Identification of Glu-120 as the catalytic nucleophile in *Streptomyces lividans* endoglucanase CelB

David L. ZECHEL*, Shouming HE*, Claude DUPONT† and Stephen G. WITHERS*¹

*Department of Chemistry, University of British Columbia, Vancouver, 2036 Main Mall, Vancouver, British Columbia, Canada, V6T 1Z1,

and †INRS Institut Armand-Frappier, Centre de recherche en microbiologie et biotechnologie, 531 boul. des Prairies, Laval, Québec, Canada, H7N 4Z3

Streptomyces lividans CelB is a family-12 endoglucanase that hydrolyses cellulose with retention of anomeric configuration. A recent X-ray structure of the catalytic domain at 1.75 Å resolution has led to the preliminary assignment of Glu-120 and Glu-203 as the catalytic nucleophile and general acid–base respectively [Sulzenbacher, Shareck, Morosoli, Dupont and Davies (1997) *Biochemistry* 36, 16032–16039]. The present study confirms the identity of the nucleophile by trapping the glycosyl-enzyme intermediate with the mechanism-based inactivator 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-cellobioside (2FDNPC). The kinetics of inactivation proceeded in a saturable fashion, yielding the parameters $k_{\text{inact}} = 0.29 \pm 0.02 \text{ min}^{-1}$ and $K_{\text{inact}} = 0.72 \pm 0.08 \text{ mM}$. Uncompetitive inhibition was observed at high concentrations of 2FDNPC ($K_i = 9 \pm 1 \text{ mM}$), a behaviour that was also observed with the substrate 2',4'-dinitrophenyl β -D-cellobioside ($k_{\text{cat}} = 40 \pm 1 \text{ s}^{-1}$, $K_m = 0.35 \pm 0.03 \text{ mM}$, $K_i = 24$

$\pm 4 \text{ mM}$). Protection against inactivation was afforded by the competitive inhibitor cellobiose. The electrospray ionization (ESI) mass spectrum of the intact labelled CelB indicated that the inactivator had labelled the enzyme stoichiometrically. Reactivation of the trapped intermediate occurred spontaneously ($k_{\text{H}_2\text{O}} = 0.0022 \text{ min}^{-1}$) or via transglycosylation, with cellobiose acting as an acceptor ligand ($k_{\text{react}} = 0.024 \text{ min}^{-1}$, $K_{\text{react}} = 54 \text{ mM}$). Digestion of the labelled enzyme by pepsin followed by LC–ESI–tandem MS (MS–MS) operating in neutral loss mode identified a labelled, singly charged peptide of m/z 947.5 Da. Isolation of this peptide by HPLC and subsequent collision-induced fragmentation by ESI–MS–MS produced a daughter-ion spectrum that corresponded to a sequence (QTEIM) containing Glu-120. The nucleophile Glu-120 and the putative acid–base catalyst Glu-203 are conserved in all known family-12 sequences.

INTRODUCTION

The degradation of cellulose into cello-oligosaccharides by bacterial and fungal microbes is achieved by the secretion of cellulases into the extracellular medium. These are glycosidases that hydrolytically cleave the β -(1,4) glycosidic bonds of cellulose with either retention or inversion of anomeric configuration [1]. The bacterial cellulase *Streptomyces lividans* CelB is a 36 kDa endoglucanase that has been assigned on the basis of sequence to Henrissat's family 12 (formerly known as family H) [2–6]. Characteristic of many cellulases, the tertiary structure of CelB consists of an N-terminal signal peptide and a catalytic domain connected to a C-terminal cellulose-binding domain by a flexible linker [2,7]. The structure of the catalytic domain was recently solved to 1.75 Å resolution, revealing a 'jelly-roll' fold that is remarkably similar to that found in the family 11 xylanases [8], thereby confirming the prediction from hydrophobic cluster analysis that these two glycosidase families comprise a structurally related 'clan' (GH-C) [5].

Although the stereochemical course of substrate hydrolysis has not been determined for CelB, another member of family 12, *Humicola insolens* endoglucanase III, is known to hydrolyse cello-oligosaccharides with retention of anomeric configuration [9]. The classical glycosidase retaining mechanism (Scheme 1) [1] involves glycosylation and deglycosylation steps, each catalysed by a pair of carboxy groups. During glycosylation a covalent glycosyl-enzyme intermediate is formed by the nucleophilic attack

of one carboxy group at the substrate anomeric centre, with the second carboxy group providing general acid assistance to the departing glycosidic oxygen. General base-catalysed attack of water on the opposite face of the anomeric centre results in deglycosylation, displacing the nucleophile and forming a product with the same anomeric configuration as the substrate. These steps are not simple S_N2 displacements; from kinetic isotope effects and linear free energy relationship studies with other retaining glycosidases it is known that the transition states for both steps have significant oxocarbenium ion character [10,11]. This has led to the development of a class of mechanism-based inactivators that possess an electronegative fluorine atom at the 2 or 5 position of the pyranosyl ring, a substitution that greatly destabilizes both transition states. A reactive aglycone leaving group (fluoride or 2,4-dinitrophenol) ensures that glycosylation is faster than deglycosylation, resulting in entrapment of the glycosyl-enzyme intermediate and specific labelling of the nucleophile. Numerous glycosidases have been successfully trapped in this fashion [12–14] and their nucleophiles have been identified through a process of proteolysis and electrospray ionization MS (ESI–MS) analysis [15].

