Crystal structure of the family 7 endoglucanase I (Cel7B) from *Humicola insolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate

Lloyd F. MACKENZIE*, Gerlind SULZENBACHER†, Christina DIVNE‡, T. Alwyn JONES‡, Helle F. WÖLDIKE§, Martin SCHÜLEIN§, Stephen G. WITHERS* and Gideon J. DAVIES†¹

*Department of Chemistry, University of British Columbia, Vancouver, V6T 1Z1, Canada, †Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K.,

‡Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, S-751 24 Uppsala, Sweden, and §Novo Nordisk A/S, Novo Allé, DK-2880 Bagsvaerd, Denmark

Cellulose is the major polysaccharide component of the plant cell wall and the most abundant naturally produced macromolecule on Earth. The enzymic degradation of cellulose, by cellulases, is therefore of great environmental and commercial significance. Cellulases are found in 12 of the glycoside hydrolase families classified according to their amino acid sequence similarities. Endoglucanase I (Cel7B), from the soft-rot fungus *Humicola insolens*, is a family 7 enzyme. The structure of the native form of Cel7B from *H. insolens* at 2.2 Å resolution has been solved by molecular replacement using the known *Trichoderma reesei* cellobiohydrolase I [Divne, Ståhlberg, Reinikainen, Ruohonen,

Pettersson, Knowles, Teeri and Jones (1994) *Science* **265**, 524–528] structure as the search model. Cel7B catalyses hydrolysis of the β -1,4 glycosidic linkages in cellulose with net retention of anomeric configuration. The catalytic nucleophile at the active site of Cel7B has been identified as Glu-197 by trapping of a 2-deoxy-2-fluorocellotriosyl enzyme intermediate and identification of the labelled peptide in peptic digests by tandem MS. Site-directed mutagenesis of both Glu-197 and the prospective catalytic acid, Glu-202, results in inactive enzyme, confirming the critical role of these groups for catalysis.

INTRODUCTION

Cellulose is a polymer of β -1,4-linked glucopyranose units which is found in a great many organisms, but most notably in plants where it plays a structural role in the cell wall. It is widely believed to be the most abundant natural polymer on earth. Cellulose-based fabrics are found in many textiles and cellulose-derived materials are the largest constituent of municipal waste. The degradation of cellulose is, therefore, of ecological and economic importance, with great potential both for the beneficial conversion of plant biomass and for selective cotton-fibre modifications [1]. Cellulose forms highly crystalline fibres which are generally quite resistant to degradation. There are, however, many fungal and bacterial species which are able to break down cellulose. These organisms produce a vast ensemble of different cellulases [endoglucanases and cellobiohydrolases (CBHs)], often acting in synergy, to break down the β -1,4 glycosidic linkages. Glycoside hydrolases have been classified into over 67 families on the basis of amino acid sequence similarities [2–5] and cellulases are found in 12 of these (5–9, 12, 26, 44, 45, 48, 60–61).

Humicola insolens is a soil-based hyphomycete which produces at least seven different cellulases, from families 5, 6, 7, 12 and 45 [6,7]. We have described, previously, the structure of the endoglucanase V from family 45 [8], Endoglucanase I (hereafter Cel7B) is from family 7. Whilst most fungal cellulases are multi-domain proteins consisting of a catalytic core linked to a cellulose-binding domain via a flexible, often glycosylated, linker peptide

[9], the Cel7B from *H. insolens* is unusual in that it is a single domain protein consisting of a catalytic core domain only. Glycoside hydrolase family 7 contains both endoglucanases (E.C 3.2.1.4) and CBHs (E.C 3.2.1.91), which hydrolyse cellulose in a predominantly *endo* or processive fashion respectively. The structure of a CBH, the CBH I from *Trichoderma reesei*, was the first structure described from this family [10]. It is a primarily β -strand structure with a similar topology to that first observed in the plant legume lectins, such as concanavalin A [11].

Cel7B is classified as an endocellulase due to its ability to break down carboxymethyl-substituted cellulose, a substrate which is unyielding to the catalytic advances of CBHs. On the natural substrate, cellulose, Cel7B is widely believed to act on the ‘amorphous’ and ‘disordered’ regions and is not able to digest the highly crystalline regions of cellulose. The inactivity of the *H. insolens* Cel7B on crystalline substrates is also, no doubt, further reduced due to the lack of a suitable cellulose-binding domain. Cel7B, in common with all the family 7 enzymes studied, catalyses glycosyl hydrolysis with a net retention of the configuration at the anomeric carbon, giving the β anomer as the reaction product [12]. Kinetic studies of *H. insolens* Cel7B reveal that Cel7B has four kinetically significant subsites for saccharide binding and displays significant transglycosylation at high substrate concentrations [6,7,12]. Indeed, the strong transglycosylating ability of Cel7B has recently been utilized for chemo-enzymic syntheses [13].

Cel7B performs glycoside hydrolysis with net retention of anomeric configuration via a double-displacement mechanism in

Abbreviations used: CBH, cellobiohydrolase; DNPC, 2,4-dinitrophenyl β -D-cellobioside; 2FDNPC, 2,4-dinitrophenyl-2-deoxy-2-fluoro- β -D-cellobioside; 2CIDNPC, 2,4-dinitrophenyl-2-deoxy-2-chloro- β -D-cellobioside; DTT, dithiothreitol; ESMS, electrospray MS.

¹ To whom correspondence should be addressed (davies@york.ac.uk).

