BIOCHEMISTRY

 Volume 44, Number 39 October 4, 2005

Articles

Crystal Structure of *Thermobifida fusca* Endoglucanase Cel6A in Complex with Substrate and Inhibitor: The Role of Tyrosine Y73 in Substrate Ring Distortion[†,‡]

Anna M. Larsson,[§] Terese Bergfors,[§] Elisa Dultz,[§] Diana C. Irwin,[||] Annette Roos,[§] Hugues Driguez,[⊥] David B. Wilson,[||] and T. Alwyn Jones*[,§]

Department of Cell and Molecular Biology, University of Uppsala, Biomedical Centre, Box 596, SE-751 24 Uppsala, Sweden, Department of Molecular Biology and Genetics, Cornell University, 458 Biotechnology Building, Cornell University, Ithaca, New York 14853-2703, and Centre de Recherches sur les Macromolécules Végétales CNRS, B.P. 53, 38041 Grenoble Cedex 9, France

Received April 13, 2005; Revised Manuscript Received July 18, 2005

ABSTRACT: Endoglucanase Cel6A from *Thermobifida fusca* hydrolyzes the β-1,4 linkages in cellulose at accessible points along the polymer. The structure of the catalytic domain of Cel6A from *T. fusca* in complex with a nonhydrolysable substrate analogue that acts as an inhibitor, methylcellobiosyl-4-thio-β-cellobioside (Glc_2-S-Glc_2), has been determined to 1.5 Å resolution. The glycosyl unit in subsite −1 was sterically hindered by Tyr73 and forced into a distorted ^2S_o conformation. In the enzyme where Tyr73 was mutated to a serine residue, the hindrance was removed and the glycosyl unit in subsite −1 had a relaxed $^4C_1$ chair conformation. The relaxed conformation was seen in two complex structures of the mutated enzyme, with cellotetrose (Glc_4) at 1.64 Å and Glc_2-S-Glc_2 at 1.04 Å resolution.

Cellulases are a diverse class of enzymes that hydrolyze the β-1,4 linkages in cellulose. The enzymes are traditionally grouped in two classes, endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91). Endoglucanases cut a cellulose chain at accessible points along its length, whereas cellobiohydrolases cleave off cellobiose predominantly from the ends of cellulose chains. Individual cellolytic microorganisms produce a range of different complementary cellulases (*1*). These enzymes act synergistically to convert insoluble cellulose into cellobiose and other soluble sugars, thereby achieving maximal degradation of the product (*2*).

The thermophilic soil bacterium *Thermobifida fusca* produces at least seven different cellulases (*2*, *3*), and these have been classified into the glycoside hydrolase (GH)[1] families 5, 6, 9, and 48 (*4*). The endocellulase Cel6A (belongs to GH 6) displays high individual activity on crystalline cellulose (*5*). It has a molecular weight of 42 kDa, an optimal activity at 55 °C, a broad pH optimum, and is composed of two domains, a catalytic one at the amino terminus and a cellulose-binding domain at the carboxy-

[†] This work was supported by a Swedish Science Research Council (VR) grant.

[‡] PDB ID codes: 2BOD, 2BOE, 2BOF, and 2BOG.

* To whom correspondence should be addressed. Telephone: +46-18-4714982. Fax: +46-18-536971. E-mail: alwyn@xray.bmc.uu.se.

[§] University of Uppsala.

[||] Cornell University.

[⊥] Centre de Recherches sur les Macromolécules Végétales CNRS.

[1] Abbreviations: CMC, carboxymethyl cellulose; GH, glycoside hydrolase; Glc_4, cellotetraose; Glc_2-S-Glc_2, methylcellobiosyl-4-thio-β-cellobioside; MU$(Glc)_2$, 4-methylumbelliferyl-β-cellobioside; MU$(Glc)_3$, 4-methylumbelliferyl-β-cellotrioside; MUXylGlc, 4-methylumbelliferyl-4-*O*-(β-D-glucopyranosyl)-β-D-xylopyranoside; PDB, Protein Data Bank; rmsd, root-mean-square deviation; SC, phosphoric acid-swollen cellulose; TIM, triose phosphate isomerase.

10.1021/bi0506730 CCC: $30.25
Published on Web 09/07/2005

terminal end (*6*). The crystal structure of the catalytic domain is determined at 1.8 Å resolution (*7*).

The GH 6 family contains both endoglucanases and cellobiohydrolases. One important difference between the two classes of cellulases is the environment of the active site. In the cellobiohydrolases, the active site is positioned inside a tunnel formed by two extensive loops, and following each catalytic event, the substrate is threaded through this tunnel (*8*). In contrast, these loops are reduced in length in endoglucanases, and the active site is a more open cleft. The importance of the loops for the nature of the enzymatic activity has been shown experimentally for the cellobiohydrolase Cel6B from *Cellulomonas fimi* (*9*), where deletion of 15 amino acids in one of the tunnel-forming loops enhances the endoglucanase activity of the enzyme.

The enzymes in GH 6 perform the hydrolysis of glycoside linkages by net inversion of anomeric configuration in a single-displacement mechanism (*10*). Inverting enzymes are generally believed to have a base that activates a water molecule, which performs a nucleophilic attack on the anomeric carbon, and an acid that donates a proton to the leaving group (*11*). The catalytic acid in GH 6 is the aspartate corresponding to Asp117 in *T. fusca* Cel6A (*8*, *12*−*14*); however, the residue that acts as a base in the reaction is still unclear. Mutational studies from Damude et al. (*15*) suggest that endoglucanase Cel6A from *C. fimi* acts via a typical acid−base mechanism, where the aspartate residue corresponding to Asp265 in *T. fusca* Cel6A is the base. Mutation of this aspartate in *C. fimi* Cel6A gives a completely inactive enzyme. In contrast, mutating Asp265 in *T. fusca* Cel6A to alanine and asparagine gives a 40- and 10-fold decrease, respectively, in activity on carboxymethyl cellulose (CMC; *13*). On the basis of crystallographic studies and kinetic data, Koivula et al. (*14*) proposed that Asp175 in *T. reesei* Cel6A acts as a base in a Grotthus-type mechanism; i.e., the proton is transferred from the nucleophilic water via a second water to Asp175. Support for this model of deprotonation of the nucleophilic water is found in the crystal structure of *Humicola insolens* cellobiohydrolase Cel6A complexed with cellobio-derived isofagomine (*16*). In this structure, a water molecule positioned close to the imino nitrogen interacts with a second water molecule that is hydrogen-bonded to the carboxylate of Asp180 (Asp175 in *T. reesei*). The corresponding aspartyl residue in *T. fusca*, Asp79, is in the wild-type structure positioned in a loop region 11 Å from the catalytic acid. Despite the great distance, mutations of Asp79 show that the residue is important for the catalytic reaction, with a 75-fold decrease in activity on CMC for the alanine mutant and a 25-fold decrease in activity on CMC for the asparagine mutant (*13*).

