The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution

Goetz Parsiegla¹, Michel Juy¹,
Corinne Reverbel-Leroy², Chantal Tardif²,
Jean-Pierre Belaïch², Hugues Driguez³ and
Richard Haser^{1,4}

¹Laboratoire d'Architecture et Fonction des Macromolécules Biologiques, ²Laboratoire de Bioénergétique et Ingénierie des Protéines, Institut de Biologie Structurale et Microbiologie, Centre National de la Recherche Scientifique, 31 Chemin Joseph-Aiguier, 13402 Marseille cedex 20, and Université de Provence, Place Victor Hugo, 13331 Marseille Cedex 03 and ³CERMAV, CNRS, Domaine Universitaire de Grenoble, 601 rue de la Chimie, 38041 Grenoble, France

¹Present address: Institut de Biologie et Chimie des Protéines, UPR 412, Passage du Vercors 7, 69367 Lyon Cedex 07, France

⁴Corresponding author
e-mail: r.haser@ibcp.fr

The mesophilic bacterium *Clostridium cellulolyticum* exports multienzyme complexes called cellulosomes to digest cellulose. One of the three major components of the cellulosome is the processive endocellulase CelF. The crystal structure of the catalytic domain of CelF in complex with two molecules of a thiooligosaccharide inhibitor was determined at 2.0 Å resolution. This is the first three-dimensional structure to be solved of a member of the family 48 glycosyl hydrolases. The structure consists of an ($\alpha\alpha$)₆-helix barrel with long loops on the N-terminal side of the inner helices, which form a tunnel, and an open cleft region covering one side of the barrel. One inhibitor molecule is enclosed in the tunnel, the other exposed in the open cleft. The active centre is located in a depression at the junction of the cleft and tunnel regions. Glu55 is the proposed proton donor in the cleavage reaction, while the corresponding base is proposed to be either Glu44 or Asp230. The orientation of the reducing ends of the inhibitor molecules together with the chain translation through the tunnel in the direction of the active centre indicates that CelF cleaves processively cellobiose from the reducing to the non-reducing end of the cellulose chain.
Keywords: cellulase/crystal-structure/family 48 glycosyl hydrolase/processive endo-action/thiooligosaccharide inhibitor

Introduction

Plants are the most common source of renewable carbon and energy on earth. They annually produce $\sim 4 \times 10^9$ tons of cellulose, a highly stable polymer consisting of β -1,4-linked glycosyl residues, along with other polysaccharides (Coughlan, 1990). The potential of these biological resources as possible substitutes for diminishing fossil energy resources is becoming increasingly important.

The biological degradation of cellulose has been studied for many years, and a number of cellulolytic enzymes, especially cellulases produced by fungi and bacteria, have been isolated and characterized (Tomme *et al.*, 1995). These enzymes, which cleave the β -1,4 bond of cellulose, belong to the large family of glycosyl hydrolases. On the basis of sequence comparisons and hydrophobic cluster analysis, the catalytic domains of all glycosyl hydrolases have been classified into 63 families of homologous folds (Henrissat and Bairoch, 1993, 1996; Henrissat, 1997). The catalytic domains of cellulases and related xylanases have been identified in 14 of these families. According to their mechanism of cellulose degradation, cellulases are subdivided into either non-processive cellulases, simply called endocellulases or processive cellulases, the latter including the different exocellulases, and the new processive endocellulases (Barr *et al.*, 1996; Reverbel-Leroy *et al.*, 1997a). Endocellulases randomly cleave the cellulose chain at exposed positions and produce new reducing ends, while processive cellulases remain firmly attached to the chain and release mainly cellobiose or cellotetraose units from one end of the chain (Sakon *et al.*, 1997). These mechanisms are similar to those already described by Robyt and French (1967) in the degradation of amylose. In the degradation of crystalline cellulose, non-processive cellulases and processive cellulases have been found to work synergistically (Creuzet *et al.*, 1983; Henrissat *et al.*, 1985; Irwin *et al.* 1993).

The three-dimensional (3D) structures of eight non-processive and three processive cellulases from different microorganisms have been solved to date. These cellulases belong to glycosyl hydrolase families 5, 6, 7, 8, 9, 10, 11, 12 and 45. The fact that processive cellulases remain attached to the substrate chain after its initial cut leads to 3D structural differences between processive and non-processive cellulases. The active centres of the processive cellulases CBHI and CBHII from *Trichoderma reesei* are covered by a tunnel in which the cellulose chain is held and translocated during the processive action (Rouvinen *et al.*, 1990; Divne *et al.*, 1994, 1998). The active centres of the non-processive cellulases EGI of *Fusarium oxysporum* and E2 of *T.fusca* belonging to the same families, 6 and 7, are located in an open cleft (Spezio *et al.*, 1993; Davies and Schülein, 1995). The tunnel-forming loops present in the processive cellulases are missing in the non-processive ones, resulting in the release of both sides of the cellulose chain after the cleavage reaction.

The active site of the processive endocellulase E4 from *T.fusca* (glycosyl hydrolase family 9) is also located in an open cleft, which is contrary to the non-processive cellulases prolonged in an attached cellulose-binding domain (Sakon *et al.*, 1997). The cellulose chain stays connected to this cellulose-binding domain after the initial

Fig. 1. Chemical structure of the thiooligosaccharide inhibitor methyl 4-S-β-cellobiosyl-4-thio-cellobioside, which is called IG4 in this article. The sugar subunits are labelled from (A) to (D) from the non-reducing to the *O*-methylated reducing end.

cut, which is in agreement with the observed endocellulase activity and the following processive action along the chain (Sakon *et al.*, 1997). The different organization of the active centre of the processive endocellulase CelF of *Clostridium cellulolyticum* (glycosyl hydrolase family 48) and its 3D structure are described in this article.

The mesophilic anaerobic bacterium *C. cellulolyticum* secretes multifunctional multienzyme complexes called cellulosomes which degrade cellulose (Madarro *et al.*, 1991; Belaich *et al.*, 1993). These secreted complexes were first described from the thermophilic bacterium *C. thermocellum* (Lamed *et al.*, 1983). They consist of a scaffolding unit (called CipC in *C. cellulolyticum*) which mainly contains a cellulose-binding domain and several hydrophobic domains called cohesins (Salamitou *et al.*, 1992; Bayer *et al.*, 1994) to which various cellulases are attached via a C-terminal docking domain (Tokatlidis *et al.*, 1991). Nine genes coding for components of the cellulosomes of *C. cellulolyticum* have been identified so far (Faure *et al.*, 1989; Shima *et al.*, 1991; Bagnara-Tardif *et al.*, 1992; Pagès *et al.*, 1996; Reverbel-Leroy *et al.*, 1996; Gal, 1997), namely the two isolated genes *celA* and *celD* and a gene cluster including *cipC*, *celF*, *celC*, *celG*, *celE*, *celH* and *celJ*. Five of the corresponding cellulases, CelA, CelC, CelD, CelF and CelG, have been cloned in *E. coli* and characterized (Fierobe *et al.*, 1991, 1993; Shima *et al.*, 1993; Gal *et al.*, 1997a; Reverbel-Leroy *et al.*, 1997a). CelA (5), CelC (8), CelD (5) and CelG (9) are non-processive cellulases (glycosyl hydrolase family in brackets), whereas CelF (48) is a processive endocellulase (Reverbel-Leroy *et al.*, 1997a). The 3D-structures of CelA (5) and CelC (8) have already been solved (Ducros *et al.*, 1995; M.Juy, G.Parsiegla, C.Gaudin, A.Belaich, J.P.Belaich and R.Haser, in preparation).