Scheme 1 The double-displacement reaction as applied to a retaining (e-e) β -glycoside hydrolase

A covalent glycosyl-enzyme intermediate is formed, with general acid/base assistance, via oxocarbenium-ion transition states.

which a covalent glycosyl-enzyme intermediate is formed and hydrolysed, with acid/base assistance, via oxocarbenium ion-like transition states, see Scheme 1. The most reliable route to identify the catalytic nucleophile involves trapping of this glycosyl-enzyme intermediate and identification of the derivatized amino acid. 2-Deoxy-2-fluoro glycosides, with good leaving groups, have proved to be excellent reagents for this purpose [14]. The fluorine at C-2 destabilizes the transition states for formation and hydrolysis of the intermediate, both inductively and through removal of potential transition-state stabilizing hydrogen-bonding interactions. The result of this is a reduction in the rates for both formation and hydrolysis of the covalent intermediate. The presence of a good leaving group, however, ensures that the first step is fast, rendering the intermediate kinetically accessible. Identification of labelled peptides in proteolytic digests of the enzymes so labelled has been enormously facilitated by the introduction of tandem MS strategies which allow both detection of the peptide of interest and its sequencing [15–19]. In the present paper we present the three-dimensional X-ray structure of the *H. insolens* family 7 endoglucanase, Cel7B, at 2.2 Å resolution and identification of the catalytic nucleophile through the trapping of the covalent intermediate. Site-directed mutagenesis confirms the importance of the catalytic nucleophile and acid/base.

EXPERIMENTAL

General procedures

Buffer chemicals and other reagents were obtained from Sigma Chemical Company, unless otherwise noted. Pepsin (from porcine gastric mucosa) and endoglycosidase F were obtained from Boehringer Mannheim. *H. insolens* Cel7B was expressed using the *Aspergillus oryzae* expression system [20]. Transformants were grown in flasks or 10 litre tank fermentors using standard conditions. The substrate 2,4-dinitrophenyl β -D-cellobioside (DNPC) and the inactivators 2,4-dinitrophenyl-2-deoxy-2-fluoro- β -D-cellobioside (2FDNPC) and 2,4-dinitrophenyl-2-deoxy-2-chloro- β -D-cellobioside (2CIDNPC) were synthesized as

previously described [21,22]. Kinetic studies were performed at 40 °C in 100 mM sodium phosphate buffer, pH 7.5, containing 0.1 % BSA. A continuous spectrophotometric assay based on the hydrolysis of DNPC was used to monitor enzyme activity by measurement of the rate of 2,4-dinitrophenolate release ($\lambda = 400$ nm, $\epsilon = 10910$ M⁻¹ cm⁻¹), using a Unicam UV4 UV/VIS spectrophotometer equipped with a circulating water bath. Michaelis–Menten parameters for the substrate, DNPC, were determined using a range of DNPC concentrations from 0.012–0.24 mM.

Inactivation kinetics

The inactivation of Cel7B by either 2FDNPC or 2CIDNPC was monitored by incubation of the enzyme (0.0098 mg/ml) under the above conditions in the presence of various concentrations of the inactivators (0–0.23 mM). Residual enzyme activity was determined at appropriate time intervals by addition of an aliquot (10 μ l) of the inactivation mixture to a solution of DNPC (0.13 mM, 800 μ l) in the above buffer and measurement of 2,4-dinitrophenolate release. Pseudo-first-order rate constants at each inactivator concentration (k_{obs}) were determined by fitting each curve to a first-order rate equation. Values for the inactivation rate constant (k_i) and the dissociation constant for the inactivators (K_i) were determined by fitting to eqn. (1) [inactivator (I) = 2FDNPC or 2CIDNPC]:

$$k_{\text{obs}} = \frac{k_i[I]}{(K_i + [I])} \quad (1)$$

Reactivation kinetics

Reactivation of the 2-fluorocellobiosyl- or 2-chlorocellobiosyl-inactivated enzyme was studied as follows. Enzyme (200 μ l, 0.66 mg/ml, inactivated as above, but in the absence of BSA) was concentrated using 10 kDa nominal cut-off centrifugal concentrators (Amicon Corporation, Danvers, MD, U.S.A.) to a volume of approx. 50 μ l, then diluted with 200 μ l of buffer without BSA. This was repeated twice, and the retentate was diluted to a final volume of 200 μ l with buffer containing 1 mg/ml BSA. Aliquots of the inactivated enzyme were then incubated at 40 °C in the presence of phosphate buffer alone or in the presence of D-cellobiose (19.5 mM). Reactivation was monitored by removal of aliquots (10 μ l) at appropriate time intervals and assaying as described above. Any activity loss due to denaturation of the enzyme was corrected for by a control experiment involving incubation of enzyme with no inactivator. The observed reactivation rate constant, $k_{\text{obs,react}}$, for each reactivator was determined from direct fit of the activity (full rate minus observed rate) versus time data to a first-order equation.

Deglycosylation of Cel7B

The enzyme is deglycosylated at two sites (Asn-89 and Asn-247). N-linked sugars, except for the terminal N-acetylglucosamine, were removed by treatment with endoglycosidase F. Endoglucanase (20 mg) was incubated in 0.1 M potassium phosphate buffer, pH 7.5, containing 20 mM EDTA with 5 units of endoglycosidase F, and incubated for 100 h at 37 °C. The deglycosylated enzyme was purified by hydrophobic-interaction chromatography on a phenyl-Sepharose column, previously equilibrated with 0.5 M ammonium acetate buffer, and the pure enzyme was eluted in water. The deglycosylated protein has a molecular mass of 45 kDa on SDS/PAGE.

Labelling of Cel7B

Cel7B (20 μ l, 3.8 mg/ml, in 100 mM sodium phosphate buffer, pH 7.5) was incubated with 2FDNPC (1–20 μ l of a 22 mM stock solution) at 40 °C for 12–24 h. This mixture was immediately used for intact protein experiments or digested with pepsin as described below.