Crystallographic studies on the *T. reesei* Cel6A enzyme show that the nonhydrolysable substrate analogue, methyl-cellobiosyl-4-thio-β-cellobioside (Glc_2-S-Glc_2), is distorted in the −1 binding site, and it was suggested that this may be a key component of the mechanism (*17*). A similar distortion is seen for *H. insolens* Cel6A in complex with the same inhibitor (*18*) and for *H. insolens* Cel6A, as well as for the *Mycobacterium tuberculosis* enzyme in complex with cellobio-derived isofagomine (*16*, *19*). The principal determinant of ring distortion is a tyrosine that is conserved in GH 6. Koivula et al. (*20*) brought up the idea that the tighter binding seen in *T. reesei* Cel6A when this tyrosine is mutated to a phenylalanine (Y169F) is due to a more relaxed conformation of the glycosyl unit in subsite −1. However, the distortion in subsite −1 is still present in the crystal structure of the Y → F mutant in complex with Glc_2-S-Glc_2 (*17*), but the hydroxy methyl group in the −1 site undergoes a small conformational change. This change relative to wild-type binding probably accounts for the observed tighter binding. Mutating the corresponding tyrosine Tyr73 in *T. fusca* Cel6A to a phenylalanine reduces the activity on CMC 10-fold, and the Y → S mutant has 5000-fold reduced activity (*21*).

In this paper, we attempt to provide the structural background for interpreting these results. We present the crystal structures of the catalytic domain of wild-type Cel6A from *T. fusca* in complex with a cellotetraose-like ligand, [WT-(Glc_2-S-Glc_2)] as well as complex structures of the Y73S mutant, Y73S-Glc_4 and Y73S-(Glc_2-S-Glc_2).

MATERIAL AND METHODS

Mutagenesis and DNA Manipulations. The wild-type *T. fusca* Cel6Acd was cloned into the pET26b+ vector (Novagen) to produce pOS12 (*22*) and transformed into DH5α. This plasmid consists of the Cel6A signal peptide, MRMSPRPLRALLGAAAAALVSAAALAFPSQAA, followed by the catalytic domain from NDSPF...EMAIAA, mature amino acid 286. The plasmid was then transformed into the *Escherichia coli* expression strain BL21-DE3 (Novagen) to give strain D1119. The Cel6A Y73S mutation was made as described (*21*). Using unique restriction sites, the portion of the gene containing the Y73S mutation was subcloned into pOS12 and the resulting plasmid was transformed into BL21-gold-DE3 (Stratagene) to give strain D1260. The Cel6Acd genes in these plasmids were sequenced to be sure there were no unwanted mutations.

Enzyme Production and Purification. The *E. coli* expression strains were grown at 30 °C overnight in 100 mL of LB with 60 μg/mL kanamycin. A total of 30 mL of the overnight culture was transferred into each liter of M9 medium containing 0.5% glucose and 60 μg/mL kanamycin. After growth at 30 °C for about 5 h to an OD_{600} of about 0.8, isopropyl thio-β-D-galactoside was added to the culture to 0.8 mM and growth continued at 30 °C for 16 h.

The cells were separated from the supernatant by centrifugation for 15 min at 9000*g*, and the periplasmic shock fluid was obtained as recommended by Novogen (*23*). The Cel6Acd proteins are typically about equally divided between the supernatant and the shock fluid. The WT Cel6A supernatant and shock fluid were combined; phenylmethylsulfonyl fluoride was added to 0.1 mM; and ammonium sulfate was added to 1 M. This material was clarified by Depth Microfiltration using a CUNO Betapure filter cartridge, 0.2 μm nominal (AU09Z13NG020), and loaded onto a CL-4B phenyl Sepharose (Sigma) column (15−20 mL/L of culture). The column was washed with three column volumes of 0.6 M $(NH_4)_2SO_4$, 0.01 M NaCl, 5 mM KP_i at pH 6 followed by three column volumes of 0.3 M $(NH_4)_2SO_4$, 0.005 M NaCl, and 5 mM KP_i at pH 6. The protein was eluted with 5 mM KP_i at pH 6 and then water. The amount and purity of Cel6Acd found in each fraction were determined by SDS−PAGE, and only the best fractions were

combined. BisTris-HCL buffer at pH 7 was added to 5 mM and adjusted to no less than pH 7. This material was diluted to a conductivity of no more than 0.2 mmho with water and loaded on a Q Sepharose column (2−5 mg of protein/mL of column) equilibrated with 5 mM BisTris at pH 7. The protein was eluted with a gradient of 0−0.2 M NaCl in 5 mM BisTris at pH 7 (20× column volume). The fractions were checked by SDS−PAGE as above, and the best ones were combined. The proteins were concentrated using Millipore PTGC polysulfone ultrafiltration 10 000 MWCO membranes and a stirred cell device. The buffer was changed to 0.01 M HEPES at pH 7 and 10% glycerol by concentration and dilution 3 times. The final concentration step was done using a Millipore Centricon 30 000 MWCO. The final yield of purified WT protein was 35 mg (77 mg/mL) from 2 L of culture.

