Crystal Structures of the Cellulase Cel48F in Complex with Inhibitors and Substrates Give Insights into Its Processive Action[†,‡]

G. Parsiegla,[§] C. Reverbel-Leroy,[||] C. Tardif,[||] J. P. Belaich,[||] H. Driguez,[⊥] and R. Haser*[,§]

Institut de Biologie et Chimie des Protéines, Laboratoire de Bio-Cristallographie, Centre National de la Recherche Scientifique UMR 5086 and Université Claude Bernard Lyon I, 7 passage du Vercors, 69367 Lyon Cedex 07, 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, and Université de Provence, Place Victor Hugo, 13331 Marseille Cedex 03, France, and Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique and Université J. Fourier de Grenoble, BP53, 38041 Grenoble Cedex 9, France

Received May 18, 2000; Revised Manuscript Received July 5, 2000

ABSTRACT: Cellulase Cel48F from *Clostridium cellulolyticum* was described as a processive endo-cellulase. The active site is composed of a 25 Å long tunnel which is followed by an open cleft. During the processive action, the cellulose substrate has to slide through the tunnel to continuously supply the leaving group site with sugar residues after the catalytic cleavage. To study this processive action in the tunnel, the native catalytic module of Cel48F and the inactive mutant E55Q, have been cocrystallized with cellobiitol, two thio-oligosaccharide inhibitors (PIPS−IG3 and IG4) and the cello-oligosaccharides cellobiose, -tetraose and -hexaose. Seven sub-sites in the tunnel section of the active center could be identified and three of the four previously reported sub-sites in the open cleft section were reconfirmed. The sub-sites observed for the thio-oligosaccharide inhibitors and oligosaccharides, respectively, were located at two different positions in the tunnel corresponding to a shift in the chain direction of about a half sugar subunit. These two positions have different patterns of stacking interactions with aromatic residues present in the tunnel. Multiple patterns are not observed in nonprocessive endo-cellulases, where only one sugar position is favored by aromatic stacking. It is therefore proposed that the aromatic residues serve as lubricating agents to reduce the sliding barrier in the processive action.

Cellulose is a highly stable polymeric component of plant matters and the most abundant renewable source of carbon and energy on earth. The β-1,4-glycosidic bonds of cellulose are hydrolyzed by various different cellulases, which have been classified according to their protein fold and can be found in at least 11 different glycosyl hydrolase families (*1*−*3*). Each cellulase degrading organism produces a set of specific cellulases which act synergistically in the digestion of this complex substrate (*4*, *5*). In general, their catalytic modules can be subdivided into two groups according to their cellulose digestion process: nonprocessive cellulases, simply called endo-cellulases, and processive cellulases, which include different exo-cellulases and processive endo-cellulases. Cel48F from *C. cellulolyticum* which was simply called CelF before, is the matter of this article. Its processive properties have been determined by measuring the release of reducing ends in a PAS-cellulose degradation, which clearly classified it as a processive cellulase (*6*). The endo activity of the enzyme was discovered by analyzing the increase in the fluidity of a CM-cellulose solution as a function of the amount of reducing sugars during hydrolysis, which followed the same slope as an endo-cellulase (*6*). Additionally, a competition between Cel48F and an endo-cellulase could be observed on cellulose sites accessible for endo-cellulases (*6*). These tests finally classified Cel48F as a member of the processive endo-cellulases, from which two 3D structures belonging to two different glycosyl hydrolase families have been solved to date. The first, E4 from *Thermonospora fusca*, belongs to family 9 (*7*) and the second, Cel48F from *C. cellulolyticum* to family 48 (*8*). These 3D-structures show a different architecture of their active site, when compared to nonprocessive endo-cellulases, to enable a processive endo-attack.

E4 from *T. fusca* is excreted into the bacterial surroundings together with various other cellulose digesting enzymes all of which possess a cellulose binding module (CBM)[1] connected by a flexible linker for attaching to cellulose. The active site of E4 from *T. fusca* is an open cleft which is extended by a second cellulose binding domain closely linked to the main catalytic domain (*7*). This domain belongs to

[†] This work has been supported by the European Commission (Biotechnology program BIO4-97-2303), the "Centre National de la Recherche" (CNRS) and the European Union TMR/LSF grant.

[‡] The coordinates and structure factors of the complexes of Cel48F with IG4 (1fce), PIPS-IG3 (1f90), cellobiitol (1fbo), cellobiose (1fae), cellotetraose (1f9d) and cellohexaose (1fbw) were deposited at the Protein Data Bank for immediate release (PDB identification number in brackets).

* To whom correspondence should be addressed.

[§] Institut de Biologie et Chimie des Protéines.

[||] Laboratoire de Bioénergétique et Ingénierie des Protéines.

[⊥] Centre de Recherches sur les Macromolécules Végétales.

[1] Abbreviations: CBM, cellulose binding module; CBD, cellulose binding domain; IG4, methyl 4-*S*-β-cellobiosyl-4-thio-β-cellobioside; PIPS−IG3, *p*-iodophenyl 4-*S*-β-cellobiosyl-1,4-dithio-β-D-glucoside; PIP, *p*-iodophenyl; PCR, polymerase chain reaction.

10.1021/bi001139p CCC: $19.00
Published on Web 08/24/2000

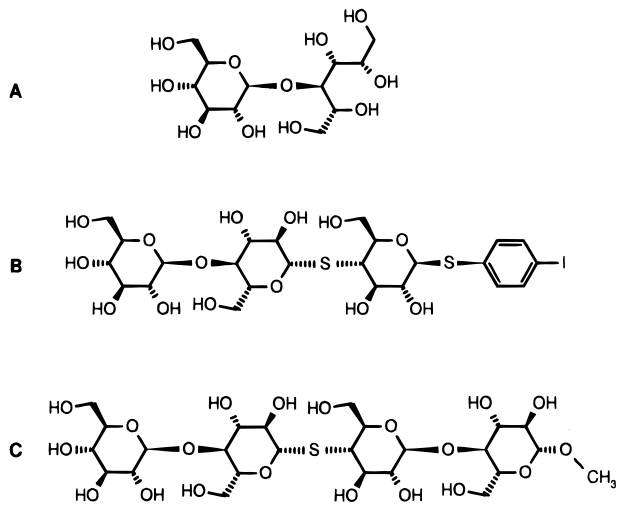

FIGURE 1: Chemical structure of the inhibitors used for cocrystallization (sugar rings are designed as flat rings). (A) Cellobiitol; (B) *p*-iodophenyl 4-*S*-*β*-cellobiosyl-1,4-dithio-*β*-D-glucoside, which is called PIPS−IG3 (unpublished); (C) methyl 4-S-*β*-cellobiosyl-4-thio-*β*-cellobioside, which is called IG4 (*19*).

the family III cellulose binding domains (CBD) and prolongs the active-site cleft located in the main catalytic domain by offering additional sugar binding sub-sites on its surface. A cellulose chain may therefore easily enter the long open cleft formed by these two domains to be cut initially in an endo manner, with the subsequent release of one part of the chain. The long binding site holding the remaining chain is composed of 11 sub-sites and is mostly located on the family III CBD. The large number of sub-sites have enough affinity for the cellulose chain to ensure that the chain remains bound while sliding through the cleft to the following exo-cut position. The second function of the family III CBD might be to offer a polar environment in order to favor the liberation of a single cellulose chain from the hydrogen-bonded network of its neighboring cellulose strands as proposed by Tormo et al. (*9*).