MS

Mass spectra were recorded using a PE-Sciex API 300 triple quadrupole mass spectrometer (Sciex, Thornhill, Ontario, Canada) equipped with an ionspray ion source. Peptide digests were separated by reverse-phase HPLC on an Ultrafast Microprotein Analyzer (Michrom BioResources Inc., Pleasanton, CA, U.S.A.) directly interfaced with the mass spectrometer. In each of the MS experiments, the proteolytic digest was loaded on to a C₁₈ column (Reliasil, 1 mm $\times$ 150 mm), then eluted with a gradient of 0–60 % solvent B over 30 min followed by 100 % B over 10 min at a flow rate of 50 μ l/min (solvent A: 0.05 % trifluoroacetic acid/2 % acetonitrile in water; solvent B: 0.045 % trifluoroacetic acid/80 % acetonitrile in water). A post-column splitter was present in all experiments, splitting off 80 % of the sample into a fraction collector and sending 20 % into the mass spectrometer. Spectra were obtained in either the single-quadrupole scan mode (LC/MS) or tandem MS neutral-loss mode.

Stoichiometry of incorporation of inactivator by electrospray MS (ESMS)

Cel7B (10 μ g, native or 2FDNPC treated) was introduced into the mass spectrometer through a microbore PLRP-5 column (1 mm $\times$ 50 mm) on a Michrom HPLC system. The quadrupole mass analyser of an API 300 triple quadrupole mass spectrometer was scanned over an m/z range of 400–2400 Da, with a step size of 0.5 Da and a dwell time of 1 ms per step. The ion source voltage (ISV) was set at 4.8 kV and the orifice energy (OR) was 50 V. The molecular masses of the enzyme species were determined by using the deconvolution software, Multiview 1.0, supplied by Sciex.

Pepsin digest conditions

Cel7B (20 μ l, native or labelled, $\sim$ 2 mg/ml) was mixed with 40 μ l of 150 mM phosphate buffer, pH 2, and 20 μ l of pepsin, 0.2 mg/ml in 150 mM phosphate buffer, pH 2). Mixtures were incubated at room temperature for 120 min, then immediately analysed by ESMS or frozen until required.

ESMS analysis of the proteolytic digest

For LC/MS experiments the quadrupole mass analyser of an API 300 triple quadrupole mass spectrometer was scanned using the same setting described above for the intact protein ESMS. In the neutral-loss scanning mode, MS/MS spectra were obtained by searching for the mass loss of m/z 489, 244.5 and 163, corresponding to the loss of a 2-fluorocellotriosyl label from a peptide ion in the singly, doubly or triply charged state. Thus, scan range, m/z 400–2200; step size, 0.5; dwell time, 1.5 ms; ion source voltage (ISV), 5.5 kV; orifice energy (OR), 45; RNG = 400; Q0 = –10; R01 = –11.3 V; R02 = –52 V; R03 = –57 V; CAD = 1 (collision gas = N₂).

Reduction of disulphide bonds

Cleavage of the disulphide bonds present in the digested protein (Cel7B contains 18 cysteine residues all of which form disulphide

bonds) was achieved by reduction with dithiothreitol (DTT) as follows. A sample (25 μ l) of the 2-fluorocellotriosyl-labelled purified peptide was treated with 10 mM DTT, 200 mM sodium phosphate, pH 7.5 (100 μ l), for 15 min at room temperature and then acidified to $\sim$ pH 2 with 50 % trifluoroacetic acid. The samples were then immediately analysed by ESMS as described earlier.

Aminolysis of the 2-fluorocellotriosyl-labelled peptide

To a sample (20 μ l, 0.10 mg/ml) of the purified 2-fluorocellotriosyl-labelled peptide was added concentrated ammonium hydroxide (5 μ l). The mixture was incubated for 15 min at 25 °C, acidified with 50 % trifluoroacetic acid and analysed by ESMS as described above.

Chemical sequencing

Purification of the 2-fluorocellotriosyl-labelled Cel7B peptide was achieved by HPLC separation of the peptic digest as described above and by collecting the appropriate fractions containing the labelled peptide via a post-column splitter. The amino acid sequence of the labelled peptide was determined by S. Perry (University of British Columbia) using standard pulsed-liquid-phase protocols and instrumentation on a Perkin-Elmer model 476A sequencer and model 120A PTH (phenylthiohydantoin) analyser (Applied Biosystems, Foster City, CA, U.S.A.).

Crystallization and data collection

Cel7B was crystallized as described previously [23]. The tetragonal crystal form, space group P₄₁22/P₄₃22 with cell dimensions $a = b = 102.9$ Å, $c = 282.0$ Å, proved to be the most suitable for high-resolution data collection. A single tetragonal Cel7B crystal was mounted in a 1 mm diam. glass capillary tube with the c^* axis roughly parallel to the spindle axis of the camera, but offset sufficiently to reduce loss of data in the 'blind region'. Data were collected on the Mar Research image-plate installed on beamline X-11 at the EMBL Hamburg outstation at a wavelength of 0.92 Å. Data were processed with the DENZO program [24] and scaled and reduced using the CCP4 programs ROTAVATA and AGROVATA. All further calculations involved the CCP4 suite of programs [25] unless otherwise stated.

