists of the characteristic $(\alpha/\alpha)_6$-barrel with the N-terminal site of the inner helices forming the active site cleft. The much smaller Ig-like module consists of the two-layered β-sheets that form the anticipated β-sandwich structure also observed for *C. thermocellum* Cel9A (*17*). The overall structures of the catalytic modules of CbhA and Cel9A are very similar with an RMSD 1.944 Å for the backbone of 444 (out of 610) overlapping residues. Another conserved structural feature is two calcium-binding sites, which correspond to sites A and B in the Cel9A structure. The residues forming the third calcium-binding site in Cel9A are not conserved in the GH9_CbhA and no enhanced electron density could be observed in the position corresponding to that calcium ion in Cel9A. This observation is of some interest, since this site stabilizes the active conformation of Cel9A (*33*), but apparently is not necessary in CbhA. The other two calcium-binding sites seem to be crucial for the proteins structural integrity as their removal with EGTA results in decreased thermostability of Cel9A (*33*). The residues involved in the calcium coordination are marked with a blue star in the structure-based sequence alignment of the four representative enzyme structures (Figure 3).

The RMSD between the structures of *C. thermocellum* Cel9A and *C. cellulolyticum* Cel9G is about 0.9 Å. Therefore, only Cel9A has been included into the subsequent comparative analysis.

Substrate Binding and Catalytically Important Residues. The active site cleft is located at the N-terminal site of the inner helices of the $(\alpha/\alpha)_6$-barrel structure. The structure of the inactive E795Q mutant in complex with cellotetraose clearly shows that the cleft is just long enough to accommodate four subsites (−2 to +2) for glucose binding. The electron density for the cellotetraose moiety was very well defined for the glucose units bound in subsites +2 and +1; however, the density in sites −1 and −2 was, albeit visible,

