Structure, Vol. 11, 855–864, July, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00124-2

Structural Basis for Ligand Binding and Processivity in Cellobiohydrolase Cel6A from *Humicola insolens*

Annabelle Varrot,[1] Torben P. Frandsen,[2,4] Ingemar von Ossowski,[2] Viviane Boyer,[3] Sylvain Cottaz,[3] Hugues Driguez,[3] Martin Schülein,[2,5] and Gideon J. Davies[1,*]

[1]York Structural Biology Laboratory
Department of Chemistry
University of York
Heslington
York Y010 5YW
United Kingdom

[2]Novozymes A/S
Smoermosevej 25
DK-2880 Bagsvaerd
Denmark

[3]Centre de Recherches sur les Macromolécules Végétales
CNRS
BP 53
38041 Grenoble cedex 9
France

Summary

The enzymatic digestion of cellulose entails intimate involvement of cellobiohydrolases, whose characteristic active-center tunnel contributes to a processive degradation of the polysaccharide. The cellobiohydrolase Cel6A displays an active site within a tunnel formed by two extended loops, which are known to open and close in response to ligand binding. Here we present five structures of wild-type and mutant forms of Cel6A from *Humicola insolens* in complex with nonhydrolyzable thio-oligosaccharides, at resolutions from 1.7–1.1 Å, dissecting the structural accommodation of a processing substrate chain through the active center during hydrolysis. Movement of ligand is facilitated by extensive solvent-mediated interactions and through flexibility in the hydrophobic surfaces provided by a sheath of tryptophan residues.

Introduction

Cellulose, a simple polysaccharide of β-1,4-linked glucose, is the main component of the plant cell wall and, consequently, the most abundant biopolymer. Natural recycling of cellulose, at a rate of around 10^{11} tons per year, is due to the action of a consortium of microbial hydrolases, including endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91), the two enzyme classes believed to act on polymeric cellulose. Plant cell wall-degrading enzymes also find widespread industrial use in the paper, textile, and detergent industries (Miettinen-Oinonen and Suominen, 2002) and are increasingly the bedrock of enzymatic routes to biomass conversion (Mielenz, 2001), one of the major industrial goals of the 21st century.

Cellulose itself is a recalcitrant material whose half-life has been estimated at over four million years (Wolfenden et al., 1998). The enzyme systems that degrade this substrate are, therefore, among the most powerful biological catalysts known. They face two problems: the chemistry of glycosidic bond cleavage, itself an energetically challenging reaction, and the necessity to extract, hydrolyze, and maintain a single polysaccharide chain without losing it back to the crystalline milieu. The latter problem appears to have been solved by cellobiohydrolases. These enzymes present their catalytic machinery enclosed within a tunnel whose loops maintain a single polysaccharide chain, detached from the crystalline surface, through numerous catalytic events in what has been termed a "processive" reaction (Davies and Henrissat, 1995; Rouvinen et al., 1990). Additionally, the vast majority of cellulases are multidomain enzymes in which the catalytic core domain is linked to one or more carbohydrate binding domains (CBMs) that further contribute to binding and catalysis on polymeric substrates (Tomme et al., 1995). Indeed, recent evidence suggests that fungal CBMs actually direct the enzyme to a specific face (110) of the cellulose crystal (Lehtio et al., 2003).

In crystalline cellulose, each successive glucoside is rotated by 180° along the chain axis relative to its neighbors (Figure 1), and, hence, a significant facet of processive hydrolysis by cellobiohydrolases is that each enzymatic subsite must not only be able to accommodate both faces of the pyranoside ring (Figure 2), but also to tolerate the C6-hydroxymethyl substituents as the polysaccharide chain slides through the substrate binding tunnel. Furthermore, during such a processive movement, only every second glycosidic bond is correctly presented to the catalytic apparatus, explaining why these enzymes processively liberate disaccharide products (Figure 2) (as revealed by the seminal structure determination of the *Trichoderma reesei* Cel6A; Rouvinen et al., 1990). Here we present a series of four thio-oligosaccharide complexes of the cellobiohydrolase Cel6A from *Humicola insolens* with wild-type and mutant enzymes. The structures reveal how Cel6A may house up to eight sugar units in a -4 to $+4$ substrate binding channel. Most significantly, the complexes have trapped Cel6A in an act of "virtual processivity," such that complexes in which the subsites see both sugar faces are observed, as well as an intermediate between the two extremes. These structures illuminate the mechanism by which a traversing polysaccharide chain is accommodated within an enclosed tunnel during catalysis.

Results and Discussion

The Catalytic Residues of Cel6A

Cellulases have traditionally been divided into two classes, termed endoglucanases and cellobiohydro-

*Correspondence: davies@ysbl.york.ac.uk
[4]Present address: Symphogen A/S, Elektrovej Building 375, DK 2800 Lyngby, Denmark.
[5]Martin Schülein passed away in 2001 and is greatly missed.

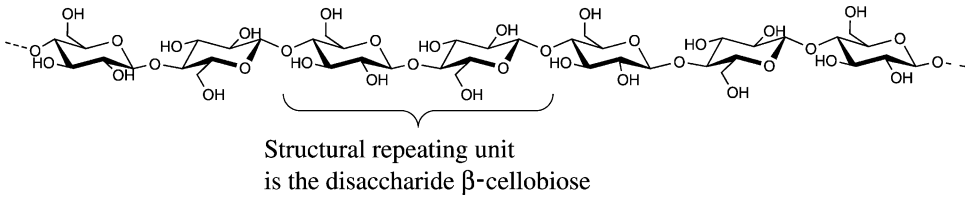


Figure 1. The Structure of a Single Cellulose Chain

While β-1,4-linked glucose is the chemical repeating unit, the structural repeat is cellobiose, and, consequently, each glucoside is orientated at 180° with respect to its neighbors, especially in crystalline forms (Gessler et al., 1994).

lases, reflecting their catalytic preference for soluble/amorphous and insoluble forms of cellulose, respectively. As reflected in its name, the latter class of enzyme releases the disaccharide cellobiose as the predominant reaction product. Glycosidic bond hydrolysis by both classes involves general acid/base chemistry, with the catalytic environment dictating whether an enzyme class performs catalysis with inversion or retention of anomeric configuration (Davies et al., 1997). Both endoglucanases and cellobiohydrolases from glycoside hydrolase family GH-6 (Coutinho and Henrissat, 1999; Henrissat and Bairoch, 1996) perform catalysis with inversion of anomeric configuration, generating the α-anomer of cellobiose as the product ("e → a" enzymes in the nomenclature proposed by Sinnott [1990]). In the classical single displacement mechanism (Koshland, 1953), this requires the presence of two catalytic carboxylate groups: a proton donor to protonate the glycosidic bond and, hence, promote departure of an otherwise poor leaving group together with a catalytic base to activate a hydrolytic water molecule for direct nucleophilic attack at the anomeric center.