Structure solution

The structure was solved by molecular replacement, using the program AMoRe [26], with the refined structure of the *T. reesei* CBH I as the search model [10]. Initial studies showed that it was not possible to solve the structure of the tetragonal crystal form of *H. insolens* Cel7B with this search model. Cel7B also crystallized in a monoclinic form in space group P2₁, with cell dimensions $a = 66.9$, $b = 75.2$, $c = 87.0$, $\beta = 102.91^\circ$. This form has two molecules of Cel7B in the asymmetric unit related by purely translational non-crystallographic symmetry as indicated by the strong peak, at $\frac{1}{2}$, 0, $\frac{1}{2}$, in the native Patterson [23]. Data collected for this crystal form were relatively poor and extended only to 3.2 Å [23], but the structure solution using the intact CBH I model proved trivial. The correctly oriented CBH I model was subject to rigid body refinement in XPLOR [27], and the resultant electron-density map was averaged according to the non-crystallographic symmetry using the RAVE program [28]. This 3.2 Å map was not of particularly high quality, but allowed the removal of those parts of the CBH I model that were clearly not present in the Cel7B structure and the assignment of the

correct Cel7B sequence into the model using the program O [29]. This partial model was used to solve the structure of the tetragonal form by molecular replacement. The molecular replacement solution revealed two molecules in the asymmetric unit in space group $P4_122$. This indicates a high solvent content of approx. 72%. Cyclic averaging was performed with the RAVE suite of programs. The resultant electron-density map was of high quality and allowed rapid rebuilding and correction of the atomic model for residues 1–398. The C-terminal 17 residues were not visible in the electron-density maps. A total of 7498 reflections, representing approx. 10% of the 74226 reflections between 10 and 2.2 Å, were set aside for cross-validation analysis [30]. Cross-validation was used throughout as a strict monitor to determine the relative merits of various refinement strategies such as the relative weight between X-ray and stereochemical terms, the B factor restraints, the non-crystallographic symmetry positional and temperature factor restraints and the inclusion of solvent molecules. The structure was refined initially with XPLOR and later using a maximum likelihood-based method with REFMAC [31].

RESULTS AND DISCUSSION

Structure solution and quality of the final model structure

X-ray data for the tetragonal form of Cel7B were collected at the EMBL Hamburg Outstation. A total of 289090 observations of 63790 reflections were measured, with 520 observations being rejected during the data reduction procedure. The data have an R_{merge} for all data between 40 and 2.20 Å resolution of 0.069. The R_{merge} in the outer resolution bin (2.32–2.20 Å) is 0.209. The data are 96.6% complete with an average multiplicity of 3.9 observations/reflection and a mean $I/\sigma I$ of 15 (mean $I/\sigma I$ of 6 in the 2.32–2.2 Å outer bin). A summary of the data and final model quality is given in Table 1. The structure of Cel7B was solved by molecular replacement using the known crystal structure of the *T. reesei* CBH I as the search model and utilizing the non-crystallographic symmetry in the $P2_1$ and $P4_122$ crystal forms. The structure was subsequently refined, with tight non-crystallographic symmetry restraints, at a resolution of 2.2 Å. The final model consists of two protein chains of 398 amino acids (the C-terminal 17 residues are either disordered or absent) with two *N*-

Figure 1 Ribbon diagrams of the structures of (top) the *H. insolens* Cel7B with the N- and C-termini indicated, and (bottom) the CBH I from *T. reesei* [10]

The extra loops in the CBH I structure that enclose the substrate-binding canyon are clear. The Figures were drawn with the MOLSCRIPT program [51].

Table 1 Statistics of the final model geometry for the *H. insolens* Cel7B

Data resolution (outer shell)	30.0–2.20 (2.32–2.20) Å
Data R_{merge} * (outer shell)	0.069 (0.209)
Data $I/\sigma I$ (outer shell)	15 (6)
Data multiplicity (outer shell)	3.8 (3.1)
Data completeness (outer shell)	96.6 (92.4)
Non-crystallographic symmetry	2 fold, tight restraints
Resolution range	30.0–2.2 Å
R_{cryst} (66728 reflections)	0.183
R_{free} (7498 reflections)	0.220
No. protein atoms	6176
No. water molecules	478
No. <i>N</i> -acetylglucosamine atoms	56
Average B protein	28 Å ² main chain/33 Å ² side-chain
Average B water	42 Å ²
Rmsd (1–2 bonds)	0.015 Å
Rmsd positional between NCS atoms	0.056 Å main chain/0.10 Å side-chain
Rmsd B value between NCS equivalent atoms	4.2 Å ²
Ramachandran plot outliers	None

* $R_{\text{merge}} = \sum_i \sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \sum_i \sum_{hkl} \langle I_{hkl} \rangle$, where I is the intensity of any given observation, i , of any reflection with indices hkl .

acetylglucosamine residues and a total of 437 solvent water molecules. The final model has good geometry and a final R_{cryst} and R_{free} of 0.183/0.220, for all data between 30 and 2.2 Å. No amino acids lie in prohibited regions of the Ramachandran plot [32] as calculated by PROCHECK [33].

Cel7B is a single-domain enzyme with overall dimensions of 60 Å × 45 Å × 40 Å (Figure 1, top). The structure is constructed primarily of two large anti-parallel β -sheets consisting of seven and eight strands respectively. These stack on top of each other forming a convex and concave surface. The structure of Cel7B is thus similar to that first described for the CBH I from *T. reesei* [10]. The major difference is that whereas CBH I has its active site enclosed in a tunnel formed by a number of surface loops (Figure 1, bottom) these loops are shortened or absent in the Cel7B structure. The active site of Cel7B is thus located in a long open groove, as expected for an enzyme known to have endo-catalytic activity (described more fully in [34]). In Cel7B, this substrate-binding cleft is approx. 50 Å long and 20 Å deep with the active site residues, Glu-197, Asp-199 and Glu-202, located about two-thirds of the way along this channel at the bottom of the cleft. The *H. insolens* Cel7B described here has recently been used to solve the structures of the related endoglucanases from *Fusarium oxysporum* [35] and *T. reesei* [34]. The high degree of sequence similarity (57%) between the *H. insolens* and *F. oxysporum* enzymes [36] is reflected in a close structural similarity, with the

Figure 2 Inactivation of *H. insolens* Cel7B by 2FDNPC

(a) Semilogarithmic plot of residual activity versus time at the indicated inactivator concentrations: (●) 0.035 mM, (○) 0.058 mM, (■) 0.115 mM, (□) 0.230 mM. (b) Inactivation with 0.058 mM 2FDNPC in the absence (○) and presence (●) of 0.45 mM D-cellobiose. (c) Replot of the first-order rate constants from (a).