In this article we present the 3D-structure of the catalytic domain of CelF, a new processive endocellulase fold and the first structure of a member of the glycosyl hydrolase family 48 to be described. The full-length enzyme CelF has a molecular mass of 77.62 kDa and is one of the three major components of the cellulosome complexes of *C. cellulolyticum*, and can therefore be expected to play a key role in cellulose degradation. Two other family 48 enzymes, CelS from *C. thermocellum* and P70 from *C. cellulovorans*, are thought to play a similar role in their cellulolytic systems (Morag *et al.*, 1991; Doi *et al.*, 1993; Morag *et al.*, 1993). CelF has been cloned and expressed in *E. coli* and its C-terminally truncated active form (71 kDa) crystallized in the presence of the thiooligosaccharide inhibitor IG4 (Figure 1) (Reverbel-Leroy *et al.*, 1997b).

Results

The crystal structure of the catalytic domain of CelF in complex with two molecules of the inhibitor IG4 (Figure

1) was determined at 2.0 Å resolution by multiple isomorphous replacement with use of the anomalous data of the derivatives (MIRAS method). The final model consists of a single chain of 629 residues forming the catalytic subunit of CelF, 334 water molecules, one calcium ion and two molecules of the inhibitor IG4. The structure was refined to a final *R*-factor of 16.3%, including all the reflections ($R_{\text{free}} = 21.0\%$). In the Ramachandran plot (Ramakrishnan and Ramachandran, 1965), 90.7% of the residues were found to be in the most favourable regions, 8.9% in the favourable regions, and only two outliers, Glu44 and Val402, in the generously allowed and disallowed regions, respectively. These two residues are part of the active centre and may therefore accommodate energetically non-favourable main chain angles due to the inhibitor–enzyme interactions.

Overall structure

The catalytic domain of CelF is a monomeric globular unit with dimensions of $\sim 70 \times 65 \times 55$ Å³. Its major characteristic is a left-handed ($\alpha\alpha$)₆-helix barrel with short loops on the C-terminal side of the inner helices and long connections on its N-terminal side (Figure 2A). The long connections form a layer of five antiparallel β-sheets covering the barrel. The other β-sheets and helices are involved in the formation of a tunnel and a cleft (Figure 2B). One short parallel β-sheet connects two of the tunnel-forming loops. This globular unit belongs mainly to one large domain (A). In addition to the main domain A, a region comprising residues Gly545–Val602 forms domain B, which is firmly attached at one side of the cleft.

The active centre is situated on the N-terminal side of the inner helices, at the end of the tunnel before the cleft. Two of the tunnel-forming loops bind one calcium ion at the solvent-exposed side above the active centre. One IG4 molecule (Inh1) is located in the tunnel, and a second one (Inh2) in the open cleft. The two cysteines present in the structure are not involved in a disulfide bridge. Three *cis*-prolines (Pro122, Pro174 and Pro406) could be detected, amounting to 9.4% of the total proline content.

The ($\alpha\alpha$)₆-barrel

The 12 helices of the ($\alpha\alpha$)₆-barrel (a1–a12) consist of 15–22 amino acids and show an alternating connection pattern between outer and inner helices, as is common in the case of ($\alpha\alpha$)₆-barrel structures (Aleshin *et al.*, 1992; Juy *et al.*, 1992). The inner helices a2, a4, a6, a8, a10 and a12 run parallel, forming a barrel. The shape of the barrel is slightly ellipsoidal due to helix a6, which is tilted. Four of the outer helices (a5, a7, a9 and a11) are tilted horizontally with respect to the axes of the inner helices. The two remaining helices (a1 and a3) run parallel to these axes. The N- and the C-termini of the catalytic domain of CelF correspond to the barrel termini and are positioned on neighbouring helices. On the C-terminal side of the inner helices, the barrel helices are connected by five short loops up to five residues in length. The six connections I–VI (see Figure 2A) on the N-terminal side vary from 29 to 138 residues in length and form the tunnel, the cleft and parts of the active centre. Domain B, between residues Gly545 and Val602, is inserted into the barrel–helix connection VI after a β-sheet. It is ~60 residues long and forms a flap on the right side of the

Fig. 2. Secondary structure plots of the catalytic subunit of CelF as assigned by the program PROMOTIF. **(A)** 2D plot: the 12 helices of the $(\alpha\alpha)_6$ -barrel are named a1–a12. Helices with an even number belong to an helix of the inner barrel. The long helix connections on the N-terminal side of the inner helices are numbered from I to VI. I, residues 27–55; II, residues 89–227; III, residues 272–324; IV, residues 364–411; V, residues 448–493; VI, residues 534–608. **(B)** 3D stereo-plot (drawn with MOLSCRIPT): the α -helices are shown in red, 3_{10} -helices orange and β -sheets blue. The two inhibitor molecules are green and the three proposed catalytic active residues, Glu55, Glu44 and Asp230, pink. The coil region of Domain B is blue with the colour of the helices and β -sheet slightly changed compared to Domain A.

cleft, which continues on the surface of this tunnel side. The first part of Domain B consists of two 3_{10} and two α -helices which are roughly arranged in the form of a square. They are followed by a long bent loop ending in a small β -sheet which is packed with one side facing the helical part. Domain B interacts with the tunnel-forming loops by a hydrophobic surface and forms hydrogen bonds (H-bonds) with three of the omega loops that close the tunnel to the solvent side (see the next section). No water molecules are involved in these interactions.

The $(\alpha\alpha)_6$ -barrel in CelF shows some uncommon features compared with the other $(\alpha\alpha)_6$ -barrels described so far. The centre of the barrel is not as densely packed as in the $(\alpha\alpha)_6$ -barrel of family 9 endoglucanases. The arrangement of the helices resembles that observed in the structures of families 8 and 15. Unlike those of the latter $(\alpha\alpha)_6$ -barrel families, the aromatic residues of the six inner helices of CelF are oriented with their hydrophilic part (OH in tyrosine or NH in tryptophan side chains)

towards the centre of the barrel. This results in a hydrophilic region of ~ 22 Å in length which runs along the central axis of the barrel. This region contains nine water molecules, begins at the C-terminal end of the inner barrel and leads to the active centre groove between residues Glu55 and Glu230. Halfway along its length, Trp62, a strictly conserved residue present in all the known family 48 glycosyl hydrolases, disrupts the chain of water molecules. This Trp side chain is surrounded by long hydrophilic residues (Glu424, Gln237 and Gln420), which are also strictly conserved (with the exception of Gln237, substituted by Glu in the CbhB of *Cellulomonas fimi*). Such a continuous hydrophilic region through an $(\alpha\alpha)_6$ -barrel is not common in $(\alpha\alpha)_6$ -barrel structures.

The tunnel

The most interesting feature of the structure is the tunnel that covers about two-thirds of the N-terminal barrel face. Its interior has an elliptical shape with dimensions of

Fig. 3. Cut through the molecular surface along the $(\alpha\alpha)_6$ -barrel axis of CelF. Regions with negative and positive potential are red and blue, respectively. The inhibitor molecule Inh1 is located in the tunnel and Inh2 in the cleft. Here, the CelF molecule is rotated $\sim 180^\circ$ around the x -axis in relation to most of the other figures in order to offer the best view of the tunnel and the inhibitor molecules. The molecular surface and potential were calculated with the program GRASP (Nicholls *et al.*, 1991); water molecules were excluded.

$\sim 25 \times 12 \times 7$ Å (Figure 3). The inner surface of the tunnel is formed by the helix connections I, II, III and IV. Connections V and VI form flaps, which cover part of the outer surface of the tunnel and form the cleft following the tunnel exit. Looking from the solvent side into the tunnel with the $(\alpha\alpha)_6$ -barrel orientated downwards, the entrance and the right side of the tunnel are provided by connections I and II and show no specific types of residues. The bottom surface consists mostly of a β -sheet formed by connections II and III. Trp310 and Trp312 of this β -sheet are involved in stacking interactions with the inhibitor, as described in detail later. The left tunnel side is made of connections III and IV, which expose hydrophobic as well as hydrophilic residues to the tunnel interior.