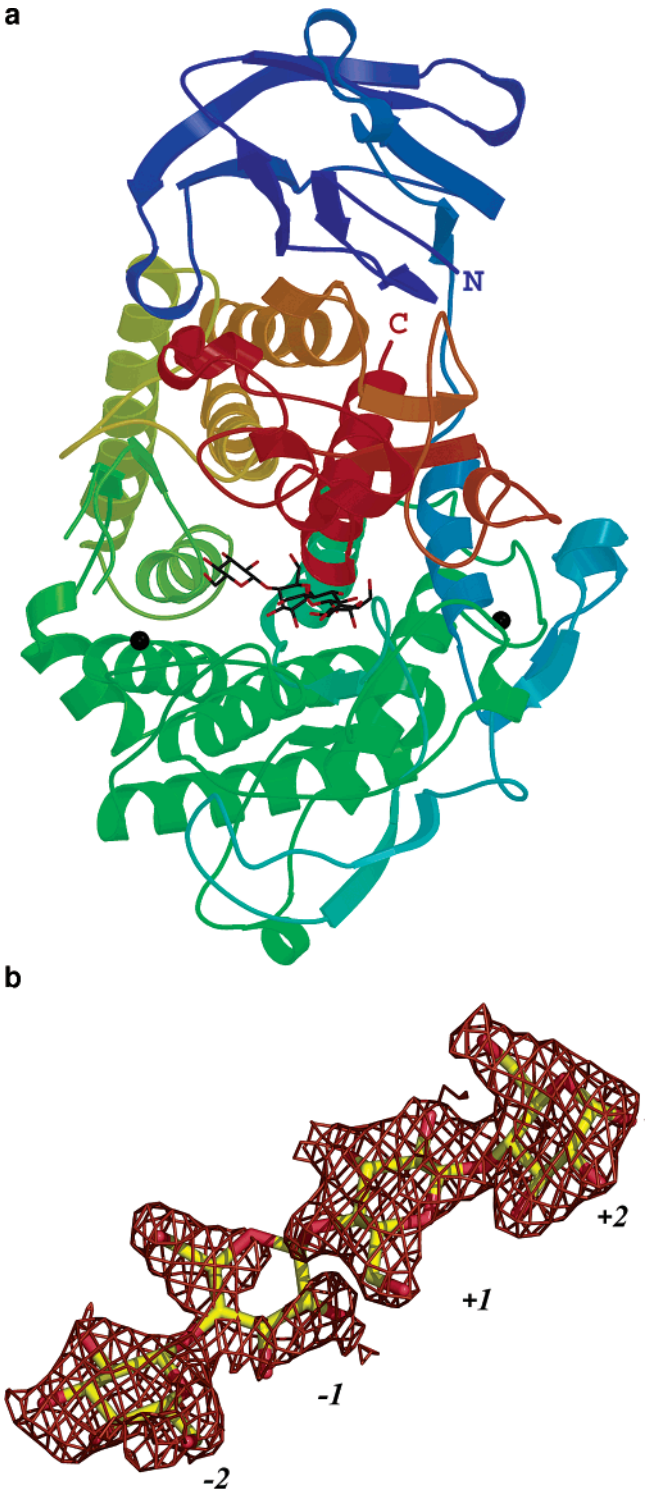

FIGURE 1: (a) Molscript (*39*) cartoon drawing of Ig-GH9_CbhA:CTT complex. The rainbow coloring starts at the N-terminus, giving the all β-sheet Ig-like module a dark blue taint. A stick presentation of the CTT ligand is visible at the heart of the $(\alpha/\alpha)_6$-barrel in the larger catalytic module. The two calcium(II) ions are displayed as dark spheres. (b) Shown is the $2F_O-F_C$ composite omit map of the ligand region contoured at 1σ around a stick model of CTT. This figure was generated with PyMOL (*40*).