392 equivalent C α atoms overlapping with a root mean square deviation of 0.69 Å (calculated using LSQMAN [37]). The similarity with the *T. reesei* enzyme is less pronounced: only 341 equivalent C α atoms can be assigned and these overlap with a root mean square deviation of 1.2 Å.

The structure is stabilized by the presence of nine disulphide bonds which are located between residues: 18 and 24, 51 and 73, 63 and 69, 140 and 365, 172 and 195, 176 and 194, 215 and 234, 223 and 228, 239 and 315. The N-terminal glutamine residue is present as the modified pyroglutamate group, as was observed in the *T. reesei* CBH I structure and which would be expected, based on sequence comparisons alone, to be present in all the family 7 enzymes. Electron density corresponding to N-linked glycosylation is visible at the two potential N-linked glycosylation sites: Asn-89 and Asn-247. These are sufficiently well ordered to allow the positioning of a single N-linked *N*-acetylglucosamine at each position. Coordinates for the *H. insolens* Cel7B have been deposited on the Brookhaven Protein Databank [38] with accession code 2A39.

Inactivation of Cel7B through trapping of a glycosyl-enzyme intermediate

In common with other β -glucosidases, a catalytically competent glycosyl-enzyme intermediate can be trapped on Cel7B by reaction of the enzyme with an activated 2-deoxy-2-fluoro glycoside. Thus incubation of Cel7B with either 2FDNPC or 2CIDNPC resulted in the time-dependent inactivation of the enzyme according to pseudo-first-order kinetics (Figure 2a). Direct fit of the data (described in the Experimental section) allowed the determination of the inactivation rate constants ($k_i = 0.13 \pm 0.01 \text{ min}^{-1}$; $k_i = 0.36 \pm 0.06 \text{ min}^{-1}$) and the dissociation constants ($K_i = 0.16 \pm 0.01 \text{ mM}$; $K_i = 0.36 \pm 0.03 \text{ mM}$) for the 2-fluoro and 2-chloro analogues respectively. Incubation with

Figure 3 ESMS experiments on intact Cel7B (a) unlabelled, (b) labelled with 0.5 mM 2FDNPC, and (c) labelled with 10.0 mM 2FDNPC

2FDNPC (0.058 mM) in the presence of a competitive inhibitor D-cellobiose (1.50 mM; $K_i = 0.45 \text{ mM}$) reduced the apparent rate constant for inactivation from $k_{\text{obs}} = 0.033 \pm 0.001 \text{ min}^{-1}$ to $0.0086 \pm 0.0004 \text{ min}^{-1}$ (Figure 2b). This is the expected degree of protection against inactivation if the two ligands compete for the same site, indicating that the inactivations are active-site directed. These results suggest, by analogy with observations from *Agrobacterium* sp. β -glucosidase [39,40], that the inactivation is a consequence of the accumulation of a relatively stable 2-deoxy-2-fluoro- α -D-cellobiosyl-enzyme intermediate.

Demonstration of the catalytic competence of the inactivated enzyme supplied further evidence that this inactivation occurs via stabilization and trapping of the normal intermediate in catalysis. After removal of excess inactivator from the labelled Cel7B, the sample was incubated in buffer at 40 °C and the return of activity associated with the regeneration of free enzyme was monitored. The reactivation of the enzyme followed a first-order process with rate constants of $k_{\text{react}} = 0.0003 \text{ min}^{-1}$ ($t_{1/2} = 40 \text{ h}$) and 0.0012 min^{-1} ($t_{1/2} = 10 \text{ h}$) for the 2-fluorocellobiosyl and 2-chlorocellobiosyl enzymes respectively. Addition of D-cellobiose (19.5 mM) to the reactivation mixtures resulted in an increased rate of regeneration of free enzyme, with pseudo-first-order rate constants of $k_{\text{trans}} = 0.015 \text{ min}^{-1}$ ($t_{1/2} = 46 \text{ min}$) and 0.173 min^{-1} ($t_{1/2} = 4 \text{ min}$) for the 2-fluorocellobiosyl and 2-chlorocellobiosyl enzymes respectively. Therefore, reactivation by transglycosylation occurs approx. 100-fold faster than by spontaneous hydrolysis, which is consistent with the ability of the intermediate to be intercepted by glycosyl acceptors, as has been seen in the reactivation of several 2-fluoroglycosyl-labelled enzymes [16,19,42].

Stoichiometry of incorporation of inactivator

The ESMS analysis of peptide N-glycosidase F-treated Cel7B, identified three major components of mass 45180, 45304 and $45973 \pm 10 \text{ Da}$ (Figure 3a). The expected molecular mass of the enzyme based on amino acid sequence is 45973 Da. The differences between the observed and expected masses can be attributed to residual glycosylation (a GlcNAc residue at positions Asn-89 and Asn-247) of the enzyme and to proteolytic cleavage of the original 415 amino acids to 407, 408 and 412 amino acids. After inactivation of the enzyme with 0.3 mM

Table 2 ESMS mass determination of intact Cel7B treated with 2FDNPC

Values given as Da. ^aMass of 2-fluorocellobiosyl label = 326 Da; ^bmass of 2-fluorocellotriosyl label = 488 Da.