The architecture of the tunnel ceiling involves four loops (106–133, 179–191, 209–219 and 400–409) belonging to the helix connections II and IV, which close the tunnel on the solvent-exposed side (Figures 4 and 5). They fit well into the classification of omega loops (Fetrow, 1995), and are stabilized by H-bonds and hydrophobic interactions (Table I). The loops are attached to each other in two pairs. The largest part of the tunnel ceiling is formed by the first pair, loops 106–133 and 209–219. The long loop 106–133 (Figure 4) is twisted about 180° around the section of residues 107–112 and 132–125, and closes the tunnel like a flap. A short parallel β -sheet with loop 209–219 is formed after the twist, while six H-bonds and salt bridges (Table I) stabilize the other side of the flap with the left tunnel side.

The second pair, loops 179–191 and 400–409, is connected by one H-bond between backbone atoms and participates in the calcium-ion-binding site that is located on the solvent-exposed side above the proposed active

centre (Figure 5). All of the seven calcium ligands are oxygen atoms. Four of them belong to loop 179–191 (Gln185O, Gln185OE1, Glu190OE1 and Glu190OE2), while the fifth is part of loop 400–409 (Asp405OD2). On the surface-exposed side, two water molecules complete the coordination sphere of the ion. The calcium oxygen distances vary from 2.4 Å, in the case of the carbonyl oxygen Gln185O, to 2.7 Å in that of Glu190OE1 and Glu190OE2. The seven ligands of the calcium ion form a pentagonal bipyramid, with a geometry disturbed by Asp405, which is slightly out of the ideal coordination plane.

Loop 400–409 displays an unusual conformation. Val402 is in a disallowed region of the Ramachandran plot and Pro406 is in its *cis*-conformation. The driving force inducing the disallowed conformation of Val402 may be the participation of Asp405 in the calcium binding site. Asp405 is hydrogen-bonded to Arg549 and therefore negatively charged, which makes it a strong ligand for the positively charged calcium ion. The observed loop conformation may stabilize the orientation of Tyr403, which points inside the tunnel, reducing its height at the beginning of a depression in the centre of the $(\alpha\alpha)_6$ -barrel (Figure 5). Tyr403 may help to push a substrate chain into the depression and to position the substrate's glycosidic bond close to the location of the presumed cleavage site.

The two pairs of tunnel-forming loops are not directly connected. Only one van der Waals interaction (Tyr121–Arg182) exists between them; elsewhere they are separated by a layer of well-defined water molecules involved in two water bridges between loops 106–133 and 179–191. The tunnel exit at the active-centre side is formed by a ring of five aromatic residues: Trp154, Phe180, Trp298, Tyr323 and Tyr403. The following depression contains residues belonging to the inner barrel helices and an omega-loop (35–50) which is part of helix connection I.

Although the calcium ion seems to be essential to the closure of the tunnel in the 3D-structure, no Ca^{2+} -dependence of the cleavage reaction has been observed (Reverbel-Leroy *et al.*, 1997a). In CelS/S₈ from *C.thermocellum* and Avicelase II of *C.stercorarium*, other family 48 cellulases, Ca^{2+} or other bivalent ions have been found to stabilize the enzyme in the higher temperature ranges (Bronnenmeier *et al.*, 1991; Morag *et al.*, 1991; Kruus *et al.*, 1995). A calcium ion was inserted during refinement in the ion binding site, even though the crystallization buffer contained 20 mM of MgCl_2 . Calcium fitted better to the electron density map than a magnesium ion as the mode of coordination, and the distances observed are typical of Ca^{2+} complexes (Glusker, 1991). The bound ion was therefore assumed to be Ca^{2+} , which was probably in a very stable complex with CelF during its production and not lost during the purification procedure.

The active centre

The family 48 glycosyl hydrolases cleave the sugar chain with inversion of the anomeric carbon (Shen *et al.*, 1994). CelF has been shown to be a processive endo-glycosyl hydrolase, performing a processive degradation of the cellulose chain after an initial endo-attack (Reverbel-Leroy *et al.*, 1997a). As described in detail hereafter, its active centre is located on the N-terminal side of the

Fig. 4. Close-up on the tunnel-forming omega loops, indicating the H-bonds between loop 106–133 and the protein residues. The parallel β -sheet between loop 106–133 and loop 209–219 is located above the inhibitor molecule Inh1.

inner helices, as is usually observed among $(\alpha\alpha)_6$ -barrel-containing glycosyl hydrolases. Unlike the $(\alpha\alpha)_6$ -barrels of the non-processive glycosyl hydrolases of family 8, 9 and 15, which have an open-cleft architecture, the active centre of CelF is located in a depression covered by one end of the tunnel. This resembles the architecture observed in processive enzymes like the cellobiohydrolases of families 6 and 7, in which the active centre is located in a tunnel (Rouvinen *et al.*, 1990; Divne *et al.*, 1994) and differs from the position of the active centre of the processive endocellulase E4 of *T.fusca*, which is located in an open cleft (Sakon *et al.*, 1997).

One of the two inhibitor molecules (Inh2) is bound in such a way that its non-reducing end penetrates in the active-site depression (Figure 6). Glu44, Glu55 and Asp230 are located close to the non-reducing end of Inh2 and are therefore likely candidates for providing the corresponding acid and base in the cleavage reaction. They are strictly conserved in all sequences of family 48 members determined so far. Glu55 is hydrogen-bonded to the non-reducing end of the Inh2 molecule, which makes it the most likely candidate for the role of the proton donor. Asp230 is located in the depression in the $(\alpha\alpha)_6$ -barrel between the two bound inhibitor molecules Inh1 and Inh2. Glu55 and Asp230 belong to two neighbouring inner helices (α_2 and α_4) and form the end of the hydrophilic region running through the inner $(\alpha\alpha)_6$ -barrel. The average distance between the four acidic oxygens of Glu55 and Asp230 is 6.6 Å, which is very short for an acid–base pair functioning in an inverting reaction (McCarter and Withers, 1994). They are bridged by a water molecule, which is further bound to Arg234. A second water bridge is formed between Asp230 and

Fig. 5. View of the calcium-binding site between loop 179–191 and loop 400–409, showing the interactions of Asp405 in detail. The positions of Val 402 (with disallowed ϕ/ψ angles), of the *cis*-proline 406 and of Tyr 403, which points into the tunnel, have also been inserted.

Arg421. All these interactions lead to an environment of Asp230 that supports a charge on its carboxylate group.

The second possible catalytic base, Glu44, is part of an omega loop (35–50) and is located between subunits A and B of Inh2 in a very basic environment. It is hydrogen-bonded to Arg544, Arg609 and His36, which are all strictly conserved in the sequences of the members of family 48. Considering this basic environment, Glu44 is probably charged. The average distance between the four acidic oxygens of Glu44 and Glu55 is 8.6 Å, which is

