Crystal Structure of the Cellulase Cel9M Enlightens Structure/Function Relationships of the Variable Catalytic Modules in Glycoside Hydrolases

G. Parsiegla,[‡,§] A. Belaïch,[§,||] J. P. Belaïch,[§,||] and R. Haser*[,⊥,#]

Institut de Biologie Structurale et Microbiologie, Laboratoire d'Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, 31 Chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France, Université de Provence, Place Victor Hugo, 13331 Marseille Cedex 03, France, Laboratoire de Bioénergétique et Ingénierie des Protéines, Institut de Biologie Structurale et Microbiologie, Centre National de la Recherche Scientifique, 31 Chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France, Institut de Biologie et Chimie des Protéines UMR 5086, Laboratoire de Bio-Cristallographie, Centre National de la Recherche Scientifique, 31 Chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France, and Université Claude Bernard Lyon I, 7 passage du Vercors, 69367 Lyon Cedex 07, France

Received March 14, 2002; Revised Manuscript Received July 12, 2002

ABSTRACT: Cellulases cleave the β-1.4 glycosidic bond of cellulose. They have been characterized as endo or exo and processive or nonprocessive cellulases according to their action mode on the substrate. Different types of these cellulases may coexist in the same glycoside hydrolase family, which have been classified according to their sequence homology and catalytic mechanism. The bacterium *C. celluloyticum* produces a set of different cellulases who belong mostly to glycoside hydrolase families 5 and 9. As an adaptation of the organism to different macroscopic substrates organizations and to maximize its cooperative digestion, it is expected that cellulases of these families are active on the various macroscopic organizations of cellulose chains. The nonprocessive cellulase Cel9M is the shortest variant of family 9 cellulases (subgroup 9_C) which contains only the catalytic module to interact with the substrate. The crystal structures of free native Cel9M and its complex with cellobiose have been solved to 1.8 and 2.0 Å resolution, respectively. Other structurally known family 9 cellulases are the nonprocessive endo-cellulase Cel9D from *C. thermocellum* and the processive endo-cellulase Cel9A from *T. fusca*, from subgroups 9_{B1} and 9_A, respectively, whose catalytic modules are fused to a second domain. These enzymes differ in their activity on substrates with specific macroscopic appearances. The comparison of the catalytic module of Cel9M with the two other known GH family 9 structures may give clues to explain its substrate profile and action mode.

Cellulose is the most abundant renewable carbon source on earth and is an excellent target for the production of energy by biomass conversion. In nature, cellulose fibers are part of plant cell walls in which they are embedded in hemicelluloses and lignin. The fibers are composed of β-1,4-connected D-glucose residues which are proposed to be organized in crystalline and amorphous regions. Cellulases, which hydrolyze the β-1,4-glycosidic bond in cellulose, may be used to convert cellulose fibers to small fragments under mild conditions.

Cellulases have been classified on the basis of their action mode on the substrate into endo or exocellulases and into processive or nonprocessive cellulases. Endo and exocellulases attack the cellulose chain randomly or at one extremity, respectively. Processive cellulases stay attached to the chain after the initial hydrolysis step and continue to successively release small cellulose fragments from the nonreducing or the reducing end of the chain. Both kinds of endocellulases, processive and nonprocessive ones, have already been characterized, whereas up to now only processive exocellulases have been reported. These different types of cellulases may act in a cooperative way on cellulose (*1*−*3*) and are simultaneously secreted by cellulose-digesting organisms to increase their efficiency.

Glycoside hydrolases (GH) cleave glycosidic bonds with either inversion or retention of the configuration at the reducing end of the substrate (*4*). They have been currently grouped into 88 families according to their structural fold and catalytic mechanism, using hydrophobic cluster analysis and sequence alignments (*5*, *6*). Cellulases are present in at least 13 of these families. It has been shown that endo and exocellulases (families 6 and 7), and that processive and nonprocessive endocellulases (family 9) may coexist in the same family (*7*−*12*). These different cellulolytic activities

[†] The coordinates and structure factors of Cel9M (1IA6) and the Cel9M/cellobiose complex (1IA7) have been deposited at the Protein data bank for immediate release (PDB identification number in brackets).

* Corresponding author. Address: IBCP-CNRS, Laboratoire de Bio-Cristallographie, 7, passage du Vercors, 69367 Lyon cedex 07, France. Phone: +33 4 72 72 26 08. Fax: +33 4 72 72 26 16. E-mail: r.haser@ibcp.fr.

[‡] Laboratoire d'Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique.

[§] Université de Provence.

[||] Laboratoire de Bioénergétique et Ingénierie des Protéines, Centre National de la Recherche Scientifique.

[⊥] Laboratoire de Bio-Cristallographie, Centre National de la Recherche Scientifique.

[#] Université Claude Bernard Lyon I.

[1] Abbreviation: GH, glycoside hydrolase.

10.1021/bi025816m CCC: $22.00 © 2002 American Chemical Society
Published on Web 08/21/2002

inside the same GH family are the result of addition of new domains or by variations in the architecture of some surface loops that change the active site topology, and therefore the interactions with the substrate.

The mesophilic bacterium *Clostridium cellulolyticum* uses multienzyme complexes called cellulosomes to digest cellulose (*13*, *14*). These complexes contain a variety of cellulases attached to a scaffolding module that support a cellulose-binding domain. The cellulases are attached to the scaffolding module by specific protein/protein interactions between their C-terminal docking domain (called dockerin) and the attachment domains (called cohesins) of the scaffolding module (*15*). Up to now, the sequences of 10 cellulases, namely, cel5A, cel5D, cel48F, cel8C, cel9G, cel9E, cel9H, cel9J, cel9M, and cel5N, have been identified in the genome of *C. cellulolyticum*. They all belong to only four GH families, namely, 5, 8, 9, and 48. The crystal structure of at least one representative of each of these GH families has been solved (Cel8A from *Clostridium thermocellum*, *16*; Cel48F from *C. cellulolyticum*, *17*; Cel5A from *C. cellulolyticum*, *18*; Cel9A from *Thermonospora fusca*, *12*; and Cel9D from *C. thermocellum*, *11*), identifying the main fold of the central domain.