Cel7B unlabelled	Cel7B + 0.5 mM 2FDNPC		Cel7B + 10.0 mM 2FDNPC	
	Mass	Δ Mass ^a	Mass	Δ Mass ^b
45180 ± 7	45501 ± 7	321	45666 ± 7	486
45304 ± 7	45629 ± 7	325	45788 ± 7	484
45973 ± 7	46296 ± 7	323	46464 ± 7	489

2FDNPC for 4 h, a sample was analysed by ESMS, and a set of three new peaks was observed in the mass spectrum (Figure 3b). The average mass difference between the labelled and unlabelled Cel7B, 323 ± 20 Da, is equal, within experimental error, to that of the 2-fluorocellobiosyl label (326 Da), indicating a 1:1 stoichiometry of inactivation. However, integration of the peak areas indicates that only 30–40% of the enzyme was 2-fluorocellobiosyl-labelled during the 6 h inactivation, whereas kinetics of inactivation clearly showed > 90% labelling of the enzyme within 1 h using a similar inhibitor concentration but at 100-fold lower enzyme concentration. The inhibitor concentration was increased to 10 mM in order to drive the inactivation to completion. This 20-fold increase in 2FDNPC resulted in approx. 80% labelling of Cel7B over a period of 24 h (Figure 3c). Interestingly, the mass increase as a result of labelling was not the expected 326 Da for 2-fluorocellobiosyl label, but rather 486 ± 20 Da, corresponding to a 2-fluorocellotriosyl label (expected mass of 488 Da) (Table 2). While the reason for this is not completely clear, two possible explanations spring to mind. One is that, in the presence of high concentrations of 2FDNPC, two molecules of the inactivator undergo a transglycosylation reaction catalysed by Cel7B, resulting in disproportionation to 2FDNPG and 2FDNPG3. The 2FDNPG3 thus formed is then a more efficient inactivator and reacts preferentially. Alternatively, 2FDNPC reacts with the enzyme in the standard fashion to generate the 2-fluorocellobiosyl enzyme. However, in the presence of high concentrations of 2FDNPC this undergoes a transglycosylation process yielding a 2,4-dinitrophenylcello-tetraoside with fluorine atoms at the 2-positions of the first and third sugar residues. If the non-reducing terminal glucose residue is hydrolysed by the enzyme it would yield a difluorinated 2,4-dinitrophenylcellotrioside inactivator which could then react to yield an inactivated enzyme with a mass difference of 490, indistinguishable from that of the 2-fluorocellotriosyl enzyme under these conditions.

Identification of the labelled active-site peptide by ESMS

Peptic digestion of a sample of Cel7B inactivated with 2FDNPC resulted in a mixture of peptides, the components of which were separated by reverse-phase HPLC using the mass spectrometer as a detector. When scanned in LC/MS mode, the total ion chromatogram displayed a large number of peaks arising from every peptide in the digest (Figure 4a). Identification of the peptide bearing the 2-fluorocellotriosyl label was achieved using the tandem mass spectrometer in neutral-loss mode. In this technique the ions are subjected to limited fragmentation by an inert gas in the collision cell of the mass spectrometer. The ester linkage between the sugar and the peptide is one of the more labile bonds present and would be expected to undergo facile homolytic cleavage with the loss of a neutral fluorosugar of

Figure 4 ESMS experiments on peptic digest of Cel7B

(a) Labelled with 2FDNPC, total ion chromatogram (TIC) in normal MS mode, (b) labelled with 2FDNPC, TIC in neutral loss mode, and (c) unlabelled, in neutral-loss mode. (d) Mass spectrum of peptide at 28.6 min.

known mass. Therefore, when the quadrupoles Q1–Q3 are scanned in a linked mode, detecting only ions differing in mass by loss of the label, the peptide can be detected. The spectrometer was scanned in neutral-loss mode looking for the losses of 489, 244.5 and 163 Da corresponding to loss of the label from the singly, doubly or triply charged peptides respectively. Only the neutral-loss experiment of 163 Da revealed any significant signal. This was a single major peak (Figure 4b), that was absent in the control experiment (Figure 4c), eluting at time 28.6 min and having a triply charged mass of 1057.5 ± 1 Da (Figure 4d). The triply charged peptide measured at m/z 1057.5 corresponds to a labelled peptide of mass 3169.5 Da [$(3 \times 1057.5) - 3H$]. Since the mass of the 2-fluorocellotriosyl label is 489, the unlabelled peptide must therefore have a molecular mass of 2681.5 Da ($3169.5 - 489 + 1$), corresponding to a large peptide of > 20 amino acids. Analysis of the amino acid sequence of Cel7B revealed that there are 18 cysteine residues, all of which form disulphide bonds, suggesting that the peptide of interest might be cross-linked via a disulphide to another peptide. A sample of the purified 2-fluorocellotriosyl-labelled peptide of mass 3169.5 Da was reduced using DTT and analysed by LC/MS. Two new peaks were detected: a singly charged species (m/z 1227.5, 1226.5 Da) and a doubly charged species (m/z 974.5, 1947 Da) which combine to give a mass of 3173.5 ± 2 Da. The mass difference between the labelled peptide (3169.5 Da) and the mass of both peptide fragments generated from DTT treatment (3173.5 Da) of 4 Da, indicates the reduction of two disulphide bonds. In order to determine which of the peptide fragments contained the label, a sample of the DTT-reduced peptide was treated with ammonium hydroxide to cleave the ester linkage between the peptide and the fluorosugar. LC/MS analysis of this sample showed one new doubly charged peak at m/z 728.5 Da, which corresponds to a peptide of mass 1455 ± 2 Da. This clearly identifies the labelled peptide fragment as that of mass 1947 Da

Table 3 Edman degradation of 2F-cellotriosyl labelled active-site peptide

PTH, phenylthiohydantoin.