Cellulase Cel48F from the mesophilic anaerobic bacteria *C. cellulolyticum* is a major component of the multi-enzyme complex called the cellulosome, which is secreted by bacteria in order to digest cellulose (*6*, *10*). The cellulosome concept has been described first in *C. thermocellum* and later found to occur in other bacteria and fungi (*3*, *11*−*16*). In vivo, Cel48F is composed of two modules, a N-terminal catalytic module and a C-terminal docking module, called dockerin, which attaches the enzyme to the scaffolding protein (*6*, *17*). The whole enzyme (M_r, 77 626 Da) has been cloned and produced in *E. coli* (*6*, *8*). The 629 amino acid long catalytic module was crystallized in the presence of the thio-oligosaccharide inhibitor IG4 (Figure 1) and the 3D-structure solved (*8*, *19*). In contrast to the open cleft architecture of E4 from *T. fusca*, the active site of the catalytic domain of Cel48F from *C. cellulolyticum* is composed of a 25 Å long closed tunnel section which is followed by an open cleft region. The tunnel enables Cel48F to hold the substrate chain during the processive part of the mechanism. This architecture resembles the tunnel in the active site of processive exo-cellulases such as CBHI from *Trichoderma reesei* (*20*, *21*). Even though no structure of Cel48F with an open tunnel section has been crystallized and solved to date, it is expected that the tunnel part will initially be open like a cleft to fix a cellulose chain at an amorphous region. After the initial binding of cellulose, the tunnel is closed and Cel48F cleaves after the first endo-cut processively cellobiose units from the reducing end of the chain. Subsequently, it would be highly interesting to understand how the cellulose chain advances through the closed tunnel and what could be the driving force behind its displacement.

To get a detailed description of the substrate interactions in the tunnel, we solved the crystal structures of Cel48F in complex with different substrates and inhibitors and compared them with the structure of the Cel48F/IG4 complex. The native Cel48F catalytic module was cocrystallized with PIPS−IG3, cellobiose, and cellobiitol under the same conditions as described for the IG4 complex (Figure 1) (*19*). Mutation of the putative proton donor Glu55 to glutamine resulted in an inactive enzyme which allowed us to crystallize the mutant enzyme E55Q in the presence of the substrates cellotetraose or cellohexaose. The corresponding 3D-structures of the complexes showed differences in the sub-site positions in the tunnel section. This gave weight to the hypothesis that the specific arrangement of the aromatic residues present in the tunnel reduces the energy barriers between the sub-site positions to produce a surface tailored to reduce the sliding resistance of the substrate chain in the tunnel section.

MATERIALS AND METHODS.

Cel48F E55Q Construction. The double PCR mutagenesis technique (*22*) was used to generate a point mutation in the *celF* gene carried by the pETFp plasmid (*18*). The site-directed mutagenesis was performed using the complementary mutagenic oligonucleotides Fm55f (5′-**GTTACAA-CAAGT*CAA*GCTATGTCCTATTAT**-3′) and Fm55r (5′-**ATAATAGGACATAGCTT*G*ACTTGTTGTAAC**-3′) to introduce the CAA Gln codon in place of the GAA Glu 55 codon (mutagenic nucleotides in italic type). Expand High Fidelity PCR system (Roche) was used in the PCR experiments. The mutated 5′part of the cassette was synthesized using the forward primer F18 (5′-CCCTATACATATG-**GCTTCAAGTCCTGCAAAC**-3′) and the mutagenic Fm55f reverse primer. F18 carrying a *Nde*I site (underlined), was located at the beginning of the mature protein encoding sequence (hybridizing region in bold-face type). The mutated 3′ part of the cassette was synthesized using the mutagenic Fm55a direct primer and the reverse primer F22 (5′-**GCCTTGAAGTTTCATCTCC**-3′) located downstream from the unique *Kpn*I site in *celF* sequence. These two purified PCR products were mixed and amplified using F22 and F18 to generate the whole mutated cassette. The amplified fragment was purified (Qiaquick, Qiagen), digested by *Nde*I and *Kpn*I and then cloned between the *Nde*I and *Kpn*I sites into pETFp to replace the wild-type sequence. The resulting plasmid carrying the mutation was called pETF55. The successful insertion of the fragment was verified by sequencing.

The mutated protein E55Q was produced from BL21-(DE3)[pETF55] culture. It was purified and its enzymatic activity was assayed using phosphoric acid-swollen cellulose (PAS-cellulose) as substrate following the procedures previously reported for Cel48F wild type (*6*).

Crystallization and 3D-Structure Solution. Crystals suitable for X-ray diffraction have been grown using vapor diffusion

Table 1: Statistics of Data Collection and Refinement

	Cel48F/ cellobiitol	Cel48F/ PIPS−IG3	Cel48F/ IG4[a]	Cel48F/ cellobiose	E55Q/ cellotetraose	E55Q/ cellohexaose
resolution (Å)	30.0−2.3	20.0−2.5	30.0−2.0	20.0−2.0	60.0−2.3	10.0−2.0
cell: *a* (Å)	61.46	61.60	61.45	61.44	61.33	61.38
b (Å)	84.82	84.57	84.54	84.69	84.74	84.92
c (Å)	121.75	121.82	121.94	121.75	121.72	121.82
completeness (%)	95.7 (90.3)	99.6 (98.3)	99.9 (100)	99.8 (100)	99.9 (99.5)	94.1 (94.9)
redundancy	3.2 (2.6)	3.6 (3.7)	4.2 (3.9)	5.8 (5.2)	7.3 (6.8)	3.5 (3.4)
$I/\sigma I$	6.3 (1.7)	7.2 (2.8)	8.0 (3.0)	7.4 (1.8)	6.3 (2.0)	5.9 (1.6)
R_{sym} (%)	9.8 (41.8)	8.2 (27.7)	7.3 (24.8)	8.6 (41.3)	11.0 (37.7)	12.2 (46.9)
no. of water molecules	263	218	334	348	223	318
avg *B*-factor protein (Å^2)	23.5	14.4	14.8	18.3	21.1	14.2
avg *B*-factor inhibitor (Å^2)	24.1	53.1	33.6	15.9	25.4	18.5
R_{free} (%)	21.0	24.3	21.0	18.8	18.9	19.6
rms bonds (Å)	0.006	0.011	0.009	0.008	0.005	0.009
rms angles (deg)	1.3	1.4	1.5	1.5	1.4	1.5
R (%)	16.6	17.9	16.3	14.4	15.1	16.1

