Structure and function of *Humicola insolens* family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 Å resolution

Gideon J. DAVIES*[1], A. Marek BRZOZOWSKI*, Miroslawa DAUTER*, Annabelle VARROT* and Martin SCHÜLEIN†

*Department of Chemistry, University of York, Heslington, York YO10 5DD, U.K., and †Novo-Nordisk A/S, Novo Allé, DK-2880 Bagsvaerd, Denmark

Cellulases are traditionally classified as either endoglucanases or cellobiohydrolases on the basis of their respective catalytic activities on crystalline cellulose, which is generally hydrolysed more efficiently only by the cellobiohydrolases. On the basis of the *Trichoderma reesei* cellobiohydrolase II structure, it was proposed that the active-site tunnel of cellobiohydrolases permitted the processive hydrolysis of cellulose, whereas the corresponding endoglucanases would display open active-site clefts [Rouvinen, Bergfors, Teeri, Knowles and Jones (1990) Science **249**, 380–386]. Glycoside hydrolase family 6 contains both cellobiohydrolases and endoglucanases. The structure of the catalytic core of the family 6 endoglucanase Cel6B from *Humicola insolens* has been solved by molecular replacement with the known *T. reesei* cellobiohydrolase II as the search model. Strangely, at the sequence level, this enzyme exhibits the highest sequence similarity to family 6 cellobiohydrolases and displays just one of the loop deletions traditionally associated with endoglucanases in this family. However, this enzyme shows no activity on crystalline substrates but a high activity on soluble substrates, which is typical of an endoglucanase. The three-dimensional structure reveals that the deletion of just a single loop of the active site, coupled with the resultant conformational change in a second 'cellobiohydrolase-specific' loop, peels open the active-site tunnel to reveal a substrate-binding groove.

Key words: cellobiohydrolase, cellulose, glycoside hydrolase, X-ray structure.

INTRODUCTION

Cellulases, the enzymes hydrolysing the β-1,4 linkages in cellulose, are traditionally classified into two groups: endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91). This differentiation, although frequently made on the basis of the enzymes' respective activity on artificial substrates is taken to reflect the different catalytic efficiencies of the two groups of enzymes in hydrolysing crystalline cellulose. Crystalline cellulose is broken down much more efficiently by the cellobiohydrolases than by the majority of endoglucanases. The first three-dimensional structure for a cellulase was reported in 1990 with the determination of the structure of cellobiohydrolase II from *Trichoderma reesei* [1], which provided many of the essential foundations of cellulase action. It was shown that cellobiohydrolases enclose their active sites within a tunnel to facilitate the processive hydrolysis of crystalline substrates, and it was predicted that endoglucanases would display deletion and truncation of the tunnel loops to generate a more open substrate-binding cleft. These seminal predictions were later borne out by the solution of the *Thermomanospora fusca* family 6 endoglucanase E2, which was indeed related to the family 6 cellobiohydrolase through the deletion of, or reduction in, a number of surface loops [2]. Further verification of this basic principle came from the structure determinations of the cellobiohydrolases and endoglucanases from glycoside hydrolase family 7 [3–6]. The family 7 enzymes again reveal that the cellobiohydrolases enclose their catalytic machinery in a tunnel, whereas the corresponding endoglucanase displays an open substrate-binding cleft.

The open substrate-binding cleft of endoglucanases, together with the enclosed tunnel of cellobiohydrolases, provides an elegant explanation for the activity of cellobiohydrolases on crystalline materials (reviewed in [7–9]). The enclosed active site of the cellobiohydrolases permits the hydrolysis of crystalline substrates by trapping a cellulose chain in a tunnel and thus preventing it from reannealing, non-productively, to the cellulose crystal [10,11]. The low activity of cellobiohydrolases on substrates such as carboxymethylcellulose (CMC), is attributed to the fact that the enclosed tunnel precludes entry of a substrate with bulky substitutions.

Elegant verification of the role of the tunnel-enclosing loops has been provided by the site-directed deletion of one of the surface loops of the *Cellulomonas fimi* cellobiohydrolase A [12]. Deletion of the 'C-terminus-proximal' loop of this enzyme was performed to try to mimic the properties of an endoglucanase. This indeed resulted in an enzyme whose activity on soluble substrates, such as CMC, was increased, but it also revealed an increase in the specific fluidity of the substrate during the course of enzyme action, indicative of enhanced internal (endo) cleavage. Thus, whereas many of the properties of cellobiohydrolases, and their role in the synergistic hydrolysis of crystalline cellulose, might remain controversial, it is clear that their active-site tunnel confers unique properties.

Cel6B (the enzyme nomenclature is described in [13]) from *Humicola insolens* displays a much higher sequence similarity to cellobiohydrolases than to the endoglucanases. Here we present the three-dimensional structure of this enzyme solved by X-ray crystallography at a resolution of 1.6 Å. The structure does indeed reveal an open substrate-binding cleft. The accessible groove results from the deletion of a single 'cellobiohydrolase-specific' surface loop and by the peeling open of a second ('N-terminal') loop whose direct equivalent in cellobiohydrolase II

Abbreviations used: Cel6B, a family 6 endoglucanase; CMC, carboxymethylcellulose.

[1] To whom correspondence should be addressed (e-mail davies@yorvic.york.ac.uk).