Table I. H-bonds of the four tunnel-closing omega loops with protein residues

Interacting residues	d (Å)	Remarks	Interacting residues	d (Å)	Remarks
Loop 106–133:			Loop 179–191:		
Asn106 N → Asn104 OD1	3.08		Thr179 N → Asn178 OD1	2.79	
Lys107 N → Asp104 O	2.97		Arg182 NH1 → Phe556 O	2.75	
Pro108 O → Ile313 N	3.16		Glu184 O → Ala404 N	3.01	
Glu114 N → Gly380 O	3.22		Ser187 OG → Asp549 NH2	3.24	
Glu114 OE1 → His318 NE2	3.03		10 internal H-bonds		
Glu114 OE2 → Lys274 NZ	2.95		Loop 209–219:		
Glu114 O → Arg381 NH1	2.94		Pro219 O → His200 NE2	3.15	
Gln116 NE2 → Asn379 O	3.02		Phe212 O → Ser123 N	2.85	β-sheet
Gln116 OE1 → Arg381 NH1	2.81		Lys214 N → Ser123 O	2.90	β-sheet
Ser123 N → Phe212 O	2.85	β-sheet	Lys214 O → Leu125 N	2.99	β-sheet
Ser123 O → Lys214 N	2.90	β-sheet	Leu211 N → Phe208 O	3.09	
Leu125 N → Lys214 O	2.99	β-sheet	Asp210 OD1 → Trp561 NE1	2.93	
Pro130 O → Arg277 NH2	2.95		3 internal H-bonds		
Pro130 O → Arg277 NE	2.82		Loop 400–409:		
Gly132 N → Tyr275 O	2.93		Ala404 N → Glu184 O	3.01	
Arg133 N → Gly286 O	3.17		Asp405 OD1 → Arg549 NH1	3.21	
11 internal H-bonds			Asp405 OD2 → Arg549 NH2	3.17	
			Gly407 O → Thr410 OG	2.78	
			Gly407 O → Thr410 N	3.12	
			Ser408 O → His320 NE2	3.08	
			Asn409 O → Trp465 NE1	3.01	
			4 internal H-bonds

Fig. 6. Stereoview of selected residues along the inhibitor molecules Inh1 (subsites –6 to –3) and Inh2 (subsites +1 to +4) and the active centre. Part of the two helices which fix Glu55 (bold) and Asp230 (bold), as well as a part of the omega loop 35–50 around residue Glu44 (bold) are traced with thin lines.

within the expected range of an acid–base pair involved in an inverting reaction.

The depression at the active centre is followed by an open V-shaped cleft. This cleft is formed by helix connections V and VI and has a length of ~16 Å. In the middle of the cleft are two acids, Asp494 and Glu542, located on opposite walls at the bottom of the cleft. As in the case of Glu55 and Asp230, their acidic oxygens are at an average distance of 6.5 Å. Glu542 forms H-bonds

with Arg544 and Arg609, while Asp494 is hydrogen-bonded to the O₆ atom of sugar unit B of Inh2. Although the arrangement of these amino acids resembles that of a protein active site designed to cleave a glycosidic bond, no H-bond to a glycosidic oxygen or any structural evidence of a nucleophilic attack could be observed from those residues, either in the IG4 complex or in CelF complexed to cellobiose (unpublished data). Asp494, Glu542, Arg544 and Arg609 are all strictly conserved in

Fig. 7. The $(2F_o - F_c)$ electron density map contoured at the 1σ level, following the inhibitor molecules Inh1 (–6 to –4) and Inh2 (+1 to +4) from the tunnel to the end of the open cleft. It shows as well the unexplained density at the gap between the inhibitor molecules.

family 48 glycosyl hydrolases. It would certainly be worth investigating these residues using site-directed mutagenesis experiments.

The inhibitor/cellulase interactions

The density corresponding to two IG4 molecules (Inh1 and Inh2) appeared clearly in the difference Fourier map calculated with the pre-refined model containing all 629 amino acids and contoured at the 3σ level. These positions were reconfirmed by the $(2F_o - F_c)$ electron density map at 1σ level. Inh1 is located in the tunnel, while Inh2 is bound directly after the proposed cleavage site in the depression and in the following cleft. In the final $(2F_o - F_c)$ electron density map contoured at the 1σ level, patches of unexplained electron density filling a spacing of at least two glucose units can be observed between the two inhibitor binding sites (Figure 7).

In agreement with the nomenclature proposal by Davies *et al.* (1997), Inh1 binds in subsites –6 to –3. The inhibitor's conformation can be roughly subdivided into two cellobiose units, the first in subsites –6 and –5, and the second in subsites –4 and –3, which are twisted along the thioglycosidic bond. The relative twists between all sugar units were calculated using the Φ/Ψ values, as defined by Divne *et al.* (1998) and comparing them with the proposed normal values. The first cellobiose has no remarkable twist (8°) along its internal glycosidic bond –6/–5 and the second is slightly twisted along –4/–3 by 22° . The thioglycosidic connection –5/–4 has a much larger twist of $\sim 48^\circ$. This torsion between the cellobiose units follows the ellipsoidal architecture of the tunnel, whose broadest dimension changes from nearly horizontal in the first part around subsites –6 and –5, to more vertical at subsite –4 and further.

The first two sugar units A and B of Inh1 in subsites –6 and –5 fit the $(2F_o - F_c)$ electron density map satisfactorily, whereas the units C and D in subsites –4 and –3 are not as well defined and show additional as well as

missing electron density (Figure 7). Sugar unit A in subsite –6 and unit D in subsite –3 are involved in stacking interactions with Trp310 and with Tyr299, respectively, while the units in subsites –5 and –4 are only partly stacked against Trp312. All of these stacking interactions are located on the same side of the inhibitor. The other side is stabilized by seven H-bonds to tunnel residues and by one water-mediated H-bond (Figure 8A).

In and around subsites –5 and –4, the difference Fourier map contoured at the 2.0σ level shows patches of density for alternative positions for the glucose units. These positions of apparent low occupancy favour stacking interactions with Trp312 at subsite –5. Although these alternative positions were not introduced in the final model, they provide an explanation for the smeared appearance of the electron density in the $(2F_o - F_c)$ electron density map around subsites –4 and –3 of Inh1.

In subsite –2, no continuous electron density region is observed and therefore only solvent molecules are inserted. Subsite –1 is filled by an ambiguous residual electron density, which may correspond to either a sugar molecule in a disturbed or unusual conformation or to a mixture of several conformations. Since no sugar model with boat, chair or open conformation gave a satisfying fit with the present electron density, the site has been left unoccupied in the refined model.

Inh2 is situated in the open cleft after the tunnel exit and occupies subsites +1 to +4. The initial $(F_o - F_c)$ electron density map at subsites +1 and +2 was so clearly defined that the direction of the sugar chain could be determined: it is oriented with its non-reducing end towards the tunnel exit. In contrast to Inh1, which is completely enclosed in the tunnel, Inh2 is mostly exposed to the solvent and therefore occupies the probable position of the leaving group in the processive cleavage reaction. These observations lead to the conclusion that CelF acts processively from the reducing to the non-reducing end of the cellulose chain.

A

B

Fig. 8 Schematic drawing of the hydrogen-bonding pattern between the enzyme residues and (A) the inhibitor molecules Inh1 and (B) Inh2, with their subsites indicated underneath. The distances in Ångstroms are given for each bond. Aromatic residues involved in stacking interactions with the inhibitor are also indicated.

Inh2 has a relative twist between each single sugar molecule ranging from 36° to 44° , which is in contrast to the twist between two cellobiose units of Inh1. In general, the twists of both molecules Inh1 and Inh2 are cumulative and follow the same right-handed screw axis in the direction of the reducing end. At subsites +1 and +2, Inh2 fits very nicely into the electron density of the ($2F_o - F_c$) map, whereas it appears to be somewhat disordered at subsites +3 and +4, where it is not completely defined at the 1.0σ electron density level (Figure 7). Five H-bonds, one water-mediated H-bond and a stacking interaction with Trp411 stabilize the sugar units A and B in subsites +1 and +2. The subsites +3 and +4 do not participate in any stacking interaction, and the unit C in subsite +3 forms only two water-mediated H-bonds with the enzyme (Figure 8B). The weaker interactions with the protein observed in subsites +3 and +4 explain the less well-defined electron density in these subsites.

The axes along the two inhibitor molecules Inh1 and Inh2 enclose an angle of $\sim 135^\circ$, which is consistent with the existence of a kink in the substrate chain at position -1. A similar kink for the cellulose chain has been proposed for the catalytic mechanism of the inverting endocellulase CelA of *C.thermocellum* (Alzari *et al.*, 1996).