The large number of different cellulases belonging to GH family 5 and family 9 present in the same organism is consistent with the ability of these families to vary their active site architectures. Presuming a cooperative digestion of cellulose, one can expect these enzymes to be designed to act on different cellulose conformations or for different attack modes. The crystal structures of the active site variants present in these families are therefore of special interest in the understanding of the cooperative attack on cellulose (*19*, *20*).

The most abundant family among the sequenced cellulases in *C. cellulolyticum* is GH family 9, which has been subdivided into four subgroups, according to the domain composition of the enzymes (*21*). Two structural variants of this family have already been solved in other organisms: Cel9D from *C. thermocellum* (*11*) and Cel9A (formerly called E4) from *T. fusca* (*12*) belonging to subgroups 9_{B1} and 9_A, respectively, and whose catalytic module is fused to a second domain. Here we report the crystal structure of the catalytic module of Cel9M from *C. cellulolyticum*, a nonprocessive cellulase which represents the first solved structure of a short GH family 9 cellulase of subgroup 9_C, which, besides its dockerin, only contains a catalytic module and is the first published cellulase of GH family 9 from *C. cellulolyticum*.

EXPERIMENTAL PROCEDURES

Enzyme preparation and crystallization. The fully active catalytic domain of Cel9M without its dockerin was obtained from the complete recombinant enzyme expressed in *E. coli*, which cleaves autocatalytically its C-terminus by removing its dockerin, as described elsewhere (*20*). The total length of the truncated molecule was analyzed using electron spray mass spectrometry, which indicated a minor sequence error and suggested two fragments differing in one amino acid corresponding to the sequence 1 to 440 or 441 (theoretical mass, 48 657 (1−441) and 48 529 Da (1−440); mass found, 48 644 and 48 516 Da).

Table 1: Statistics of Data Collection and Refinement[a]

	Cel9M	Cel9M/Cellobiose[b]
resolution [Å]	30−1.80	16−2.0
space group	$P2_1$	$P2_1$
cell: *a*, *b*, *c* [Å]	52.35, 53.01, 71.68	51.85, 52.38, 70.97
α, β, γ [deg]	90, 108.70, 90	90, 108.46, 90
completeness [%]	97.8 (88.7)	98.9 (94.3)
redundancy	5.0 (3.0)	4.6 (3.2)
$I/\sigma I$	9.0 (2.9)	7.1 (2.2)
R_{meas}[c] [%]	8.3 (29.5)	12.0 (41.2)
no. water molecules	262	326
$B_{iso,ave}$ [Å^2]	16.3	14.1
R_{free} [%]	18.9	22.0
rms bonds [Å]	0.004	0.005
rms angles [deg]	1.20	1.20
R [%]	15.8	16.8

[a] Each data set was collected from a single crystal. The values in parentheses correspond to the highest shell of resolution (1.86−1.80 or 2.07−2.00 Å, respectively). [b] Collected at 100 K. [c] R_{meas} according to (*46*).

Initial crystals were obtained by hanging drop vapor diffusion techniques at 18 °C, using a solution of the C-terminally truncated Cel9M at a concentration of 10 mg/mL. Small plate-like crystals grew under the following conditions: 1.6 M ammonium sulfate, 0.1 M sodium chloride, and 0.1 M HEPES at pH 7.5. Crystals suitable for X-ray diffraction studies were obtained using seeding techniques with 1.4 M ammonium sulfate, 20 mM calcium chloride, and 0.1 M HEPES at pH 7.5. Crystals of the complex with cellobiose were grown in a buffer composed of 1.6 M ammonium sulfate, 10 mM calcium chloride, 0.1 M HEPES, pH 7.5, 10% ethylene glycol, and 10 mM cellobiose. The crystal was washed with the same buffer containing 20% ethylene glycol immediately before mounting in a cryo loop.

Data collection. All diffraction data were collected using copper K_α radiation produced by a Nonius FR591 rotating Anode generator, equipped with Osmic confocal mirrors, an Oxford cryo-system, and a MAR345 image plate detector. The data of native Cel9M were obtained at room temperature, while the data of the cellobiose complex were collected under cryo conditions at 100 K. Two data sets from two fragments of the same native crystal were collected and merged to obtain a higher completeness and redundancy. The intensities were integrated using the program DENZO (*22*). The data were scaled and reduced, and an R_{free} set (*23*) was selected using the CCP4 program suite (*24*). All statistics for both datasets are summarized in Table 1.

Phasing and refinement. An initial model of Cel9M containing all the side chains was constructed with the program Modeler4 (*25*), using the $(\alpha\alpha)_6$-barrel domain of the catalytic module of the cellulase Cel9A (former called E4) from *T. fusca* (*12*) as the structural template. We used this model for performing the molecular replacement with the program AMORE (*26*) from the CCP4 program package, using the reflections in the range from 4 to 8 Å resolution of the native data set collected at room temperature. Only a single solution appeared in the rotation and translation search, which was further refined with the program CNS (*27*). We performed a torsion angle dynamic at 5000° using the data from 30 to 2.5 Å with subsequent positional refinement, which decreased the *R*-factor and the R_{free} to 33.8% and 44.8%, respectively.

This pre-refined structure was then used as the starting model for running the program ARP_WARP V5.1 (*28*) with all the data up to 1.8 Å. Exactly 403 of the residues could be fitted by this method, which decreased the R and R_{free} to 28.0% and 33.0%, respectively. Missing residues and ions were then introduced using the program O (*29*) and refined with the program CNS. Water molecules were assigned and kept if they were still present at 1.0 σ density level in the $(2F_o-F_c)$ electron density map and if they made at least one H-bond with a non water atom of the model. In the last refinement steps, multiple side-chain conformations were introduced. The model of the native Cel9M catalytic module was refined to a final R-factor of 15.8% with a R_{free} of 18.9% and contained 423 of the 441 expected residues, 1 zinc ion, 1 calcium ion, 1 nickel ion, 1 sulfate ion, and 285 water molecules. Residues 240−247 and 374−375 could not be detected, as they are located in two obviously flexible loops (residues 239−248 and 371−379, respectively). The C-terminus could be traced up to residue G431 at the end of an α-helix.