[a] Data included for comparison from Parsiegla et al. (*8*). Each data set was collected with a single crystal. The Space group of all crystals was $P2_12_12_1$. The values in parentheses correspond to the highest shell of resolution. $R_{sym} = \sum_{hkl}\sum_{i}|\langle I_{hkl}\rangle - I_{hkl,i}|/\sum_{hkl}\sum_{i}|I_{hkl}|$ where *i* is the number of reflections *hkl*.

and micro-seeding techniques using microcrystals of the Cel48F/IG4 complex which has been described previously (*19*)**.** All crystals were grown in the presence of 4 mM of substrate/inhibitor under conditions as described for the Cel48F/IG4 complex, with the exception of the PIPS−IG3 complex where due to the poor solubility of the compound a saturated solution was used. The diffraction datasets of the Cel48F complexes (PIPS−IG3, cellobiitol and cellobiose) were collected using CuKα radiation from a Rigaku RU-200 rotating anode generator with a graphite monochromator and a MAR 300 image plate scanner at 15 °C. The E55Q/cellotetraose complex dataset was collected using CuKα radiation from a Nonius FU 581 rotating anode generator equipped with a graphite monochromator and a MAR 345 image plate scanner at 15 °C, whereas the E55Q/cellohexaose complex dataset was collected at the EMBL Outstation of the DESY synchrotron, Hamburg, using radiation at a wavelength of 0.9 Å and a MAR 180 image plate scanner at 15 °C. The crystals were mounted in glass capillaries and were sufficiently stable to allow the collection of an entire dataset from one single crystal. All crystals were isomorphous to the previously reported Cel48F/IG4 complex (*8*) and belonged to the space group $P2_12_12_1$. The crystal data of all complexes are summarized in Table 1.

The data were processed and reduced using the DENZO program (*23*) and the CCP4 package (*24*). All 3D-structures of the complexes were solved using the XPLOR program (*25*) except for the E55Q/cellotetraose structure in which case the CNS program package was used (*26*). The refinement procedure began with 20 cycles of rigid body refinement, using the protein chain and the Calcium-ion position of Cel48F/IG4 complex as the starting model. The bulk-solvent correction (*27*) and the weighting factor WA were recalculated after each step of the refinement. The rigid body refinement was followed by a simulated annealing procedure at 3000 K, a manual fitting step for residues which were still out of density and the insertion of the mutation of Glu55 to glutamine if necessary. The Warp/Arp program (*28*, *29*) was then used to insert water molecules (except for the Cel48F/PIPS−IG3 complex where water positions were obtained using the "watpick" procedure of the XPLOR package). Water molecules were kept if their corresponding electron density reached a one sigma level in the $(2F_o - F_c)$ electron density map and they did not occupy map regions which probably belonged to substrate/inhibitor moieties. This selection significantly reduced the difference between the *R*-factor and the R_{free}-factor (*30*). Finally, the substrate/inhibitor molecules were inserted consecutively in the model, beginning with the most obvious positions at the leaving group side (+1, +2) and followed by a positional refinement. Accurate bond parameters for the thio-oligosaccharide PIPS−IG3 were taken from the Engh and Huber parameters (*31*) of the XPLOR package and the thioglycosidic bond parameters from the structure of 2,3,4,6-tetra-*O*-acetyl-1-*S*-benzhydroximoyl-α-D-glucopyranose (*32*). The program O (*33*) was used for all model building stages. After the completion of the model a restrained individual *B*-factor refinement was performed, followed by a final correction of the water molecule positions and an overall positional refinement.

The refined structures of the various complexes were superimposed, with the least squares function of the program O, by comparing the atoms of the protein backbone with the Cel48F/IG4 structure (*8*), which served as the reference, to elucidate differences in the positions of the substrate/inhibitor molecules and the side-chain orientations. Figures3 and 4 were calculated using the programs MOLSCRIPT and RASTER-3D (*34*, *35*).

RESULTS

Cel48F was cocrystallized with six different substrates/inhibitors, i.e., cellobiose, cellotetraose, cellohexaose, cellobiitol, PIPS−IG3, and IG4 (Figure 1) to visualize, analyze and compare substrate binding in the active center of Cel48F. The Cel48F/IG4 structure has already been described elsewhere (*8*), but is used and discussed here in order to compare inhibitor positions and interactions with the five new complexes. A summary of the data statistics and the refinement results are given in Table 1. Crystals of all complexes were obtained using a start concentration of 4 mM sugar/inhibitor in the droplet, with the exception of PIPS−IG3 where a saturated solution was used owing to its

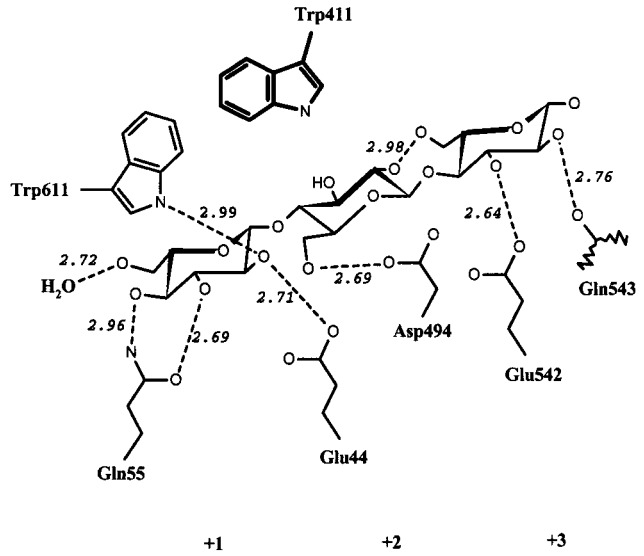

FIGURE 2: Schematic representation of H-bonds between the sugar subunits in the subsites +1 to +3 in the cellohexaose complex. All H-bond partners of the first coordination shell are shown, water molecules are only shown if they were directly in contact with a protein residue. The length of each H-bond is indicated in angstroms, aromatic residues which were involved in stacking interaction are shown in bold.