This labelling strategy was used to identify the nucleophile in the family-11 *Bacillus circulans* xylanase as Glu-78 [16], allowing the only other Glu residue in this enzyme (Glu-172) to be assigned as the acid–base catalyst. Because *B. circulans* xylanase and CelB belong to the same structurally related clan (GH-C), a direct comparison of the xylanase structure [17] with that of CelB

Abbreviations used: DNPC, 2',4'-dinitrophenyl β -D-cellobioside; ESI–MS, electrospray ionization MS; 2FDNPC, 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-cellobioside; MS–MS, tandem MS; TIC, total ion current.

¹ To whom correspondence should be addressed (e-mail withers@chem.ubc.ca).

Scheme 1 Hydrolytic mechanism for a retaining cellulase

permitted the assignment of Glu-120 and Glu-203 as the nucleophile and acid–base catalysts respectively [8]. These residues are located 7 Å apart on opposite sides of a long, open active-site cleft that can accommodate five glucopyranoside moieties (–3 to +2). This paper describes the steady-state kinetic behaviour of CelB with 2',4'-dinitrophenyl β -D-cellobioside (DNPC), as well as trapping of the covalent intermediate with the corresponding inactivator, 2',4'-dinitrophenyl 2-deoxy-2-fluoro- β -D-cellobioside (2FDNPC). Proteolysis of the labelled enzyme and analysis by ESI–MS confirms the identity of the nucleophile as Glu-120, a residue that is conserved in all known family-12 sequences.

EXPERIMENTAL

General

The syntheses of DNPC and 2FDNPC were performed as described previously [18]. The expression and purification of CelB have also been published elsewhere [8]. This is a truncated form of the enzyme, lacking a signal peptide and cellulose-binding domain, corresponding to Asp-41 to Gly-274 of the sequence listed in the National Center for Biotechnology Information (NCBI) database (accession number 2462718). Enzyme concentrations were determined by using the molar absorption coefficient ϵ_{280} 54.6 mM⁻¹·cm⁻¹ ($\epsilon_{0.1\%}^{280}$ 2.2), which was calculated from the content of tyrosine and tryptophan residues [19].

Steady-state kinetics

Steady-state kinetic studies were performed on a Unicam 8700 UV/Vis spectrometer equipped with a circulating-water bath. Appropriate concentrations of DNPC were diluted in 200 μ l of 50 mM sodium citrate/0.1% BSA (pH 6.5) and equilibrated at 37 °C in quartz microcuvettes. An aliquot of CelB [10 μ l, 0.81 μ g/ml, diluted in 50 mM sodium citrate/0.1% BSA (pH 6.5)] was added to initiate hydrolysis; the release of 2,4-dinitrophenol was monitored continuously at 400 nm ($\Delta\epsilon_{400}$ 10910 M⁻¹·cm⁻¹) [10]. Enzymic hydrolysis rates were measured for DNPC concentrations ranging from 0.2 to 40 times the estimated K_m . These enzymic hydrolysis rates were corrected for the significant spontaneous hydrolysis rate of DNPC (measured before the addition of CelB). The resulting velocity curve was fitted to eqn. (1), by using the program Fig.P [20]. This equation describes an uncompetitive substrate inhibition scheme, which involves the binding of a second substrate molecule to the active site with a dissociation constant K_i to form an unproductive ternary complex (E·S·S):

$$V_0 = \frac{V_{\max}[S]}{K_m + [S] + [S]^2/K_i} \quad (1)$$

Inactivation kinetics

CelB [100 μ l, 0.02 mg/ml in 50 mM sodium citrate/0.1% BSA (pH 6.5)] was incubated at 37 °C in the presence of various

Scheme 2 Mechanism for the inactivation of CelB by 2FDNPC and subsequent reactivation of the trapped intermediate

The inactivator 2FDNPC and the reactivation acceptor are denoted by I and L respectively.

concentrations of 2FDNPC (0.057–11.38 mM). Aliquots (10 μl) were removed from the inactivation mix at appropriate intervals and added to cuvettes pre-equilibrated at 37 °C with 0.5 mM solutions of DNPC [900 μl ; 50 mM sodium citrate/0.1 % BSA (pH 6.5)], whereupon the hydrolysis rates were measured as described above. Each inactivation reaction was monitored for a minimum of three half-lives and the resulting curve of fractional activity (V/V_0) over time was fitted to single-exponential kinetics [$V/V_0 = \exp(-k_{\text{obs(inact)}}t) + C$] by using GraFit. The resulting values of $k_{\text{obs(inact)}}$ were subsequently plotted against the corresponding concentrations of 2FDNPC and fitted to eqn. (2) by using Fig.P, which describes an inactivation scheme in which the inactivator also acts as an uncompetitive inhibitor (Scheme 2):

$$k_{\text{obs(inact)}} = \frac{k_{\text{inact}}[\text{I}]}{K_{\text{inact}} + [\text{I}] + [\text{I}]^2/K_i} \quad (2)$$

A control reaction (no inactivator) indicated that the spontaneous loss of enzyme activity was negligible within the time range considered.