The structure of the cellobiose complex was solved using the program CNS with the free molecule as a template. The cellobiose molecule could be detected in the $(2F_o-F_c)$ map at 0.9 σ level after a rigid body refinement and was therefore inserted in the model. The residues of the missing loops could be successfully constructed in the $(2F_o-F_c)$ map at 0.7 σ density level. The completed model was then refined and multiple side-chain conformations inserted. The final model contained the complete sequence of Cel9M up to Gly431, 325 water molecules, the same ions as observed in the free model, one cellobiose, and three ethylene-glycol molecules. The final R-factor was 16.8%, and the R_{free} was 22.0%.

Characterization of the metal ions. The metal ions occupying three sites were determined on the basis of the anomalous signals of these sites and their coordination spheres. We used a data set collected with Cu Kα radiation at 1.54 Å wavelength and a data set collected with synchrotron radiation at 0.92 Å wavelength at the FIP beamline at the ESRF Grenoble (data not shown).

One site in Cel9M is a putative Ca site which is highly conserved in the entire family. In the anomalous map calculated from the data collected with Cu Kα radiation, the signal at the putative Ca site was higher than 4 σ, whereas the signal in the two other ion sites was very weak or completely absent. In the anomalous map calculated with the data collected at 0.92 Å wavelength, the signal at the putative Ca site was significantly weaker, but the two other sites showed a good signal greater than 4 σ. The presence of one zinc atom per protein molecule was detected earlier by plasma emission spectroscopy, so we assigned a zinc ion in view of the ligands and its coordination sphere. The anomalous signal of the ion in the third site corresponded well to an absorption at the nickel K-edge, and the observed square pyramidal coordination is common for nickel ions (*30*). Furthermore, nickel incorporation due to contact of the enzyme with a nickel-affinity column during the purification procedure appears quite plausible.

Computer graphics. All atomic model figures in this article were created using the programs MOLSCRIPT (*31*) and RENDER 3D (*32*), whereas the surface representations were calculated with the program GRASP (*33*).

RESULTS AND DISCUSSION

The entire cellulase Cel9M of *C. cellulolyticum* is composed of a catalytic module and a dockerin. We crystallized the active catalytic module without its dockerin, corresponding to a chain of about 441 amino acids or 48.6 kDa. Data sets of the free truncated enzyme and of its cellobiose complex were collected and their structures refined to 1.8 and 2.0 Å resolution, respectively. The C-terminus in both structures could be traced only up to residue 431, leaving the positions of the last 10 residues undefined.

Overall structure. The overall globular form of the catalytic module of Cel9M resembles a flat cone with the active site cleft located at its base. Its main structural motif is an $(\alpha\alpha)_6$ barrel, typical for the central domain in family 9 glycoside hydrolases (*11*, *12*). The six barrel helices are connected by short sequences of about 3 residues at the C-terminal side of the inner helices and long sequences at their N-terminal side. The length of the N-terminal connections varies between 14 and 21 residues on the half of the barrel formed by the helix connections 195−208, 238−259, and 291−305 and between 43 and 80 residues on the other half. This asymmetric architecture results in a tilt of the surface covering the N-terminal side of the inner helices by an angle of about 40° relative to the central axis of the $(\alpha\alpha)_6$ barrel (Figure 1). This architecture allows the extension of the active site cleft, which is located here, further than in a noninclined $(\alpha\alpha)_6$-barrel surface. The two extremities are bound by two flexible loops ranging from L239 to M248 and T371 to K379. These loops could not be traced completely in the electron density of the native, free structure but were entirely built in the cellobiose complex. It should be stressed that the B-factors of these loop residues were still more than two times greater than the average in the complex.

Metal ions. The presence of three different types of metal ions, namely, calcium, zinc, and nickel, could be detected. The calcium ion is bound in a cavity close to the active site cleft near the putative subsite −4. Its coordination sphere is a capped trigonal biprism, formed by the ligands O−Asp257 (2.4 Å), O−Ser208 (2.4 Å), OG−Ser208 (2.6 Å), OD1−Asp211 (2.8 Å), OD2−Asp211 (2.5 Å), and two water molecules each, at a distance of 2.4 Å, both H-bonded by other amino acids (Figure 2). Even though its role is probably only structural, this calcium ion is highly conserved in the GH family 9.

The zinc ion is coordinated in a tetrahedral manner by two cysteine and two histidine residues, namely, via SG−Cys22 and SG−Cys38, each at 2.3 Å, and ND1−His39 and NE1−His55, each at 2.1 Å from the metal ion (Figure 2). This kind of coordination is typically observed in zinc fingers from transcription factors or other proteins interacting with nucleotides where the zinc site stabilizes a rigid arrangement of secondary structure elements. In Cel9M the zinc site appears as an anchor point to stabilize the coil formed by residues 21−63, one of the three long helix connections which extends the active site further away from the $(\alpha\alpha)_6$-barrel axis. It also stabilizes, via His55, the loop containing Asp56 and Asp59 which is involved in catalysis.

The nickel site is located at the N-terminus of the protein, with a square pyramidal coordination formed by NE2−His4 (2.1 Å), OD2−Asp343 (2.0 Å), N−Ala1 (2.2 Å), O−Ala1