Co-ordinates for the structure described in this paper have been deposited with the Protein Data Bank under the accession number 1DYS.PDB.

contributes to tunnel enclosure. Comparison of the *H. insolens* Cel6B structure with the equivalent Cel6A structures reveals a spectrum of potential loop conformations for this class of enzyme.

MATERIALS AND METHODS

Purification and characterization of Cel6B

The gene for *H. insolens* Cel6B was first identified and isolated by using a novel expression cloning system in the *Saccharomyces cerevisiae* system [14]. The cDNA for the gene encoding *H. insolens* Cel6B was cloned into an expression plasmid between the fungal amylase promoter and the amyloglucosidase terminator from *A. niger*. A catalytic core construct was made by the insertion of a stop codon after the catalytic core domain. This was transformed into an *Aspergillus oryzae* expression system, essentially as described previously [15]. After fermentation, the protein was purified by anion-exchange chromatography at pH 8.0 in 20 mM Tris/HCl buffer and eluted with a 0–0.5 M NaCl gradient. The Cel6B core was eluted as a single peak with an apparent molecular mass of 38 kDa, as judged by SDS/PAGE.

Crystallization and data collection

The *H. insolens* Cel6B catalytic core domain was desalted and concentrated to 40 mg/ml in distilled water. The protein was crystallized by the hanging-drop method with 25% (w/v) poly(ethylene glycol) 4000 as a precipitant and the addition of 0.2 M Li_2SO_4 buffered with 0.1 M Tris/HCl, pH 8.5. Similar crystals also grew with monomethylether poly(ethylene glycol) 5000 as precipitant and with 0.1 M sodium acetate buffer, pH 8.5. Crystals were mounted in a rayon fibre loops and placed in a boiling nitrogen stream at 120 K. A cryoprotectant solution was made that consisted of 28% (w/v) poly(ethylene glycol) 4000, 0.2 M Li_2SO_4 and 0.1 M Tris/HCl, pH 8.5, with the inclusion of glycerol to a final concentration of 25% (v/v). Data were collected with a MAR Research image plate system together with a copper rotating anode and using long, focusing, mirror optics (Yale/Molecular Structure Corporation). Data were processed and reduced with the DENZO/SCALEPACK programs [16]. All further calculations used the CCP4 suite of programs unless otherwise stated.

Structure solution and refinement

The structure was solved by molecular replacement by using the AMoRe [17,18]. The *T. reese*i CBH II coordinates (Y169F mutant, PDB code 1CB2) were used as a search model with data from 15 to 4 Å resolution and an outer radius of Patterson integration of 25 Å. Initial electron density maps were averaged with RAVE [19] and the model was built into the averaged map with program O [20]; 5% of the reflections were then set aside for cross-validation analysis [21], and were used to monitor various refinement strategies such as geometric and temperature-factor restraint values and the insertion of solvent water and as the basis for the maximum likelihood refinement by using the REFMAC program [22]. Because all observed data from 15 Å resolution were employed in the refinement, a low-resolution bulk solvent correction was applied. Water molecules were added in an automated manner with ARP [23] and verified by manual inspection. Coordinates and observed structure-factor amplitudes for the protein structure described in this paper have been deposited with the Protein Databank [24].

RESULTS AND DISCUSSION

Cel6B from *H. insolens* hydrolyses soluble and amorphous substrates such as CMC and phosphoric-acid-swollen cellulose with great efficiency, leading to its classification as an endoglucanase [25,26]. It displays extremely high activity towards soluble cellodextrins such as reduced cellohexaose, with a k_{cat} of 86 s^{-1} compared with only 4 s^{-1} for the corresponding cellobiohydrolase, Cel6A, from this family. This enhanced catalytic activity towards soluble substrates is typical of endoglucanases in comparison with their cellobiohydrolase counterparts [25,26]. It might reflect both the problem of product departure from enclosed-tunnel active sites and the preferential use of binding energy for crystalline cellulose degradation (as opposed to hydrolysis) by the cellobiohydrolases [27,28]. Surprisingly, the sequence of Cel6B is most similar to family 6 cellobiohydrolases (Figure 1). Only one of the active-site-enclosing loops, the 'C-terminal proximal loop', is truncated. This loop corresponds to the loop artificially deleted by Meinke et al. [12] in their engineering of the Cel6A from *C. fimi*. Sequence database searches reveal that no other family 6 enzymes possess this single deletion of the C-terminal loop. However, there are sequences that closely resemble cellobiohydrolase II sequences, which possess the C-terminal tunnel loop, but lack elements of the N-terminal loop, examples being CelA from *Neocallimastix patriciarum* [29] and the *Orpinomyces* enzymes CelA and CelC [30]. These three additional enzymes, like Cel6B, show significant activity on CMC in contrast with the true cellobiohydrolases from this family.