Protection against inactivation was demonstrated by reacting CelB with 2FDNPC (2.84 mM) in the presence and in the absence of 47 mM cellobiose, a competitive inhibitor. When the concentration of 2FDNPC is sufficiently low to neglect uncompetitive inhibition, the ratio of observed rate constants with and without a competitive inhibitor present is given by eqn. (3) (Scheme 2):

$$\frac{k_{\text{obs(inact)}}^{\text{prot}}}{k_{\text{obs(inact)}}} = \frac{K_{\text{inact}} + [\text{I}]}{K_{\text{inact}}(1 + [\text{P}]/K_d) + [\text{I}]} \quad (3)$$

where P is the competitive inhibitor with a dissociation constant K_d .

Reactivation kinetics

CelB was diluted to 0.1 mg/ml in 50 mM sodium citrate, pH 6.5 (without BSA) to a total volume of 120 μl and incubated for 60 min at 37 °C with 0.5 mM 2FDNPC. This resulted in more than 95 % inactivation of the enzyme, as determined by the residual activity assay described above. Excess inactivator was removed by dilution with BSA free buffer (to 400 μl) and subsequent ultrafiltration (to a final volume of approx. 100 μl) with a Millipore Ultrafree-MC centrifugal concentrator (10 kDa nominal molecular mass cut-off). This was performed at 4 °C to minimize premature reactivation. After three such cycles, the final retained material was diluted to approx. 0.025 mg/ml inactivated CelB with buffer containing 0.1 % BSA (650 μl total volume). The inactivated enzyme was then incubated at 37 °C in the presence of various concentrations of cellobiose (0–43 mM) and assayed for activity as described above. To compensate for loss of activity through denaturation of the enzyme, a control

reaction involving the incubation of CelB in the absence of 2FDNPC was taken through the same steps. The increase in fractional activity (V/V_0) was monitored over 20 h and the resulting time courses were fitted to single-exponential kinetics [$V/V_0 = A[1 - \exp(-k_{\text{obs(react)}}t)] + C$] by using GraFit. The rate constants ($k_{\text{obs(react)}}$) so derived were plotted against the corresponding concentrations of cellobiose in double-reciprocal form and fitted to the reciprocal form of eqn. (4), which was derived from Scheme 2:

$$k_{\text{obs(react)}} = \frac{k_{\text{react}}[\text{cellobiose}]}{K_{\text{react}} + [\text{cellobiose}]} \quad (4)$$

Labelling of CelB with 2FDNPC

CelB (120 μl , 3.7 mg/ml in 50 mM sodium citrate without BSA) was incubated with 0.5 mM 2FDNPC for 60 min at 37 °C. Residual activity was decreased to 5 % by this time. The sample was then used immediately for proteolysis or for ESI-MS analysis of the intact protein. Alternatively the samples were stored at –78 °C until required.

Proteolysis

The labelled CelB was highly resistant to proteolysis, as judged by HPLC profiles of peptic digests over time; it was therefore necessary to denature the enzyme (by boiling) before digestion. Immediately after incubation with 2FDNPC, CelB (100 μl , 3.7 mg/ml labelled or native) was mixed with 200 μl of 200 mM sodium phosphate, pH 2, and boiled for 3 min. The resulting enzyme precipitates were rapidly clarified on the addition of 2 μl of pepsin [5.6 mg/ml in 200 mM sodium phosphate (pH 2)]. Labelled and native digests were incubated overnight at room temperature and were either analysed immediately by ESI-MS or frozen until required.

Stoichiometry of incorporation of the inactivator determined by ESI-MS

All mass spectra [intact LC-MS, tandem MS (MS-MS) neutral loss, and MS-MS daughter-ion spectra] were recorded with a PE-Sciex API 300 triple quadrupole mass spectrometer (Sciex, Thornhill, Ontario, Canada) equipped with an ESI source. For the intact mass spectra, CelB (4 μg , 2F-cellobiosyl labelled or unlabelled) 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 was scanned over an m/z range of 600–2400 Da, with a step size of 0.5 Da and a dwell time of 1 ms per step. The ion source potential was set at 5 kV; the orifice energy was 50 V. The molecular masses of the CelB species were determined by using the deconvolution software Multiview 1.1, supplied by Sciex.