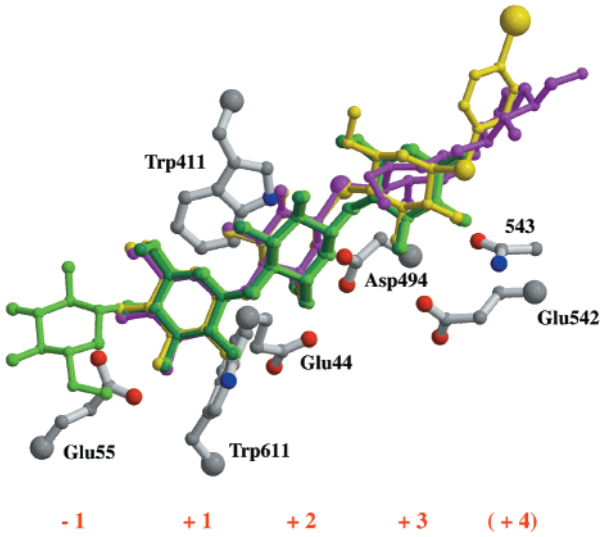

FIGURE 3: Overview of the bound substrate/inhibitor moieties in the open cleft part of the active center at subsites −1 to pseudo-subsite (+4). Cellotetraose (light green) is the only substrate which occupied subsite −1. Only a cellotriose could be found at positions +1 to +3 in the cellohexaose complex (dark green) which occupies the same positions as cellotetraose. PIPS−IG3 (yellow) and IG4 (magenta) occupy subsites +1 to pseudo-subsite (+4). Cellobiose and cellobiitol are in subsites +1 and +2 at similar positions as the cello-oligosaccharides and are not shown.

poor solubility. The observed positions for inhibitor/substrate molecules were subdivided in two sections: a tunnel section ranging from position −7 to −1 and an open cleft section from positions +1 to +4 [the nomenclature proposed by Davies et al. (*36*) is used]. First, the open cleft section from sub-site +1 to +4 will be discussed.

Subsites in the Leaving Group Site. Inspection of the initial ($F_o - F_c$) electron density maps showed that not all of the complexes had newly bound sugar moieties in both sections of the active site. In the cellobiose complex sub-sites +1 and +2 were clearly occupied by a cellobiose molecule, but no additional electron density could be found in the tunnel part. A similar result was obtained in the case of the cellobiitol complex, where only the density of one cellobiitol molecule covering positions +1 and +2 was found. These two sub-sites +1 and +2 were occupied by subunits of the inhibitor/substrate molecules in all the structures discussed here. In all complexes the position of the glucosyl unit in sub-site +1 was approximately the same as previously reported for the CelF/IG4 complex [root-mean-square deviations (rmsd) ranging from 0.10 to 0.40 Å]. The second glucosyl unit in sub-site +2 was slightly tilted in the thio-oligosaccharide complexes (Cel48F/PIPS−IG3 and Cel48F/IG4) compared to all complexes with a natural substrate (cellobiose, cellotetraose, and cellohexaose). This difference in the orientation in sub-site +2 was a consequence of the thio-bond between sub-sites +2 and +3 in the thio-oligosaccharide compounds where the carbon−sulfur bond is about 0.4 Å longer and the CSC-angle about 20° smaller than in a normal *O*-glycosidic bond (*32*). The positions of the glucosyl units in sub-site +3 of the cellotetraose and cellohexaose complexes were very similar (rmsd of 0.22 Å) and formed two new H-bonds with the protein (Figure 2). The thio-oligosaccharides, however, were less well-defined here in terms of electron density and, compared to the natural substrate positions, were rotated around the thio-bond between sub-sites +2 and +3 (dihedral angles C1′−S4−C4−C3, PIPS−IG3 = 98.34° and IG4 = 134.22°, and C1′−O4−C4−C3, cellotetraose = 87.42° for a comparison) giving them a different orientation in sub-site +3 (rmsd of 1.21 Å with PIPS−IG3 and 2.20 Å in the case of IG4, both compared to the cello-oligosaccharides). These orientations reduced the energy gain from the formation of H-bonds with the protein in sub-site +3, which were longer in the PIPS−IG3 complex than in the natural substrate complex and had disappeared completely in the IG4 complex. The chemical constitution of the subunits in sub-site +4 is different. In PIPS−IG3 it is a para-iodo-phenyl (PIP) ring which is connected via a thio bond to the inhibitor and in IG4 it is a methyl β-D-glucosyl residue connected via a glycosidic bond. These two inhibitor moieties permit different interactions and have therefore a different orientation in the subsite. No electron density indicating a glucosyl residue in sub-site +4 could be observed in the complexes with natural substrates. A summary of the comparison between all substrate/inhibitor positions in the open cleft section of the active center is given in Figure 3.

Subsites in the Tunnel Part of the Active Center. Additional electron density corresponding to sugar moieties could be observed in the tunnel section of the active center (sub-sites −1 to −7) only in the case of the cocrystallization experiments with ligands with at least three sugar subunits. In view of the location of the sub-sites in the tunnel one can subdivide the structures into two groups: the thio-oligosaccharide inhibitor complexes with the native enzyme and the cello-oligosaccharide complexes with the mutant E55Q. To distinguish between the different sub-site locations, inhibitor sub-sites are marked with a prime (Figure 4). In both complexes with the thio-oligosaccharide inhibitors, sub-sites −3′ to −6′ in the tunnel are occupied. The overall position of PIPS−IG3 was, compared to the IG4 complex, shifted about 0.8 Å along the tunnel axis in the direction of the active center resulting in a rmsd of 0.98 Å calculated for the first