Structure of Cel6B

Crystals of Cel6B are in space group $P2_12_12$ with cell dimensions $a = 109.93$ Å, $b = 104.45$ Å, $c = 53.79$ Å and $\alpha = \beta = \gamma = 90°$. There are 2 molecules of Cel6B in the asymmetric unit with a

```
                3
Cel6B (Hi)      -----GNPFSGRTLLVNSDYSSKLDQTRQAFLSRGDQTNAAKVKYVQEKVGTFYWIS-NI
Cel6A (Hi)      ----NGNPFEGVQLWANNYYRSEVHTLAIPQITDP---ALRAAASAVAEVPSFQWLDRNV
Cel6A (Tr)      TATYSGNPFVGVTPWANAYYASEVSSLAIPSLTG----AMATAAAAVAKVPSFMWLD-TL

               57
Cel6B (Hi)      FLLRDIDVAIQNARAAKARGENP--IVGLVLYNLPDRDCSAGESSGELKLSQNGLNRYKN
Cel6A (Hi)      TVDTLLVQTLSEIREANQAGANPQYAAQIVVYDLPDRDCAAAASNGEWAIANNGVNNYK-
Cel6A (Tr)      DKTPLMEQTLADIRTANKNGGN--YAGQFVVFDLPDRDCAALASNGEYSIADGGVAKYK-

                                        ‡
              115                      139
CEL6B (Hi)      EYVNPFAQKLKAASDVQFAVILEPDAIGNMVTGT-SAFCRNARGPQQEAIGYAISQLQAS
Cel6A (Hi)      AYINRIREILISFSDVRTILVIEPDSLANMVTNMNVPKCSGAASTYRELTIYALKQLDLP
Cel6A (Tr)      NYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLP

              174
Cel6B (Hi)      HIHLYLDVANGGWLGWADKLEPTAQEVATILQKAGNNAKIRGFSSNVSNYNPYSTSNPPP
Cel6A (Hi)      HVAMYMDAGHAGWLGWPANIQPAAELFAKIYEDAGKPRAVRGLATNVANYNAWSVSSPPP
Cel6A (Tr)      NVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPS

              234
Cel6B (Hi)      YTSGSPSPDESRYATNIANAMRQRGLP-TQFIIDQSRVALS-GARSEWGQWCNVNPAGFG
Cel6A (Hi)      YTSPNPNYDEKHYIEAFRPLLEARGFP-AQFIVDQGRSGKQPTGQKEWGHWCNAIGTGFG
Cel6A (Tr)      YTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFG

                                    *
              292                  316
CEL6B (Hi)      QPFTTNTNNPNVDAIVWVKPGGESDGQ-----------CGMG----GAPAAGMWFDAYAQ
Cel6A (Hi)      MRPTANTGHQYVDAFVWVKPGGECDGTSDTTAARYDYHCGLEDALKPAPEAGQWFNEYFI
Cel6A (Tr)      IRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFV

              337
CEL6B (Hi)      MLTQNAHDEIA
Cel6A (Hi)      QLLRNANPPF-
CEL6A (Tr)      QLLTNANPSFL
```

Figure 1 Sequence alignment of Cel6B from *Humicola insolens* (Hi) with the two family 6 cellobiohydrolases of known three-dimensional structure, Cel6A and CBH II from *H. insolens* and *T. reesei* (Tr) respectively

The 'catalytic' acid, Asp-139, is indicated with a double dagger and one candidate proposed as the catalytic base, Asp-316, is indicated with an asterisk. Numbering for *H. insolens* Cel6B refers to the mature protein.

Table 1 Refinement and structure quality statistics for native *Humicola insolens* Cel6B structure

$R_{merge} = \Sigma_{hkl}\Sigma_i|I_{hkli} - < I_{hkl} > |/\Sigma_{hkl}\Sigma_i < I_{hkl} >$ (where I_{hkl} is the intensity for a given reflection with indices hkl). Abbreviation: r.m.s., root-mean-square.

(a) Data quality

Resolution of data (outer shell) (Å)	15–1.60 (1.66–1.60)
R_{merge} (outer shell)	0.065 (0.317)
Mean $I/\sigma I$ (outer shell)	30.4 (8.0)
Completeness (outer shell) (%)	99 (98)
Multiplicity (outer shell)	3.7 (4.0)

(b) Refinement

No. of protein atoms/molecule (residues 3–347)	2641
No. of protein molecules in asymmetric unit	2
No. of solvent water molecules	901
Resolution used in refinement (Å)	15–1.60
R_{cryst}	0.18
R_{free}	0.24
R.m.s. deviation, 1–2 bonds (Å)	0.017
R.m.s. deviation, 1–3 angles (Å)	0.035
R.m.s. deviation, chiral volumes (Å^3)	0.150
Average main-chain B (Å^2) (A/B)	20/20
Average side-chain B (Å^2) (A/B)	24/23
Average solvent B (Å^2)	35
Non-crystallographic deviation (main-chain)	0.19
Main-chain ΔB, bonded atoms (Å^2)	4.3

packing density of 2.03 Å^3/Da, corresponding to a solvent content of 39 % [31]. X-ray data to 1.6 Å are 99 % complete with an R_{merge} of 0.065, a mean $I/\sigma I$ of 17 and a mean multiplicity of 3.7 observations per reflection (Table 1). Structure solution by molecular replacement, using the known *T. reesei* CBH II coordinates as the search model [1], revealed two significant solutions, as expected. The final model structure contains 2 molecules of Cel6B (residues 3–347), each of 2641 non-hydrogen atoms and a total of 901 solvent water molecules. It has a crystallographic R of 0.18, with a corresponding R_{free} of 0.24 for observed data between 15 and 1.62 Å resolution. This model has deviations from stereochemical target values of 0.017 and 0.034 Å (corresponding to approx. 1.1°), for 1–2 and 1–3 bonds respectively (Table 1). With the exception of Ala-182, discussed below, all the non-glycine residues have conformational angles (ϕ, ψ) in permitted regions of the Ramachandran plot [32], with none of them in the 'generously allowed' or 'disallowed' regions as defined by PROCHECK [33]. Met-329 is, unusually, present oxidized in both molecules in the asymmetric unit. The proton donor, Asp-139, is present in the left-handed turn region of the Ramachandran plot with $\phi = 64.2°$, $\psi = 1.9°$.