The Cel6Acd Y73S protein was prepared in essentially the same way, except only the shock fluid was used and the Q Sepharose column was run first followed by phenyl Sepharose. The final yield of the Y73S protein was 150 mg from 6 L of culture at 75 mg/mL. The protein concentrations were determined by absorbance at 280 nm using extinction coefficients calculated from the predicted amino acid compositions.

Crystallization and Data Collection. Crystallization experiments were performed using the hanging-drop vapor diffusion method at room temperature. Streak seeding with wild-type crystals was routinely employed with the mutants and complexes.

The crystallization conditions for the catalytic domain of the wild-type protein (amino acids 1−286) and its mutant, Y73S, were similar. The proteins were dissolved at 20−40 mg/mL in HEPES buffer at pH 7.0 and 5−10% glycerol. Crystals appeared within a few days from 20 to 30% poly(ethylene glycol) 4000 and 0.15−0.34 M sodium malonate at pH 4.0. These crystals were successfully flash-cooled in liquid nitrogen with a cryoprotectant consisting of 35% PEG 4000 and 1.0 M sodium malonate.

The complex of Y73S with the nonhydrolysable substrate analogue (Glc_2-S-Glc_2; *24*) was obtained by cocrystallization in the presence of a 5-fold molar excess of the inhibitor. The complex of the wild-type protein with (Glc_2-S-Glc_2) was produced by a 4-h soak with the ligand, followed by flash-cooling. The Y73S mutant were soaked in saturated $(Glc)_4$ solution to obtain the complex. The saturated $(Glc)_4$ solution was prepared by drying drops of 20 mM stock solution and then dissolving it (75 mM) in cryoprotectant, into which the crystal was transferred and soaked for 10 min.

X-ray diffraction data were collected from a single crystal at beamlines ID14-1 at ESRF (Grenoble, France) and I711 at MAX-lab (Lund, Sweden). All data sets were collected at 100 K. The data were reduced and scaled using the CCP4 suit (*25*).

Model Building and Refinement. All structures were solved by molecular replacement with AMoRe (*26*) using the wild-type structure (PDB code 1TML) as a starting search model. The model was improved by alternating cycles of refinement with REFMAC5 (*27*) and rebuilding with O (*28*). Water molecules were automatically added with ARP/wARP (*29*), and their behavior was manually monitored during refinement. The ligands were incorporated after inspection of the SigmaA-weighted (*30*) maps.

The superpositions were performed with lsqman (*31*). All figures were prepared with O and rendered with Molray (*32*). Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 2BOD, 2BOE, 2BOF, and 2BOG.

RESULTS

Crystallization, Data Collection, and Structure Refinement. Crystals of the catalytic domain of Cel6A variants from *T. fusca* were obtained in space group $P2_12_12_1$ with one molecule per asymmetric unit. The structures were solved with molecular replacement using the catalytic domain of the wild-type *T. fusca* Cel6A (*7*) as the template. Molecular replacement was used because the space group was different from the original $P2_1$. A summary of the statistics from the data collection and structure refinement is found in Table 1.

Structural Changes between the Y73S and Wild-Type Apo Structures. The structure of *T. fusca* Cel6A is a seven-stranded TIM (triose phosphate isomerase) α/β fold with the active site located in the C-terminal end of the β barrel (*7*). The Y73S apo structure confirmed previous results from circular dichroism (*21*) that the Y73S mutation did not affect the overall fold of the protein compared to the wild-type structure [PDB code 1TML; root-mean-square deviation (rmsd) 0.48 for the 281 Cα], although some local differences were seen. The most striking difference between the wild-type structure and the mutant one was the flexibility in residues 76−87. These residues belong to the amino-proximal loop that, in GH 6 cellobiohydrolases, participates in the tunnel over the active site. The three glycine residues, Gly81, Gly86, and Gly87, found in this region, contribute to the high flexibility. No interpretable density for residues 82−85 was present in any of the structures that we present herein, and the *B* factors were high for the flanking residues compared to the rest of the protein. A disulfide bridge between Cys80 and Cys125 is stabilizing the residues adjacent to the flexible loop. In the original wild-type structure, it had defined density and was bent away from the active site, although the *B* factors were higher than on average (*7*). rmsd values between the Y73S and wild-type apo structures for residues 268−273 in the carboxy-proximal loop on the opposite side of the active site were up to 3 times higher than average, although the density was well-defined. In the GH 6 cellobiohydrolases, this loop is considerably longer and contributes to the tunnel shape of the active site, whereas our loop bends away from the active-site cleft. A disulfide bridge between Cys232 and Cys267, which is identical to the Cys368 and Cys415 disulfide bridge in *T. reesei* cellobiohydrolase Cel6A, stabilizes this loop.

Structural Changes between the Y73S Apo and Ligand Structures. A comparison of the Y73S apo structure with the Y73S-Glc_4 ligand complex reveals no change in the overall fold of the protein (rmsd 0.45 Å for 278 equivalent Cα atoms). The only notable differences were seen in the residues flanking the flexible amino-proximal loop and around Cys125, which forms a disulfide bond to Cys80, in this loop. The change in position of Asp79 and Cys80 was almost 2 Å, and the residues around Cys125 and Ser123 to Met126, had moved 1.5−3 Å. A comparison of the Y73S apo structure with the Y73S-(Glc_2-S-Glc_2) ligand complex (rmsd 0.30 Å for 279 equivalent Cα atoms) showed smaller