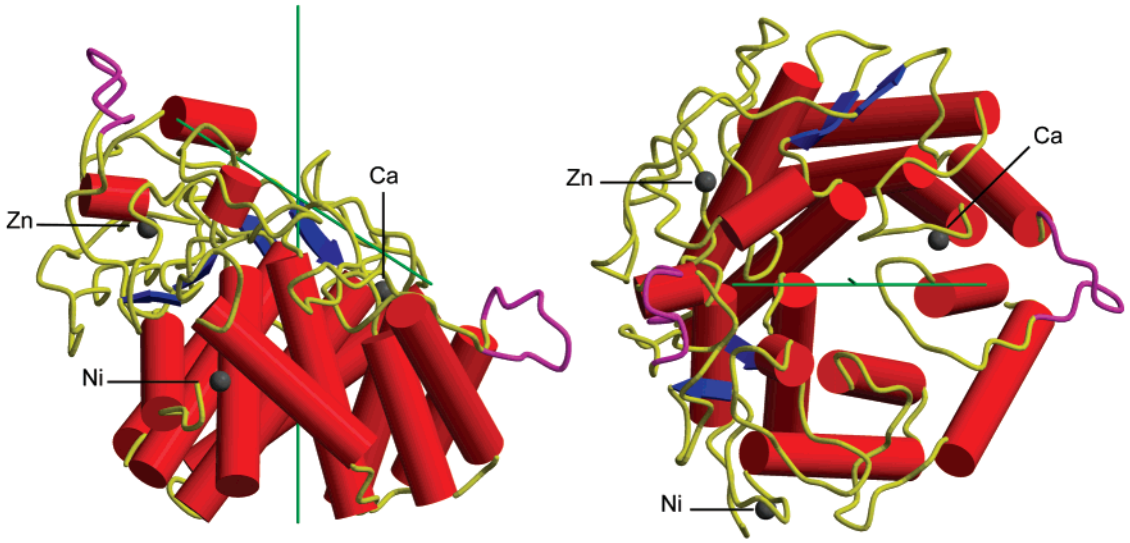

FIGURE 1: Sketch from the side and the top of Cel9M showing the secondary-structure elements (α-helix, red cylinders; β-sheet, blue flash) and the ion positions (gray). A virtual axis through the helix barrel and one along the active site are colored green.

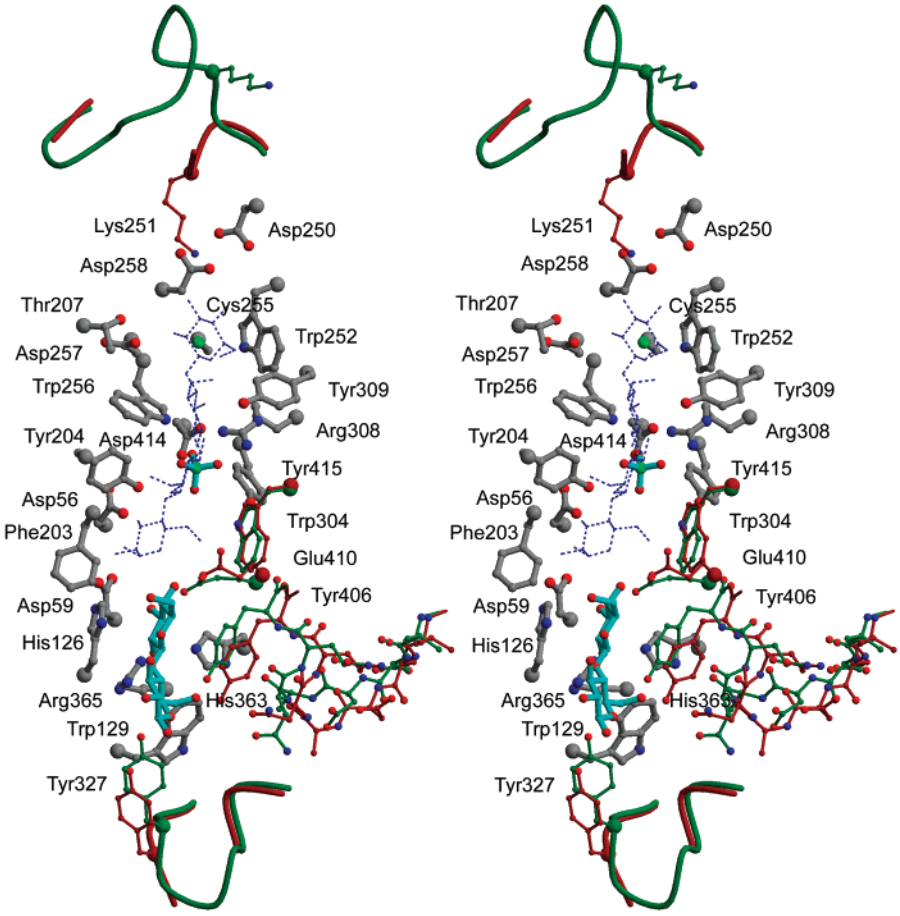

FIGURE 2: Stereoview of the active site cleft of Cel9M. The residues interacting in the active site are indicated, as well as the position of the bound cellobiose and the sulfate ion (cyan). Different residue positions in the free enzyme (red) and in the cellobiose complex (green) are shown. The partial (red) and complete (green) flexible loops 370−378 and 239−259 are represented as main-chain tracings. The modeled cellotetraose from Cel9A is represented in broken blue lines.

(2.4 Å), and a water molecule at a distance of 2.1 Å, which forms a H-bond with Asp338 (Figure 2). As the full-length protein was initially purified by an affinity chromatographic step on a nickel-containing column, it is possible that the bound nickel ion is an artifact of the purification procedure.

Active site architecture and cellobiose binding. The active site is formed by a cleft of about 35 Å long and 8−10 Å broad, which is located on the N-terminal site of the inner $(\alpha\alpha)_6$-barrel helices. Several aromatic residues are present along the cleft to serve as possible partners for stacking interactions with the substrate (Figure 3), as it is often found in active site clefts of glycoside hydrolases and other carbohydrate binding proteins (*34*−*38*). The similarity between the active site clefts of Cel9A and Cel9M allowed

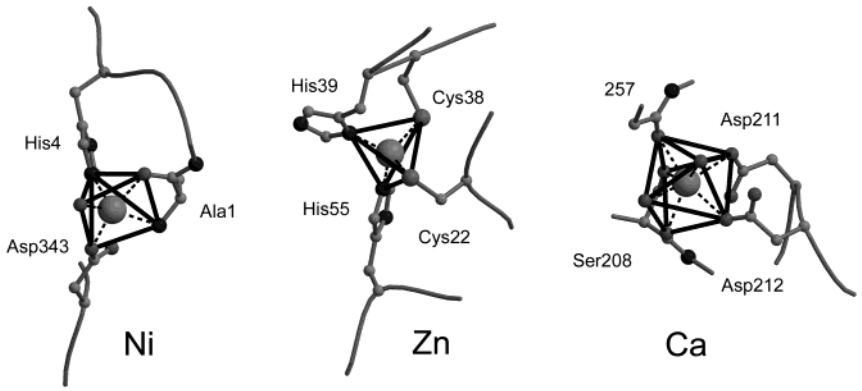

FIGURE 3: Representation of the coordination spheres of their bound nickel (Ni), zinc (Zn), and calcium (Ca) ions.