Glycoside hydrolase family 6 contains 59 (statistics as of April 1, 2003) members, both endoglucanases and cellobiohydrolases. There is general agreement as to the identity of the catalytic acid from both crystallographic (Koivula et al., 2002; Varrot et al., 1999a, 1999b; Zou et al., 1999) and elegant solution studies (Damude et al., 1995; Wolfgang and Wilson, 1999). In the case of Cel6A from *H. insolens*, this residue is Asp^{226}, which lies on the loop between strand III and helix 5′ of a distorted $(\beta/\alpha)_7$ barrel fold (Varrot et al., 1999b). The role and existence of a catalytic base in this family has long been the cause of much contention, with strong evidence pointing to the residue equivalent to Asp^{405} (Damude et al., 1995), but equally convincing data arguing against such a role for this residue and instead indicating a Grotthus deprotonation of water via a chain of solvent molecules (Koivula et al., 2002; Varrot et al., 2003; Wolfgang and Wilson, 1999; Zou et al., 1999). In the case of the *Humicola* cellulolytic system, mutation of this putative base to alanine or asparagine led to an inactive endoglucanase Cel6B (M.S., unpublished data), consistent with the mutagenesis work of Damude and colleagues (Damude et al., 1995). Confusingly, however, the D405N and D405A mutants of the cellobiohydrolase Cel6A retained approximately 0.5%–1% activity (lecture by M.S. to the 2001 Carbohydrate Bioengineering Meeting, Stockholm; Varrot et al., 2002). Asp^{405} is involved in an intimate ion pair with an arginine residue, and mutations at this position might destabilize the structure and generate potentially misleading losses of activity. The structure of the D405N mutant was therefore studied, both to dissect the role of this residue in binding and catalysis and to facilitate subsequent complex determination with partially nonhydrolyzable substrate mimics (Figure 3).

The structure of the D405N mutant was determined at 1.5 Å resolution. The final model consisting of residues 87–450 was refined to a crystallographic R value of 0.113 and a corresponding R_{free} of 0.149 (Table 1). In the catalytic center, the structure is isomorphous with native Cel6A, with the only structural change being small rigid-body movements (~0.5 Å) of some helices, with helices α_1 and α_2 (residues 101–133) moving 1.3 Å. The rms deviation between the native and D405N structures is 0.58 Å for the 360 equivalent CA positions, but only 0.37 Å if residues 100–133 are not taken into account (calculated with LSQMAN [Kleywegt and Jones, 1994]). These small changes may originate from differences in crystal packing due to the incorporation of two additional residues at the N terminus, resulting from a change in the expression construct. There are few structural consequences of this mutation within the close proximity of Asp^{405}. The side chain of Lys^{399} rotates slightly but still makes a similar water-mediated interaction with Asn^{405} (Figure 4). The OD1 atom of Asn^{405} remains salt-linked to Arg^{357} with slightly greater mobility for ND2, reflected in a temperature factor of 14 $Å^2$ compared to a temperature factor of approximately 7 $Å^2$ for OD1. We conclude that decreases in activity, in this case 100- to 300-fold, resulting from the mutation must have their origins in catalytic efficiency and are not caused by local structural unfolding. One clear role for Asp^{405} is in binding of the 3-OH substituent in the −1 subsite, perhaps stabilizing an unusual $^{2,5}B$ (boat) transition state for catalysis (Zou et al., 1999; Varrot et al., 2002, 2003), and this, or a possible contribution to base catalysis, is likely to be the cause of reduced activity.

Productive Complexes: The Eight Subsites of Cel6A

Early work on cellobiohydrolase function, both using modified fluorescent ligands and analyzing the amount of reducing sugar released during hydrolysis, revealed the potential for loop conformational changes to allow internal endo catalytic attack (Ståhlberg et al., 1993; Amano et al., 1996; Boisset et al., 1998). One such ligand, a fluoresceinylthioureido-derivatized tetrasaccharide, was synthesized in partially S-linked form for crystallographic studies. The crystal structure of the nonhydrolyzable thio-oligosaccharide form of this ligand

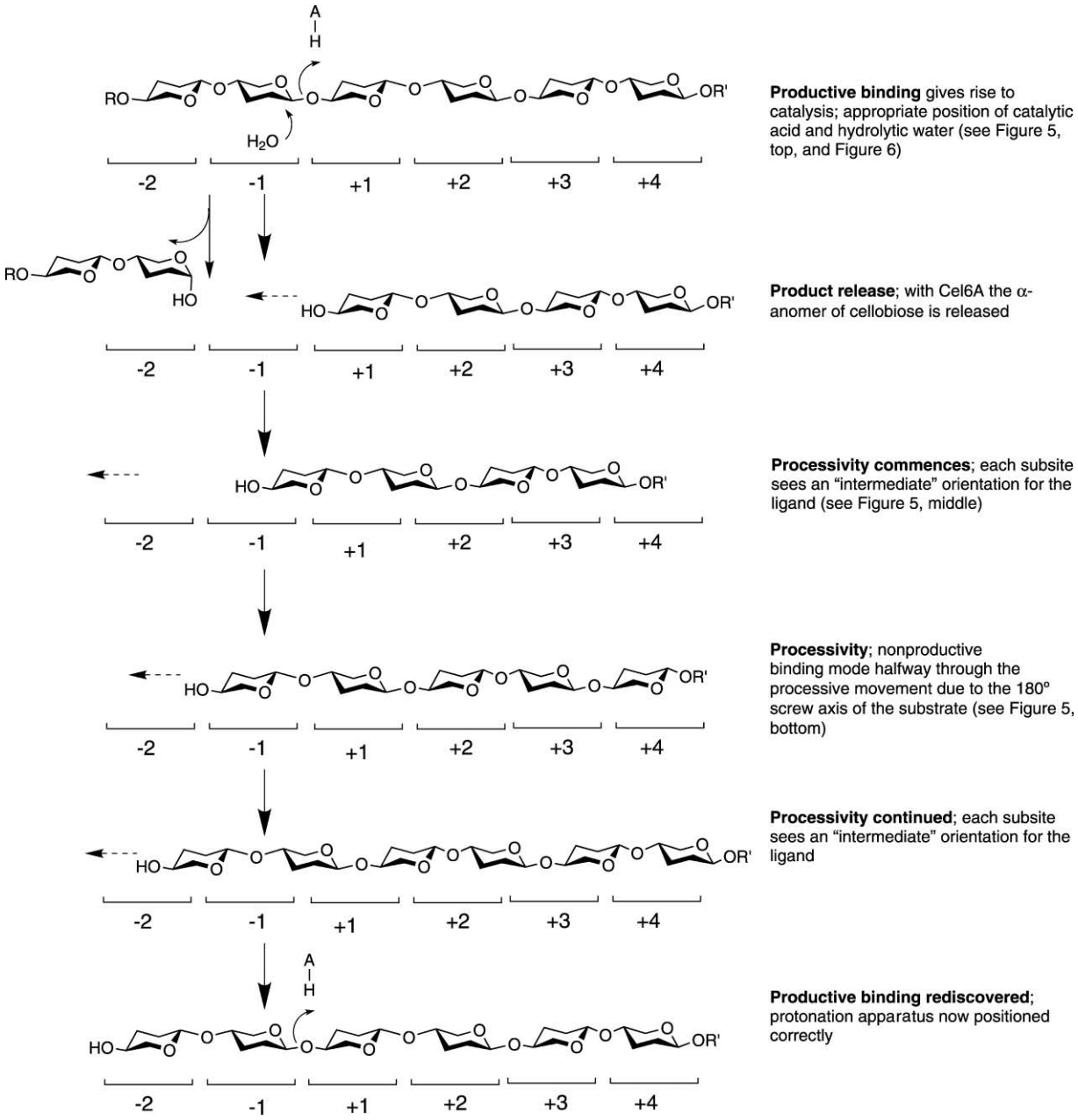
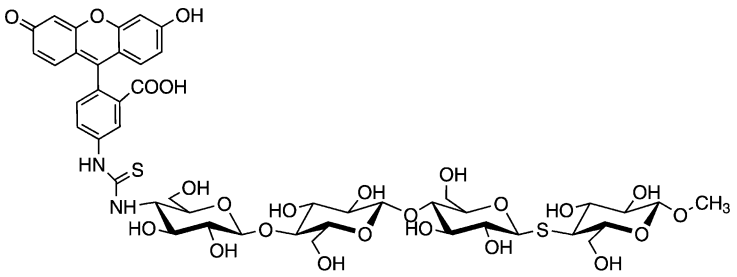