Discussion

The structure of CelF of *C.cellulolyticum* is the first structure of a family 48 glycosyl hydrolase to be described and is shown to have a structure different from other processive endocellulases. CelF starts the digestion of cellulose by performing an endo-attack at an exposed chain, like an endocellulase, but then continues to cleave processively cellobiose, acting from one of the newly produced cellulose chain ends. Its peculiar mode of cellulolytic action is now revealed by the enzyme's 3D-structure.

Unfortunately, no structural data are available on the uncomplexed form of CelF, which could show the conformation of the free enzyme. Therefore, a proposal for the initial endo-activity of the enzyme has to be based on the available structural details of the complex. The cellulose subsites in all known non-processive cellulases and in the processive endocellulase E4 of *T.fusca* are situated in an open cleft in order to be freely accessible for the substrate (for a review see Davies and Henrissat, 1995; Sakon *et al.*, 1997). In CelF, only subsites +1 to +4 are located in the open-cleft region, while the others are hidden in the tunnel. Two of these cleft subsites, namely +1 and +2, are occupied by the inhibitor units with the lowest B-factor and the clearest electron density. They are the most probable subsites observed in the structure to bind the cellulose chain during the endo-attack. Trials to insert a model of an uncleaved cellulose chain into subsites +1 or +2 were hindered by its collision with the tunnel residues; therefore an opening of the tunnel seems to be necessary for the performing of the initial endo-attack.

At the solvent-exposed side, four omega loops close the tunnel of CelF. Omega loops are typically found to occur in flexible regions of enzyme structures where they interact with substrates in an induced-fit mechanism (Fetrow, 1995). In CelF, these loops are connected in pairs to close the tunnel. The omega loop pair 106–133 and 209–219 form most of the ceiling of the tunnel. The sequence of the long loop 106–133 contains six prolines and five serine/threonine residues, and thus resembles in its composition that of flexible, proline-rich linker regions observed in glycanases (Gilkes *et al.*, 1991). This loop may therefore act as a flap that opens and closes the tunnel. It must be stressed that this crystal structure corresponds to the recombinant catalytic domain of CelF which is not glycosylated. The entire CelF expressed in *C.cellulolyticum* possibly has glycosylated residues (Gal *et al.*, 1997b), which could enhance the flexibility of certain regions. The interactions of this flap with the protein are mostly centred on two interfaces: one accumulation of H-bonds around residue 114 and the parallel β -sheet with loop 209–219 (Table I and Figure 5). These limited interactions with the protein may facilitate the attachment of the loop in the induced-fit mechanism. The observed twist of the loop in the closed tunnel conformation may furthermore allow a more relaxed conformation when the tunnel is open.

The most important connection in the loop pair 179–191 and 400–409 is the interaction of Asp405 with the calcium ion. Loop 179–191 forms 10 internal H-bonds (Table I), resulting in a very well-stabilized conformation in order to serve as a calcium-binding site. In an open

structure, Asp405 may be replaced by a water molecule, which would not disturb the stability of the loop conformation but would allow an ideal pentagonal bipyramidal coordination of the calcium ion involving one single loop. This complex would be consistent with the observed strong complexation of calcium during the purification procedure.

Although no open structure has been crystallized so far, the hypothesis of an open tunnel prior to substrate binding seems to be highly plausible. This hypothesis is supported by the fact that crystals of CelF could only be obtained in the presence of an inhibitor, probably due to the flexibility of loop 106–133, which might have prevented CelF from crystallizing in an open conformation.

After the initial endo-cut of the enzymatic reaction, the processive degradation process occurs and cellobiose units are released from the end of the cellulose chain. The processive action has to proceed in cellobiose steps through the tunnel up to the proposed cellobiose binding site (+1 and +2). The tunnel has enough space to contain at least six subsites up to the non-reducing end of the cellobiose binding site. Two of the four subunits of Inh1 detected at subsites –6 to –3 participate in stacking interactions with Trp310 and Tyr299, which are located on one side of the inhibitor chain. The third aromatic residue along the tunnel wall Trp312, which is located between the others, is not favourably positioned for performing a stacking interaction with Inh1 (Figure 7). The arrangement of the aromatic residues in the tunnel of CelF may stabilize multiple-inhibitor or substrate positions by stacking interactions. The alternative binding positions in the tunnel would smoothen the profile of the translocation energy required to transport a sugar chain and favour the processive action. A comparable arrangement of aromatic residues reducing the 'sliding energy' has been reported in studies on maltoporin (Dutzler *et al.*, 1996; Meyer and Schulz, 1997), where maltose chains are transported through a membrane channel, and recently in CBHI of *T.reesei* in which a similar sliding mechanism is proposed to translocate the cellulose chain along the tunnel of this processive cellulase (Divne *et al.*, 1998).

Family 48 cellulases cleave cellulose with inversion of the anomeric configuration, as it is known from the CbhB (previously known as CenE) of *C.fimi* (Shen *et al.*, 1994). In this mechanism, an acid protonates the glycosidic oxygen, while a water molecule performs a nucleophilic attack on the anomeric carbon. A base is needed to activate the water molecule. The base can be expected to be present at an average distance of ~10 Å to the acid according to the observations of McCarter and Withers (1994). Glu55 is located at the end of α -helix a1 and appears to be the best proton donor candidate, but the identity of the corresponding base is not so obvious.

One of two acidic residues (Glu44 and Asp230) is likely to serve as the base. The first candidate, Asp230, is situated at the exit of the hydrophilic region running through the $(\alpha\alpha)_6$ -barrel, on the opposite side of Glu55, to which it is connected via a water bridge. The hydrophilic region may serve as a possible water supply for the active centre during the processive cleavage reaction, after the closure of the tunnel. Contrary to the water channel observed in the inverting cellulase cellobiohydrolase II of *T.reesei* (Rouvinen *et al.*, 1990), the water molecules

which might move through the $(\alpha\alpha)_6$ -barrel of CelF would arrive at the bottom of the active-centre groove between the proton donor and Asp230 and not on the proton-donor-facing side. This is not the expected location for a water channel supplying an inverting reaction, but this function is not excluded.

Another clue for the choice of base in the active centre may be the presence of Arg234 between Glu55 and Asp230, a configuration which resembles the arrangement of two acids and an arginine in the active centre of other inverting glycosyl hydrolases, such as in CelC of *C.cellulolyticum* (M.Juy, G.Parsiegla, C.Gaudin, A.Belaich, J.P.Belaich and R.Haser, in preparation), CelA of *C.thermocellum* (Alzari *et al.*, 1996) or the retaining α -amylases (Kadziola *et al.*, 1994; Qian *et al.*, 1994; Aghajari *et al.*, 1998). In these glycosyl hydrolases, an arginine is positioned between the proton donor and the proposed nucleophile of the cleavage reaction.

The second candidate for the corresponding base is Glu44, whose average acidic oxygen's distance from Glu55 is in agreement with the expected base–acid distance for an inverting reaction (McCarter and Withers, 1994). Glu44 is not located in the active centre depression, but shifted along the inhibitor chain to a position between subsites +1 and +2 (Figure 7). The H-bonds to Arg544, Arg609 and His36 probably make it charged. His36 is located on the opposite site of Glu55 in the active centre depression in a good position to serve as an intermediate in a proton acceptor chain to activate a water molecule. Additionally, Glu44 is part of a small loop (41–49) which shows some geometric disturbances, such as unfavourable Φ/Ψ angles. This disturbance might act as the driving force, moving the loop towards a more ideal position when an uncleaved substrate is present in subsite –1. Crucial for the selection of the corresponding base in the active centre will be the orientation of the sugar residue in subsite –1 during the cleavage reaction. It determines the position of the water molecule that attacks the sugar from the rear. Thus further data are needed to discriminate between Glu44 and Asp230.