us to model a cellotetraose into the active site of Cel9M based on the position of this cellooligosaccharide in Cel9A. The model shows that there is enough space in Cel9M for at least four glucose residues occupying subsites −1 to −4 (Figure 3). This modeling study is supported by data indicating that digestion of crystalline cellulose leads to cellotetraose as the main product (*20*). Three aromatic residues Tyr406, Trp304, and Trp252 are located at the same side of the cleft, ready to stack with sugar moieties in the putative subsites +1, −1, and −4, and one aromatic residue, namely, His126 at subsite +1, is located at the opposite side. The identity of the proton donor responsible for the cleavage of the glycosidic bond could be predicted by sequence comparison with other family 9 GH and was subsequently confirmed by the Cel9M/cellobiose complex, where Glu410 is interacting with the nonreducing end of the cellobiose molecule. A water molecule is bound between Asp56 and Asp59 and is in a favorable position to attack the reaction intermediate from the other side of the active site cleft, as expected for the inverting mechanism. These observations in Cel9M are in perfect agreement with the previously proposed cleavage mechanism described for Cel9A from *T. fusca* (*12*).

One cellobiose molecule was found in the active site cleft of the complex (Figure 3). The electron density of the sugar was clear enough to permit the determination of the orientation of the molecule. It occupies subsites +2 and +1 according to the nomenclature established for glycoside hydrolases (*39*) and follows the same orientation as that observed in the processive endocellulase Cel9A. The glucose unit in subsite +1 is well stabilized by stacking interactions with His126 and Tyr406 and by several H-bonds: the nonreducing end (Glc+1)O4 by an H-bond with the putative proton donor (Glu410)OE1, which forms with (Glu410)OE2 a second H-bond with (Glc+1)O3; (Glc+1)O2 is in H-bond distance with (His363)NE2, (Arg365)NH1, and (Arg365)NH2. The glucose unit in subsite +2 is involved in a crystal contact with a neighboring Cel9M* molecule, forming a H-bond with its (Glc+2)O3 to (Lys251)NZ*, which is the only obvious stabilization for the sugar in this subsite. The weaker interaction in subsite +2 explains the observed less well-defined electron density for the sugar in this subsite.

A sulfate ion could be identified in the middle of the active site cleft in the putative subsite −2 and is bound to (Arg308)NH1 and NH2 and (Trp256)NE1. This ion which is present in both structures is probably an artifact induced by the crystallization conditions and may have prevented the binding of a second cellobiose molecule around subsite −2 in the cellobiose complex. Its presence indicates a possible stabilization of a negative charge in subsite −2.

Binding of the cellobiose molecule in the active center induces a movement of the region comprising residues 400−410. The largest shift in the peptide backbone concerns the α-carbon of Ala404 which moves 3.1 Å deeper in the cleft, while the α-carbons of other residues such as Tyr406 and Glu410 are only shifted by 1.0 or 0.8 Å, respectively. The H-bond which existed between (Glu410)OE1 and (Trp304)NE1 in the free enzyme is broken, and Glu410 is shifted by 2.3 Å (CD-atom position taken as a reference) to interact with (Glc+1)O4. Trp304 is more flexible in the complex, as indicated by the observed electron density. Tyr403 is shifted by at least 1.5 Å (CG-atom position as a reference), and its phenolic ring is very flexible, making it difficult to model the residue at 0.7 σ level in the $(2F_o-F_c)$ electron density map. Another movement concerns the loop T371 to K379 in the active site cleft where Tyr372 is displaced by 2.4 Å (OH atom taken as a reference). This loop is less flexible in the complex structure and could be entirely modeled.

Moreover, the movement of the region 400−410 creates a new crystal contact between Asn404 and Thr242* located in the flexible loop L239* to M248* of a neighboring molecule. This contact stabilizes and probably improves the electron density of the flexible loop in the cellobiose complex, which allowed us to model the principal conformation of this loop, even if not all of its residues are completely defined in the final $(2F_o-F_c)$ electron density map at the 1.0 σ level.

Comparison with known GH family 9 structures. The crystal structure of the catalytic module of the nonprocessive cellulase Cel9M of *C. cellulolyticum* is the first solved structure of a cellulase from GH family 9_C, in which the catalytic module is only an $(\alpha\alpha)_6$ barrel which is not fused to a second, nondockerin domain as observed in the two other known crystal structures from GH family 9 cellulases (Figure 4A). A comparison of their activities on substrates with different macroscopic organizations and their proposed action mode are summarized in Table 2. The coexistence of processive and nonprocessive cellulases in the same family has already been shown in GH families 6 and 7, where the processive action is modulated by the length of some loops, which form a tunnel in the processive enzymes and an open cleft in the nonprocessive ones. This tunnel attaches the processive cellulase to the cellulose chain during all catalysis steps (*7*−*10*).