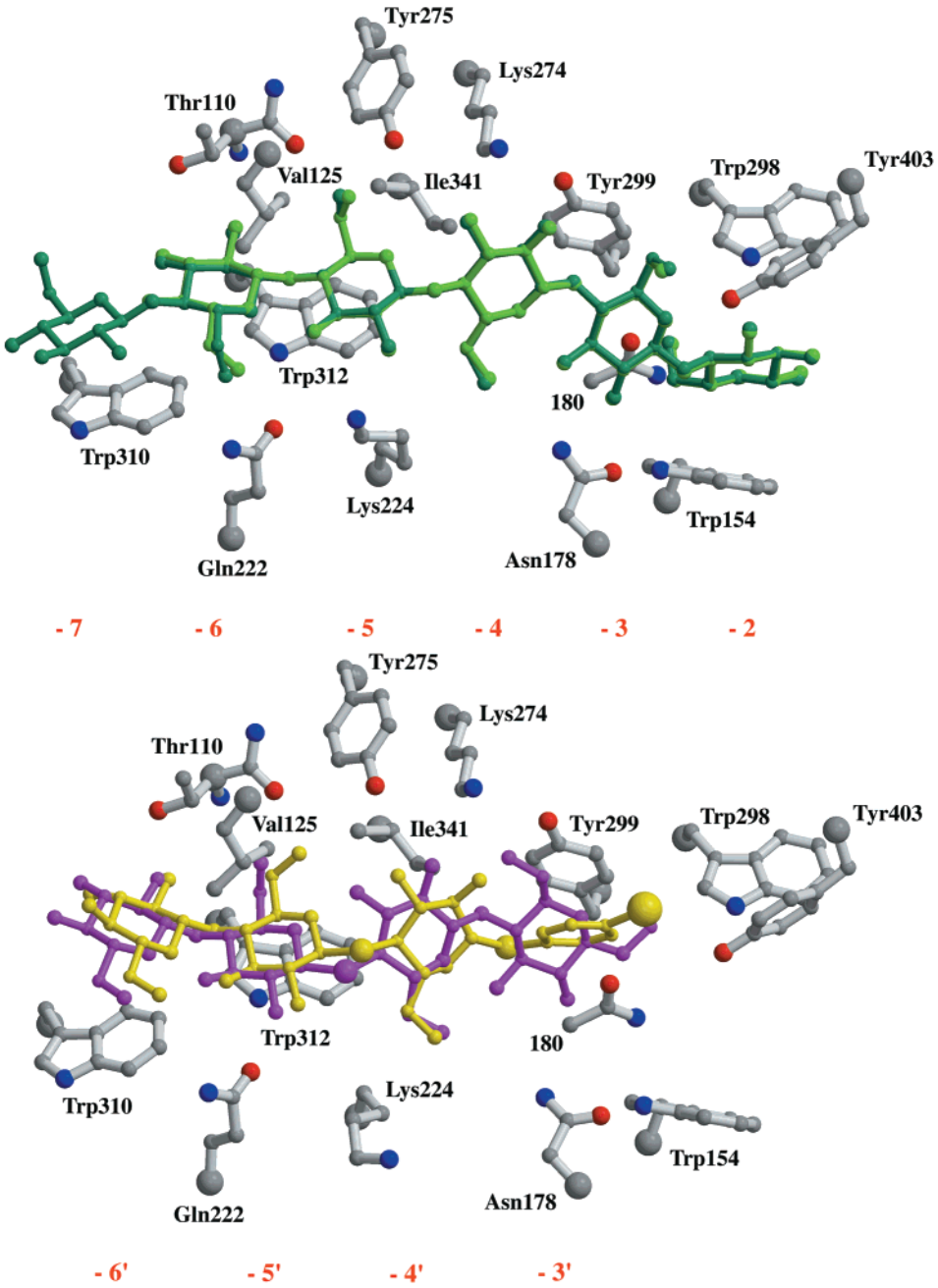

FIGURE 4: (A) The occupied subsite positions in the cello-oligosaccharide complexes cellotetraose (light green) and cellohexaose (dark green) in the tunnel part of the active center. A full cellohexaose molecule covers subsites −7 to −2, whereas five sugar subunits are detected in the cellotetraose complex at subsites −6 to −2. Both cello-oligosaccharides were positioned very equally. **B.** The subsites −6′ to −3′ were occupied by the thio-oligosaccharide inhibitors PIPS−IG3 (yellow) and IG4 (magenta). These subsites are located at about a half glucose closer to the tunnel exit as the subsites −6 to −2 in panel A. Both inhibitors are slightly out of phase of one another compared to the perfect superposition of the cello-oligosaccharides.

three subunits. The subunits of both inhibitors were stabilized by hydrogen bonds and stacking interactions with aromatic residues (see also ref *8*).

The stacking interactions with aromatic residues of the sugar subunits are now considered in detail. The first sugar subunits in sub-site 6′ of both inhibitors have stacking interactions of their more hydrophilic α-face with the phenyl part of the indole ring of Trp310. The second sugar subunit partially stacks with its more hydrophobic β-face on the pyrrole part of the indole ring of Trp312 and the third subunit was located after Trp312 without any stacking interaction. The stacking interactions in subsite −3 were influenced by the chemical structure of the fourth subunit in PIPS−IG3 which was different from IG4 as explained above. The sp3 hybridization of the C4 carbon of the methyl-glucoside of IG4, gave the sugar-ring the correct orientation to perform a stacking interaction between its β-face and Tyr299. The corresponding carbon atom of the PIP-ring of PIPS−IG3 which is connected to the thio-bond is sp2 hybridized and therefore orientating the ring differently, in our case in the direction of the tunnel walls. This orientation allowed hydrophobic interactions with Tyr299, but no stacking interactions, and gave the phenyl ring more freedom to rotate about its axis. The observed loss of electron density for the ring with increasing distance from this rotation axis is consistent with its increasing disorder. The position of the

iodine atom was well determined by an intense peak in the $F_o - F_c$ electron density map, once again confirming the orientation of the inhibitor in the tunnel section and also the overall orientation of the PIP-ring. The more bulky behavior of the PIP-ring at the end of the inhibitor PIPS−IG3 may explain its small shift relative to the position of IG4.

The positions of the sub-sites −1 to −7 of the cello-oligosaccharides (cellotetraose or cellohexaose) in the inactive mutant E55Q differed remarkably from the complexes of the inhibitors PIPS−IG3 or IG4 with the native enzyme (Figure 4). In the E55Q/cellohexaose complex, sub-sites −2 to −7 were occupied, representing one entire cellohexaose molecule. In the E55Q/cellotetraose complex the situation was not so obvious. Here sub-sites −1 to −6 were occupied by at least three different positions of cellotetraose molecules. Sub-site −1 was occupied with the glucosyl unit at the non-reducing end of the cellotetraose molecule located in sub-sites +1 to +3 (Figure 3). It was not covalently connected to the following occupied sub-sites −2 to −6. These five sub-sites showed a continuous electron density, which may be the result of two overlapping cellotetraose molecules occupying sub-sites −2 to −6. The glucosyl units in these sub-sites all fitted very well to the electron density map, giving only one unambiguous orientation. The glucosyl unit in sub-site −1 was completely defined by the electron density, only the hole in its sugar-ring having a less excellent fit. As the cellotetraose molecule entered from the leaving group side into subsite −1 against the normal sense of the catalytic reaction, the orientation of the glucosyl unit in this subsite might be an artifact of the experimental conditions and does not correspond to that of an intermediate in the catalysis. For this reason we will not discuss its orientation and interactions here. The glucosyl units in sub-sites −2 to −6 of both oligocellulose complexes were perfectly superposable with a rmsd of 0.17 Å and were all in a $^4C_1$-chair conformation.

Compared to the inhibitor complexes, the sub-sites in the cellodextrin-complexes were shifted by about 2.2 Å (PIPS−IG3) to 3.0 Å (IG4) in the direction of the cleavage site, which is located between sub-sites −1 and +1. The shift-distance corresponds to half of a glucosyl unit and places the cello-oligosaccharide sub-sites precisely between the sub-sites of the inhibitor complexes. This shift changes the pattern of stacking interactions with the aromatic residues in the tunnel. Located further outside the tunnel, the glucosyl unit in sub-site −7 in the cellohexaose complex was placed with its β-face over the pyrrole part of Trp310. Sub-site −6 was then between the two tryptophan residues without any possibility of stacking. The glucosyl unit in sub-site −5 stacked with its β-face against the phenyl part of Trp312 but there were no longer any stacking interactions with Tyr299. In addition to the interactions known from the above-described inhibitor complexes, a new stacking interaction between the β-face of the glucosyl unit in sub-site −2 and the phenyl part of Trp154 could be observed. As a consequence, both components with four glucosyl residues, the thio-oligosaccharide IG4 and the cellotetraose, were involved in three stacking interactions. Two were with different parts of Trp310 and Trp312 and one was with an additional aromatic partner, either Tyr299 or Trp154. In general, the sugars in the sub-sites of the cello-oligosaccharide complexes were all stacking with their more hydrophobic β-face of the glucosyl units whereas the sugars of the thio-oligosaccharide complexes were stacking with two β-faces and one more hydrophilic α-face.