Conclusion

The 3D structure of the catalytic domain of CelF of *C.cellulolyticum* shows that family 48 cellulases have a new fold based on an $(\alpha\alpha)_6$ -barrel, with very long helix connections on one side of the barrel, which form a tunnel and a cleft. The binding pattern observed between CelF and two molecules of the inhibitor IG4 does not completely account for the processive endo-mechanism involved, but sheds light on some of its features. The organization of the tunnel-forming loops suggests an induced-fit mechanism with an open tunnel as the starting structure, even if no direct proof for this mechanism is available yet. The energy barrier on preventing the processive action through the tunnel seems to be reduced by several energetically favourable stacking positions spaced at short intervals. The orientation of the inhibitor residues is consistent with a processive action from the reducing end towards the non-reducing end of the cellulose chain. Although the proton donor involved in the cleavage reaction seems to have been identified, it still remains to be established if one of the two residues, Glu44 or Asp230, provides the

Table II. Statistics on native and heavy atom data used in structure determination

Data set	Native	HgCl ₂	KPtCl ₄
Resolution range (Å)	30.0–2.0	30.0–3.0	30.0–3.5
Completeness (%)	99.9 (100.0)	96.3 (97.8)	99.6 (98.6)
Redundancy	4.2 (3.9)	5.1	6.0
<i>R</i> _{sym} (%)	7.3 (24.8)	16.3	7.1
<i>R</i> _{iso} (%)		27.4	23.3
No. of heavy atom sites		2	1
<i>R</i> _{Cullis}		0.58	0.76
Phasing power			
Centric		1.42	0.82
Acentric		1.68	1.18
Mean overall figure of merit	0.47 at 3.0 Å		

Each data set was collected with a single crystal. The values in parentheses correspond to the highest shell of resolution (2.05–2.00 Å for the native data set).

$R_{\text{sym}} = \sum_{hkl} \sum_i |I_{hkl}| - I_{hkl,i} / \sum_{hkl} \sum_i |I_{hkl}|$ where *i* is the number of reflections *hkl*.

$R_{\text{iso}} = \sum_{hkl} |F_{ph} - \bar{F}_p| / \sum_{hkl} |F_{ph}|$ where *Fph* are the structure factors of the derivative and *Fp* those of the native crystal.

$R_{\text{Cullis}} = \sum_{hkl} |F_{ph} - \bar{F}_p + \bar{F}_h| / \sum_{hkl} |F_{ph} - F_p|$ for the centric terms only, *Fh* are the structure factors of the heavy atom.

corresponding base. Further investigations using mutational analysis in combination with a crystallographic approach are now under way.

Materials and methods

Crystallization and crystal packing

The methods used to express and purify the truncated CelF have been described elsewhere (Reverbel-Leroy *et al.*, 1996). Crystals of truncated CelF used for the structure determination were grown by means of the vapour diffusion method with macro-seeding techniques in the presence of the thiooligosaccharide inhibitor IG4, as described previously (Reverbel-Leroy *et al.*, 1997b). The crystals are orthorhombic and belong to space group P2₁2₁2₁ with cell dimensions *a* = 61.45 Å, *b* = 84.54 Å, *c* = 121.94 Å. One molecule is present per asymmetric unit, yielding a solvent content of ~45%. Crystals of the heavy atom derivatives used for phasing were obtained by cocrystallizing CelF with HgCl₂ or KPtCl₄ (1 mM in the droplet).

Only H-bonds to the six neighbouring molecules along the *a*- and *b*-axes can be observed in the crystal packing and none along the long *c*-axis. This may be the reason for the plate-like morphology of the crystals observed. Neither the inhibitor molecules nor any of the four tunnel-closing omega loops at the solvent-exposed tunnel ceiling are involved in any crystal contacts. A short sequence at the tunnel entrance of one omega loop (106–133) forms an intermolecular interface with two of the neighbouring molecules. This might explain why the correct formation of the tunnel is essential to the crystallization and why only crystals cocrystallized with an inhibitor and a closed-tunnel conformation could be obtained.

Data collection

All the data sets were collected using CuK_α radiation from a Rigaku RU-200 rotating-anode generator, with a graphite monochromator. The diffraction data on the native and the derivative crystals were collected on a MARresearch image plate scanner, at 15°C. The data sets were integrated using the DENZO software program (Otwinowski, 1993) and were scaled and further processed with programs from the CCP4 package (Collaborative Computational Project Number 4, 1994). The statistics on all the diffraction data sets are given in Table II.

Phase calculation

The positions of one platinum and two mercury sites were easily obtained by interpreting the Harker sections of the Patterson maps, and were

Table III. Statistics on the refinement and the quality of the final model

Resolution range (Å)	30.0–2.0
No. of unique reflections	43593
Number of non-hydrogen atoms	5442
Number of water molecules	334
<i>R</i> factor (%)	16.3
Free <i>R</i> factor (%)	21.0
Mean B of buried atoms (Å ²)	12.9
Mean B of water atoms (Å ²)	24.3
Mean B of inhibitor atoms (Å ²)	33.6
R.m.s. bond length (Å)	0.009
R.m.s. bond angle (°)	1.465
R.m.s. improper angle (°)	1.234
Ramachandran plot:	90.7
Residues in most favoured regions (%)	
Residues in generously allowed regions (%)	0.2 (Glu44)
Residues in disallowed regions (%)	0.2
	(Val402)

confirmed by inspecting the anomalous difference Patterson maps. A common origin was determined by calculating a difference Fourier map using the phases based on the single platinum site. Heavy atom positions were refined and the phases were calculated using the MLPHARE software program (Otwinowski, 1991). Anomalous scattering data on all the derivatives were included to ensure that the correct enantiomorph was chosen and to enhance the quality of the phases. The data on each derivative were used to a different resolution in order to obtain a reasonable phasing power. The mean figure of merit (FOM) of the acentric/centric phases at 3.0 Å was 0.47. Heavy-atom phases were used at up to 3.0 Å resolution and phase extension was performed at up to 2.4 Å resolution using the native diffraction data. Solvent flattening, histogram matching and phase extension were performed to 2.4 Å with the DM software program (Cowtan, 1994). The resulting phases were used to calculate the initial MIR map. All the statistics on the phase calculation are summarized in Table II. The mercury sites were subsequently found to be positioned next to Cys196 and Cys497 close to the inhibitor-binding sites. The single platinum site is located on the surface next to Met148 and Asp92.

Model building and refinement

The initial MIR-map showed a monomeric protein well-separated from the solvent. The BONES program and the O graphics program (Jones *et al.*, 1991) were used to trace the initial backbone skeleton. Interpretable regions of the bones were used to build strands of polyaniline chains into the density map, in which the (αα)₆-helix barrel was the first structural motif to emerge. The polyaniline chains were adapted to the density using energy minimization routines of the X-PLOR program (Brünger & Kurkowski, 1990). The stepwise combination of the phases from the growing polyaniline model with the MIR phases continuously improved the FOM and enhanced the quality of the electron density map. When ~85% of the backbone positions had been traced by the polyaniline chain, the side chains were inserted using the bulky aromatic residues in the helices as starting points for the fitting of sequence fragments to the electron density. The model was then stepwise completed by monitoring the *R*_{free} of the model (Brünger, 1992) and the FOM of the combined model/MIR phases. The final stages in the sequence adaptation were refined using simulated annealing at 1000 K. Water molecules were inserted using the 'watpick' script from the X-PLOR program package. Finally, the calcium ion and the inhibitor molecules were fitted to the density map and the entire structure refined using simulated annealing and bulk-solvent correction (Jiang & Brunger, 1994). The final refinement procedure consisted of running 30 cycles of prestage followed by a simulated annealing procedure at 3000 K and 25 cycles of restrained individual B-factor refinement followed by 30 steps of positional refinement. The bulk-solvent correction and the overall weighting of the reflection data (WA value) were recalculated after each step in the procedure. A starting B-factor of 15.0 Å² was used for all the protein atoms and the calcium ion, while the solvent and inhibitor atoms were set at an initial value of 20.0 Å². All the water molecules were finally checked to ensure that a 1.0 σ electron density

level was present in the calculated ($2F_o - F_c$) electron density map and at least one H-bond with a non-solvent molecule was formed. The Engh and Huber parameter sets were used for all the refinement procedures (Engh and Huber, 1991). To construct the inhibitor, the standard Weiss parameters for carbohydrates delivered with the X-PLOR program and the bond parameters from the structure of 2,3,4,6-tetra-*O*-acetyl-1-*S*-benzhydroximoyl- α -D-glucopyranose (Durier and Driguez, 1992) were used. Ten per cent of the reflection data were put aside for the R_{free} cross-validation procedure.