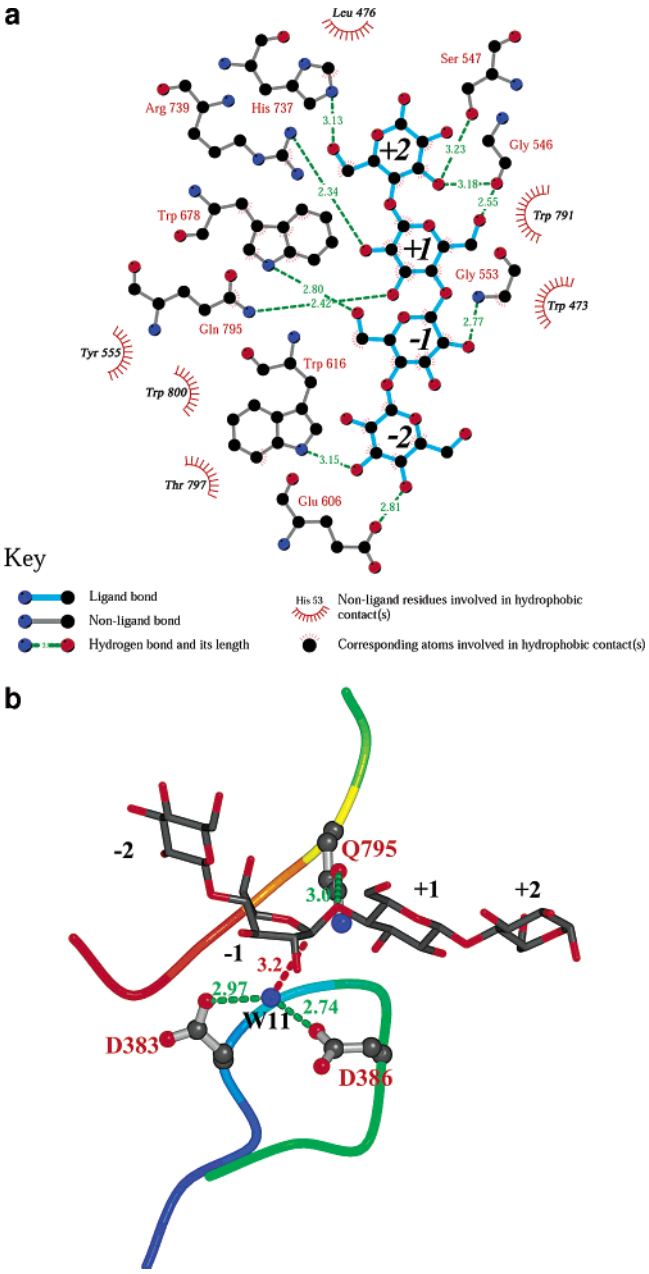

FIGURE 2: (a) Schematic drawing of the hydrophilic and hydrophobic protein ligand interactions in the GH9_CbhA:CTT complex generated by LIGPLOT (*41*). Particularly striking are the many aromatic residues that line the active site. Gln795 is the mutated residue, which would be a glutamate in the native structure. (b) Arrangement of the catalytic residues in the active site with respect to CTT. Also shown is the catalytic water molecule (W11) that is positioned to carry out the nucleophilic attack.

considerably weaker (Figure 1b). The weaker binding on the product site is consistent with the required release of the cleaved cellobiose prior to initiation of another round of catalysis. While the glucose units in subsites +2, +1, and −2 all fit their respective electron density in the common chair conformation, the glucose unit in the −1 subsite is clearly distorted. This observation is consistent with a recent report of the 0.94 Å structure of endoglucanase CelA (*34*), where the glucose unit in the −1 site assumed a distorted boat conformation that seemed to facilitate oxicarbenium ion formation and subsequent substrate cleavage. The conformation modeled for this glucose unit in GH9_CbhA allows placing of the following unit into its density at the −2 site but the electron density at the −1 is too weak to be certain of the molecules conformation there. Figure 2a shows a schematic representation of the intermolecular contacts between protein and substrate. The abundance of aromatic residues, which stabilize the carbohydrate moiety through