ESI-MS analysis of the proteolytic digest

Peptides were separated by reverse-phase HPLC on an Ultra-fast Microprotein Analyser (Michrom BioResources, Pleasanton, CA, U.S.A.) directly interfaced with the mass spectrometer. In each of the MS experiments, the proteolytic digest was loaded on a C_{18} column (Reliasil; 1 mm $\times$ 150 mm), then eluted with a gradient of 0–60 % solvent B over 60 min, followed by 100 % B over 2 min at a flow rate of 50 $\mu\text{l}/\text{min}$ [solvent A was 0.05 % (v/v) trifluoroacetic acid/2 % (v/v) acetonitrile in water; solvent B was 0.045 % (v/v) trifluoroacetic acid/80 % (v/v) acetonitrile in water]. A post-column splitter sent 80 % of the sample into a fraction collector and 20 % into the mass spectrometer. The total

ion current (TIC) was initially detected in single-quadrupole scan mode (LC-MS) to produce an HPLC chromatogram for the whole digest. An m/z range of 300–2200 Da was scanned by using mass spectrometer settings identical with those used for the intact protein. The TIC for a second HPLC run was subsequently collected in neutral loss mode, scanning for the loss of the 2-fluorocellobiosyl label (327 Da) from a singly charged peptide. The collision gas was argon/ N_2 (9:1). HPLC fractions corresponding to the labelled peptide were saved for sequencing.

Sequencing of the labelled peptide by ESI-MS-MS

The MS-MS daughter-ion spectrum of the labelled peptide isolated above was obtained in the triple-quadrupole scan mode by selectively introducing the m/z 947.5 Da peptide from the first quadrupole (Q1) into the collision cell (Q2) and observing the daughter ions in the third quadrupole (Q3). The following settings were used: Q1 was locked on m/z 947.5 Da; the Q3 scan range was m/z 50–970 Da; the step size was 0.5 Da; the dwell time was 1 ms per step; the ion source potential was 5.5 kV; the orifice potential was 50 V; the focusing ring voltage was 400 V; the Q0 potential was -10 V; the Q2 potential was -55 V; the collision gas was N_2 .

RESULTS AND DISCUSSION

CelB readily hydrolysed DNPC to produce a biphasic velocity profile (Figure 1) characteristic of uncompetitive substrate inhibition. A direct fit of the plot of V_0 against $[S]$ to eqn. (1) gave $k_{cat} = 40 \pm 1 \text{ s}^{-1}$, $K_m = 0.35 \pm 0.03 \text{ mM}$ and $K_i = 24 \pm 4 \text{ mM}$ ($\pm$ S.D.). The long, open binding cleft observed by X-ray crystallography suggested that CelB could accommodate five or six glucopyranosyl units (-3 to $+2$ or $+3$); the simultaneous binding of two molecules of DNPC to form an unproductive ternary complex is therefore certainly possible. This mode of inhibition has also been observed in *Bacillus licheniformis* 1,3-1,4- β -endoglucanase (family 16) with various β -D-glucan oligosaccharides ($K_i = 30$ – 60 mM) [21], as well as in *Agrobacterium* sp. β -glucosidase (family 1), an exoglycosidase, with β -glucosyl fluoride, 4-nitrophenyl β -D-glucopyranoside and 4-nitrophenyl β -D-galactopyranoside ($K_i = 60$ – 160 mM) [22,23]. Both enzymes have extended binding clefts (-4 to $+1$ and -1 to $+2$ respectively). It is noteworthy that the structure of CelB includes

Figure 1 Steady-state kinetics for the hydrolysis of DNPC by CelB

The dotted lines indicate the 95% confidence level for the curve fits to eqn. (1).

Figure 2 Inactivation of CelB by 2FDNPC

(a) Representative data for the loss of activity (V/V_0) over time in the presence of various concentrations of 2FDNPC: $\square$, 2.84 mM; $\bullet$, 11.38 mM; $\circ$, 0.17 mM; $\triangle$, 0.087 mM; $\blacksquare$, 0.057 mM. (b) Plot of $k_{obs(inact)}$ against $[2FDNPC]$. The dotted lines indicate the 95% confidence level for the curve fitted to eqn. (2).

a surface loop that might close over the active-site cleft on substrate binding, and that a decrease in the distance separating the putative nucleophile and acid-base residues (7 Å) to the typical retaining glycosidase span of 5.5 Å seems to be necessary for catalysis [8,24]. Such conformational adjustments might inadvertently stabilize a non-productive ternary complex.