Table 1: Data Collection Statistics

	WT-(Glc_2-S-Glc_2)	Y73S	Y73S-(Glc_2-S-Glc_2)	Y73S-Glc_4
beamline	I711, MAX	ID14 EH1, ESRF	ID14 EH1, ESRF	I711, MAX
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
cell parameters (Å)	43.3, 66.8, 81.2	43.0, 66.2, 81.0	43.0, 66.5, 81.4	42.9, 66.9, 81.5
resolution (Å)[a]	18.4−1.50 (1.58−1.50)	38.1−1.15 (1.16−1.15)	40.8−1.04 (1.10−1.04)	52.0−1.64 (1.73−1.64)
number of unique reflections	38 138	82 912	104 597	28 960
redundancy[a]	3.9 (3.3)	6.9 (6.3)	8.1 (6.7)	4.1 (3.9)
completeness (%)[a]	99.3 (98.2)	100.0 (100.0)	93.9 (91.0)	99.6 (97.1)
R_{sym} (%)[a]	11.5 (26.3)	5.7 (31.8)	10.5 (34.6)	6.1 (17.4)
mean $I/\sigma I$[a]	11.2 (2.7)	18.7 (4.9)	17.2 (4.5)	15.6 (6.3)
R_{merge} (%)[a,b]	12.9 (31.5)	6.1 (34.7)	10.9 (37.4)	7.0 (20.2)
model refinement statistics				
resolution (Å)	18.38−1.50	38.07−1.15	40.82−1.04	51.99−1.64
number of reflections	36 229	78 766	100 450	27 911
number of reflection in test set	1909	4147	5308	1495
R_{work} (%)	18.5	16.4	14.9	15.7
R_{free} (%)	21.4	19.2	16.5	18.6
number of atoms				
protein	2094	2096	2069	2080
ligand	46	-	46	45
solvent	225	367	404	273
rmsd from ideal values[c]				
bond distance (Å)	0.010	0.010	0.010	0.010
bond angle (deg)	1.4	1.4	1.5	1.3
$\langle B \rangle$ (Å^2)				
protein	16	13	11	14
ligand	28	−	10	14
solvent	26	25	23	25

[a] Values in paranthesis refer to the outer resolution shell. [b] $R_{merge} = [\sum_{hkl}(N/N-1))^{1/2}\sum_i|I - \langle I\rangle|/\sum_{hkl}\sum_i|I|] \times 100\%$. [c] Ideal values from Engh and Huber (*39*).

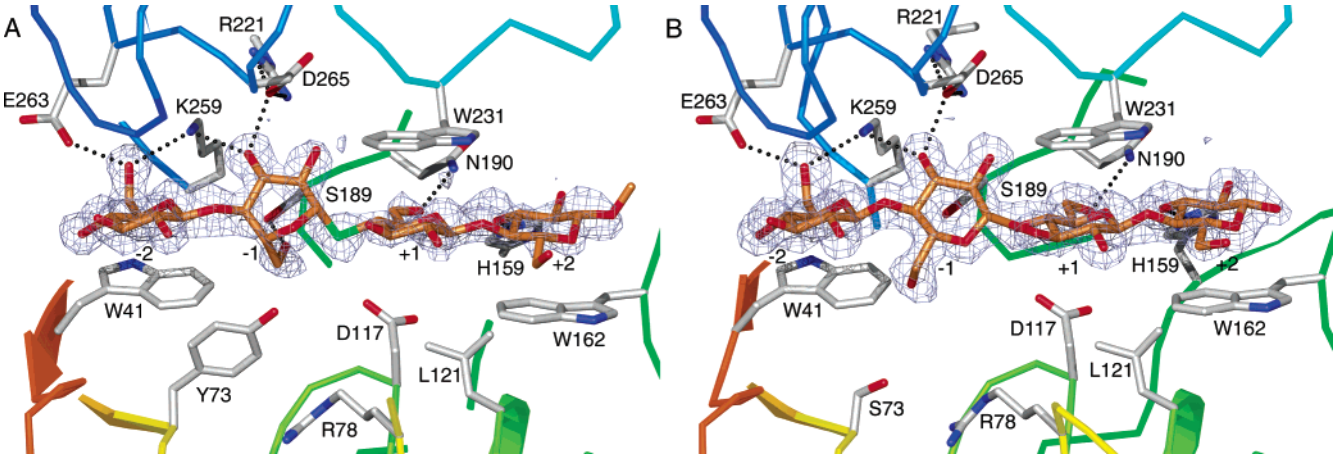

FIGURE 1: Inhibitor Glc_2-S-Glc_2 bind in the active site of wild-type Cel6A with a distortion in subsite −1 (A). The mutation Y73S allows the substrate Glc_4 to bind with all glycosyl residues in the relaxed chair conformation (B). The densities are σ_A-weighted $2F_o - F_c$ maps (*30*) at 0.3 electrons/Å^2 in A and 0.4 electrons/Å^2 in B.

changes in this region. Only Cys80 and Gly87, on either side of the missing loop residues, had large shifts in position, 1.6 and 2.7 Å, respectively. A comparison of the WT-(Glc_2-S-Glc_2) structure with the Y73S-(Glc_2-S-Glc_2) (rmsd 0.36 Å for 279 equivalent Cα atoms) showed the same pattern. Again, the only significant changes were found around the missing loop residues and the disulfide bond.

Ligand Interaction. In all three ligand complex structures, WT-(Glc_2-S-Glc_2), Y73S-$(Glc)_4$, and Y73S-(Glc_2-S-Glc_2), glucose units were bound from subsite −2 to subsite +2 in the nomenclature of Davies et al. (*33*). The major difference in the sugar units of the three complexes was the presence of ring distortion in subsite −1 for the WT ligand complex (Figure 1A). In the Y73S mutant, the glycosyl unit adopted the energetically more favorable chair conformation (Figure 1B). The relative positions of the three ligands in superimposed structures are shown in Figure 2.

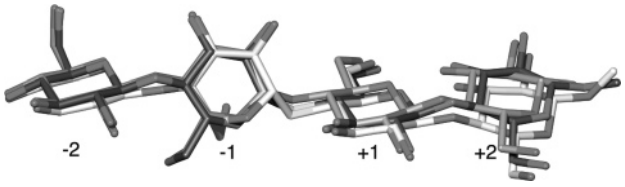

FIGURE 2: Inhibitor Glc_2-S-Glc_2 and substrate Glc_4 binds in a very similar way in the active site. Here, the three ligands are shown after superimposing the structures, WT-(Glc_2-S-Glc_2) in white, Y73S-$(Glc)_4$ in gray, and Y73S-(Glc_2-S-Glc_2) in dark gray.