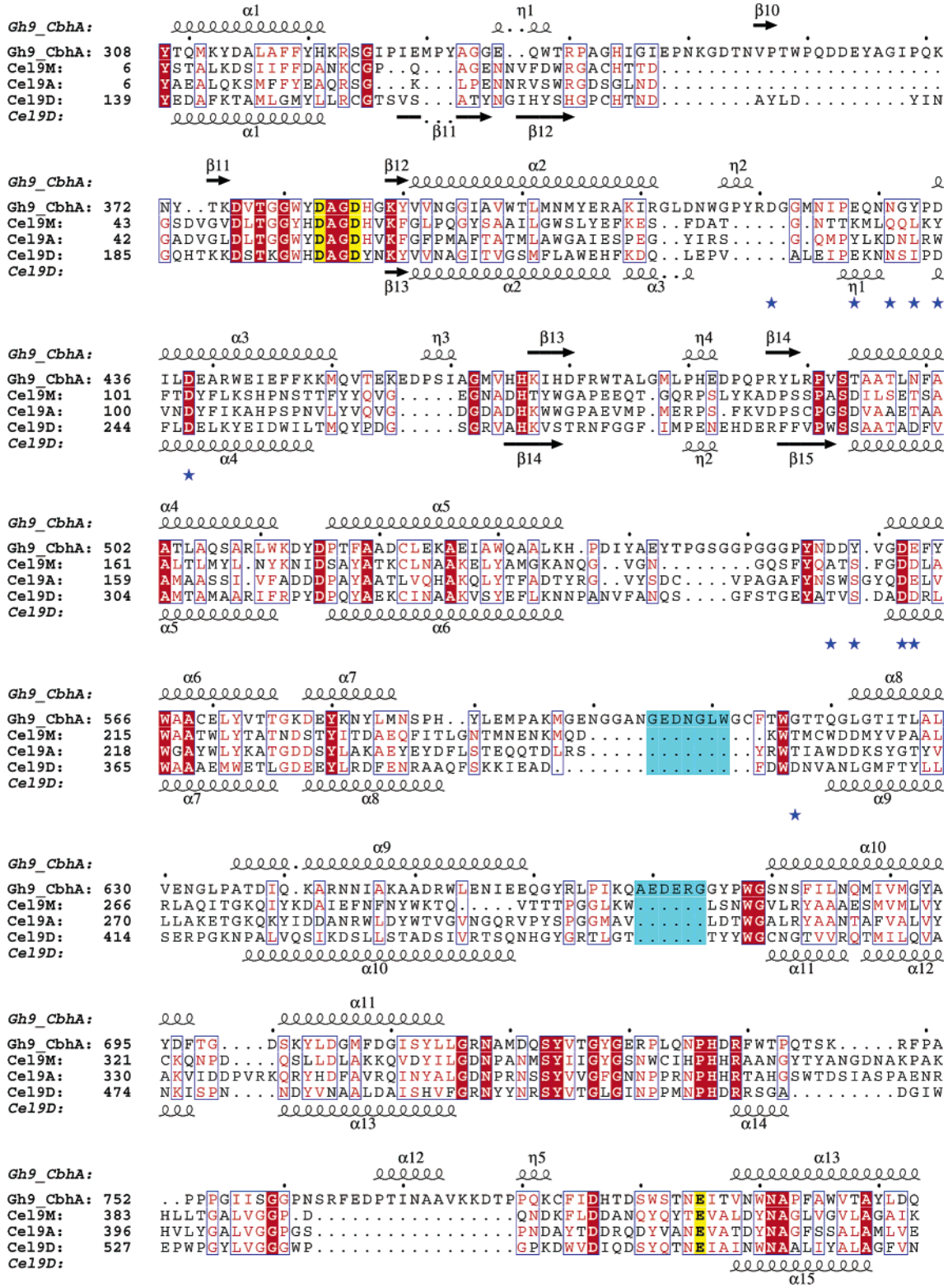

FIGURE 3: ESPript (*42*) structure-based sequence alignment of the catalytic modules of four family 9 glycoside hydrolases for which structures have been determined. Respectively, above and below the alignment the secondary structures of the closely related GH9_CbhA and Cel9D are shown. The catalytically important residues are highlighted yellow; residues involved in calcium binding in GH9_CbhA are marked with a blue star and the loop regions responsible for the exo activity of GH9_CbhA are highlighted turquoise. Abbreviations: GH9_CbhA, *C. thermocellum* CbhA; Cel9M, *C. cellulolyticum* Cel9M; Cel9A, *T. fusca* Cel9A; Cel9D, *C. thermocellum* Cel9A (former CelD).

stacking interactions, is a common feature of carbohydrate binding proteins. Residues Y555, W616, W678, and H737 are strictly conserved among the four structures of family 9 enzymes, while W791 and W800 are homologous in the other proteins. Residue R739 that forms a short 2.3 Å hydrogen bond to the +1 glucose unit seems to be also important as it is conserved throughout.

The three catalytically important residues are D383, D386, and E795. While E795 is believed to stabilize the transition state, D383 and D386 are thought to deprotonate the water

molecule that carries out the nucleophilic attack on the C1 carbon of the targeted glucose unit. In fact, in the complex structure of GH9_CbhA water molecule 11 is in a good position to carry out such an attack as it is hydrogen-bonded to D383 and D386, located within 3.2 Å of the target C1 atom of the glucose unit in subsite +1 and at an angle of almost 180° to the cleaved C1−O1 bond (Figure 2b).

Structural Basis for the Exocellulase Activity of GH9_CbhA. The structural elements discussed to this point are well conserved among the family 9 glycoside hydrolases. Even the bound substrates overlap quite well from subsites +2 to −2, when the different available complex structures are superimposed. Merely the number of subsites is variable. Cel9A from *T. fusca*, for instance, has additional subsites −3 and −4, which results in cellotetraose being the main cleavage product of this enzyme (*9*). However, these features do not account for the observed purely exocellulolytic activity of GH9_CbhA. Figure 4 shows a comparative view of the surface topologies for the active site clefts of GH9_CbhA and of three other enzymes for which structures are available. It immediately becomes apparent that the cleft in GH9_CbhA is considerably narrower and even forms a tunnel upon substrate binding. Responsible for this tunnel formation is a conformational shift of the loop region constituted by residues 540 to 560, which allows S547 to form a hydrogen bond with the O3 atom in the +2 subsite and thus "close" the active site. This loop region is present in all four enzymes but extended by about four residues in GH9_CbhA. This observation might be suggestive for this enzyme's mode of action as tunnel formation was proposed to be a hallmark of processive cellulases (*17*).