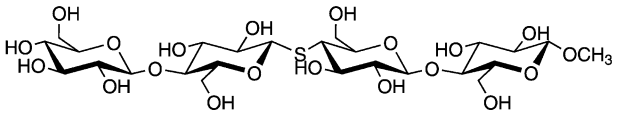
Figure 2. Catalysis and Processivity by Family 6 Cellobiohydrolases

After catalysis and disaccharide release, the nonreducing end of the cellulose chain must slide through the active-center tunnel via various "intermediate" and nonproductive binding modes until reoccupying subsites −2 and −1, with the glycosidic oxygen correctly presented to the catalytic acid.

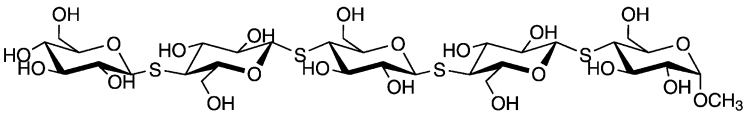
(hereafter "Fluo") (Figure 3) with wild-type Cel6A was determined at 1.75 Å resolution (Table 1). Initial difference electron density maps revealed two molecules of the thio-tetrasaccharide occupying the −4 to −2 and +1 to +4 subsites and, thus, revealed two new subsites for Cel6A, −3 and −4 (Figure 5, top). In contrast to a number of complexes with other thio-oligosaccharides, this compound does not span the active center, perhaps reflecting the steric barrier (initially) posed by the fluoresceinylthioureido moiety. However, there is no evidence that the fluoresceinylthioureido moiety remains covalently linked to either ligand, although there is poor density for a single molecule of fluorescein, displaced approximately 4 Å away from the −4 subsite. We presume that the ligand must have decayed during prolonged incubation, and the fluoresceinyl moiety migrated. The overall structure, reflecting the productive binding mode of Cel6A in which the catalytic acid is hydrogen bonded to the glycosidic oxygen at the active site (Figure 2), is almost identical to that observed for the previously reported glucose-cellotetraose complex (Varrot et al., 1999b) (Cα rmsd of 0.25 Å). The sugar units



Fluo Methyl 4-S-(4^{III}-FTU-β-cellotriosyl)-4-thio-β-D-glucopyranoside



IG4 Methyl 4^{II}-thio-β-cellotetraoside



SDP5 Methyl $4,4^{II},4^{III},4^{IV}$-tetrathio-α-cellopentaoside

Figure 3. Thio-Oligosaccharide Ligands Used in This Study

from both molecules and both structures superpose well in all subsites, apart from the apparent +4 subsite, where there are very few interactions with the protein and the glucose moiety is free to undergo slight movements. A key feature of this complex is that the ligand occupies the −3 and −4 subsites, with a rotation of 125° around the CA-CB bond of Tyr^{104}, opening up the −3 subsite.

The demonstration that there is no insurmountable steric barrier at the nonreducing extremity of the substrate binding cavity is significant. Early work had assumed that cellobiohydrolases were exo (chain-end specific) enzymes, in which single glucan chains enter and thread into the tunnel from one end only, with specific chain-end recognition occurring in the −2 subsite. It is now increasingly clear that the substrate-enclosing loops display breathing and conformational change in response to ligand binding (for example Varrot et al., 1999b, 2002; Zou et al., 1999; Koivula et al., 2002) that may occasionally facilitate internal, or endo, cleavage (Henrissat, 1998). Here we additionally show that there is no absolute steric barrier in the −2 subsite. Such observations explain many of the inconsistencies of the early published work in this field suggesting that cellobiohydrolases showed facets of both exo and endo hydrolysis. It remains unclear, however, whether occasional internal cleavage by cellobiohydrolases plays any significant role in the degradation of cellulose in a biological context.

An overlap with the distorted Cel6A-isofagomine and

Table 1. Data and Structure Quality Statistics for the *Humicola insolens* Cellobiohydrolase Cel6A Mutants and Complex Structures

	wt Fluo	D405N	D405N IG4	D405N SDP5	D416A SDP5
Data Quality					
Resolution of data (outer shell) (Å)	40–1.75 (1.81–1.75)	20–1.50 (1.55–1.50)	50–1.70 (1.76–1.70)	20–1.10 (1.13–1.1)	30–1.30[a] (1.35–1.30)
R_{merge} (outer shell)	0.041 (0.128)	0.070 (0.358)	0.055 (0.157)	0.053 (0.242)	0.068 (0.287)
Mean I/σI (outer shell)	17.3 (6)	17.2 (3.4)	17.0 (5.8)	23.7 (5.6)	16.7 (3)
Completeness (outer shell) (%)	97.5 (91.3)	99.5 (95)	93.9 (88.7)	95 (91.5)	90 (50.2)
Multiplicity (outer shell)	2.2 (2.0)	3.7 (3.5)	2.9 (2.7)	4.3 (4.3)	3.6 (1.6)
Structure Quality					
R_{cryst} (%)	13.5	11.3	12.7	10.5	14.6
R_{free} (%)	17.2	14.9	16.8	12.4	17.5
Rms deviation 1–2 bonds (Å)	0.017	0.015	0.014	0.017	0.018
Rms angle deviation (°)	1.62	1.70	1.65	1.9	1.76
Ramachandran outliers[b]	8	7	6	8	7
Protein Data Bank code	1ocb	1oc6	1oc5	1oc7	1ocj

[a] Poor data completeness in the outer resolution shell reflects data from the corners of a square detector. The 1.4 Å shell data are 90% complete.

[b] Calculated with the Uppsala Ramachandran server (see Kleywegt and Jones, 1996).