Side-Chain Movements and H-Bonding Patterns. Some side-chain orientations in the cellodextrin complexes with cellotetraose or cellohexaose had changed compared to their positions in the inhibitor-complexes with PIPS−IG3 or IG4 and in the structures with an unoccupied tunnel as in the case of cellobiose or cellobiitol. The most obvious changes concerned Lys224, whose γ-amino-group had moved by about 4.0 Å, forming a new H-bond with the O2-atom of the sugar in sub-site −5, and Tyr403 whose hydroxyl group had moved by about 1.0 Å, inducing new H-bonds with the O6-atom of the sugar in sub-site −3 and with the O3-atom of the sugar in sub-site −2. In the cello-oligosaccharide complexes a region involving three peptide backbones had moved slightly from the active center to the enzyme surface, concerning residues 51 (0.38 Å) to 53 (0.25 Å) and 155 (0.35 Å) to 160 (0.2 Å) and 172 (0.2 Å) to 175 (0.25 Å) (movements of the Cα-atoms in brackets). The sugar occupying sub-site −2 had pushed the indole ring of Trp154 by 0.42 Å out of the active center inducing the movement of these sections. Other residues located in the tunnel showing some small variations of their side-chain orientations were Asn178, Gln181, and Thr213.

The hydrogen-bonding network between cellohexaose and Cel48F in the tunnel region is summarized in Figure 5. The stabilization of the inhibitor/substrates via direct H-bonds with the protein in the two different sub-site patterns of the IG4 and the cellohexaose complexes can be compared using the sub-site quadruplets −7 to −4 and −6 to −3 in the cellohexaose and the sub-site quadruplet −6′ to −3′ in the IG4 complex. In the cellohexaose complex the quadruplets were stabilized by nine H-bonds in the first and 12 H-bonds in the latter case, whereas the IG4 molecule was stabilized by seven direct H-bonds (see in ref *8*). The improved binding of the cello-oligosaccharides was also indicated by an overall lower *B*-factor. The IG4 inhibitor was internally stabilized by three internal H-bonds compared to one in the cello-hexaose which might be due to its reduced number of interactions with the protein.

Looking in the tunnel from subsite −7 to the cleavage site, the geometry of the cello-oligosaccharide and the thio-oligosaccharides chain followed an overall right-handed helical twist. Two glycosidic bonds in the cello-oligosaccharide were remarkably twisted [all twists calculated according to Divne et al. (*21*)]: those between sub-sites −5/-4 with relative twist angles of 28° and between −3/−2, 23°. Other bonds were weakly twisted with relative twist angles of 12° (sub-sites −7/−6), 8° (−6/−5), and −3° (−4/−3). These twists were smaller than those observed in the thio-oligosaccharide inhibitor IG4 but they followed exactly the same overall curvature of the chain.

In total, we could establish 10 sub-sites with the natural substrate, seven of them in the tunnel section before the proposed cleavage site (−7 to −1) and three at the positions of the leaving group (+1 to +3). This is in good agreement with the observed cleavage pattern of the cellohexaose digestion assay, giving about 10% of cellotriose and 90% of cellobiose (*6*). A fourth sub-site +4 present in the IG4 complex could not be confirmed. Even the cellotetraose molecule in the E55Q/cellotetraose complex, which occupied the leaving group side of the active center, covered the sub-

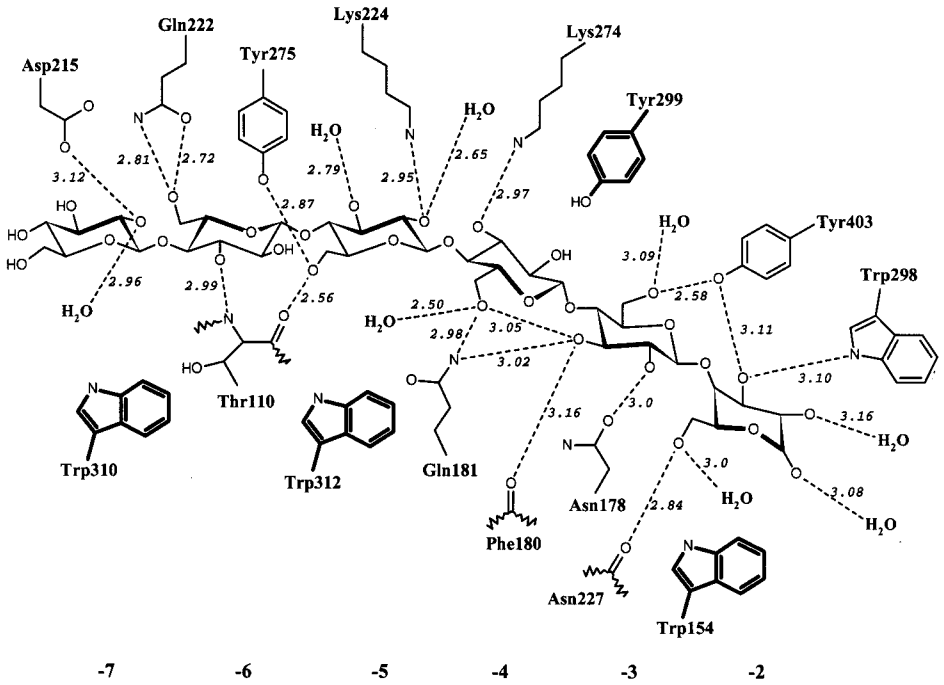

FIGURE 5: Schematic representation of H-bonds between the sugar subunits in the subsites −7 to −2 in the tunnel part of the cellohexaose complex. All H-bond partners of the first coordination shell are shown, water molecules are only shown if they are directly in contact with a protein residue. The length of each H-bond is indicated in angstroms, aromatic residues which were involved in stacking interaction are shown in bold.

sites −1 to +3 instead of +1 to +4, entering the inactive catalytic center from the "wrong" direction. This may be regarded as proof of the poor stabilization of sugars in sub-site +4, whose presence is therefore assumed to be an artifact related to the properties of the uncleavable IG4 molecule and considered to be of no importance for the catalytic reaction.

DISCUSSION.