Uncompetitive substrate inhibition seems to be rare among retaining glycosidases; indeed, a high concentration of substrate frequently leads to the reverse effect, an acceleration in rate, owing to transglycosylation. No transglycosylation products were formed during the reaction of CelB with DNPC (15 mM) when analysed by TLC over the course of 20 h. This might be a substrate-specific effect, however, as illustrated by the switch to transglycosylation activity by *Agrobacterium* sp. β -glucosidase with aryl β -xylobiosides [10]. Similarly, cellohexitol is rapidly transglycosylated into higher oligomers by another family-12 glycosidase, *H. insolens* endoglucanase III [9]. Presumably the greater binding energy in the aglycone subsites realized with longer substrates favours transglycosylation. The importance of aglycone binding energy in this family is further illustrated by the inability of *H. insolens* endoglucanase III to hydrolyse cellotriitol, and also of *Erwinia carotovora* CelS, another family-12

Figure 3 Protection against inactivation with cellobiose

Inactivation of CelB in the presence of 2.84 mM 2FDNPC (●) or 2.84 mM 2FDNPC with 47 mM cellobiose (○).

endoglucanase, to hydrolyse 4-methylumbelliferyl β -cellobioside [25]. The weakened aglycone binding interactions available in these substrates, in combination with the poor leaving-group ability of the aglycones ($pK_a \approx 16$ for glucitol and $pK_a = 7.53$ for 4-methylumbelliferone), can be expected to hinder enzymic hydrolysis. In contrast, the high reactivity of DNPC ($pK_a = 3.96$ for 2,4-dinitrophenol) obviates the need for energetic gains through aglycone binding interactions.

Uncompetitive substrate inhibition is also observed with the mechanism-based inactivator 2FDNPC. The initial time dependence of inactivation of CelB through the accumulation of a covalent 2-fluorocellobiosyl-enzyme intermediate followed first-order kinetics (Figure 2a). The corresponding dependence of $k_{\text{obs(inact)}}$ on [2FDNPC] (Figure 2b) parallels the velocity curve observed for the hydrolysis of DNPC. Fitting this curve directly with the analogous eqn. (2) (derived from Scheme 2) gave $k_{\text{inact}} = 0.29 \pm 0.02 \text{ min}^{-1}$, $K_{\text{inact}} = 0.72 \pm 0.08 \text{ mM}$ and $K_i = 8 \pm 1 \text{ mM}$. This inactivation is active-site-specific, as shown by the decrease in $k_{\text{obs(inact)}}$ with 2.84 mM 2FDNPC from 0.18 ± 0.02 to $0.048 \pm 0.007 \text{ min}^{-1}$ in the presence of 47.4 mM cellobiose (Figure 3), a competitive inhibitor of CelB ($K_d \approx 5 \text{ mM}$). The ratio (0.27) of these rate constants is in reasonable agreement with the protection ratio (0.34) predicted by eqn. (3). The catalytic competence of the trapped 2-fluorocellobiosyl-enzyme intermediate was demonstrated by the spontaneous time-dependent regaining of activity observed when the trapped enzyme was freed of excess inactivator ($k_{\text{H}_2\text{O}} = 0.0022 \text{ min}^{-1}$, $t_{1/2} = 314 \text{ min}$). Furthermore reactivation was accelerated in a saturable fashion (Figure 4) through a transglycosylation mechanism (Scheme 2) when cellobiose was added as an acceptor ligand, yielding the kinetic parameters $k_{\text{react}} = 0.027 \text{ min}^{-1}$ and $K_{\text{react}} = 54 \text{ mM}$ [eqn. (4)].

A comparison of the inactivation and reactivation kinetics with the steady-state kinetics of DNPC can provide an estimate of the kinetic consequences of the 2-fluoro substitution. The deglycosylation rate, reflected by k_{cat} and $k_{\text{H}_2\text{O}}$, is reduced 1.1×10^6 -fold, a value comparable with those observed for *Agrobacterium* sp. β -glucosidase [26], *Escherichia coli*, lac Z β -galactosidase [27], *Fusarium oxysporum* endoglucanase I [28] and *Clostridium thermocellum* endoglucanase CelC [29] on comparison of the equivalent kinetic parameter. The glycosylation rate, reflected by k_{cat}/K_m and $k_{\text{inact}}/K_{\text{inact}}$, is reduced to a lesser

Figure 4 Turnover of the 2-deoxy-2-fluorocellobiosyl-CelB intermediate via hydrolysis or transglycosylation

(a) Reactivation of CelB with time in the presence of various concentrations of cellobiose: ●, 0 mM; ■, 11 mM; ◇, 28 mM; □, 56 mM; ○, 84 mM, ◆, 125 mM. (b) Double-reciprocal plot of $1/k_{\text{obs(react)}}$ against $1/[\text{cellobiose}]$.

degree, 1.7×10^4 -fold. This is presumably due to the greater leaving-group ability of 2,4-dinitrophenol than that of the enzymic carboxy group of the covalent intermediate, although the deglycosylation transition state typically has greater oxocarbenium ion character than that for the glycosylation step [10,30] and might also contribute to this rate difference. In contrast with these large transition-state effects, the similarity of K_{inact} and K_i to the corresponding K_m and K_i values for DNPC suggests that the 2-fluoro substitution has a minimal impact on substrate binding.