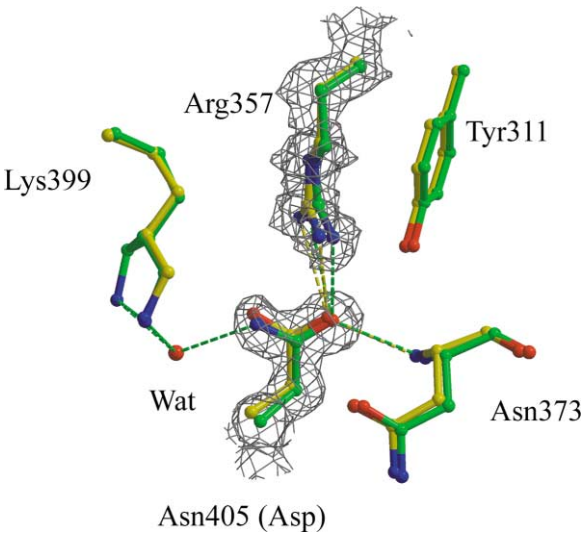


Figure 4. Electron Density for the Asn[405]-Arg[357] Ion Pair at the Catalytic Center of the Cel6A D405N Mutant

Electron density is a maximum likelihood/σ_A (Read, 1986)-weighted $2F_{obs} - F_{calc}$ synthesis at 0.5 electrons/Å^3.

wild-type IG4 complexes reveals that the protein-ligand interactions of the −2 to +4 subsites of the "productive" mode are similar to that described previously (Varrot et al., 1999b, 2002, 2003) (Figures 6 and 7). The salient feature of these interactions is that, in addition to stacking of the pyranoside rings with tryptophan residues in the +4, +2, +1, and −2 subsites, there are approximately 15 direct hydrogen bonds between protein and ligand, with a similar number mediated by solvent. In the newly revealed −3 subsite, the O-2 hydroxyl is directly hydrogen bonded to Glu[403] OE1 and interacts with the main chain nitrogen of Trp[137] through a water molecule. The O-3 hydroxyl is hydrogen bonded directly to the hydroxyl of the displaced fluoresceinyl ring and via solvent to Glu[403] OE2 and Glu[108] OE1. The O-6 hydroxyl makes a direct hydrogen bond to both Asp[139] OD2 and His[140] NE. In the −4 subsite, The O-2 hydroxyl makes water-mediated interactions with Glu[108] OE2 and the main chain oxygen of Gly[432]. The O-3 and O-2 hydroxyls are hydrogen bonded via solvent to Gln[433] OE1. The O-6 hydroxyl makes a direct hydrogen bond to Arg[140] NH2.

Processivity Commences: the D416A-SDP5 Complex

The controversies surrounding the identity of the putative catalytic base of family GH-6 enzymes, particularly the discrepancies between mutant kinetics of cellobiohydrolases and related endoglucanases, led us to speculate whether phenotypic differences could be rescued by other appropriately positioned carboxylates (Varrot et al., 2002). One candidate in a position to rescue potential catalytic base mutants is Asp[416], located on the C-terminal loop of cellobiohydrolases, which is absent in the endoglucanases from this family. Gross structural deletions of the C-terminal loop of Cel6A, such as residues 407–417, produce a totally inactive enzyme, but the D416A mutant retains approximately 10% of wild-type activity (M.S., unpublished data). The D416A mutant does, however, facilitate the serendipitous trapping of a third class of complex, whose sugar units are located halfway between the two extremes described above/below. The 1.3 Å structure of D416A-SDP5 (Table 1) is very similar to that of the native wild-type enzyme (rmsd of 0.35 Å for 360 equivalent CA positions) but again shows small movements of some secondary structural elements. There is a rigid-body movement of ~0.7 Å for the α_1 and α_2 helices and the C-terminal loop (397–435), with up to 1.3 Å for the Cα of Thr[411]. The β_{VI}-β_{VII} loop (363–375) is displaced approximately 0.9 Å for Gly[369]. In the N-terminal loop, there is the movement of 1 Å from Ala[184] CA. All these movements optimize the new interactions between the protein and the substrate. The five glucosyl moieties of the ligand are well defined in density in classical $^4C_1$ chair conformation. The binding mode of SDP5 here reveals, in terms of sugar position, the start of a processive event, with the pyranosides intermediate between the position observed in productive complexes and the "processed" mode described below: the sugars lie between subsites. The complex occupies the −2.5 to +2.5 subsites, although only subsites −0.5 to 2.5 are shown in the figure, for clarity (Figure 5, middle). Refined with an occupancy of 1.0, the resultant temperature factors are 13.7, 16.7, 19.8, 12.8, and 14.3 Å^2. In this intermediate binding mode, most of the interactions are similar to those in the productive mode and are not described further.

Nonproductive, or Processed, Binding Mode

The D405N mutant with greatly reduced activity was used as a template with which to obtain complexes with two thio-oligosaccharides (IG4) and (SDP5) (Figure 3). The IG4-D405N complex structure was solved at 1.7 Å and refined to a crystallographic R value of 0.127 and a corresponding R_{free} of 0.168 (Table 1), with the four pyranosides from the +1 to +4 subsites displaying temperature factors of 11.9 Å^2, 12.8 Å^2, 12.2 Å^2, and 15.3 Å^2, respectively. It is significant that, while wild-type enzyme complexes with IG4 revealed active-center-spanning complexes occupying the −2 to +2 subsites, with distorted ^2S_O-conformed glucosyl moieties in the −1 subsite (Zou et al., 1999; Varrot et al., 2002), the D405N complex with this ligand instead binds only in the +1 to +4 subsites. The reasons are unclear but may reflect local changes in the electrostatics of the active-center and substrate binding tunnel upon mutation. The SDP5-D405N complex structure was solved at 1.1 Å resolution (Table 1). The ligand was refined with an estimated occupancy of 0.7 and consequent B factors of 14.3 Å^2, 12.5 Å^2, 13.6 Å^2, 14.8 Å^2, and 20.7 Å^2 for the −1 to +4 subsites, respectively (Figure 5, bottom). As with all structures described in this paper, sugar rings were observed in undistorted $^4C_1$ chair conformation.

The backbone structures for the D405N-IG4 and D405N-SDP5 structures are essentially identical to the unliganded D405N structure described above, with rms deviations of 0.14 Å and 0.10 Å for the 364 equivalent CA positions. Difference $F_o - F_c$ maps reveal, without ambiguity, electron density for the IG4 and SDP5 substrate molecules spanning the +1 to +4 and −1 to +4 subsites, respectively. The two thio-oligosaccharides are bound in a nonproductive mode (Figure 2), such that each glucosyl moiety is found flipped by 180° around