Alanine-182 is a forbidden region of the Ramachandran plot (Figure 2), with conformational torsion angles of $\phi = 56.8°$, $\psi = -123.9°$. In the related family 6 structures, and almost all family 6 sequences, this residue is glycine and hence its importance might have been overlooked. The main-chain amide nitrogen of Ala-182 makes a 2.8 Å hydrogen bond to the carboxylate OD1 of Asp-180. Asp-180 is the residue implicated in the modulation of the pK_a of the acid, Asp-139 [34], by virtue of a solvent-mediated hydrogen bond with the OD1 atom of the catalytic acid, Asp-139. In the *H. insolens* Cel6A oligosaccharide complex, the acid OD2 atom makes a 2.5 Å hydrogen bond to the 'glycosidic' oxygen of the substrate [35]. Hence the energetically unfavoured main-chain geometry of Ala-182 functions to orientate Asp-180 and the catalytic acid Asp-139 in appropriate positions for catalysis.

Cel6B displays the distorted β/α-barrel (Figure 3a), as observed in the previous family 6 structure determinations: the *T. reesei* CBH II [1], the *T. fusca* endoglucanase E2 [2] and the *H. insolens* Cel6A [35,36]. Cel6B is structurally most similar to the two cellobiohydrolases, a feature that was exploited in the molecular replacement structure determination. Pairwise comparison of Cel6B with each of the other structures by using LSQMAN [37] gives an overlap, for equivalent Cα atoms, of 1.25 Å with Cel6A from *H. insolens* (313 equivalent atoms), 1.34 Å with CBH II from *T. reesei* (321 equivalent atoms) and 1.47 Å (only 223 equivalent atoms) for comparison with the endoglucanase E2 from *T. fusca*. Again, these results confirm the close structural similarity of Cel6B to the cellobiohydrolases from family 6.

Comparison with cellobiohydrolase II

The differences between Cel6B and Cel6A native structures take place principally in the surface loops (Figure 3b). This is in contrast with the endoglucanase E2 from *T. fusca*, whose α-helices also show a different orientation from that of the family 6 cellobiohydrolases. As described previously, the main feature of Cel6B is the deletion of the C-terminal active-site loop and the peeling open of the N-terminal loop, resulting in an open substrate-binding cleft (Figure 4). The open N-terminal loop (residues 86–106 in Cel6B) is well conserved throughout family 6 in terms of sequence. Several residues, such as Tyr-86, Arg-91, Asp-92, Cys-93, Ser-98 and Gly-101, are invariant. Furthermore, Asp-92 and Ser-98 are known to interact with oligosaccharide in a complex of *H. insolens* Cel6A [35]. In the Cel6A complex, a closure of the two active-site loops was observed on substrate binding. It is now clear that these loops are flexible and it seems likely that, if the invariant residues have the same role in different members of this family, the N-terminal loop of Cel6B must also close on substrate binding. This would permit the conserved serine residue, Ser-98, to interact with the substrate in the -1 and/or $+1$ subsites [38] as observed in other family 6 structures [35,39]. Conformational flexibility in the N-terminal loop of family 6 enzymes is characterized by a change in the N_i, $C\alpha_i$, C_i, N_{i+1} dihedral angle (residues 99–100 in Cel6B). For the most closed conformation, the ligand complex of Cel6A, a value of $+141.7°$ is observed, which decreases to $+10$ and $-4°$ in the native Cel6A and Cel6B structures respectively. Cel6B displays a glutamic residue at position 97, corresponding to an alanine residue in Cel6A. On a loop closure, analogous to that seen with the cellobiohydrolase, this glutamate might interact with the substrate in the -1 and $+1$ subsites. Other endoglucanases possess a histidine residue at this position. Interestingly, all cellobiohydrolases have an alanine residue in place of this histidine. It is possible that alanine in this position helps to avoid steric clashes with the extended C-terminal loop of the cellobiohydrolases.

Asp-92 is conserved structurally with its equivalent in the two cellobiohydrolase structures [1,35,36]. It is believed to have a minor role in catalysis, because its deletion decreased activity to 20 % of the wild-type enzyme in Cel6A from *T. reesei* [1]. A role in the modulation of the pK_a of the acid Asp-139 has been proposed in which Asp-92 would assist in elevating the pK_a of the acid Asp-139. The equivalent of Asp-92, Asp-79 in the endoglucanase E2 of *T. fusca*, is at 5 Å from the position observed in all other family 6 structures. However, it is relevant that in the structure of E2 the conformation of the N-terminal active-site loop is stabilized by the presence of a sulphate. Wilson and colleagues have speculated that this conformation might not be