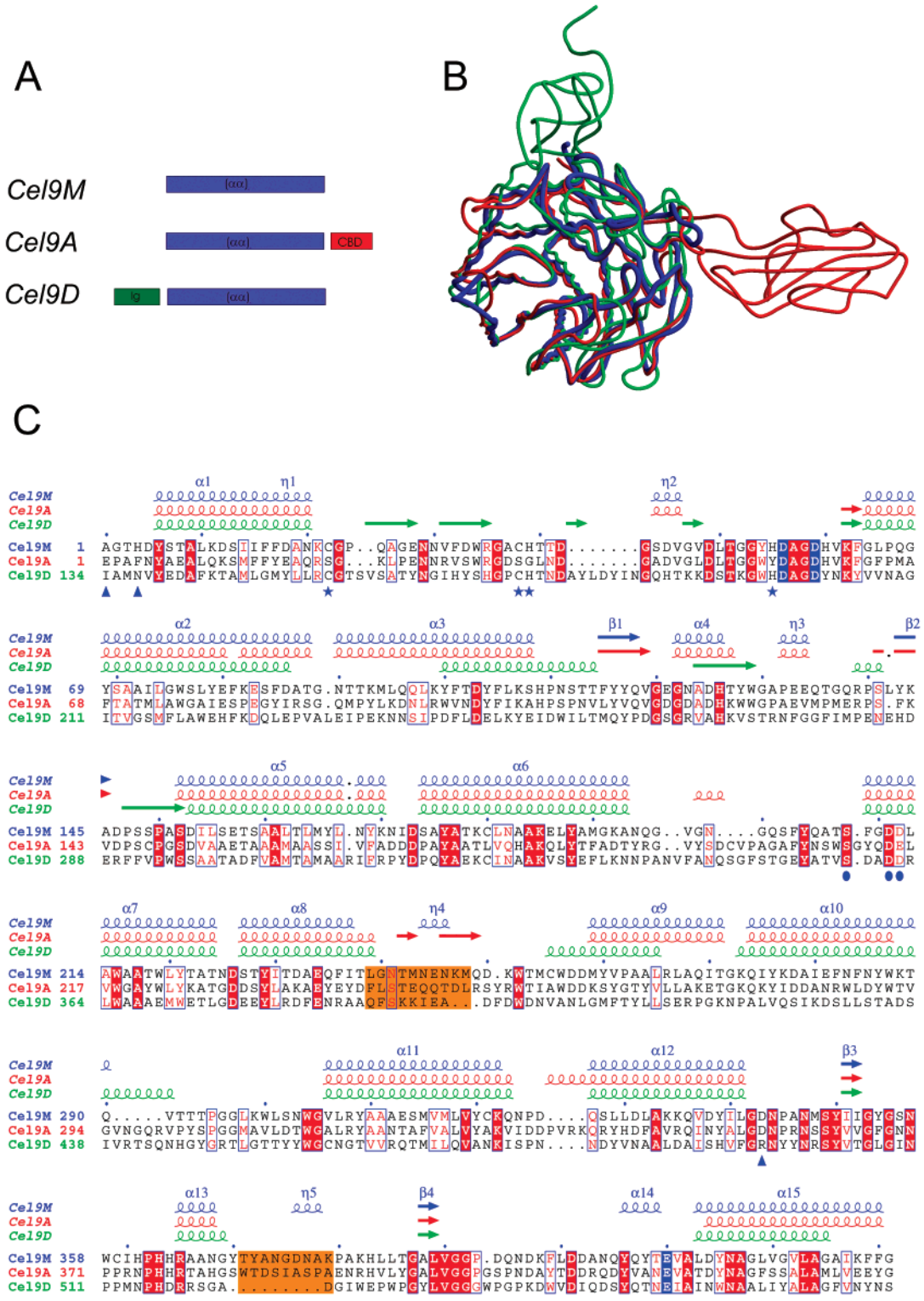

FIGURE 4: (A) Scheme of the locations of domains fused to the catalytic modules of Cel9A and Cel9D compared to Cel9M. (B) Comparison of the domain compositions in known structures of GH family 9: Cel9A from *T. fusca* in red, Cel9D from *C. thermocellum* in green, and Cel9M from *C. cellulolyticum* in blue. For clarity, only Cα atoms are connected. C. Sequence comparison of the central $(\alpha\alpha)_6$-barrel of Cel9M (blue) with Cel9A (red) and Cel9D (green), as analyzed by the program ESPript (*48*). Metal ion position in Cel9M are indicated [nickel (▲), zinc (☆), and calcium (●)], as well as the catalytic residues (blue font) and the sequences of the two flexible loops (golden font).

The processive action of Cel9A from *T. fusca* has been proposed to be controlled by a class III cellulose binding domain of about 150 residues in length which is fused to the C-terminus of the central $(\alpha\alpha)_6$ barrel domain. It extends the active site cleft, and its additional sugar subsites may support the cellulose chain during its progression through the active site (Figure 4B) (*12*, *40*). The involvement of a second domain in GH family 9 is different from the variation of the length of the loops in GH families 6 and 7; both factors would allow a processive attack by the enzyme on the substrate. In Cel9A the second domain creates a flat surface with the catalytic module which seems to favor interactions

Table 2: Qualitative Comparison of Cellulases from GH Family 9: Proposed Action Mode and Activities on Substrates with Different Macroscopic Organizations[a]

		PASC	AVICEL	BMCC
Cel9D of *C. thermocellum*[b]	nonprocessive endo	high	low	very low
Cel9M of *C. cellulolyticum*[c]	nonprocessive endo	high (57)	good (2.5)	good (2)
Cel9A of *T. fusca*[d]	processive endo	high (54)	n.d.	good (6)

[a] PASC ≪ AVICEL < BMCC: the amount of crystalline cellulose in the substrate is increasing from left to right. [b] Activities observed by Carrard et al. (*47*) are not expressed in tabulated numbers but plots. The release of reducing sugars reaches a plateau when all cleavable (amorphous) regions are digested indicating a very low activity on the crystalline parts of AVICEL and BMCC. [c] Activities (U/μmol in parentheses) of Cel9M attached to its dockerin observed by Belaich et al., in contrast to CelD continues the release of reducing sugars in the digestion of AVICEL or BMCC after all amorphous sites are digested. The significantly lower values are typical for the activity on crystalline substrates (*20*). [d] Activities (U/μmol in parentheses) of the catalytic module of Cel9A fused to its family IIIc cellulose binding module as observed by Irwin et al. (*40*).