The hydroxyl group of Tyr73 in the WT-(Glc_2-S-Glc_2) structure interacted (3.0 Å) with a water molecule that was not within hydrogen-bond distance to any other atom. The C6 hydroxyl group in the distorted glycosyl unit in subsite −1 was within hydrogen-bond distance of the hydroxyl group of Ser189. This hydrogen bond was not present in the Y73S ligand complexes because, in the relaxed conformation, the

C6 hydroxyl group was more than 5 Å away from Ser189. Instead the hydroxyl group of Ser189 was within hydrogen-bond distance (2.8 Å) of a water molecule that also interacts (3.1 Å) with the carboxylate of the catalytic acid Asp117. This water molecule was not present in the WT-(Glc_2-S-Glc_2) structure, and the C6 hydroxyl group did not interact with Asp117 (3.8 Å). The C6 hydroxyl group in subsite −1 was also in a nonbonded contact (2.7 Å) with the sulfur atom in the glycosidic linkage.

The catalytic acid Asp117 was present in the same orientation in all structures. The distance between the closest Asp117 carboxylate oxygen and the glycosidic linkage was 3.5, 3.6, and 3.3 Å for Y73S-Glc_4, WT-(Glc_2-S-Glc_2), and Y73S-(Glc_2-S-Glc_2), respectively. The other carboxylate oxygen of Asp117 formed a hydrogen bond (2.6−2.9 Å) to a water molecule.

In the other subsites, similar interactions could be seen between the ligand and protein for all three complexes. Three tryptophans, Trp41, Trp231, and Trp162, are present in subsites −2, +1, and +2, respectively. In the +1 and +2 sites, the α face of the glycosyl unit stacks on the indole ring of the tryptophan. In subsite −2, the tryptophan is positioned on the *β*-face side of the glycosyl unit but without forming full stacking interactions. The C6 hydroxyl group in subsite −2 was within hydrogen-bond distance to the carboxylate of Glu263 and the amino group of Lys259. The amino group of Lys259 was also within hydrogen-bond distance to the C3 hydroxyl group of the glycosyl unit in subsite −1, which also was within hydrogen-bond distance to a carboxylate of Asp265. The only direct contacts linking the protein and the +1 glucosyl unit were between the C6 hydroxyl group and amide nitrogen of Asn190 and between the C3 hydroxyl group and an imidazole nitrogen of His159 in subsite +2. Besides the direct hydrogen bonds, there were also three water-mediated hydrogen bonds, one in each of the subsites −2, −1, and +2. Three internal hydrogen bonds are present in all ligands; the C2 hydroxyl group in subsite −2 and +1 are within hydrogen-bond distance to the glycosidic oxygen in subsite −1 and −2, respectively, and in subsite +1, the ring oxygen is within hydrogen-bond distance to the C3 hydroxyl group in subsite +2.

DISCUSSION

Extensive structural studies on ligand interactions with the Cel6A cellobiohydrolases from both the fungi *T. reesei* and *H. insolens* have been done, and several complex structures are available (*8*, *16*−*18*, *20*, *34*). The sequence similarity between the catalytic domains of these two cellobiohydrolases is 64%, while Cel6A from *T. fusca* is more distantly related with only about 30% sequence identity (using Tcoffee) to *T. reesei* and *H. insolens*. In *H. insolens* cellobiohydrolase Cel6A, five glycosyl units are observed binding to subsite −2 and +1 to +4 (*35*). In the structures presented in this paper, four binding sites, from −2 to +2, have been identified for *T. fusca* Cel6A. Superimposing the *T. fusca* and *H. insolens* Cel6A structures reveals a shorter binding cleft in *T. fusca* Cel6A. The glycosyl unit in *H. insolens* Cel6A subsite +4 is stacked against Trp277, for which there is no corresponding tryptophan in *T. fusca* Cel6A. The loop formed by residues 316−329 in *H. insolens* Cel6A is also absent in *T. fusca* Cel6A (Figure 3).

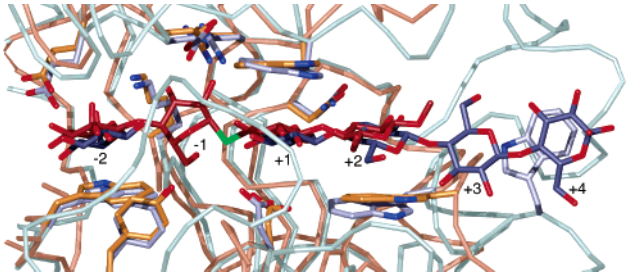

FIGURE 3: Superimposing *T. fusca* Cel6A (in gold and red) and *H. insolens* cellobiohydrolase Cel6A (in gray and purple) reveals the shorter binding site in *T. fusca* Cel6A.

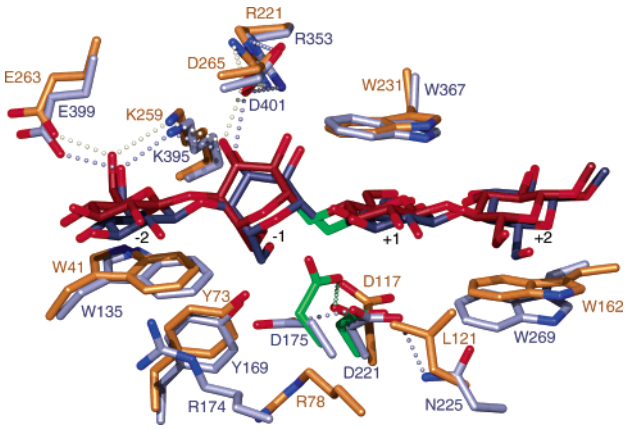

FIGURE 4: Inhibitor Glc_2-S-Glc_2 is shown bound to wild-type *T. fusca* Cel6A (in gold and red) and *T. reesei* cellobiohydrolase Cel6A (in purple and green). The major differences in the active site between the two enzymes are found around the catalytic acid, Asp117 in *T. fusca* Cel6A and Asp221 in *T. reesei* Cel6A.