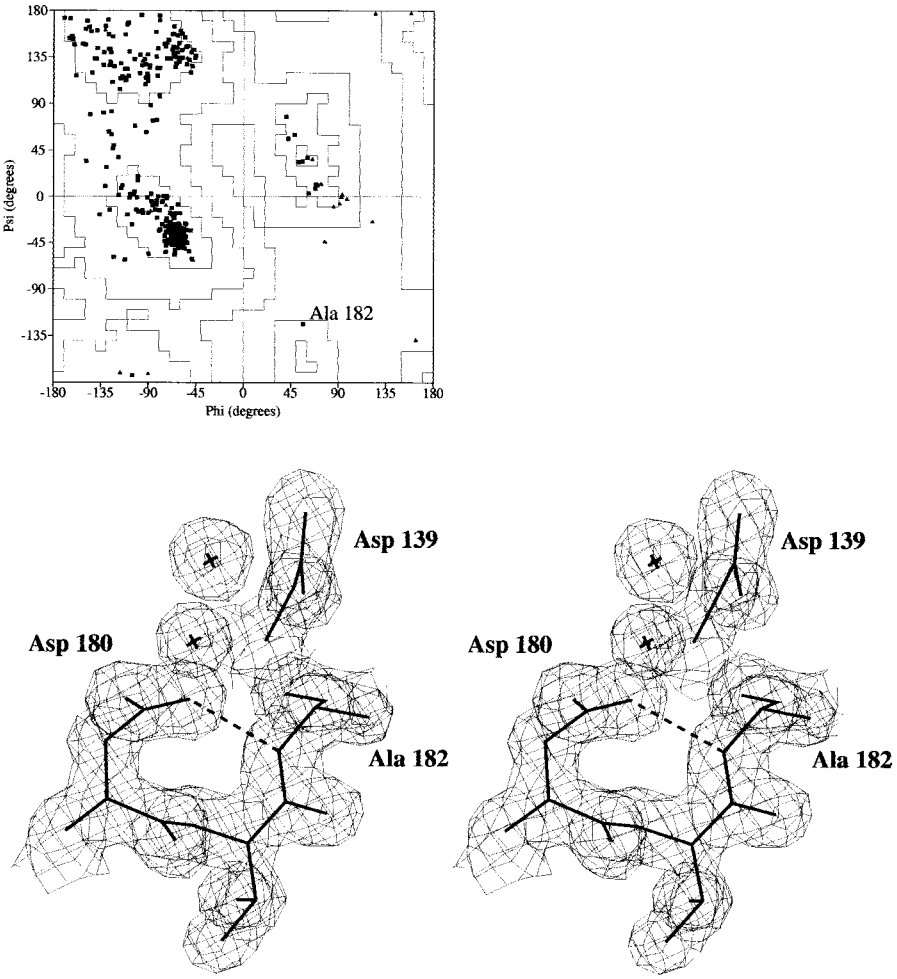

Figure 2 Conformational outliers in Cel6B

Ramachandran plot for Cel6B (calculated with PROCHECK [33]) (upper panel); electron density for Ala-182 (lower panel). The map shown is a maximum-likelihood weighted $2F_{obs} - F_{calc}$ (where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes respectively) synthesis at 0.44 electron/Å^3 and is in divergent ('wall-eyed') stereo.

representative of the true conformation owing to the hydrogen bond between the main-chain nitrogen of Asp-79 to the sulphate [2].

Small conformation changes have been observed in the active-site residues. The acid Asp-139 has the conformation previously observed in the *T. fusca* E2 and the *H. insolens* Cel6A structures. We believe that, for the protonation step, this conformation is likely to be the catalytically relevant conformation, because in complex structures it forms a hydrogen bond to the glycosidic oxygen. In the *T. reesei* CBH II structures the catalytic acid does not form a hydrogen bond to the interglycosidic atom and therefore seems to be inappropriately positioned for its protonation role: a conformational change of substrate or enzyme seems to be necessary for catalysis [1,39].

Apart from the catalytic acid, for which evidence is clear, identification of the additional catalytic apparatus in family 6 enzymes remains a matter of some debate. The role, identity, and even the existence of a classical Brønsted base in the family 6 enzyme mechanism remain controversial [34,35,39–41]. This problem is exacerbated by the potential for contaminating wild-type enzyme in kinetics determinations, by the scope for buffer-assisted catalysis, by the fact that both removal of the base and depletion of structural integrity would result in an enzyme that was completely inactive. Furthermore, this family contains numerous additional carboxylate functions that could 'rescue' a base mutant to some extent.

Damude et al. [34,40] have proposed candidates for acid and base on the basis of an elegant kinetic analysis of a series of active-site variants of the family 6 endoglucanse CenA from *C. fimi*. This enzyme showed a bell-shaped pH profile indicative of both acid and base contributions to catalysis. They concluded that Asp-252, corresponding to Asp-139 in *H. insolens* Cel6B, was the catalytic acid and Asp-392, corresponding to Asp-316, was the base. The catalytic acid assignment was in agreement with proposals based on various three-dimensional structure determinations, both available at that time [1,2] and determined later [35,36]. More recent structural [39] and kinetic [41] work has cast doubt on these proposals. All we can say as a result of the Cel6B structure is that the residue proposed in some previous studies, Asp-316 [2,34], displays a conformation similar to that in all other family 6 enzymes in which it is salt-linked to the conserved Arg-269. When compared with the *T. reesei* Cel6A complexes recently determined [39], it is clear that Asp-316 in Cel6B is unlikely to be correctly positioned to