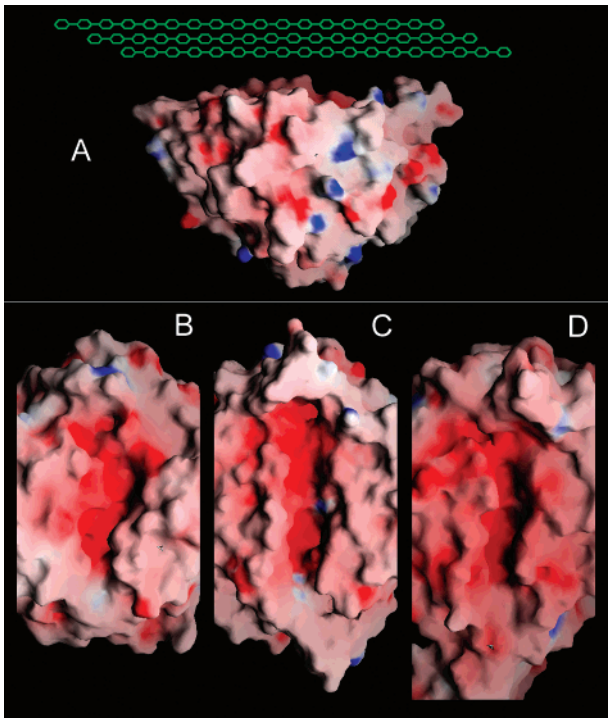

FIGURE 5: (A) Representation of the entire catalytic module of Cel9M. The flat surface of the active site is oriented to the top, highlighting the proposed interaction with a layer of crystalline cellulose (schematic). (B−D) Representation of the molecular surface of the active site clefts for endocellulase Cel9D (B), Cel9M (C), and the processive endocellulase Cel9A (D). Positive and negative potential are colored in red and in blue, respectively. The active sites are oriented with positive subsites to the bottom and negative ones to the top. The class III cellulose binding domain of Cel9A prolongs the active site cleft at the bottom but is not shown.

with crystalline cellulose and might explain its high activity on this substrate (*12*, *41*). The absence of the class III cellulose binding domain in Cel9M is therefore consistent with the nonprocessive action of this cellulase. The surface around the active site of Cel9M is very flat, like that of Cel9A, which explains its good activity on crystalline cellulose (*20*) (Figure 5A).

The nonprocessive endocellulase Cel9D from *C. thermocellum* contains a second domains appended to the catalytic module like the processive cellulase Cel9A, but in this case it is an immune globulin like folded domain of about 130 residues, which is fused to the N-terminal side of the $(\alpha\alpha)_6$-barrel domain (*11*). This domain is nearly perpendicular to the active site cleft and has an unknown function (Figure 4B). The resulting surface around the active site of Cel9D is not flat at all, which might be a good explanation for its poor activity on crystalline cellulose. It has been shown recently that the presence of a cellulose binding domain does not necessarily lead to a processive cellulase. The addition of a cellulose binding domain to Cel9D via a linker, far away from the active site cleft, does not change the nonprocessive character of this protein (*42*).

Despite its action mode, the amino acid sequence of the nonprocessive Cel9M $(\alpha\alpha)_6$ barrel is much closer to that of the processive endocellulase Cel9A (40% identity) than to that of the corresponding barrel of the nonprocessive Cel9D (23% identity). Sequence comparison of the three enzymes highlights the conservation of the proposed catalytic residues Asp56, Asp59, and Glu410 (Cel9M numbering) and that of the calcium site in all three structures (Figure 4C). However, the zinc site is only present in Cel9D and Cel9M. The additional calcium sites present in Cel9D have been reported to be important for its thermostability (*43*, *44*) and are not conserved in the structures of Cel9A and Cel9M from mesophilic bacteria.

All residues interacting with substrates in subsites +1, −1, and −2 are strictly structurally conserved in Cel9A and Cel9M. In Cel9D these residues are also conserved or are mimicked by homologous residues. Trp209 (Cel9A counting), which performs a stacking interaction with the substrate in subsite −3 in Cel9A, has been deleted in Cel9M and Cel9D. Interestingly, the stacking interaction of the substrate in subsite −4 with Trp252 is structurally conserved in Cel9M and in Cel9A, but not in Cel9D. The deletion of Trp209 possibly destabilizes subsite −3, which might therefore destabilize the nonproductive complex formation in the active site cleft of Cel9M and Cel9D. The stabilization of this nonproductive complex is regarded as a prerequisite for a processive action. It has been shown that a single mutation which destabilizes this non productive complex may change completely the processive character of an active site cleft (*45*). The absence of Trp209 in both nonprocessive cellulases may therefore point to another explanation for their action mode. This hypothesis has to be confirmed by further mutagenesis experiments.

Loop 370−378, located at the opposite extremity of the active site cleft, is entirely deleted in Cel9D. This loop closes the active site at subsite +2 in Cel9M and forms the junction with the cellulose binding domain in Cel9A. Thus, the active site of the endocellulase Cel9D has open boundaries on each side of its cleft (Figure 5B) and resembles a channel that widens toward the nonreducing end of the bound sugar chain. This part of the active site resembles more a trough than a cleft in Cel9A (Figure 5D) and Cel9M (Figure 5C). It begins with a steep slope, then continues such as a channel which

is just wide enough to hold the sugar chain, and finally is blocked by a loop. The active site of Cel9M is slightly wider than that in Cel9A but more closed than that in Cel9D. A less restrained access to the active site is observed in other endo acting glycosyl hydrolases and may possibly favor this kind of attack. The flexible loops 239−249 and 370−378, blocking both extremities of the active site cleft, may be involved in the binding of a long cellulose chain during an endo attack in Cel9M. Loop 370−378 contains also an exposed tyrosine, whose phenylic ring is well orientated to stack with a sugar residue. The two flexible loops may therefore serve as "catcher regions" to feed the active site with substrate. In conclusion, the structure of Cel9M explains very well its nonprocessive activity on crystalline cellulose.

ACKNOWLEDGMENT

We thank Dr. Michel Becchi (IBCP, Lyon, France) for the mass spectrometric analysis of the truncated Cel9M and Dr Louis Germanique (CEREGE, Marseille, Plateau de l′Arbois, France) for the plasma emission spectrometry experiments.