The ESI mass spectrum of intact CelB (results not shown) provided a mass of 24567 Da for the catalytic domain, identical to the average $[M+H]^+$ mass predicted for the sequence. This suggests that the sulfoxide form of Met-122 observed in the crystal structure [8] is a by-product of X-ray diffraction or the conditions used for crystal growth. Surprisingly, the corresponding ESI mass spectrum of labelled CelB (results not shown) indicated that only approx. 50% of the enzyme had incorporated a single 2-fluorocellobiosyl moiety (327 Da), resulting in a mass shift of $326 \pm 6 \text{ Da}$. According to an assay of activity with DNPC before ESI-MS, the labelled enzyme was estimated to be more than 95% inactivated. The origin of this discrepancy is not clear, but it is most probably due to a loss of label on the HPLC

Figure 5 Detection of the 2-deoxy-2-fluorocellobiosyl labelled peptide by ESI-MS-MS

(a) Total ion current (TIC) for the peptic digest of CelB labelled with 2FDNPC. (b) TIC for the labelled CelB digest, with ESI-MS-MS operating in neutral loss mode. Only those peptides that lose a neutral fragment of m/z 327 (the mass of the label) are detected. The labelled peptide of interest is indicated with an asterisk. (c) TIC for the peptic digest of the unlabelled enzyme in neutral loss mode. (d) Mass spectrum of the labelled peptide eluted at 23 min in (b).

column before MS analysis. Nevertheless this degree of labelling was sufficient to identify the nucleophile, as described below.

The labelled CelB was markedly resistant to digestion with pepsin relative to the unlabelled enzyme, as judged by HPLC profiles of the digests over time. This resistance to proteolysis has been observed with other trapped glycosidases and is attributed to a stabilization of the glycosidase structure on the formation of the covalent intermediate [31]. However, it was possible to denature the labelled CelB (by boiling) before proteolysis without significant loss of the label. Figure 5(a) shows the peptic digest of the labelled enzyme resolved by reverse-phase HPLC with ESI-MS as a detector. Labelled peptide(s) were detected in a second HPLC run with the spectrometer operating in neutral loss mode. This involved scanning the HPLC eluate with Q1 and Q3 while subjecting the gas-phase peptide ions to mild fragmentation with a neutral gas in the collision cell Q2. The α -acylal bond of the label is significantly more labile than any other bond in the peptide and readily cleaves in a homolytic fashion, leaving the peptide with its original charge but with decreased mass. With Q3 offset from Q1 by the mass of the label that is lost in the collision cell, peptides originally bearing this label can be detected. As shown in Figure 5(b), scanning for the loss of the 2-fluorocellobiosyl label (327 Da) from a singly charged peptide produced a fragment of m/z 947.5 Da that was not present in the unlabelled digest (Figure 5c). Subtraction of the mass of the label ($947.5 - 327 + 1\text{H}$) yields an unlabelled peptide of m/z 621.5 Da. A computer-aided search [32] of the CelB sequence for a peptide with an average m/z of 621.5 ± 1 Da produced only eight sequences, including QTEIM, which corresponds to the region of

Figure 6 Sequencing the 2-deoxy-2-fluorocellobiosyl-labelled peptide (m/z 947.5) by ESI-MS-MS

Daughter-ion mass spectrum for the labelled peptide at m/z 947.5 Da ($621.5 + 327 - 1\text{H}^+$). The 'b' and 'y' ions corresponding to the sequence QTEIM are indicated, as well as the corresponding 2-fluoroglucopyranosyl (164 Da)- and 2-fluorocellobiosyl (327 Da)-labelled forms.

the putative nucleophile Glu-120. Examination of the total ion chromatogram for the unlabelled digest confirmed the presence of a peptide of m/z 621.5 with a retention time essentially identical with that of the labelled peptide.

The sequence QTEIM was confirmed by ESI-MS-MS analysis of the labelled peptide isolated by HPLC. Q1 was set to m/z 947.5 to selectively introduce the peptide to the collision chamber (Q2), with the resulting daughter-ion spectrum detected by Q3. As shown in Figure 6, the spectrum contained a series of 'b' ions (N-terminal fragments produced by cleavage between CO-NH bonds) corresponding to the fragments QT (m/z 229.5), QTE (m/z 358.5), QTEI (m/z 472.5) and QTEIM (m/z 603.5). A series of 'y' ions (C-terminal fragments produced by cleavage between CO-NH bonds) were also observed, corresponding to M (m/z 149), IM (m/z 262) and EIM (m/z 392.5) fragments. Unequivocal evidence for the position of the label at Glu-120 was provided by fragmentation of the label itself (see Figure 6). Cleavage of the internal glycosidic bond to release the terminal glucosyl moiety gave rise to a series of 'b' ions bearing a 2-fluoroglucopyranosyl residue (164 Da) corresponding to QTE (m/z 523.5), QTEI (m/z 636.5) and QTEIM (m/z 767.5). Full 2-fluorocellobiosyl-labelled 'b' ions were also observed: QTE (m/z 685), QTEI (m/z 798.5) and QTEIM (m/z 930). However, a labelled 'b' or 'y' ion corresponding to QT was not observed, thereby eliminating the unlikely possibility of threonine acting as the nucleophile.