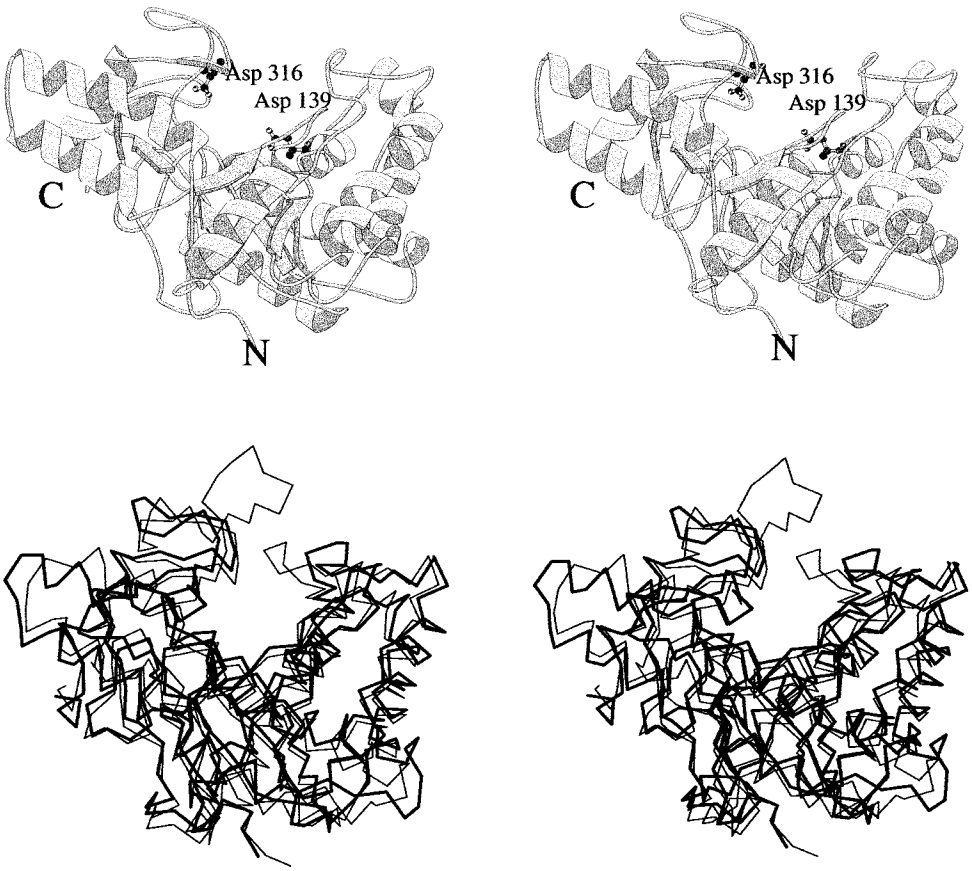

Figure 3 Structure of Cel6B

Upper panel: topology; catalytic aspartates mentioned in the text are indicated in ball-and-stick representation. Lower panel: α-carbon trace with the structure of Cel6A overlaid in faint lines. This Figure, in divergent ('wall-eyed') stereo, was drawn with MOLSCRIPT [43].

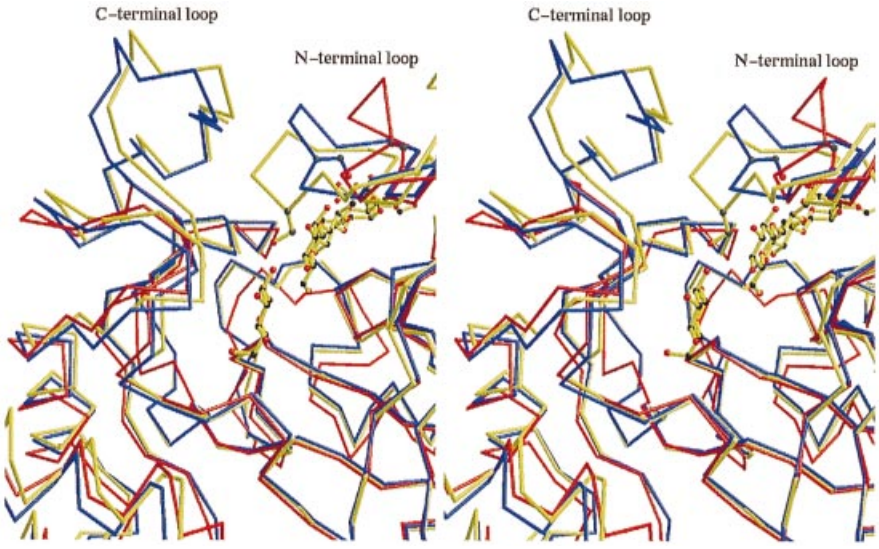

Figure 4 Active-centre loops of the family 6 *H. insolens* cellulases Cel6B (red), Cel6A-native (blue) [36] and the Cel6A glucose/cellotetraose complex (yellow) [35]

Glucose/cellotetraose in the Cel6A complex are shown for reference. The Figure is in divergent ('wall-eyed') stereo.