Comparison of sub-site occupations in the various inhibitor/substrate complex structures allows us to get an overview of the tunnel/side-chain interactions in both parts of the active center. Two factors seem to be of particular importance for the binding of cello-oligosaccharides: the minimum length of the cellulose chain and interactions with the aromatic residues in the tunnel. At the protein to inhibitor/substrate ratio used of about 1:25, the short compounds cellobiose or cellobioitiol only bound in sub-sites +1 and +2 in the cleft part after the cleavage site, which is the proposed leaving group side of the active center, but not in the tunnel region. Cellobiitol was the smallest compound which bound in the active center. Single D-glucose molecules neither bound in the tunnel part nor in the leaving group side of the active center, which was confirmed from the 3-D structures obtained from cocrystallization experiments under similar conditions with D-glucose (data not shown). The low affinity for a single D-glucose molecule at the leaving group site may additionally favor the cellobiose-like progression of the substrate chain (to a productive binding position) versus a glucose-like progression. This is in good agreement with the biochemical data, where, besides cellobiose, initially smaller amounts of cellotriose but not of glucose were detected in the digestion assays (*6*).

The limits for the binding of thio-oligosaccharides in cello-oligosaccharide sub-sites can be observed in the leaving group side, where the interactions seem to be weaker after sub-site +2. This is caused by the different geometry of the thio bond which moves the glucosyl unit in sub-site +3 further away from its H-bonded partners than in the natural substrates, thereby inducing a certain disorder. This is a good example of the formation of H-bonds in the active site being sensitive to geometric constraints.

The sub-site positions in the tunnel can be subdivided into two groups, which are occupied by two kinds of complexes: the thio-oligosaccharide inhibitors PIPS−IG3 or IG4 and the cello-oligosaccharides cellotetraose or cellohexaose. The shift distance between the corresponding sub-sites of the two groups is about the length of a half sugar unit. Stacking interactions involved four aromatic residues at different sub-sites, Trp310 in sub-sites −7 and −6′ (−6.5), Trp312 in −5′(−5.5) and −5, Tyr299 in −3′ (−3.5), and Trp154 in −2 (Figure 6). Additional hydrophobic interactions with non aromatic tunnel residues can be observed in sub-sites −6 with Thr110 and Val125, in −4′ (−4.5) and −4 with Ile314 and in −3 with Thr226 leading to a continuous hydrophobic interaction strap all along the sugar pathway between sub-sites −7 to −3. This architecture stabilizes intermediate positions and may reduce the sliding energy between sub-sites using the aromatic residues as a form of lubricating agents. Aromatic stacking interactions at intermediate positions to facilitate the transport of a sugar chain have as well been reported for maltoporins where aromatic residues were located at varying distances to form a greasy slide (*37*, *38*).

Stacking interactions with phenyl rings are formed in each group of sub-sites with two pairs of stacking partners at alternate positions, Trp310/Tyr299 or Trp312/Trp154. The internal distance of the sub-sites in each pair is equivalent to four sugar units, corresponding to two cellobiose shifts and a shift of one and a half sugar units between the pairs,

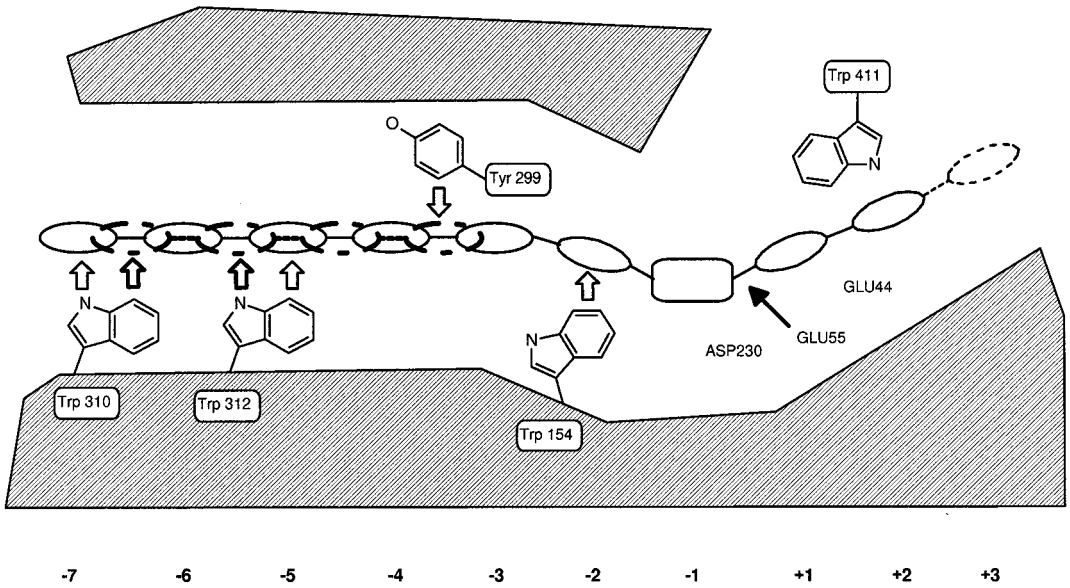

FIGURE 6: Scheme of the active site of Cel48F with the positions of inhibitor (bold broken lines) and substrate (normal lines) positions presented as ellipsoids or in subsite −1 as a box. The possible aromatic stacking partners are shown and their interactions at the subsites in the tunnel part of the active center are indicated by flashes. The position of the cleavage site is marked by a filled arrow with the relative positions of the proton donor Glu55 and the possible catalytic bases Asp230 and Glu44.

which finally offers a continuous pathway with at least one phenylic stacking interaction in the tunnel for cellotriose or longer substrates. Such an architecture may help to reduce possible product inhibition by cellobiose in the tunnel.

The glucosyl units of the cello-oligosaccharides have the lowest *B*-factors, the clearest electron density and form more H-bonds with the protein than the corresponding thio-oligosaccharides. The chain ends of the glucosyl units in sub-sites −2 and +1 are in a distance of 5.5 Å which is the right length to connect them with a glucosyl unit in sub-site −1. We propose them now to be the sub-sites occupied during initiation of the cleavage reaction. The thio-oligosaccharides have higher *B*-factors and are less well-defined by the electron density. They are therefore likely to occupy only metastable positions which correspond to an intermediate phase in the sliding process through the tunnel. A schematic view of the observed sub-site position and the related stacking partners is shown in Figure 6.

Even though hydrophobic interactions between the substrate and aromatic and nonaromatic residues reduce the energy barriers between sub-sites −7 to −2, it appears that stacking interactions with phenyl rings mark especially stable positions during the processive action through the tunnel. The thio-oligosaccharides occupy the first position which offers two stacking interactions with phenyl rings at positions −6.5 to −3.5 which are less well stabilized by H-bonds than the integral number sub-sites. The occupancy of these thermodynamically metastable positions is probably an effect of the geometry of the thio-bond (see above) which may increase the kinetic energy barrier for the movement to the sub-sites −5 to −2 which again offer two stacking interactions with phenylic rings and a more favorable stabilization by H-bonds. Indeed, in a Cel48F/IG4 complex structure from a crystal obtained under the same conditions but after a year of crystallization time, we could clearly detect two positions occupied by inhibitor molecules in the tunnel. One at the sub-site quadruplet −6′ to −3′ and the other at −5 to −2 (data not shown).