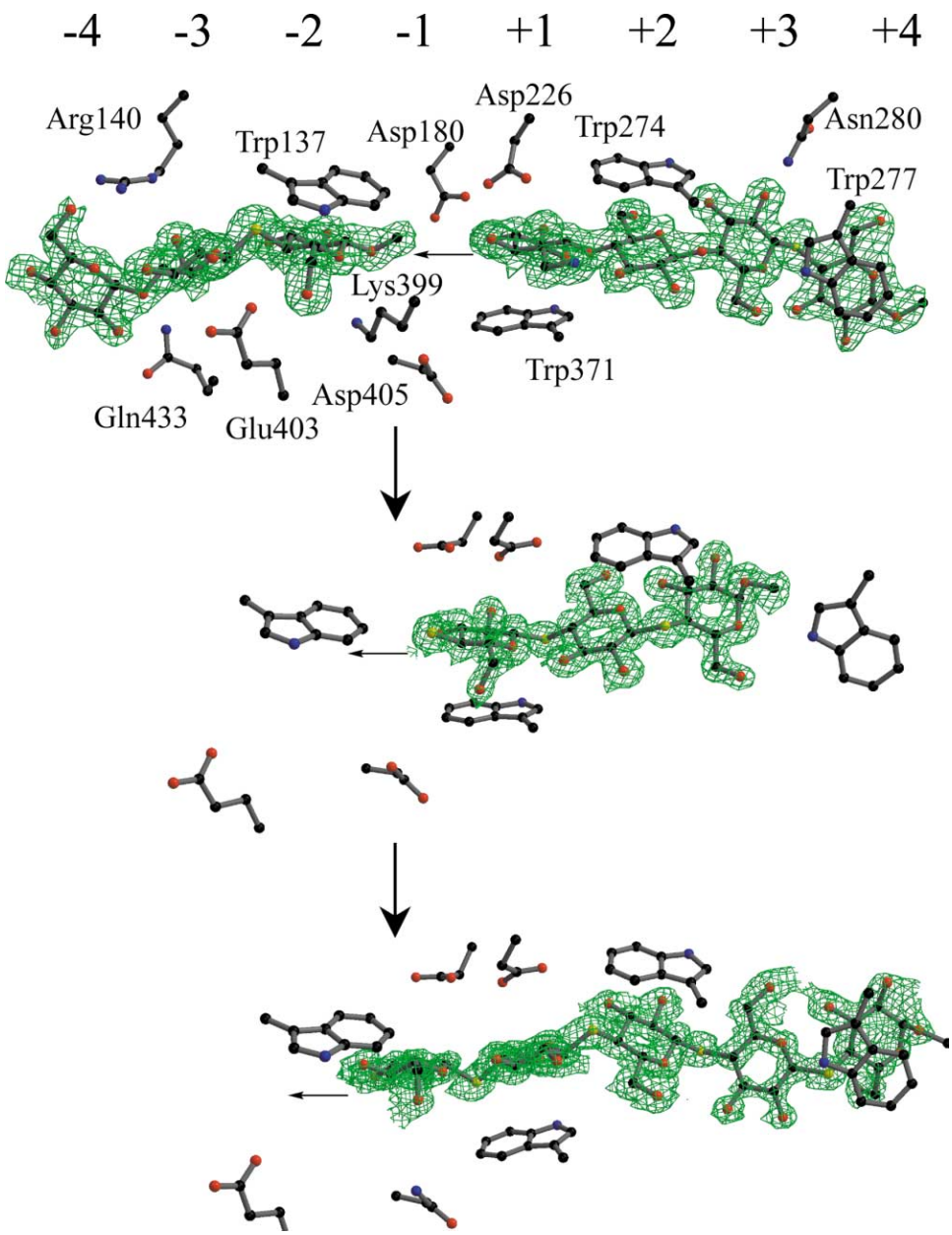


Figure 5. Crystallographic Observation of Virtual Processivity

(Top) Productive binding (see Figure 2); the catalytic acid is appropriately placed for protonation of the glycosidic oxygen between the −1 and +1 subsites, with attack by water from below (wild-type Cel6A with Fluo ligand).

(Middle) Processivity commences. The ligand is located in an "intermediate" position between subsites (the –0.5 to +2.5 subsites only from the Cel6A D416N complex with DP5)

(Bottom) Nonproductive, or processed, binding. Each subsite sees the opposite face of the ligand because of the 180° rotation between each adjacent sugar unit (Cel6A D405N complex with IG4).

The maps are maximum likelihood/σ_A-weighted $2F_{obs} - F_{calc}$ syntheses contoured at 0.39, 0.56, and 0.52 eÅ^{-3}, respectively.

the chain axis compared with the orientation observed in the glucose-cellotetraose complex (Varrot et al., 1999b), the wild-type with IG4 (Zou et al., 1999; Varrot et al., 2002), and the complex with the Fluo ligand described above. These observations could be serendipitous, reflecting slight changes in the thio-oligosaccharide bond lengths and angles (similar effects have been observed with a family 48 processive endoglucanase [Parsiegla et al., 2000]) or, as described above, may reflect changes in the electrostatic environment after mutation of Asp^{405} to Asn. Either way, such complexes mimic the processed glucan chain exactly halfway through its catalytic cycle. The unusually positioned sugars are accommodated by the reorientation of the catalytic acid Asp^{226} and the tryptophan residues involved in stacking interactions and through the harnessing of solvent-mediated

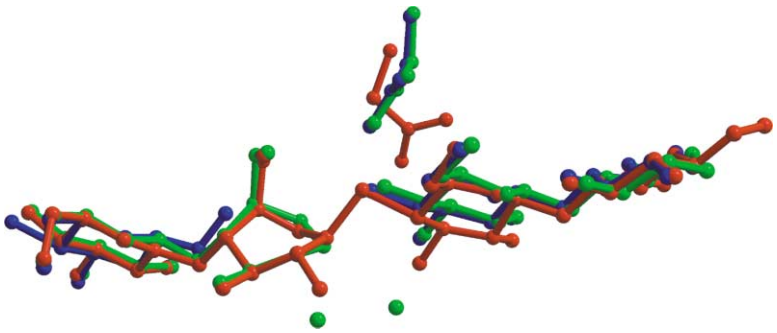


Figure 6. Overlay of Various Productive Complexes of *H. insolens* Cel6A

The structures shown are the wild-type Cel6A-Fluo complex (blue; this work), the Cel6A-isofagomine complex (green; Varrot et al., 2003), and the active-center-spanning wild-type Cel6A-IG4 complex (red; Varrot et al., 2002).

hydrogen bonds, described further below. Both complexes show similar interactions with the protein in the +1 to +4 subsites.

Virtual Processivity Revealed by Crystallography: Flexible Tryptophans and Solvent-Mediated Interactions Contribute to the Accommodation of a Processing Ligand

In contrast to the productive mode, the vast majority of the hydrogen-bonding interactions between the protein and the ligand in the one-subsite-displaced nonproductive binding modes are mediated by water. In the +1 to +4 subsites, where direct comparison is possible, the nonproductive complex shows just four direct hydrogen bonds to protein compared with eight for the productive mode, and this is reflected in a concomitant increase in solvent-mediated interactions for the nonproductive mode. Similar effects are observed in subsite −1, where productive complexes normally make three direct H bonds (Varrot et al., 2002, 2003) compared with just one in the nonproductive mode. These effects, coupled to the hydrophobic platforms afforded by tryptophan residues all along the substrate tunnel, small rigid-body movements of secondary-structural elements, and flexibility in the position of the catalytic proton donor, Asp^{226}, allows Cel6A to slide a ligand through the catalytic channel without need for liberation of the polysaccharide chain.