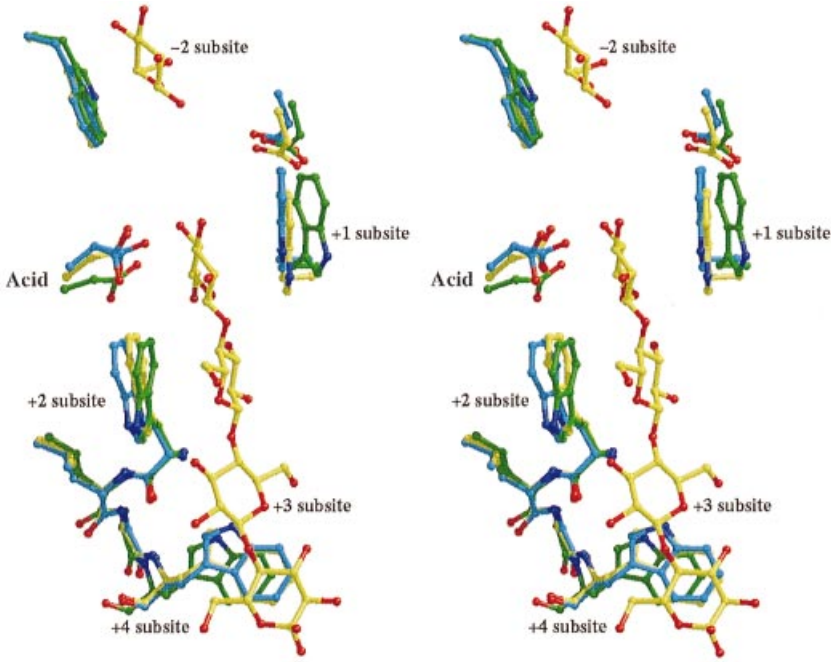

Figure 5 The six subsites of the family 6 *H. insolens* cellulases Cel6B (green), Cel6A-native (blue) [36] and the Cel6A glucose/cellotetraose complex (yellow) [35]

The Figure is in divergent ('wall-eyed') stereo.

act as base. A small conformational rearrangement of the −1 subsite sugar ring would first be necessary if Asp-316 were to act as base. Such a conformational change in the substrate cannot be ruled out. In the *T. reesei* Cel6A the acid is inappropriately positioned for protonation, as described above, yet if it were modelled as observed in the Cel6B and E2 structures it would make a steric clash with the −1 subsite pyranoside; small conformational readjustments are therefore likely.

Significant side-chain conformational changes along the substrate-binding cleft are observed for the four tryptophan residues involved in substrate stacking interactions in comparison with the Cel6A complex: Trp-52 in the −2 subsite, Trp-282 in +1, Trp-186 in +2 and Trp-189 in +4 (Figure 5). In *T. reesei* Cel6A, Trp-272 was proposed to form the base of the +4 subsite, where it is additionally implicated in the degradation of crystalline cellulose [42]. The proposal that this residue binds to sugars in the +4 subsite was confirmed by the structure determination of an oligosaccharide complex of the *H. insolens* Cel6A [35]. Cel6B possesses this residue, Trp-189, but shows insignificant activity on crystalline cellulose, indicating that the residue must have a different role in Cel6B. It is possible that the conformation and environment of this residue in Cel6A have evolved to aid both the degradation of crystalline materials and the processive movement of cellulose chains along the substrate-binding cleft. Most of the other residues involved in substrate bonding are well conserved in family 6: Lys-310, Glu-314, Arg-91 and Ser-98 in the −2 subsite, Asp-316 and Tyr-86 in −1, and Asp-139, Asp-92 and Asn-225 in +1. Cel6B contains at least six kinetically significant subsites [25,26], and in the structure we would estimate that there is space for seven subsites from −3 to +4. There is less steric constraint in the −3 subsite than is observed in the cellobiohydrolase Cel6A structure, as would be expected. Despite the fact that Cel6B shows extensive structural similarity to cellobiohydrolases, there is no evidence to suggest a more processive action by this endocellulase. Cel6B is unable to digest crystalline cellulose [25,26] and there is no evidence that cellobiose is the predominant reaction product. It seems that, as with Cel45 [25,26], the flexible loops generate a powerful endocellulase with high activity on amorphous but insoluble substrates. The only hint that Cel6B might be more processive comes from the ratio of the activities towards CMC and phosphoric-acid-swollen cellulose. Cel6B shows a slightly lower ratio than Cel45, indicating either that it might be more processive or that it might find the substituents of CMC more of a hindrance.

Endoglucanase Cel6B from *H. insolens* provides the first example of an endoglucanase related to its cellobiohydrolase counterparts by reduction in just a single surface loop. It is a naturally occurring version of the cellobiohydrolase mutant constructed by Meinke et al. [12]. The enzyme presents much greater sequence and structural similarities to the cellobiohydrolases from family 6 than to the endoglucanases. A fully open active-site cleft is further achieved though the peeling open of the N-terminal active-site loop. This N-terminal loop, however, contains a number of invariant residues known to interact with substrate in other family 6 structures, such as the *H. insolens* and *T. reesei* Cel6A enzymes [35,39]. This strongly suggests that conformational change is also likely in Cel6B so as to provide the optimal substrate-binding environment. Further work to probe the role of base catalysis in this family is being devised.