The Flexible Loop. The two tunnel-forming loops present in the GH 6 cellobiohydrolases are considerably shorter in *T. fusca* Cel6A. The openness of the active site is important for the endoglucanase activity (*9*). Endoglucanase Cel6B from *H. insolens* shows high structure similarity to the cellobiohydrolases, but the carboxyl-terminal loop of the cellobiohydrolases is missing (*35*). In ligand complex structures, endoglucanase Cel6 from *M. tuberculosis* shows an amino-proximal loop that is folded back over the binding cleft (*19*). In *T. fusca*, the flexible amino-proximal loop (residues 76−89) is shortened 5 residues compared to the one in *T. reesei* and the carboxyl-proximal loop (residues 260−275), which is 20 residues shorter. In the WT structure, the amino-proximal loop is fixed by a sulfate ion in a position bent away from the active site. In all of the ligand-bound structures presented here, the density from Gly81 to Gly86 was too poorly defined to allow the building of this loop. In addition, the residues before and after had badly defined density and higher temperature factors than the average for the structure. The movement of the loop could be an important feature for the endo activity of the enzyme on the substrate, but a folding back over the active site might be needed to bring Asp79 closer into position for cleavage. In the structures available to date, Asp79 is not close enough to the active site to participate in the reaction as a base. However, from its location in the region with high flexibility, further movements could be possible that allow it to get closer to the active site. To make Asp79 undertake a similar position as Asp175 in *T. reesei*, the sterical hindrance by Arg78 must be removed (Figure 4). It is positioned by

Table 2: Enzyme Activities on Different Substrates

mutant\substrate	CMC (%)	SC (%)	$MU(Glc)_2$ (M)	$MU(Glc)_3$ (M)	cellotriose (M)	MUXylGlc (M)
WT	100	100	5.41×10^{-7}	4.18×10^{-8}	1.00×10^{-4}	$\sim 2 \times 10^{-9}$
Y73F	8.4	5.7	2.07×10^{-7}	2.10×10^{-8}	1.38×10^{-4}	$\sim 5 \times 10^{-9}$
Y73S	0.022	0.088	5.67×10^{-8}	1.64×10^{-8}	8.3×10^{-5}	1.38×10^{-8}
D117A	0.03	0.02	2.7×10^{-7}	5.7×10^{-8}	1.2×10^{-4}	

hydrogen bonds to the carboxylate of Glu115 and the carbonyl oxygens of Ala75 and Pro116 with both the primary and secondary guanidinium nitrogens. The position of Arg78 was well-defined in all structures presented here, but in a mutant structure where the catalytic acid D117 is mutated to an alanine, the Arg78 showed less well-defined density for two conformations (unpublished results): one in a similar position as the structures presented herein and another position similar to the one for the corresponding residue in *T. reesei* Cel6A. Because these two conformations of Arg78 are not seen in any other *T. fusca* Cel6A structure, one interpretation is that they are just the result of the changes in the active site caused by the mutation of the catalytic acid.

The Catalytic Acid. Although it is clear which amino acid functions as a proton donor in GH 6, the crystallographic studies have shown different arrangements of the protonation machinery. In the ligand structures of *T. reesei* wild type and a Y169F mutant with Glc_2-S-Glc_2 (*17*) the catalytic acid, Asp221, is flipped away from the glycosidic oxygen and not in position to act as a proton donor. In these structures, two conformations have been seen for Asp175 that interacts closely with the catalytic acid (Figure 4). Different conformations, both for Asp175 and Asp221, are suggested to depend on the movements of the loop regions when the substrate is threaded through the tunnel, and their interaction is thought to be important for the protonation of Asp221 (*17*). Flipping of the catalytic acid is observed also in *H. insolens* cellobiohydrolase Cel6 (*16*, *18*). In *T. fusca* Cel6A, the catalytic acid was observed in only one orientation, pointing toward the glycosidic bond. The residues surrounding the catalytic acid differ remarkably compared to *T. reesei* Cel6A (Figure 4), while the residues interacting with the substrate are remarkably well-conserved.

The catalytic acid Asp117 in *T. fusca* Cel6A was about 3.5 Å from the glycosidic bond in all of the complexed structures presented here, and the carboxylates of Asp117 did not interact with any other residues. Both Asn225 and Asp175 in *T. reesei* Cel6A are at hydrogen-bond distance from Asp221 when the residue is flipped away from the glycosidic bond (Figure 4). In *T. fusca* Cel6A, Leu121 replaces Asn225. The position of the hydrophobic Leu121 could prevent the catalytic acid in *T. fusca* from adopting a conformation flipped away from the glycosidic bond that is seen in *T. reesei* Cel6A. In *T. fusca* Cel6A, Arg78 (Arg174 in *T. reesei* Cel6A) is filling up the space that Asp175 partly occupies in *T. reesei* Cel6A (Figure 4), but Arg78 might be able to change location with changes in the amino-terminal loop that covers the active site (see above).