REFERENCES

1. Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996) Identification of two functionally different classes of exocellulases, *Biochemistry 35*, 586−592.
2. Henrissat, B., Driguez, H., Viet, C., and Schulein, M. (1985) Synergism of cellulases from Trichoderma reesei in the degradation of cellulose, *Bio/Technology 3*, 722−726.
3. Kim, E., Irwin, D. C., Walker, L. P., and Wilson, D. B. (1998) Factorial optimization of a six-cellulase mixture, *Biotechnol Bioeng 58*, 494−501.
4. McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis, *Curr. Opin. Struct. Biol. 4*, 885−892.
5. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J. 293*, 781−788.
6. Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases, *Biochem J. 316*, 695−696.
7. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990) Three dimensional structure of Cellobiohydrolase II from Trichermoderma reesei, *Science 249*, 380−86.
8. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Petterson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1994) The three-dimensional structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei, *Science 265*, 524−528.
9. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase, *Biochemistry 32*, 9906−9916.
10. Davies, G. J., and Schülein, M. (1995) in *Carbohydrate Bioengeneering* (Petersen, S. B., Svensson, B., and Pedersen, B., Eds.) pp 232−235, Elsevier Science, Amsterdam.
11. 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.
12. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from Thermospora fusca, *Nature Struct. Biol. 4*, 810−818.
13. Bayer, E. A., Morag, E., and Lamed, R. (1994) The cellulosome−a treasuretrove for biotechnology, *Trends Biotechnol. 12*, 379−386.
14. Belaich, J. P., Tardif, C., Belaich, A., and Gaudin, C. (1997) The cellulolytic system of Clostridium cellulolyticum, *J. Biotechnol. 57*, 3−14.
15. Fierobe, H. P., Pages, S., Belaich, A., Champ, S., Lexa, D., and Belaich, J. P. (1999) Cellulosome from Clostridium cellulolyticum: molecular study of the Dockerin/Cohesin interaction, *Biochemistry 38*, 12822−32.
16. Alzari, P. M., Souchon, H., and Dominguez, R. (1996) The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum, *Structure 4*, 265−275.
17. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution, *Embo J. 17*, 5551−62.
18. Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H. P., Belaich, J. P., Davies, G. J., and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5, *Structure 3*, 939−949.
19. Gaudin, C., Belaich, A., Champ, S., and Belaich, J. P. (2000) CelE, a multidomain cellulase from Clostridium cellulolyticum: a key enzyme in the cellulosome? *J. Bacteriol. 182*, 1910−5.
20. Belaich, A., Parsiegla, G., Gal, L., Villard, C., Haser, R., and Belaich, J.-P. (2002) Cel9M, a new family 9 cellulase of the Clostridium cellulolyticum cellulosome, *J. Bacteriol. 184*, 1378−1384.
21. Ding, S. Y., Bayer, E. A., Steiner, D., Shoham, Y., and Lamed, R. (1999) A novel cellulosomal scaffoldin from Acetivibrio cellulolyticus that contains a family 9 glycosyl hydrolase, *J. Bacteriol. 181*, 6720−9.
22. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., and Sweet, R. M., Eds.) pp 307−326, Academic Press, London.
23. Brünger, A. T. (1992) The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472−474.
24. Collaborative, C. P. N. (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr., Sect. D 50*, 760−763.
25. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
26. Navaza, J. (1994) AMoRe: an automated package for molecular replacement, *Acta Crystallogr. 50*, 157−163.
27. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D 54*, 905−921.
28. Lamzin, V. S., and Wilson, K. S. (1997) in *Methods in Enzymology* (Carter, C. W., and Sweet, R. M., Eds.) pp 269−305, Academic Press, London.
29. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building models in electron density maps and the location of errors in the models, *Acta Crystallogr., Sect. A 47*, 110−119.
30. Glusker, J. P. (1991) Structural aspects of metal liganding to functional groups in proteins, *Adv. Prot. Chem. 42*, 1−76.
31. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
32. Merrit, E. A., and Murphy, M. E. P. (1994) Raster3D version 2.0 a program for photorealistic molecular graphics, *Acta Crystallogr., Sect. D 50*, 869−837.
33. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins Struct. Funct. Gen. 11*, 282.
34. Kadziola, A., Abe, J. I., Svensson, B., and Haser, R. (1994) Crystal and molecular structure of barley a-Amylase, *J. Mol. Biol. 239*, 104−121.
35. Knegtel, R. M., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L., and Dijkstra, B. W. (1995) Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from Bacillus circulans strain 251 with natural substrates and products, *J. Biol. Chem. 270*, 29256−64.
36. Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg, G., Stahlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995) Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I, *Protein Sci 4*, 1056−64.
37. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from Trichoderma reesei, *J. Mol. Biol. 275*, 309−325.
38. Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (2000) Crystal structures of the

cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action, *Biochemistry 39*, 11238−46.
39. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases, *Biochem. J. 321*, 557−559.
40. 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*, 1709−14.
41. Gal, L., Gaudin, C., Belaich, A., Pages, S., Tardif, C., and Belaich, J. P. (1997) CelG from Clostridium cellulolyticum: a multidomain endoglucanase acting efficiently on crystalline cellulose, *J. Bacteriol. 179*, 6595−601.
42. Carrad, C., Koivula, A., Söderlund, H., and Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose, *PNAS 97*, 10342−10347.
43. Chauvaux, S., Beguin, P., Aubert, J. P., Bhat, K. M., Gow, L. A., Wood, T. M., and Bairoch, A. (1990) Calcium-binding affinity and calcium-enhanced activity of Clostridium thermocellum endoglucanase D, *Biochem. J. 265*, 261−5.
44. Chauvaux, S., Souchon, H., Alzari, P. M., Chariot, P., and Beguin, P. (1995) Structural and functional analysis of the metal-binding sites of Clostridium thermocellum endoglucanase CelD, *J. Biol. Chem. 270*, 9757−62.
45. Pages, S., Kester, H. C., Visser, J., and Benen, J. A. (2001) Changing a single amino acid residue switches processive and non-processive behavior of Aspergillus niger endopolygalacturonase I and II, *J. Biol. Chem. 276*, 33652−6.
46. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography, *Nat. Struct. Biol. 4*, 269−75.
47. Carrard, C., Koivula, A., Söderlund, H., and Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose, *PNAS 97*, 10342−10347.
48. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPRIT: multiple sequence alignments in Post Script, *Bioinformatics 15*, 305−308.

BI025816M