Stacking interactions at the intermediate positions (−6′ to −3′) occur twice with the β-face and once with the α-face of the sugar, whereas all sugar subunits in the more stable cello-oligosaccharide positions are oriented to interact with the hydrophobic β-side. The better hydrophobic interaction of the aromatic residues with the β-face supports the cellobiose-like displacement, and therefore, the productive binding in the active center, compared to the glucose-like displacement, where the sugars are stacking each second step with their less favorable α-face to the aromatic residues and the sugar chain binds at the cleavage site with the wrong orientation of the glycosidic oxygen (nonproductive binding).

The tunnel architecture in the processive cellobiohydrolase CBHI from *T. reesei* resembles in some aspects the tunnel in Cel48F, though the arrangement of the aromatic residues differs. There as well, the two tryptophane residues Trp38 and Trp40 offer a varying fit for stacking interactions with glucosyl units of the substrate chain in the tunnel to allow its sliding (*21*). The sugar/aromatic residue interactions in many static sugar binding sites such as the starch binding sites in barley α-amylase (*39*) and the CGTase from *B. circulans* (*40*) have nearly perfect stacking interactions with all aromatic residues in one specific position. Even more important is the perfect interaction on at least one side of the active center in the non processive endoglucanases such as CelA of *C. thermocellum* (*41*) and barley α-amylase to only allow the fit of one specific position of the substrate chain, a necessary feature for their catalytic mechanism. This difference between stacking interactions in static or non-processive sugar binding sites and in the active center of processive enzymes may further stress the importance of the layout of aromatic residues in the processive mechanism.

The different structures of complexes of Cel48F with inhibitors and substrates show the positions of productive-binding and intermediate sub-site positions in the tunnel part of the active center. This allows us to visualize the variable stacking interactions with aromatic residues which we consider to be of major importance to reduce the energy

barrier during the sliding of the substrate chain through the tunnel in the processive action and to stabilize the substrate at the productive-binding positions.

ACKNOWLEDGMENT

We thank the staff of the EMBL X31 beamline at the DORIS storage ring, DESY, Hamburg for their assistance during the data collection of the cellohexaose complex and Dr. Michel Juy for the valuable discussions and the attentive reading of the manuscript. We also thank Vincent Moreau for the gift of PIPS-IG3.

REFERENCES

1. Henrissat, B., and Bairoch, A. (1993) *Biochem. J. 293*, 781−788.
2. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
3. Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) *Curr. Opin. Struct. Biol. 8*, 548−557.
4. Henrissat, B., Driguez, H., Viet, C., and Schulein, M. (1985) *Biotechnology 3*, 722−726.
5. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) *Biotechnol. Bioeng. 42*, 1002−1013.
6. Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J. P., and Tardif, C. (1997) *J. Bacteriol. 179*, 46−52.
7. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) *Nat. Struct. Biol. 4*, 810−818.
8. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) *EMBO J. 17*, 5551−5562.
9. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.
10. Gal, L., Pagès, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C., and Belaich, J. P. (1997) *Appl. Environ. Microbiol. 63*, 903−909.
11. Lamed, R., Setter, E., and Bayer, E. A. (1983) *J. Bacteriol 156*, 828−836.
12. Beguin, P. (1990) *Annu. Rev. Microbiol. 44*, 219−248.
13. Belaich, J. P., Gaudin, C., Belaich, A., Bagnara-Tardiff, C., Fierobe, H. P., and Reverbel, C. (1993) *Abstracts of the MieBioforum 93: genetics, biochemistry and ecology of lignocellulose degradation*, pp 53−62, Toba, Japan.
14. Doi, R. H., Goldstein, M., Hashida, S., Park, J. S., and Tagaki,, M. (1994) *Crit. Rev. Microbiol. 20*, 87−93.
15. Fanutti, C., Ponyi, T., Black, G. W., Hazlewood, G. P., and Gilbert, H. J. (1995) *J. Biol. Chem. 270*, 29314−29322.
16. Bayer, E. A., Morag, E., and Lamed, R. (1994) *Trends Biotechnol. 12*, 379−386.
17. Pagès, S., Belaich, A., Tardif, C., Reverbel-Leroy, C., Gaudin, C., and Belaich, J. P. (1996) *J. Bacteriol. 178*, 2279−2286.
18. Reverbel-Leroy, C., Belaich, A., Bernadac, A., Gaudin, C., Belaich., J. P., and Tardif, C. (1996) *Microbiology 142*, 1013−23.
19. Reverbel-Leroy, C., Parsiegla, G., Moreau, V., Juy, M., Tardif, C., Driguez, H., Belaich, J. P., and Haser, R. (1998) *Acta Crystallogr., Sect. D 54*, 114−118.
20. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, A. (1994) *Science 265*, 524−528.
21. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. A. (1998) *J. Mol. Biol. 275*, 309−325.
22. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res. 16*, 7351−7367.
23. Otwinowski, Z., and Minor, W. (1997) *Methods Enzmol. 276*, 307−326.
24. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
25. Brünger, A. T., and Kurkowski, A. (1990) *Acta Crystallogr., Sect. A 46*, 46−57.
26. 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) *Acta Crystallogr., Sect. D 54*, 905−921.
27. Jiang, J. S., and Brünger, A. T. (1994) *J. Mol. Biol. 243*, 100−115.
28. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr., Sect. D 49*, 129−147.
29. Perrakis, A., Sixma, T. K., Lamzin, V. S., and Wilson, K. S. (1997) *Acta Crystallogr., Sect. D 53*, 448−455.
30. Brünger, A. T. (1992) *Nature 355*, 472−474.
31. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr., Sect. A 47*, 392−400.
32. Durier, V., Driguez, H., Rollin, P., Duee, E. ang G. Buisson (1992) *Acta Crystallogr., Sect. C 48*, 1791−1794.
33. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
34. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
35. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystalllogr., Sect. D 50*, 869−873.
36. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J. 321*, 557−559.
37. Dutzler, R., Wang, Y.-F., Rizkallah, P. J., Rosenbusch, J. P., and Schirmer, T. (1996) *Structure 4*, 127−134.
38. Meyer, J. E. W., and Schulz, G. E. (1997) *Protein Sci. 6*, 1084−1091.
39. Kadziola, A, Søgaard, M., Svensson, B., and Haser, R. (1998) *J. Mol. Biol. 278*, 205−217.
40. Lawson, C. L., van Montfort, R., Strokopytov, B, Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L., and Dijkstra, B. W. (1994) *J. Mol. Biol. 236*, 590−600.
41. Alzari, P. M., Souchon, H., and Dominguez, R. (1996) *Structure 4*, 265−275.

BI001139P