Aromatic residues, in particular, tryptophan residues, are often implicated in carbohydrate-protein interactions (Vyas, 1991). In Cel6A, tryptophan residues, invariant in family GH-6 cellobiohydrolases, align with the glucose moieties in −2, +1, +2, and +4 subsites. The tryptophan residue in the −2 subsite, Trp^{137}, is particularly important for catalysis. It interacts with the α face of the glucose ring, and this subsite consequently has the highest affinity for glucose (Rouvinen et al., 1990; Varrot et al., 1999b). Mutations in *Trichoderma* enzyme demonstrate that tight and specific binding in the –2 subsite is required for activity (Ruohonen et al., 1993). Here we have demonstrated that Trp^{137}, in addition to tryptophan residues in the +1, +2, and +4 subsites, all display the ability to interact with either face of the glucosyl moiety to allow the polysaccharide chain to process down the active cavity (Figure 6). Trp^{277} forms the base of the +4 subsite, implicated in the degradation of crystalline cellulose (Koivula et al., 1998); it displays subtle changes, such that the aromatic plane aligns with the sugar ring in the productive and nonproductive binding modes, in which the aromatic plane tilts approximately 5° to accommodate a sliding ligand. More substantial changes upon binding are observed for Trp^{371} in the +1 subsite. Its position is shifted both as a result of the rigid-body movement of the $\beta_{VI}-\beta_{VII}$ loop described previously and through orientational changes in the intermediate and productive binding modes through rotation around the CA-CB bond of 2° and –5° and the CB-CG bond of –4° and –10°, respectively. Together they result in 0.4 Å–1.0 Å changes in position depending upon the state of ligation. So, in addition to well-defined roles for these residues in the twisting of the substrate and subsequent destabilization of the intramolecular hydrogen bonds (Varrot et al., 1999b; Zou et al., 1999), the indole moieties of these residues provide a flexible and hydrophobic sheath through which the substrate may penetrate during catalysis (Figure 7).

A well-documented feature of family 6 cellulases is that the catalytic acid Asp^{226} has been observed in two different, but well-defined, orientations, which has often been taken to represent different protonation states for this residue and its immediate neighbors (Koivula et al., 2002; Rouvinen et al., 1990; Varrot et al., 1999a, 2002; Zou et al., 1999). In only one of these orientations does this residue hydrogen bond to the O4 atom of the +1 sugar corresponding to the glycosidic oxygen of a Michaelis complex. In the second orientation this residue is swung out away from the interglycosidic linkage into a position that may not allow protonation, bringing into question direct structural interpretation of resultant complexes (Rouvinen et al., 1990; Varrot et al., 1999b, 2002; Zou et al., 1999). The complexes presented here suggest that this movement may not simply represent ambient pH but may also play an integral role in the accommodation of the chain as it is processed through the active center during catalysis; this residue swings out of the way in the nonproductive and intermediate complexes to avoid steric clashes with the (misplaced) O-6 hydroxyl in the +1 subsite (Figures 6 and 7).

Biological Significance

The enzymatic degradation of the polysaccharide cellulose remains one of the key reactions in maintenance of the biosphere. It is catalyzed by a troop of enzymes whose key cellulolytic players are the endoglucanases, active on soluble single glucan chains, and the cellobiohydrolases, which may digest the crystalline insoluble substrate. Dissection of the catalytic mechanism of bond cleavage performed by these enzymes is becom-

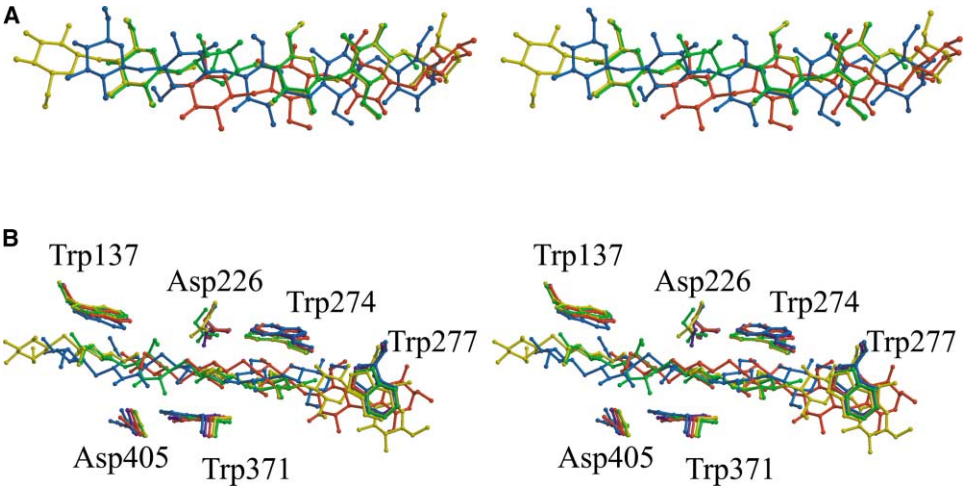


Figure 7. Overlays of the Three Binding Modes Observed for *H. insolens* Cel6A

(A) Top view showing the wild-type native structure (1BVW.pdb) (purple) as a reference. Productive binding, D416A-IG4 complex (green) and wild-type Fluo complex (yellow); intermediate binding, D416A-SDP5 complex (blue); nonproductive binding, D405N-SDP5 complex (red).
(B) Side view also showing the positions of the substrate channel tryptophan residues and the catalytic acid Asp^{226}. Both figures are presented in divergent (wall-eyed) stereo.

ing increasingly well understood. The action of these enzymes on polymeric insoluble substrates, however, remains a challenging and poorly understood area of biochemistry. Unfortunately, it is this action on real polymeric substrates that is at the heart of environmentally, indeed legislatively demanded, routes to enzymatic biomass conversion. One particular aspect, the means by which cellobiohydrolases obtain, maintain, and process a single glucan chain from the crystalline solid is central to both the fundamental understanding of these enzymes and their application for cost-effective conversion of waste biomass. Here we show, through a series of five high-resolution structures of the *H. insolens* cellobiohydrolase Cel6A, how a glucan chain may process through the substrate binding cleft from its productive binding mode, reflected in appropriately positioned protonation and solvation apparatus, through positionally intermediate and nonproductive binding modes, in which each subsite sees the wrong face of the sugar. Such movement is presumably driven by the strength of the –2 subsite of Cel6A, which allows numerous processive catalytic events without dissociation of the enzyme from the polymer. While the productive binding mode is characterized by hydrogen bonding complementarity between ligand and protein, sliding of the chain involves disruption of this network, which is instead replaced by a complex mesh of solvent-mediated interactions. The use of a hydrophobic sheath of tryptophan residues, many of which accommodate small positional changes during ligand binding, mirrors their observation in carbohydrate binding domain complexes (see for example Boraston et al., 2002; Charnock et al., 2002; Xie et al., 2001), in both cases contributing to necessarily fluid binding of a polysaccharide ligand. Structural observation of a series of ligand complexes representing the extremes and intermediates of a traversing oligosaccharide ligand contributes to the emerging picture of how an enzyme works on these recalcitrant, but significant, substrates.