A BLAST search [33] retrieved 12 other endoglucanase sequences that share a highly conserved nucleophile sequence, E(I/L)M(I/V)W (Figure 7). Additionally, two *Thermotoga* endoglucanase sequences were aligned with CelB with the use of hydrophobic cluster analysis and, as shown in Figure 7, also contain the distinctive family-12 nucleophile sequence. In contrast, the sequence around the putative general acid-base catalyst (Glu-203) is less strictly conserved, suggesting that the structure

		↓			↓	
<i>S. lividans</i> CelB	118	QTEIMWFNVRVGPIQPIG	135	194	YLTSVQAGFEPWQNG	208
<i>S. rochei</i> A2	155	RTEIMWFNVRVGPIQPIG	172	231	YLTSIQAGFEPWQNG	245
<i>S. halstedii</i>	155	RTEIMWLKNKVGPIQPIG	172	231	YLTSVQAGFEPWQNG	245
<i>M. tuberculosis</i>	48	QBEIMWFNHNQSGIQPVG	65	124	YLTSIRAGLEPWSGD	138
<i>R. marinus</i>	157	GABEIMWFLNNGGVMGPP	174	233	YLHAVETGFELWEGG	247
<i>A. aculeatus</i>	132	DYELMIWLARYGGVQPIG	149	209	YLITLQFGTEPFTGG	223
<i>T. reesei</i> EGIII	130	DYELMIWLKGKYGDIQPIG	147	207	YVLSYQFGTEPFTGS	221
<i>A. awamori</i>	130	DYELMIWLARYGSVQPIG	147	211	HLITLQCGTEPFTGG	225
<i>A. oryzae</i>	130	DYELMIWLARYGTQPIG	147	211	YLINMQFGTEPFTGG	225
<i>E. carotovora</i> CelS	156	TDELMWLNDNTNAG-PAG	172	237	YISSVFEFTGTEIF-GG	250
<i>E. carotovora</i> CelB	156	TDELMWLNDNTNAG-PAG	172	237	YISSVFEFTGTEIF-GG	250
<i>T. neapolitana</i> endoA	132	DVEIMVWFYFNELT-PGG	148	222	YFTVWEIGTE-FGSP	235
<i>T. neapolitana</i> endoB	150	DVEIMVWFYNNILM-PGG	166	241	YFCVWEIGTE-FGDP	254
<i>T. maritima</i> CelA	132	DVEIMVWFYFNVLTP-PGG	148	222	YFTVWEIGTE-FGSP	235
<i>T. maritima</i> CelB	150	DAEIMVWFYNNVLM-PGG	166	241	YFCVWEIGTE-FGDP	254

Figure 7 Alignment of the sequence regions containing the labelled nucleophile and the putative general acid–base catalyst of CelB with other family-12 endoglucanases

Accession numbers from NCBI are indicated in parentheses. The sequences shown are *S. lividans* CelB (2462718), *Streptomyces rochei* (strain A2) endoglucanase (1076114), *Streptomyces halstedii* endoglucanase (2209260), *Mycobacterium tuberculosis* 'hypothetical' endoglucanase (2896727), *Rhodothermus marinus* endoglucanase (2304961), *Aspergillus aculeatus* endoglucanase I (FI-CMCASE) (121835), *Hypocrea jecorina* (*Trichoderma reesei*) endoglucanase III (2116583), *Aspergillus awamori* (var. *kawachi*) endoglucanase A (CMCASF-1) (2494338), *Aspergillus oryzae* endoglucanase (2467373); *Erwinia carotovora* endoglucanase S (121830), *E. carotovora* endoglucanase B (2570837), *Thermotoga neapolitana* endoglucanase A (1934673), *T. neapolitana* endoglucanase B (1934674), *T. maritima* endoglucanase A (1297061) and *T. maritima* endoglucanase B (1297062). The *T. maritima* cellulase sequences were aligned by hydrophobic cluster analysis (courtesy of B. Henrissat CNRS-CERMAV, Grenoble, France). The sequence numbering includes the signal peptide sequences, with the exception of CelB, which begins at Asp-41 of the NCBI entry. Absolutely conserved residues are shown in bold. The proposed nucleophiles and acid–base catalysts are indicated by arrows (↓ and ↓ respectively).

of the acid–base region is subject to a reduced evolutionary pressure. As observed in other retaining glycosidases [34–38], this is consistent with a lesser contribution of the general acid–base residue to catalysis relative to that of the nucleophile. As noted previously, the nucleophile in CelB occupies a structurally equivalent position to the nucleophile of family-11 xylanases [8]. Interestingly, despite the similar folds and the fact that the sequence after the nucleophile is highly conserved in each family [EYY(V/DV) in family 11], there is little resemblance between the two families in this region, beyond the fact that it is relatively hydrophobic.