The key structural element, however, that distinguishes GH9_CbhA as an exocellulase is the abrupt blockage of the active site cleft after the −2 subsite. This blockage creates a steep slope on the product site of the cleft as opposed the more gradual incline in the active sites of the other enzymes with endo activity that permits exit of a bound cellulose chain. Responsible for the formation of this steep slope in GH9_CbhA is a loop region consisting of residues 590−620. This loop, while present, is about 10 residues shorter in the other enzymes resulting in a different conformation and an open active site. In GH9_CbhA amino acids 605−611 constitute the residues that actually block the active site exit. Residue E606, forming a 2.8 Å hydrogen bond to atom O4 of the carbohydrate bound at the −2 site, is involved in substrate binding. This is a further indication that this subsite is intended for binding of a *terminal* glucose unit. The conformation of the described loop region is stabilized through interactions with yet another loop formed by residues 660−680 that is extended and conformationally distinct in GH9_CbhA. That loop length plays a role in accounting for endo or exo activity of an enzyme has been considered previously for other enzyme families. In cellobiohydrolase II a family 6 exo-β-(*1*−*4*)-glucanase from *T. reesei* deletion of surface loops converted the active site tunnel of the native enzyme into a groove and led to increased endo activity of the enzyme (*35*). It can be noted that in CelA and CelC from *Orpinomyces* strain PC-2 belonging to family 6 glycoside hydrolases and function both as exo/endo enzymes, the surface loops seem to be shortened or missing. This may explain that these enzymes have both activities (*36*). Clearly, the loop mechanism by which the exo/endo conversion