We thank Lene Odgaard Kanstrup and Torben Frandsen for preparation of the Cel6B core enzyme purified by ion-exchange chromatography. This work was funded in part by the Biotechnology and Biological Sciences Research Council, Novo-Nordisk A/S and the European Union (contract no. BIO4-CT97-2303). G.J.D. is a Royal Society University Research Fellow.

REFERENCES

1 Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. and Jones, T. A. (1990) Science **249**, 380–386
2 Spezio, M., Wilson, D. B. and Karplus, P. A. (1993) Biochemistry **32**, 9906–9916
3 Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T. and Jones, A. (1994) Science **265**, 524–528
4 Divne, C., Ståhlberg, J., Teeri, T. T. and Jones, T. A. (1997) J. Mol. Biol. **275**, 309–325
5 Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. and Davies, G. J. (1996) Biochemistry **35**, 15280–15287
6 Sulzenbacher, G., Schülein, M. and Davies, G. J. (1997) Biochemistry **36**, 5902–5911
7 Teeri, T. T. and Koivula, A. (1995) Carbohydr. Europe **12**, 28–33
8 Teeri, T. T. (1997) Trends Biotechnol. **15**, 160–167
9 Teeri, T. T., Koivula, A., Linder, M., Wohlfahrt, G., Divne, C. and Jones, T. A. (1998) Biochem. Soc. Trans. **26**, 173–178
10 Davies, G. and Henrissat, B. (1995) Structure **3**, 853–859
11 Henrissat, B. (1998) Cellulose Commun. **5**, 84–90
12 Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, Jr., R. C., Warren, R. A. J. and Gilkes, N. R. (1995) J. Biol. Chem. **270**, 4383–4386
13 Henrissat, B., Teeri, T. T. and Warren, R. A. (1998) FEBS Lett. **425**, 352–354
14 Dalboge, H. and Heldthansen, H. P. (1994) Mol. Gen. Genet. **243**, 253–260
15 Christensen, T., Wöldike, H., Boel, E., Mortensen, S. B., Hjortshøj, K., Thim, L. and Hansen, M. T. (1988) Bio/Technology **6**, 1419–1422
16 Otwinowski, Z. and Minor, W. (1997) Methods Enzymol. **276**, 307–326
17 Navaza, J. (1994) Acta Cryst. A **50**, 157–163
18 Navaza, J. and Saludijan, P. (1997) Methods Enzymol **276**, 581–594
19 Kleywegt, G. J. and Jones, T. A. (1994) in From First Map to Final Model (Bailey, S., Hubbard, R. and Waller, D., eds.), pp. 59–66, EPSRC, Daresbury
20 Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) Acta Cryst. A **47**, 110–119
21 Brünger, A. T. (1992) Nature (London) **355**, 472–475
22 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Acta Cryst. D **53**, 240–255
23 Lamzin, V. S. and Wilson, K. S. (1993) Acta Cryst. D **49**, 129–147
24 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. T., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. **112**, 535–542
25 Schülein, M., Tikhomirov, D. F. and Schou, C. (1993) in Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and Other Hydrolases, vol. 8 (Suominen, P. and Reinikainen, T., eds.), pp. 109–116, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland
26 Schülein, M. (1997) J. Biotechnol. **57**, 71–81
27 Konstantinidis, A. K., Marsden, I. and Sinnott, M. L. (1993) Biochem. J. **291**, 883–888
28 Sinnott, M. L. and Sweilem, N. S. (1998) in Carbohydrases from *Trichoderma reesei* and Other Microorganisms (Claeyssens, M., Piens, K. and Nerinckx, W., eds.), pp. 13–20, Royal Society of Chemistry, London
29 Denman, S., Xue, G. P. and Patel, B. (1996) Appl. Environ. Microbiol. **62**, 1889–1896
30 Li, X. L., Chen, H. and Ljungdahl, L. G. (1997) Appl. Environ. Microbiol. **63**, 4721–4728
31 Matthews, B. W. (1968) J. Mol. Biol. **33**, 491–497
32 Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) J. Mol. Biol. **7**, 95–99
33 Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) J. Appl. Cryst. **26**, 282–291
34 Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, Jr., R. C. and Warren, R. A. J. (1995) Biochemistry **34**, 2220–2224
35 Varrot, A., Schülein, M. and Davies, G. J. (1999) Biochemistry **38**, 8884–8891
36 Varrot, A., Hastrup, S., Schülein, M. and Davies, G. J. (1999) Biochem. J. **337**, 297–304
37 Kleywegt, G. J. and Jones, T. A. (1994) ESF/CCP4 Newsl. **31**, 9–14
38 Davies, G. J., Wilson, K. S. and Henrissat, B. (1997) Biochem. J. **321**, 557–559
39 Zou, J.-Y., Kleywegt, G. J., Ståhlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T. T. and Jones, T. A. (1999) Structure **7**, 1035–1045
40 Damude, H. G., Ferro, V., Withers, S. G. and Warren, R. A. J. (1996) Biochem. J. **315**, 467–472
41 Wolfgang, D. E. and Wilson, D. B. (1999) Biochemistry **38**, 9746–9751
42 Koivula, A., Kinnari, T., Harjunpaa, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T. A. and Teeri, T. T. (1998) FEBS Lett. **429**, 341–346
43 Kraulis, P. J. (1991) J. Appl. Cryst. **24**, 946–950

Received 6 December 1999/25 January 2000; accepted 3 March 2000