Experimental Procedures

Mutagenesis and Protein Production

Purifications of the wild-type enzyme and D416A mutant have been described previously (Varrot et al., 1999a, 2002). For the D405N mutant, the mutational changes required to construct the Cel6A catalytic domain truncation together with the D405N substitution were introduced into the *Humicola insolens* Cel6A gene by the inverted PCR method (Imai et al., 1991). The DNA sequence encoding the N-terminal cellulose binding domain and linker region was first deleted from the Cel6A gene with the two phosphorylated oligonucleotide primers 5′-TACAACGGCAACCCCTTCGAGG and 5′-GGG AGCGGCGAGAGC. A BamHI-XbaI DNA fragment carrying the Cel6A catalytic domain region fused directly to its own secretion signal peptide was subcloned into an *Escherichia coli-Aspergillus* shuttle expression vector, which served as the template plasmid for another inverted PCR step to incorporate the D405N mutation (GAC → AAC) with the two phosphorylated oligonucleotide primers 5′-GTGAGTG CAACGGTACCAGCGACACG (Asn codon underlined) and 5′-CGC CGGGCTTGACCCAGACG. All introduced changes were confirmed by DNA sequencing. For the expression of D405N mutant protein, the resulting mutant plasmid, in which transcription of the truncated and D405N-mutated Cel6A gene was now controlled by a fungal α-amylase promoter and glucoamylase terminator (Christensen et al., 1988), was cotransformed with an acetamidase selection plasmid (pTOC202) into *Aspergillus oryzae* JaL228 as described previously (Kelly and Hynes, 1985). All DNA manipulations were performed by established procedures as described (Sambrook et al., 1989).

Protein Crystallization

Crystals were obtained by the hanging drop vapor diffusion method. Crystals of the native Cel6A D405N mutant grew over a period of 3–7 days from a solution containing 100 mM HEPES (pH 7.0) as buffer, 200 mM calcium acetate, and 18% PEG 8000 as precipitant. They belong to space group $P2_12_12_1$, with approximate cell dimensions of a = 57.5 Å, b = 60.2 Å, and c = 97.2 Å. The complex of D405N with SDP5 or IG4 involved 1 hr incubation of protein with a 1 mM excess of the compound of interest prior to cocrystallization. Crystals with methyl-4,4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside (SDP5; synthesis described in Schou et al. [1993]) were obtained from a solution containing 100 mM sodium acetate (pH 4.6), 100 mM magnesium acetate, 5% DMF, and 21% monomethylether PEG 5000. Crystals with methyl 4^{II}-thio-β-cellotetraoside (IG4; synthesis described in Reverbel-Leroy et al. [1998]) were obtained from a solution

containing 100 mM HEPES (pH 7.5), 100 mM magnesium acetate, and 16% monomethylether PEG 5000. Wild-type Cel6A was cocrystallised with 5 mM of methyl 4-thio-4^{IV}-fluoresceinylthioureido-cellotetraoside (Fluo; synthesis described in Boyer [1999]) from a solution containing 100 mM sodium acetate (pH 4.6) as buffer, 100 mM calcium acetate, and 16% monomethylether PEG 5000 as precipitant. The crystals grew over a month, form multilayer plates, and belong to the monoclinic space group $P2_1$ with two molecules in the asymmetric unit and a = 49.8 Å, b = 155.7 Å, c = 51.2, and β = 118.4° as the unit cell. The D416A mutant was cocrystallized with 5 mM SDP5 in a solution containing 100 mM sodium acetate (pH 4.6), 200 mM magnesium acetate, and 20% monomethylether PEG 5000.

Data Collection and Processing

All X-ray diffraction data were collected from single crystals, which frequently involved dissection of a single fragment away from a group of clustered plates. Glycerol was added to 20% (v/v) to the crystallization buffer prior to freezing in a gaseous N_2 stream at 100–110 K.

Data for the D405N-IG4 complex and the wild-type Fluo complex were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble) on beamline ID14-4, and data for the D416A-SDP5 complex were collected on ESRF beamline ID14-1, all with an ADSC Quantum 4 CCD detector. Data for the unliganded D405N mutant and its complex with SDP5 were collected at beamline BW7B (λ = 0.8445 Å) of the EMBL Hamburg outstation with a MarResearch image plate detector. All data were processed and reduced with the HKL suite of programs (Otwinowski and Minor, 1997). All further computing was performed with the CCP4 suite (Collaborative Computational Project, Number 4, 1994), unless otherwise stated.

Model Building and Refinement

The structure of the wild-type Cel6A-Fluo complex was solved by molecular replacement with the A molecule of the wild-type glucose/cellotetraose complex structure (Protein Data Bank code 2BVW) as a search model (Varrot et al., 1999a). The program AMoRE (Navaza and Saludjian, 1997) was used in conjunction with data in the resolution range 20–4 Å and an outer radius of Patterson integration of 33 Å. AMoRe was also used to accommodate rigid-body movements prior to refinement of the D405N-IG4 and D416A-SDP5 complexes. The native and the SDP5 complex structure of the Cel6A D405N mutant were solved by the difference Fourier method with the D405N-IG4 complex coordinates without substrate and water molecules as the starting model.

For the refinement of each structure, 5% of the observations were immediately set aside for crossvalidation analysis (Brünger, 1992) and were used to monitor various refinement strategies, such as the weighting of geometrical and temperature factor restraints and the insertion of solvent water during maximum likelihood refinement with the REFMAC program (Murshudov et al., 1997). Manual corrections of the model with the X-FIT routines of the program QUANTA (Accelrys) were interspersed with cycles of maximum likelihood refinement. "Riding" hydrogen atoms were only included when their positions were definable. Anisotropic refinement of the atomic displacement parameters was used for data better than 1.5 Å with a spherical restraint whose magnitude was governed by the behavior of the crossvalidation reflections. TLS refinement (Winn et al., 2001) was used for the Cel6A-Fluo, reflecting the correlated movement of entire molecules in-crystal, and its behavior was again governed by the crossvalidation reflections. Water molecules were added in an automated manner with REFMAC/ARP (Lamzin and Wilson, 1993; Murshudov et al., 1997) and verified manually prior to coordinate deposition. The incorporation of the ligand was performed after inspection of the $mF_o - DF_c$ weighted maps. Ligand occupancy was normally set to 1.0 but, in the case of the D405N-SDP5 (3) complex, was estimated as 0.7, such that the ligand then displayed similar temperature factors to the surrounding protein atoms. The stereochemical quality of the model was assessed with the program PROCHECK (Laskowski et al., 1993). Details of the data and model quality are given in Table 1. Figures 4 and 5 were created with BOBSCRIPT (Esnouf, 1997), and Figure 7 was created with MOLSCRIPT (Kraulis, 1991).

Acknowledgments

The authors would like to thank the European Synchrotron Radiation Facility, Grenoble, and the European Molecular Biology Laboratory, Hamburg, for provision of data collection facilities and financial assistance. This work was funded, in part, by the European Union (under contract BIO4-CT97-2303) and Novozymes A/S. The Structural Biology Laboratory at York is supported by the BBSRC, and CERMAV is supported by the CNRS. G.J.D. is a Royal Society University Research Fellow.

Received: February 7, 2003
Revised: April 10, 2003
Accepted: April 17, 2003
Published: July 1, 2003

References

Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E., and Kanda, T. (1996). Fine substrate specificities of four exo-type cellulases produced by *Aspergillus niger*, *Trichoderma reesei*, and *Irpex lacteus* on (1,3), (1,4)-beta-D-glucans and xyloglucan. J. Biochem. *120*, 1123–1129.