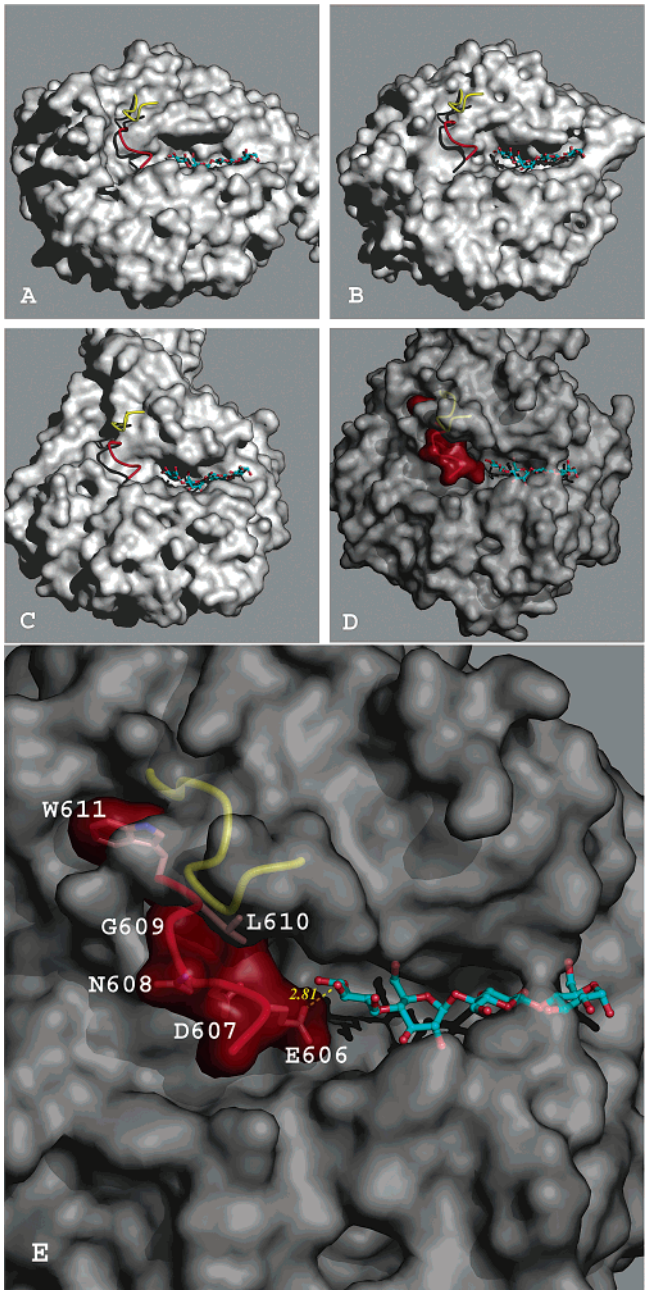

FIGURE 4: (A−D) Show the active site surface topologies of the *T. fusca* Cel9A, *C. cellulolyticum* Cel9M, *C. thermocellum* Cel9A, and *C. thermocellum* GH9_CbhA, respectively. The positions of the CTT ligand and the two distinctive loop regions from GH9_CbhA are superimposed in all structures to emphasize the absence of these structural elements in the other enzymes and highlight the distinct active site topology of the exocellulase. (E) Close-up view of the binding cleft in GH9_CbhA. This figure was generated with Molscript and PyMOL.

occurs for family 6 enzymes is quite different from what is observed for GH9_CbhA. A case much closer to that of GH9_CbhA was reported for Exg, a family 5 exo-β-(1,3)-glucanase (*37*). In this family of enzymes, which is not related to family 9 enzymes, the active site groove is located at the heart of a $(\beta/\alpha)_8$-barrel. In Exg the groove's exit is blocked by a combination of loops and helical regions that are absent in the structurally homologous endoglucanase CelC from C. *thermocellum* (*38*). Although these structural differences are more dramatic than in the case of GH9_CbhA, they underline the universal principle by which the exo/endo

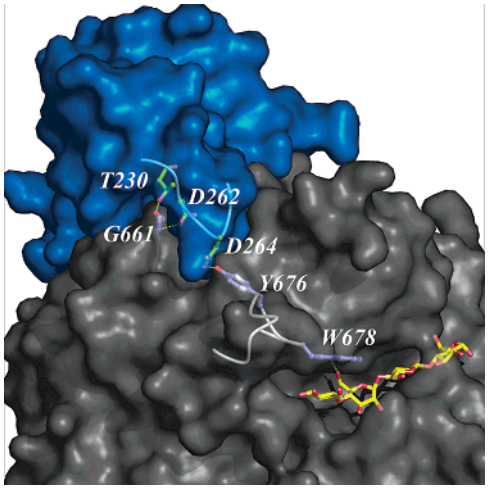

FIGURE 5: Surface plot showing those intermodular hydrogen-bonding interactions that are strictly conserved between GH9_CbhA and *C. thermocellum* Cel9A. The close proximity of Y676 to the highly conserved W678 sitting in the active site suggests the importance of the D264−Y676 interactions.

activities can be modulated through modification in loop regions.

Importance of the Ig-Like Module. The N-terminal immunoglobin-like module, which is made up of a β-sandwich, represents another interesting aspect of the structure of CbhA. While its function is still unknown, studies of truncation mutants have revealed that the presence of this 100-residue module is required for enzymatic activity (*32*, and Kataeva, unpublished observations). This is not surprising considering the extensive interface that exists between these two modules of the enzyme. This interface is maintained through a large number of hydrophobic and hydrophilic interactions, involving 41 residues from both domains (*17*). Of the other four enzymes, only Cel9A from *C. thermocellum* possesses an equivalent module of virtually identical structure (*32*). The sequence identity between GH9_CbhA and the catalytic module of Cel9A is about 20%, with a higher concentration of conserved residues at the module interface of Ig-GH9_CbhA. This finding seems to lend support to the hypothesis that the function of Ig-like module might lie in the structural stabilization of the catalytic module. Only three of 10 intermodule hydrogen bonds in Ig-GH9_CbhA are strictly conserved in Cel9A. As depicted in Figure 5 residues D262 and D264 form two of these bonds with G661 and Y676, respectively. The latter bond is particularly interesting since it stabilizes the position of the highly conserved W678, which in turn directly interacts with the substrate near the cleavage site at the −1 position, suggesting that its presence is essential for the integrity of the active site.