Distortion in the −1 Subsite. Kinetic studies on Cel6A and the Y169F cellobiohydrolase mutant from *T. reesei* (*20*) were interpreted to suggest that the enzyme distorts the glycosyl unit bound in the −1 site. This distortion could be relieved by a Y → F mutation of a conserved tyrosyl residue (residue Y169) in the site. Subsequent crystallographic studies with a nonhydrolysable ligand (Glc_2-S-Glc_2; *17*) showed that the −1 glycosyl unit is distorted in both the wild type and Y169F ligand complex. The extra space in the mutant enzyme allows a reorientation of the C5 hydroxyl group in the −1 glycosyl unit to a more energetically favorable conformation and thereby provides a structural explanation for the earlier binding results. A similar structure was later reported for the closely related cellobiohydrolase from *H. insolens* (*18*) and, recently, for *M. tuberculosis* Cel6 (*19*). The distantly related Cel6A from *T. fusca* shows essentially an identical ring distortion in the WT-(Glc_2-S-Glc_2) (Figure 4). After least-squares alignment with the WT *T. reesei* structure, the rmsd for the ligands was 0.9 Å and the ligand structure showed no significant deviations (rmsd of 0.4 Å when aligning the ligands), thus maintaining the hydroxy methyl orientation. The protein ligand interactions are highly conserved between the two structures except for one hydrogen bond in each subsite −2 and −1. In the −2 site, the hydrogen-bond interaction between the C3 hydroxyl group and the carboxylate in Asp137 in *T. reesei* Cel6A is not present in *T. fusca* Cel6A because the aspartate is replaced by an alanine. In the −1 site, the hydrogen bond between the C6 hydroxyl group and Ser189 present in *T. fusca* Cel6A is absent in *T. reesei* Cel6A because in the position of the serine it is an alanine.

A more drastic mutation of the conserved tyrosyl, Y73S, in *T. fusca* replaces the tyrosine with a much smaller serine residue. This mutation caused only minor changes in the active-site residues of the *T. fusca* enzyme (parts A and B of Figure 1) but dramatically affected the conformation of the −1 glycosyl residue. The cellotetrose and thiol derivative had identical structures, and both formed the energetically more favorable $^4C_1$ conformation in the −1 glycosyl unit in the Y73S complexes. The adoption of this conformation is a direct consequence of the mutation because modeling the relaxed ligand conformation into the WT enzyme produces severe clashes with the tyrosine side chain (see also ref *17*). The fact that the glycosyl units in subsites −2 and +1 had essentially identical positions in the WT-(Glc_2-S-Glc_2) and Y73S-(Glc_2-S-Glc_2) structures is further evidence that the tyrosine alone is responsible for the ring distortion.

Kinetic studies on *T. fusca* Cel6A (*13*, *21*) show that the activities of the Y73S mutant on CMC and phosphoric acid-swollen cellulose (SC) (0.022 and 0.088% of WT activity, respectively) are comparable to the corresponding activity of *T. fusca* Cel6A when the catalytic acid D117 is mutated to an alanine (0.03 and 0.02% of WT activity, respectively; Table 2). Thus, the Y73S mutation, resulting in the nondistorted glycosyl unit in subsite −1, strongly decreases the activity. The Y73F mutant shows 8.4% of WT activity on CMC and 5.7% on SC. The higher activity in Y73F compared to Y73S can be explained by the size of the phenylalanine ring, which is likely to force a distortion of the the −1 subsite sugar as seen in ligand complexes with

the corresponding mutation in *T. reesei* Cel6A (*17*). The Y73S mutant also shows a tighter binding to shorter substrates compared to the WT and Y73F mutant, except for the binding to MUXylGlc (Table 2). The energetically favorable $^4C_1$ chair conformation seen in the two Y73S complex structures can explain the tighter binding of MU-$(Glc)_2$, MU$(Glc)_3$, and cellotriose to this mutant. The same explanation can be used for the tight binding of MUXylGlc to the WT and Y73F mutant. A xylosyl unit differs from a glucosyl unit by the absence of a hydroxymethyl group at the C5 atom. In a *T. reesei* Cel6A wild-type structure in complex with *m*-iodobenzyl-β-D-glucopyranosyl-β-(1,4)-D-xylopyranoside, the xylosyl unit in subsite −1 adopts the energetically favorable $^4C_1$ chair conformation (*17*).

The distortion of the −1 glycosyl unit appears to be an essential feature for the efficient catalysis in GH 6. The inverting glycoside hydrolases act by a single-displacement mechanism that passes through an oxocarbenium-ion-like transition state. The transition state is stabilized by a partial double bond between the ring oxygen and C1, which requires a planarity of C5, O5, C1, and C2. The planar conformation is seen in four different ring conformations, two-half chairs ($^4H_3$ and $^3H_4$), and two boats (^{2,5}B and $B_{2,5}$; *35*). The ^2S_o distortion presented here and seen in several nonhydrolysable ligand complexes in GH 6 is close to the ^{2,5}B boat conformation. In the complex of a nonhydrolysable cellobio-derived isofagomine and Cel6A from *H. insolens*, the isofagomine in subsite −1 is distorted to a conformation between a ^2S_o skew-boat and a ^{2,5}B boat conformation (*16*). The ^{2,5}B boat conformation is also seen at atomic resolution for an unrelated inverting endoglucanase in GH 8, *Clostridium thermocellum* Cel8A (*37*). Other glycoside hydrolase families have examples of distortion close to one of the four planar conformations, and all four distortions are represented among the families (*38*).

Consistent with the observations from other GH family 6 ligand complex structures, we have shown that a distortion of the glycosyl unit in subsite −1 is present also in *T. fusca* Cel6A. On the basis of earlier kinetic studies and the structures presented here, we suggest that Tyr73 plays a key role in the distortion of the glycosyl unit in subsite −1 and that this distortion is highly significant for the enzymatic activity of *T. fusca* Cel6A.