Quality of the final model structure

The final model consists of 629 residues, beginning with the first residue after the predicted signal peptide cleavage site and ending seven residues before the predicted start of the dockerin domain. In addition, 334 water molecules, one calcium ion and two inhibitor molecules were inserted. The final R -factor containing all the reflections is 16.3% with an R_{free} of 21.0% (Table III). The overall accuracy of the geometry and the check for statistical outliers were performed with the program WHATCHECK (Vriend, 1990) from the WHATIF program package and the program PROCHECK (Laskowski *et al.*, 1993; Rullmann, 1996). The parameters of Morris *et al.* (1992) were applied to define the zones in the Ramachandran plot. The secondary structure was assigned by the program PROMOTIF (Hutchinson and Thornton, 1996). The 3D-structure plots were calculated with the programs MOLSCRIPT (Kraulis, 1991) and RASTER-3D (Merritt and Murphy, 1994) or with the graphics visualization program O (Jones *et al.*, 1991). The coordinates and structure factors have been submitted to the Brookhaven protein data bank, identification code 1fce, and are on hold for 1 year after publication.

Acknowledgements

We acknowledge Dr David Hulmes for his critical reading of the manuscript. This work was supported by the European Union Eurocell Project (Bio4-CT97-2303) and the Centre National de la Recherche Scientifique (CNRS).

References

- Aghajari, N., Feller, G., Gerday, Ch. and Haser, R. (1998) Crystal structures of the psychrophilic α -amylase from *Alteromonas haloplanctis* in its native form and complexed with an inhibitor. *Protein Sci.*, **7**, 564–572.
- Aleshin, A., Golubev, A., Firsov, L.M. and Honzatko, R.B. (1992) Crystal structure of glucoamylase from *Aspergillus awamori* var. X100 to 2.2 Å. *J. Biol. Chem.*, **267**, 19291–19298.
- Alzari, P.M., Souchon, H. and Dominguez, R. (1996) The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure*, **4**, 265–275.
- Bagnara-Tardif, C., Gaudin, C., Belaich, A., Hoest, P., Citard, T. and Belaich, J.P. (1992) Sequence analysis of a gene cluster encoding cellulases from *Clostridium cellulolyticum*. *Gene*, **119**, 17–28.
- Barr, B.K., Hsieh, Y.L., Ganem, B. and Wilson, D.B. (1996) Identification of two functionally different classes of exocellulases. *Biochemistry*, **35**, 586–592.
- Bayer, E.A., Morag, E. and Lamed, R. (1994) The cellulosome—a treasure trove for biotechnology. *Trends Biotechnol.*, **12**, 379–386.
- Belaich, J.P., Gaudin, C., Belaich, A., Bagnara-Tardif, C., Fierobe, H.P. and Reverbel, C. (1993) Molecular organization and biochemistry of the cellulolytic system of *Clostridium cellulolyticum*. In *Abstracts of the Microforum 93: Genetics, Biochemistry and Ecology of Lignocellulose Degradation*. Toba, Japan, pp. 53–62.
- Bronnenmeier, K., Rucknagel, K.P. and Staudenbauer, W.L. (1991) Purification and properties of a novel type of exo-1,4-beta-glucanase (Avicelase II) from the cellulolytic thermophile *Clostridium stercorarium*. *Eur. J. Biochem.*, **200**, 379–385.
- Brünger, A.T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Brünger, A.T. and Kurkowski, A. (1990) Slow cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. Sect. A*, **46**, 46–57.
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
- Coughlan, M. (1990) Cellulose degradation by Fungi. In Fogarty, W.M. and Kely, C.T. (eds), *Microbial Enzymes and Biotechnology*. Elsevier Applied Science, London, UK, pp. 1–36.
- Cowtan, K. (1994) 'Dm': an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett.*, **31**, 34–38.
- Creuzet, N., Berenger, J.F. and Frixon, C. (1983) Characterization of exoglucanase and synergistic hydrolysis of cellulose of *Clostridium stercorarium*. *FEMS Microbiol. Lett.*, **20**, 347–350.
- Davies, G.J. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.
- Davies, G.J. and Schülein, M. (1995) Structural studies on fungal endoglucanases from *Humicola insolens*. In Petersen, S.B., Svensson, B. and Pedersen, S. (eds), *Carbohydrate Bioengineering*. Elsevier Science B.V., Amsterdam, The Netherlands, pp. 232–235.
- Davies, G.J., Wilson, K.S. and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.*, **321**, 557–559.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T. and Jones, T.A. (1994) The three-dimensional structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science*, **265**, 524–528.
- Divne, C., Ståhlberg, J., Teeri, T.T. and Jones, T.A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.*, **275**, 309–325.
- Doi, R.H. *et al.* (1993) Structure and function of the subunits of the *Clostridium cellulovorans* cellulosome. In *Abstracts of the Microforum 93: Genetics, Biochemistry and Ecology of Lignocellulose Degradation*. Toba, Japan, pp. 43–52.
- Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H.P., Belaich, J.P., Davies, G.J. and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure*, **3**, 939–949.
- Durier, V. and Driguez, H. (1992) Stereochemical investigation of 2,3,4,6-Tetra-*O*-acetyl-1-*S*-benzhydroximoyl- α -D-glucopyranose. *Acta Crystallogr. Sect. C*, **48**, 1791–1794.
- Dutzler, R., Wang, Y.-F., Rizkallah, P.J., Rosenbusch, J.P. and Schirmer, T. (1996) Crystal structure of various maltooligosaccharides bound to maltoporin reveal a specific translocation pathway. *Structure*, **4**, 127–134.
- Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr. Sect. A*, **47**, 392–400.
- Faure, E., Belaich, A., Bagnara, C., Gaudin, C. and Belaich, J.P. (1989) Sequence analysis of the *Clostridium cellulolyticum* endoglucanase-A-encoding gene, *celCCA*. *Gene*, **84**, 39–46.
- Fetrow, J.S. (1995) Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.*, **9**, 708–717.
- Fierobe, H.P., Gaudin, C., Belaich, A., Loutfi, M., Faure, E., Bagnara, C., Baty, D. and Belaich, J.P. (1991) Characterization of endoglucanase A from *Clostridium cellulolyticum*. *J. Bacteriol.*, **173**, 7956–7962.
- Fierobe, H.P., Bagnara-Tardif, C., Gaudin, C., Guerlesquin, F., Sauve, P., Belaich, A. and Belaich, J.P. (1993) Purification and characterization of endoglucanase C from *Clostridium cellulolyticum*—catalytic comparison with endoglucanase A. *Eur. J. Biochem.*, **217**, 557–565.
- Gal, L. (1997) Etude du cellulosome de *Clostridium cellulolyticum* et l'un de ses composants: la cellulase CelG. PhD Thesis, Université de Provence, Aix-Marseille I, France.
- Gal, L., Gaudin, C., Belaich, A., Pagès, S., Tardif, C. and Belaich, J.P. (1997a) CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. *J. Bacteriol.*, **179**, 6595–6601.
- Gal, L., Pagès, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C. and Belaich, J.P. (1997b) Characterization of the cellulolytic complex (cellulosome) produced by *C. cellulolyticum*. *Appl. Environ. Microbiol.*, **63**, 903–909.
- Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Jr and Warren, A.J. (1991) Domains in microbial β -1,4-glycanases: sequence conservation, function and enzyme families. *Microbiol. Rev.*, **55**, 303–315.
- Glusker, J.P. (1991) Structural aspects of metal liganding to functional groups in proteins. *Adv. Prot. Chem.*, **42**, 1–76.
- Henrissat, B. (1997) A new cellulase family. *Mol. Microbiol.*, **23**, 848–849.
- Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, **293**, 781–788.
- Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.*, **316**, 695–696.
- Henrissat, B., Driguez, H., Viet, C. and Schulein, M. (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Biotechnology*, **3**, 722–726.

- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.
- Irwin, D.C., Specio, M., Walker, L.P. and Wilson, D.B. (1993) Activity studies of eight purified cellulases: specificity, synergism and binding domain effects. *Biotechnol. Bioeng.*, **42**, 1002–1013.
- Jiang, J.S. and Brunger, A.T. (1994) Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.*, **243**, 100–115.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building models in electron density maps and the location of errors in the models. *Acta Crystallogr. Sect. A*, **47**, 110–119.
- Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeysens, M., Béguin, P. and Aubert, J.P. (1992) Three-dimensional structure of a thermostable bacterial cellulase. *Nature*, **357**, 89–91.
- Kadziola, A., Abe, J.I., Svensson, B. and Haser, R. (1994) Crystal and molecular structure of barley α -amylase. *J. Mol. Biol.*, **239**, 104–121.
- Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Kruus, K., Wang, W.K., Ching, J. and Wu, J.H. (1995) Exoglucanase activities of the recombinant *Clostridium thermocellum* CelS, a major cellulosome component. *J. Bacteriol.*, **177**, 1641–1644.
- Lamed, R., Setter, E. and Bayer, E.A. (1983) Characterization of a cellulose binding cellulase containing complex in *Clostridium thermocellum*. *J. Bacteriol.*, **156**, 828–836.
- Laskowski, R.A., McArthur, N.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.
- Madarro, E., Pena, J.P., Lequerica, J.L., Vallés, S., Gay, R. and Flors, A. (1991) Partial purification and characterization of the cellulases from *Clostridium cellulolyticum* H10. *J. Chem. Tech. Biotech.*, **52**, 393–406.
- McCarter, J.D. and Withers, S.G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.*, **4**, 885–892.
- Merri, E.A. and Murphy, M.E.P. (1994) Raster3D version 2.0, a program for photorealistic molecular graphics. *Acta Crystallogr. Sect. D*, **50**, 869–837.
- Meyer, J.E.W. and Schulz, G.E. (1997) Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide–protein interactions. *Protein Sci.*, **6**, 1084–1091.
- Morag, E., Halevy, I., Bayer, E.A. and Lamed, R. (1991) Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. *J. Bacteriol.*, **173**, 4155–4162.
- Morag, E., Bayer, E.A., Hazlewood, G.P., Gilbert, H.J. and Lamed, R. (1993) Cellulase Ss (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of *Clostridium thermocellum*. *Appl. Biochem. Biotechnol.*, **43**, 147–151.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345–364.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.*, **11**, 282.
- Otwinowski, Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf, W., Evans, P.R. and Leslie, A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Science and Engineering Research Council, Warrington, UK, pp. 80–86.
- Otwinowski, Z. (1993) Oscillation data reduction program. In Sawyer, L., Isaacs, N. and Bailey, S. (eds), *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. Science and Engineering Research Council, Warrington, UK, pp. 56–62.
- Pagès, S., Belaich, A., Tardif, C., Reverbel-Leroy, C., Gaudin, C. and Belaich, J.P. (1996) Interaction between the endoglucanase CelA and the scaffolding protein CipC of the *Clostridium cellulolyticum* cellulosome. *J. Bacteriol.*, **178**, 2279–2286.
- Qian, M., Haser, R., Buisson, G., Duée, E. and Payan, F. (1994) The active center of a mammalian α -amylase with a carbohydrate inhibitor refined to 2.2 Å resolution. *Biochemistry*, **33**, 6284–6294.
- Ramakrishnan, C. and Ramachandran, G.N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.*, **5**, 909–933.
- Reverbel-Leroy, C., Belaich, A., Bernadac, A., Gaudin, C., Belaich, J.P. and Tardif, C. (1996) Molecular study and overexpression of the *Clostridium cellulolyticum* celF cellulase gene in *Escherichia coli*. *Microbiology*, **142**, 1013–1023.
- Reverbel-Leroy, C., Pagès, S., Belaich, A., Belaich, J.P. and Tardif, C. (1997a) The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome. Purification and characterization of the recombinant form. *J. Bacteriol.*, **179**, 46–52.
- Reverbel-Leroy, C., Parsiegla, G., Moreau, V., Juy, M., Tardif, C., Driguez, H., Belaich, J.P. and Haser, R. (1997b) Crystallization of the catalytic domain of *Clostridium cellulolyticum* CelF cellulase in the presence of a newly synthesized cellulase inhibitor. *Acta Crystallogr. Sect. D*, **54**, 114–118.
- Robyt, J.F. and French, D. (1967) Multiple attack hypothesis of α -amylase action: action of porcine pancreatic, human salivary and *Aspergillus oryzae* α -amylase. *Arch. Biochem. Biophys.*, **122**, 8–16.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. and Jones, T.A. (1990) Three dimensional structure of Cellobiohydrolase II from *Trichoderma reesei*. *Science*, **249**, 380–386.
- Rullmann, J.A.C. (1996) AQUA, computer program. Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, The Netherlands.
- Salamitou, S., Tokatlidis, K., Béguin, P. and Aubert, J.P. (1992) Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. *FEBS Lett.*, **304**, 89–92.
- Sakon, J., Irwin, D., Wilson, D.B. and Karplus, P.A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermospora fusca*. *Nature Struct. Biol.*, **4**, 810–818.
- Shen, H., Tomme, P., Meinke, A., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J. and Miller, R.C., Jr (1994) Stereochemical course of hydrolysis catalysed by *Cellulomonas fimi* CenE, a member of a new family of β -1,4-glucanases. *Biochem. Biophys. Res. Commun.*, **199**, 1223–1228.
- Shima, S., Igarashi, Y. and Kodama, T. (1991) Nucleotide sequence analysis of the endoglucanase-encoding gene, *celCCD*, of *Clostridium cellulolyticum*. *Gene*, **104**, 33–38.
- Shima, S., Igarashi, Y. and Kodama, T. (1993) Purification and properties of two truncated endoglucanases produced in *Escherichia coli* harboring *Clostridium cellulolyticum* endoglucanase gene *celCCD*. *Appl. Microbiol. Biotechnol.*, **38**, 750–754.
- Spezio, M., Wilson, D.B. and Karplus, P.A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry*, **32**, 9906–9916.
- Tokatlidis, K., Salamitou, S., Béguin, P., Dhurjati, P. and Aubert, J.P. (1991) Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Lett.*, **291**, 185–188.
- Tomme, P., Warren, A.J. and Gilkes, N.R. (1995) Cellulose hydrolysis by bacteria and fungi. *Adv. Microb. Physiol.*, **37**, 1–81.
- Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 2577–2637.

Received May 5, 1998; revised July 24, 1998;
accepted August 3, 1998.