Cycle no.	PTH derivative(s)		Yield (pmol)		cDNA sequences	
1	Y	L	45	48	Y	L
2	G	G	87		G	G
3	T	N	18	32	T	N
4	G	I	41	47	G	I
5	Y	E	35	30	Y	E
6	C	G	—	49	C	G
7	D	K	28	34	D	K
8	A	G	21	40	A	G
9	Q	S	17	18	Q	S
10	C	C	—		C	C
11	F	C	8	—	F	C
12	—	N	—	29	—	N
13	—	—	—	5	—	E
14	—	M	—	2	—	M

and the other unlabelled peptide fragment as that of mass 1226.5 Da.

Computer analysis of the amino acid sequence of Cel7B revealed six candidate peptides for the mass of 1455 Da. The peptide sequence had to contain either an aspartate or glutamate residue, since these are the potential enzymic nucleophiles for glycoside hydrolases [44], plus one or two cysteine residues as deduced from DTT reduction experiments. This resulted in just two candidate peptide sequences: 140-CGMNSALYLSEMH-152, and 185-LGNIEGKGSCCNEM-198. Similar analysis for the second portion of the peptide, mass 1226.5 Da, resulted in 12 candidate peptides, all of which could be eliminated, except for four: 46-LGPGGCGDWGNPP-58, 53-DWGNPPKDV-63, 66-VESCAKNCIME-76, and 167-YGTGYCDAQCF-177, from the requirement of one or two cysteine residues. The crystal

structure of this enzyme (see below) indicates that of all the cysteine pairs present, the only ones that overlap between the candidate peptides determined from the MS experiments are C-172–C-195 and C-176–C-194. This allows for the identification of the two peptide fragments as 167-YGTGYCDAQCF-177 and 185-LGNIEGKGSCCNEM-198 and implicates either Glu-189 or Glu-197 as the active-site nucleophile of Cel7B.

The sequence of the labelled peptide was confirmed using standard chemical sequencing. Edman degradation of the 2-fluorocellotriosyl-labelled peptide yielded the sequence information shown in Table 3, confirming the earlier identification of the 2-fluorocellotriosyl peptide. Evidence for Glu-197 rather than Glu-189 being the active-site nucleophile comes from the sequencing data. Chemical sequencing of a modified glutamic acid residue would result in a greatly reduced yield for that amino acid. However, the yield for the phenylthiohydantoin-derivative for Glu-189 was comparable with that of the flanking residues, while that of Glu-197, admittedly at the end of the sequencing run, was dramatically lower than that of its predecessors.

Covalent attachment of the 2-fluorocellobiosyl label to Glu-197 through an ester linkage was confirmed by treatment of the labelled peptide digest with ammonium hydroxide. After treatment, the labelled peptide of molecular mass 3169.5 was replaced by one new peptide having molecular mass 2680 Da. This would indicate that the labelled glutamate reacted by aminolysis, generating Gln. Glu-197 is indeed found in the enzyme active site, at the expected distance [44] of 5.5 Å from Glu-202, the presumed acid/base catalyst, as is discussed in more detail below.

Catalysis by Cel7B

Cel7B, in common with all family 7 enzymes, catalyses glycosyl-group hydrolysis with net retention of the anomeric configuration, giving the β -anomer as the reaction product [12]. The mechanism is widely accepted to involve a double-displacement reaction in which a covalent glycosyl-enzyme intermediate is

Figure 5 Maximum likelihood weighted 2.2 Å resolution 2Fo – Fc electron density, contoured at approx. 0.4 electrons Å⁻³ in the region of the active site

Residues mentioned in the text are indicated.

formed and hydrolysed, with general acid/base assistance, via oxocarbenium-ion-like transition states (Scheme 1). The active-site groups of Cel7B are located at the base of the deep substrate-binding cleft. The carboxylate functions of the proposed acid/base (Glu-202) and nucleophile (Glu-197), separated by approx. 5.5 Å, are consistent with a catalytic mechanism leading to a net retention of configuration at the anomeric carbon [44] (Figure 5). The identity of these residues was first proposed as a result of the CBH I structure determination [10,45]. Support for these proposals also came from analysis of the related *F. oxysporum* Cel7B with non-hydrolysable thio-oligosaccharide substrate analogues [35], epoxides [46] and 2-fluoroglycosides [16] and through site-directed mutagenesis of the proposed catalytic residues in the *T. reesei* CBH I enzyme [47]. Labelling of Glu-197, described here, is fully consistent with the roles for the catalytic residues proposed by this ensemble of techniques. Glu-202 functions as the catalytic acid/base, with Glu-197 the nucleophile. A third carboxylate, Asp-199, may serve to maintain the nucleophile in the appropriate charged state for catalysis. Mutation of either Glu-202 or Glu-197, to alanine, results in an enzyme with no detectable catalytic activity using either reduced oligosaccharides or *p*-nitrophenylcellobioside as substrates (M. Schülein, unpublished work), confirming their critical role in catalysis. This is in contrast to analogous site-directed mutagenesis experiments on the *T. reesei* CBH I, in which mutations to glutamine still retained significant catalytic activity [47].

Whereas the catalytic mechanisms of family 7 enzymes are now reasonably well documented, at the macroscopic level the differences between CBHs and endoglucanases from this family are more difficult to explain. The exo-versus endo-model is clearly untenable, since CBHs are able to hydrolyse chain-end modified substrates [48] and xyloglucans [49] in a clean endo-manner. These results may not be rationalized in terms of the static X-ray structures. One possibility is that one or more of the loops of the CBH I structure opens occasionally to permit endocatalytic activity [50]. Loop movements, involving 8 Å changes in the main-chain position, to enclose the active site are a feature of substrate-binding in a number of unrelated cellulases, such as the *H. insolens* endoglucanase V [8] and the *Bacillus agaradhearens* Cel5A (A. Varrot, G. Davies and M. Schülein, unpublished work). Such a mechanism may account for the apparent anomalous properties of the various family 7 enzymes.