REFERENCES

1. Coughlan, M. P. (1992) Enzymatic hydrolysis of cellulose—An overview, *Bioresour. Technol. 39*, 107−115.
2. Irwin, D. C., Spezio, M., Walker, L., and Wilson, D. B. (1993) Activity studies of 8 purified cellulases—Specificity, synergism, and binding domain effects, *Biotechnol. Bioeng. 42*, 1002−1013.
3. Wilson, D. B. (1988) Cellulases of *Thermomonospora fusca*, *Methods Enzymol. 160*, 314−323.
4. Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J. 316*, 695−696.
5. Walker, L. P., Wilson, D. B., Irwin, D. C., McQuire, C., and Price, M. (1992) Fragmentation of cellulose by the major *Thermomonospora fusca* cellulases, *Trichoderma reesei* CBHI, and their mixtures, *Biotechnol. Bioeng. 40*, 1019−1026.
6. Ghangas, G. S., and Wilson, D. B. (1988) Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*—Affinity purification and functional domains of the cloned gene-product, *Appl. Environ. Microbiol. 54*, 2521−2526.
7. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase, *Biochemistry 32*, 9906−9916.
8. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K., and Jones, T. A. (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*, *Science 249*, 380−386.
9. Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., and Gilkes, N. R. (1995) Enhancement of the endo-β-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop, *J. Biol. Chem. 270*, 4383−4386.
10. Knowles, J. K. C., Lentovaara, P., Murray, M., and Sinnott, M. L. (1988) Stereochemical course of the action of the cellobioside hydrolase I and hydrolase II of *Trichoderma reesei*, *J. Chem. Soc., Chem. Commun. 21*, 1401−1402.
11. Koshland, P. J. (1953) Stereochemistry and the mechanism of enzymatic reactions, *Biol. Rev. 28*, 416−436.
12. Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, R. C., and Warren R. A. (1995) Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*, *Biochemistry 34*, 2220−2224.
13. Wolfgang, D. E., and Wilson, D. B. (1999) Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2, *Biochemistry 38*, 9746−9751.
14. Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M. L., Zou, J. Y., Kleywegt, G. J., Szardenings, M., Stahlberg, J., and Jones, T. A. (2002) The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: The roles of aspartic acids D221 and D175, *J. Am. Chem. Soc. 124*, 10015−10024.
15. Damude, H. G., Ferro, V., Withers, S. G., and Warren, R. A. (1996) Substrate specificity of endoglucanase A from *Cellulomonas fimi*: Fundamental differences between endoglucanases and exoglucanases from family 6, *Biochem. J. 315*, 467−472.
16. Varrot, A., Macdonald, J., Stick, R. V., Pell, G., Gilbert, H. J., and Davies, G. J. (2003) Distortion of a cellobio-derived isofagomine highlights the potential conformational itinerary of inverting β-glucosidases, *Chem. Commun. 8*, 946−947.
17. Zou, J., Kleywegt, G. J., Stahlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T. T., and Jones, T. A. (1999) Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Ce16A from *Trichoderma reesei*, *Struct. Fold. Des. 7*, 1035−1045.
18. Varrot, A., Frandsen, T. P., Driguez, H., and Davies, G. J. (2002) Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio-β-cellobioside, at 1.9 Å, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 2201−2204.
19. Varrot, A., Leydier, S., Pell, G., Macdonald, J. M., Stick, R. V., Henrissat, B., Gilbert, H. J., Davies, G. J. (2005) *Mycobacterium tuberculosis* strains possess functional cellulases, *J. Biol. Chem. 280*, 20181−20184.
20. Koivula, A., Reinikainen, T., Ruohonen, L., Valkeajarvi, A., Claeyssens, M., Teleman, O., Kleywegt, G. J., Szardenings, M., Rouvinen, J., Jones, T. A., and Teeri, T. T. (1996) The active site of *Trichoderma reesei* cellobiohydrolase II: The role of tyrosine 169, *Protein Eng. 9*, 691−699.
21. Barr, B. K., Wolfgang, D. E., Piens, K., Claeyssens, M., and Wilson, D. B. (1998) Active-site binding of glycosides by *Thermomonospora fusca* endocellulase E2, *Biochemistry 37*, 9220−9229.
22. Salminen, O. (2002) Ph.D. Thesis, Cornell University, Ithaca, New York.
23. *Novagen* (2000) pET system manual, pp 57, Novagen, Darmstadt, Germany
24. Reverbel-Leroy, C., Parsiela, G., Moreau, V., Juy, M., Tardif, C., Driguez, H., Belaich, J. P., and Haser, R. (1998) Crystallization of the catalytic domain of *Clostridium cellulolyticum* cellulase CelF in the presence of a new cellulase inhibitor, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 114−118.
25. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
26. Navaza, J. (1994) AMoRe: An automated package for molecular replacement, *Acta Crystallogr., Sect. A: Found. Crystallogr. 50*, 157−163.
27. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 240−255.

28. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr., Sect. A: Found. Crystallogr. 47*, 110−119.
29. Lamzin, V. S., and Wilson, K. S. (1993) Automated refinement of protein models, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 49*, 129−147.
30. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors, *Acta Crystallogr., Sect. A: Found. Crystallogr. 42*, 140−149.
31. Kleywegt, G. J., and Jones, T. A. (1997) Model building and refinement practice, *Methods Enzymol. 277*, 525−545.
32. Harris, M., and Jones, T. A. (2001) Molray—A web interface between O and the POV-Ray ray tracer, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1201−1203.
33. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases, *Biochem. J. 321* (part 2), 557−559.
34. Varrot, A., Schulein, M., and Davies, G. J. (1999) Structural changes of the active site tunnel of *Humicola* insolens cellobiohydrolase, Cel6A, upon oligosaccharide binding, *Biochemistry 38*, 8884−8891.
35. Davies, G. J., Brzozowski, A. M., Dauter, M., Varrot, A., and Schulein, M. (2000) Structure and function of *Humicola insolens* family 6 cellulases: Structure of the endoglucanase, Cel6B, at 1.6 Å resolution, *Biochem. J. 348*, 201−207.
36. Davies, G., Sinnot, M. L., and Withers, S. G. (1997) Glycosyl transfer, in *Comprehensive Biological Catalysis* (Sinnot, M. L., Ed.) Vol. 1, pp 119−209, Academic Press, London, U.K.
37. Guerin, D. M., Lascombe, M. B., Costabel, M., Souchon, H., Lamzin, V., Beguin, P., and Alzari, P. M. (2002) Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate, *J. Mol. Biol. 316*, 1061−1069.
38. Davies, G. J., Ducros, V. M., Varrot, A., and Zechel, D. L. (2003) Mapping the conformational itinerary of β-glycosidases by X-ray crystallography, *Biochem. Soc. Trans. 31*, 523−527.
39. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement, *Acta Crystallogr., Sect. A: Found. Crystallogr. 47*, 392−400.

BI0506730