ACKNOWLEDGMENT

Use of the Advanced Photon Source was supported by the U.S. department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-ENG-38.

REFERENCES

1. Himmel, M. E., Ruth, M. F., and Wyman, C. E. (1999) Cellulase for commodity products from cellulosic biomass. *Curr. Opin. Biotechnol. 10*, 358−364.
2. Kataeva, I. A., and Ljungdahl, L. G. (2003) The *Clostridium thermocellum* cellulosome: a multi-protein complex of module-composed components in Protein Structures: Kaleidoscope of Structural Properties and Functions. (Uversky, V. N., Ed.) pp 1−16, Research Singpost, Kerala, India, in press.
3. Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (1998) Characterization and affinity applications of cellulose-binding domains. *J. Chromatogr. B., Biomed. Sci. Appl. 715*, 283−296.
4. Kataeva, I., Li, X.-L., Chen, H., Ljungdahl, L. G. (1998) CelK- a new cellobiohydrolase from *Clostridium thermocellum* cellulosome: role of N-terminal cellulose-binding domain in *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmiya, K., Sakka, K., Kobayashi, Y., and Karita, S., Ed.) pp 456−460, UniPublishers Co., Tokyo, Japan.
5. Matuschek, M., Sahm, K., Zibat, A., and Bahl, H. (1996) Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolases with S-layerlike domains. *Mol. Gen. Genet. 252*, 493−496.
6. Kataeva, I. A., Seidel, R. D., III, Shah, A., West, L. T., Li, X. L., and Ljungdahl, L. G. (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Appl. Environ. Microbiol. 68*, 4292−4300.
7. Tokatlidis, K., Dhurjati, P., and Béguin, P. (1993) Properties conferred on *Clostridium thermocellum* endoglucanase CelC by grafting the duplicated segment of endoglucanase CelD. *Protein Eng. 6*, 947−952.
8. Bayer, E. A., Shoham, Y., and Lamed, R. (2000) Cellulose-decomposing prokaryotes and their enzyme systems in *The Prokaryotes: An Involving Electronic Resource for the Microbial Community*, 3rd ed. (latest update release 3.7, September 2001) (Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., Eds.) Springer-Verlag, New York.
9. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J. Bacteriol. 180*, 1709−1714.
10. Tomme, P., Kwan, E., Gilkes, N. R., Kilburn, D. G., and Warren, R. A. (1996) Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo- and exoglucanase activities. *J. Bacteriol. 178*, 4216−4223.
11. Kataeva, I., Li, X. L., Chen, H., Choi, S. K., and Ljungdahl, L. G. (1999) Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum:* Evidence for gene duplication and recombination. *J. Bacteriol. 181*, 5288−5295.
12. Kleywegt, G. J., Zou, J. Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T., and Jones, T. A. (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6Å resolution, and a comparison with related enzymes. *J. Mol. Biol. 272*, 383−397.
13. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T., and Jones, T. A. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science 265*, 524−528.
14. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. (1998) High-resolution crystal strutures reveal how a cellulose chain is bound in a 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol. 275*, 309−325.
15. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanisms of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol. 4*, 810−818.
16. Mandelman, D., Belaich, A., Belaich, J. P., Aghajari, N., Driguez, H., and Haser, R. (2003) X-ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J. Bacteriol. 185*, 4127−4135.
17. Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Béguin, P., and Aubert, J.-P. (1992) Three-dimensional structure of a thermostable bacterial cellulase. *Nature 357*, 89−91.
18. Parsiegla, G., Belaich, A., Belaich, J. P., and Haser, R. (2002) Crystal structure of the cellulase Cel9M enlightens structure/function relationships of the catalytic modules in glycoside hydrolases. *Biochemistry 41*, 11134−11142.
19. Zverlov, V. V., Velikodvorskaya, G. V., Schwarz, W. H., Bronnenmeier, K., Kellermann, J., and Staudenbauer, W. (1998) Multidomain structure and localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *J. Bacteriol. 180*, 3091−3099.