Boisset, C., Armand, S., Drouillard, H., Chanzy, H., Driguez, H., and Henrissat, B. (1998). Structure-function relationships in cellulases: the enzymatic degradation of insoluble cellulose. In Carbohydrases from *Trichoderma reesei* and Other Microorganisms, M. Claeyssens, K. Piens, and W. Nerinckx, eds. (London: Royal Society of Chemistry), pp. 124–132.

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., and Davies, G.J. (2002). Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan binding domains. J. Mol. Biol. *319*, 1143–1156.

Boyer, V. (1999). Synthèses d'analogues de substrat pour l'étude des cellulases. PhD Thesis, University of Grenoble, France.

Brünger, A.T. (1992). Free R value—a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J., and Davies, G.J. (2002). Promiscuity in ligand binding: the three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. Proc. Natl. Acad. Sci. USA *99*, 14077–14082.

Christensen, T., Wöldike, H., Boel, E., Mortensen, S.B., Hjortshøj, K., Thim, L., and Hansen, M.T. (1988). High level expression of recombinant genes in *Aspergillus oryzae*. Biotechnology *6*, 1419–1422.

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Coutinho, P.M., and Henrissat, B. (1999). Carbohydrates-active enzymes: an integrated database approach. In Proceedings of the 3rd Carbohydrate Bioengineering Meeting on Recent Advances in Carbohydrate Bioengineering, H.J. Gilbert, G.J. Davies, B. Henrissat, and B. Svensson, eds. (Cambridge: Royal Society of Chemistry), pp. 3–12.

Damude, H.G., Withers, S.G., Kilburn, D.G., Miller, R.C., Jr., and Warren, R.A. (1995). Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. Biochemistry *34*, 2220–2224.

Davies, G.J., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.

Davies, G., Sinnott, M.L., and Withers, S.G. (1997). Glycosyl transfer. In Comprehensive Biological Catalysis, M.L. Sinnott, ed. (London: Academic Press), pp. 119–209.

Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. J. Mol. Graph. *15*, 133–138.

Gessler, K., Krauss, N., Steiner, T., Betzel, C., Sandmann, C., and Saenger, W. (1994). Crystal structure of beta-D-cellotetraose hemi-

hydrate with implications for the structure of cellulose II. Science *266*, 1027–1029.

Henrissat, B. (1998). Enzymatic cellulose degradation. Cellulose Commun. *5*, 84–90.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. *316*, 695–696.

Imai, Y., Matsushima, Y., Sugimura, T., and Terada, M. (1991). A simple and rapid method for generating a deletion by PCR. Nucleic Acids Res. *19*, 2785.

Kelly, J.M., and Hynes, M.J. (1985). Transformation of *Aspergillus niger* by the amdS gene of *Aspergillus nidulans*. EMBO J. *4*, 475–479.

Kleywegt, G.J., and Jones, T.A. (1994). A super position. ESF/CCP4 Newsletter *31*, 9–14.

Kleywegt, G.J., and Jones, T.A. (1996). Phi/psi-chology: Ramachandran revisited. Structure *4*, 1395–1400.

Koivula, A., Kinnari, T., Harjunpaa, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T.A., and Teeri, T.T. (1998). Tryptophan 272: an essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. FEBS Lett. *429*, 341–346.

Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T.T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., et al. (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.

Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. *28*, 416–436.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Lamzin, V.S., and Wilson, K.S. (1993). Automated refinement of protein models. Acta Crystallogr. D Biol Crystallogr. *49*, 129–147.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck—a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T.T. (2003). The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc. Natl. Acad. Sci. USA *100*, 484–489.

Mielenz, J.R. (2001). Ethanol production from biomass: technology and commercialization status. Curr. Opin. Microbiol. *4*, 324–329.

Miettinen-Oinonen, A., and Suominen, P. (2002). Enhanced production of *Trichoderma reesei* endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. Appl. Environ. Microbiol. *68*, 3956–3964.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Navaza, J., and Saludjian, P. (1997). AMoRe: an automated molecular replacement program package. Methods Enzymol. *276*, 581–594.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J.P., Driguez, H., and Haser, R. (2000). Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. Biochemistry *39*, 11238–11246.

Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A *42*, 140–149.

Reverbel-Leroy, C., Parsiegla, G., Moreau, V., Juy, M., Tardif, C., Driguez, H., Bélaich, J.-P., and Haser, R. (1998). Crystallisation of the catalytic domain of *Clostridium cellulolyticum* CelF cellulase in the presence of a newly synthesized cellulase inhibitor. Acta Crystallogr. D Biol. Crystallogr. *54*, 114–118.

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.

Ruohonen, L., Koivula, A., Reinikainen, T., Valkejärvi, A., Teleman, A., Claeyssens, M., Szardenings, M., Jones, T.A., and Teeri, T.T. (1993). Active site of *T. reesei* cellobiohydrolase II. In Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and Other Hydrolases, P. Suominen and T. Reinikainen, eds. (Espoo, Finland: Foundation for Biotechnical and Industrial Fermentation Research), pp. 87–96.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning, A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schou, C., Rasmussen, G., Schülein, M., Henrissat, B., and Driguez, H. (1993). 4-thiocellooligosaccharides—their synthesis and use as inhibitors of cellulases. J. Carbohydr. Chem. *12*, 743–752.

Sinnott, M.L. (1990). Catalytic mechanism of enzymic glycosyl transfer. Chem. Rev. *90*, 1171.

Ståhlberg, J., Johansson, G., and Pettersson, G. (1993). *Trichoderma reesei* has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. Biochim. Biophys. Acta *1157*, 107–113.

Tomme, P., Warren, R.A., and Gilkes, N.R. (1995). Cellulose hydrolysis by bacteria and fungi. Adv. Microb. Physiol. *37*, 1–81.

Varrot, A., Hastrup, S., Schülein, M., and Davies, G.J. (1999a). Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution. Biochem. J. *337*, 297–304.

Varrot, A., Schülein, M., and Davies, G.J. (1999b). Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. Biochemistry *38*, 8884–8891.

Varrot, A., Frandsen, T., 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. D Biol. Crystallogr. *58*, 2201–2204.

Varrot, A., Macdonald, J., Stick, R.V., Pell, P., 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. 946–947.

Vyas, N.K. (1991). Atomic features of protein-carbohydrate interactions. Curr. Opin. Struct. Biol. *1*, 732–740.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. *D57*, 122–133.

Wolfenden, R., Lu, X., and Young, G. (1998). Spontaneous hydrolysis of glycosides. J. Am. Chem. Soc. *120*, 6814–6815.

Wolfgang, D.E., and Wilson, D.B. (1999). Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2. Biochemistry *38*, 9746–9751.

Xie, H.F., Gilbert, H.J., Charnock, S.J., Davies, G.J., Williamson, M.P., Simpson, P.J., Raghothama, S., Fontes, C., Dias, F.M.V., Ferreira, L.M.A., et al. (2001). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. Biochemistry *40*, 9167–9176.

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 Cel6A from *Trichoderma reesei*. Structure *7*, 1035–1045.

Accession Numbers

Coordinates and observed structure factor amplitudes have been deposited with the Macromolecular Structure Database (http://www.ebi.ac.uk/msd/) under accession numbers 1ocb, 1oc6, 1oc5, 1oc7, and 1ocj.