This work was supported, in part, by Novo-Nordisk A/S, the Biotechnology and Biological Sciences Research Council, the Protein Engineering Network of Centres of Excellence and by the Universities of York, British Columbia and Uppsala. We thank Eleanor Dodson and Gerard Kleywegt for useful discussions and the referees for pointing out an alternative interpretation for the transglycosylation of the inhibitor. Data collection at the EMBL Hamburg outstation was assisted by the European Union HCMP Large Installations program (contract number CHGE-CT93-0040). G.J.D. is a Royal Society University Research Fellow.

REFERENCES

- Gilbert, H. J. and Hazlewood, G. P. (1993) *J. Gen. Microbiol.* **139**, 187–194
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316
- Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788
- Henrissat, B. and Bairoch, A. (1996) *Biochem. J.* **316**, 695–696
- Henrissat, B. and Davies, G. J. (1997) *Curr. Opin. Struct. Biol.* **7**, 637–644
- 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
- Schülein, M. (1997) *J. Biotech.* **57**, 71–81
- Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C. and Schülein, M. (1995) *Biochemistry* **34**, 16210–16220
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, Jr., R. C. and Warren, R. A. J. (1991) *Microbiol. Rev.* **55**, 303–315
- 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
- Edelman, G. M., Cunningham, B. A., Reeke, Jr., G. N., Becker, J. W., Waxdal, M. J. and Wang, J. L. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2580–2584
- Schou, C., Rasmussen, G., Kalltoft, M.-B., Henrissat, B. and Schülein, M. (1993) *Eur. J. Biochem.* **217**, 947–953
- Armand, S., Drouillard, S., Schülein, M., Henrissat, B. and Driguez, H. (1997) *J. Biol. Chem.* **272**, 2709–2713
- Withers, S. G., Street, I. P., Bird, P. and Dolphin, D. H. (1987) *J. Am. Chem. Soc.* **109**, 7530–7531
- Mackenzie, L. F., Brooke, G. S., Cutfield, J. F., Sullivan, P. A. and Withers, S. G. (1997) *J. Biol. Chem.* **272**, 3161–3167
- Mackenzie, L. F., Davies, G. J., Schülein, M. and Withers, S. G. (1997) *Biochemistry* **36**, 5893–5901
- Tull, D., Miao, S., Withers, S. G. and Aebersold, R. (1995) *Anal. Biochem.* **224**, 509–514
- Miao, S., McCarter, J. D., Grace, M., Grabowski, G., Aebersold, R. and Withers, S. G. (1994) *J. Biol. Chem.* **269**, 10975–10978
- Miao, S., Ziser, L., Aebersold, R. and Withers, S. G. (1994) *Biochemistry* **33**, 7027–7032
- Christensen, T., Wöldike, H., Boel, E., Mortensen, S. B., Hjortshøj, K., Thim, L. and Hansen, M. T. (1988) *Bio/Technol.* **6**, 1419–1422
- Tull, D. and Withers, S. G. (1994) *Biochemistry* **33**, 6363–6370
- Mackenzie, L. F., Wang, Q., Warren, R. A. J. and Withers, S. G. (1998) *J. Am. Chem. Soc.* **120**, 5583–5584
- Davies, G., Tolley, S., Wilson, K., Schülein, M., Wöldike, H. F. and Dodson, G. (1992) *J. Mol. Biol.* **228**, 970–972
- Otwinowski, Z. and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Collaborative Computational Project Number 4. (1994) *Acta Cryst.* **D50**, 760–763
- Navaza, J. (1994) *Acta Cryst.* **A50**, 157–163
- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987) *Science* **235**, 458–460
- 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
- Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) *Acta Cryst.* **A47**, 110–119
- Brünger, A. T. (1992) *Nature (London)* **355**, 472–475
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) *Acta Cryst.* **D53**, 240–255
- Ramakrishnan, C. and Ramachandran, G. N. (1965) *Biophys. J.* **5**, 909–933
- Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 282–291
- Kleywegt, G. J., Zou, J.-Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Srisodsuk, M., Teeri, T. T. and Jones, T. A. (1997) *J. Mol. Biol.* **272**, 383–397
- Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. and Davies, G. J. (1996) *Biochemistry* **35**, 15280–15287
- Sheppard, P. O., Grant, F. J., Oort, P. J., Cindy, A. S., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L. and O'Hara, P. J. (1994) *Gene* **150**, 163–167
- Kleywegt, G. J. and Jones, T. A. (1994) *ESF/CCP4 Newsl.* **31**, 9–14
- 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
- Withers, S. G. and Street, I. P. (1988) *J. Am. Chem. Soc.* **110**, 8551–8553
- Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B. and Aebersold, R. (1990) *J. Am. Chem. Soc.* **112**, 5887–5889
- Reference deleted
- McCarter, J. D. and Withers, S. G. (1996) *J. Biol. Chem.* **271**, 6889–6894
- Reference deleted
- McCarter, J. D. and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4**, 885–892
- Divne, C., Ståhlberg, J., Teeri, T. T. and Jones, T. A. (1997) *J. Mol. Biol.* **275**, 309–325
- Sulzenbacher, G., Schülein, M. and Davies, G. J. (1997) *Biochemistry* **36**, 5902–5911
- Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeysens, M., Teeri, T. T. and Jones, T. A. (1996) *J. Mol. Biol.* **264**, 337–349
- Boisset, C., Armand, S., Drouillard, H., Chanzy, H., Driguez, H. and Henrissat, B. (1998) In *Carbohydrases From Trichoderma reesei and Other Microorganisms* (Claeysens, M., Piens, K. and Nerinckx, W., eds.), pp. 124–132. The Royal Society of Chemistry, London
- Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E. and Kanda, T. (1996) *J. Biochem. (Tokyo)* **120**, 1123–1129
- Divne, C. (1994) Ph.D. Thesis, Uppsala University, Sweden
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950