20. Chayen, N. E. (1997) The role of oil in macromolecular crystallization. *Structure 5*, 1269−1274.
21. Otwinowski, Z., and Minor. W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276, 307−326.*
22. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D 55* (*Pt 4*), 849−861.
23. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr. 56* (*Pt 8*), 965−972.
24. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110−119.
25. Laboratory, S. D. (1994) The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D50*, 760−763.
26. McRee, D. E. (1999) Xtalview/Xfit- -A versatile program for manipulation atomic coordinate, and electron density. *J. Struct. Biol. 125*, 156−165.
27. Brunger, A. T., Karplus, M., and Petsko, G. A. (1989) Crystallographic refinement by simulated annealing: application to crambin. *Acta Crystallogr. A45*, 50−61.
28. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with interactive structure refinement. *Nat. Struct. Biol. 6*, 458−463.
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzel, T. F., and Weng, J. (1987) in *Crystallographic Databases − Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., and Sievers, R., Eds.) pp 107−132, Data Commission of the International Union of Crystallography, Bonn.
31. Tuka, K., Zverlov, V. V., Bumazkin, B. K., Velikodvorskaya, G. V., Strongin, A. Ya. (1990) Cloning and expression of *Clostridium thermocellum* genes coding for thermostable exoglucanases (cellobiohydrolases) in *Escherichia coli* cells. *Biochem. Biophys. Res. Commun. 169*, 1055−1060.
32. Béguin, P., and Alzari, P. (1998) The cellulosome of *Clostridium thermocellum. Biochem. Soc. Trans. 26*, 178−185.
33. Chauvaux, S., Souchon, H., Alzari, P. M., Chariot, P., and Béguin, P. (1995) Structural and functional analysis of the metal-binding sites of *Clostridium thermocellum* endoglucanase CelD. *J. Biol. Chem. 270*, 9757−9762.
34. Guerin, D. M., Lascombe, M. B., Costabel, M., Souchon, H., Lamzin, V., Béguin, P., and Alzari, P. M. (2002) Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate. *J. Mol. Biol. 316*, 1061−1069.
35. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science 249*, 380−386.
36. Li, X.-L., Chen, H., and Ljungdahl, L. G. (1997) Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase-hemicellulase complex. *Appl. Environ. Microbiol. 63*, 4721−4728.
37. Cutfield, S. M., Davies, G. J., Murshudov, G., Anderson, B. F., Moody, P. C. E., Sullivan, P. A., and Cutfield, J. F. (1999) The structure of the exo-β-(1,3)-glucanase from *Candida albicans* in native and bound forms: Relationship between a pocket and groove in family 5 glycosyl hydrolases. *J. Mol. Biol. 294*, 771−783.
38. Dominguez, R., Souchon, H., Spinelly, S., Dauter, Z., Wilson, K. S., Chauvaux,S., Béguin, P., and Alzari, P. M. (1995) A common fold and similar active site in two distinct families of β-glycanases. *Nat. Struct. Biol. 2*, 569−576.
39. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model. 15*, 132−134, 112−113.
40. DeLano, W. L. (2001) DeLano Scientific LLC, San Carlos, CA. http://www.pymol.org.
41. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein−ligand interactions. *Protein Eng. 8*, 127−34.
42. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multible sequence alignments in PostScript. *Bioinformatics 15*, 305−308.

BI030202I