In conclusion, the trapping of the covalent intermediate formed by *S. lividans* endoglucanase CelB with the mechanism-based inactivator 2FDNPC has confirmed the identity of Glu-120 as the nucleophile. The uncompetitive inhibition behaviour displayed by both 2FDNPC and the substrate DNPC contrasts with the transglycosylation activity more commonly observed with retaining glycosidases and might have a unique structural origin. Moreover the nucleophile sequence E(I/L)M(I/V)W is highly conserved in family-12 endoglucanases, which is consistent with the important role of this residue in catalysis.

We thank Bernard Henrissat (CNRS-CERMAV, Grenoble, France) for his assistance in compiling and aligning the family-12 sequences, and Gideon Davies (University of York, Heslington, York, U.K.) for his assistance in preparing the manuscript. This work was supported by grants from the Protein Engineering Network Centres of Excellence of Canada (PENCE), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Fonds pour la formation de chercheurs et l'aide à la recherche du Québec (FCAR).

REFERENCES

- Koshland, D. E. (1953) *Biol. Rev.* **28**, 416–436
- Wittmann, S., Shareck, F., Kluepfel, D. and Morosoli, R. (1994) *Appl. Environ. Microbiol.* **60**, 1701–1703
- 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
- Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K. and Bott, R. (1993) in *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Cellulases*, vol. 8 (Suominen, P. and Reinikainen, T., eds.), pp. 153–158, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland
- Törrönen, A., Kubicek, C. P. and Henrissat, B. (1993) *FEBS Lett.* **321**, 135–139
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. and Warren, R. A. J. (1991) *Microbiol. Rev.* **55**, 303–315
- Davies, G. and Henrissat, B. (1995) *Structure* **3**, 853–859
- Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. and Davies, G. J. (1997) *Biochemistry* **36**, 16032–16039
- Schou, C., Rasmussen, G., Kalltoft, M.-B., Henrissat, B. and Schülein, M. (1993) *Eur. J. Biochem.* **217**, 947–953
- Kempton, J. B. and Withers, S. G. (1992) *Biochemistry* **31**, 9961–9969
- Sinnott, M. L. (1990) *Chem. Rev.* **90**, 1171–1202
- Howard, S., He, S. and Withers, S. G. (1998) *J. Biol. Chem.* **273**, 2067–2072
- McCart, J. D., Yeung, J. D., Chow, J., Dolphin, D. and Withers, S. G. (1997) *J. Am. Chem. Soc.* **119**, 5792–5797
- McCart, J. D. and Withers, S. G. (1996) *J. Am. Chem. Soc.* **118**, 241–242
- Tull, D., Miao, S., Withers, S. G. and Aebersold, R. (1995) *Anal. Biochem.* **224**, 509–514
- Miao, S., Ziser, L., Aebersold, R. and Withers, S. G. (1994) *Biochemistry* **33**, 7027–7032
- Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J. and Yaguchi, M. (1994) *Protein Sci.* **3**, 467–475
- Mackenzie, L. F., Wang, Q., Warren, R. A. J. and Withers, S. G. (1998) *J. Am. Chem. Soc.* **120**, 5583–5584
- Gill, S. C. and von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326
- Fig.P Software Corporation (1995) Fig.P, version 2.7, Biosoft, Cambridge
- Malet, C. and Planas, A. (1997) *Biochemistry* **36**, 13838–13848
- Day, A. G. and Withers, S. G. (1985) *Biochem. Cell Biol.* **64**, 914–922
- Day, A. G. (1985) M.Sc. Thesis, Department of Chemistry, University of British Columbia, Vancouver
- McCart, J. and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4**, 885–892
- Saarialhti, H. T., Henrissat, B. and Palva, E. T. (1990) *Gene* **90**, 9–14
- Namchuk, M. N. and Withers, S. G. (1995) *Biochemistry* **34**, 16194–16202
- Gebler, J. C., Aebersold, R. and Withers, S. G. (1992) *J. Biol. Chem.* **267**, 11126–11130
- Mackenzie, L. F., Davies, G. J., Schülein, M. and Withers, S. G. (1997) *Biochemistry* **36**, 5893–5901
- Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aebersold, R. and Withers, S. G. (1993) *J. Biol. Chem.* **268**, 14096–14102
- Sinnott, M. L. (1978) *FEBS Lett.* **94**, 1–9
- Street, I. P., Kempton, J. B. and Withers, S. G. (1992) *Biochemistry* **31**, 9970–9978
- Vemuri, S. and Lee, T. D. (1992) *MacProMass*, version 2.0, Beckman Research Institute of the City of Hope, Duarte, CA
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
- Lawson, S. L., Wakarchuk, W. W. and Withers, S. G. (1997) *Biochemistry* **36**, 2257–2265
- Lawson, S. L., Wakarchuk, W. W. and Withers, S. G. (1996) *Biochemistry* **35**, 10110–10118
- MacLeod, A. M., Tull, D., Rupitz, K., Warren, R. A. J. and Withers, S. G. (1996) *Biochemistry* **35**, 13165–13172
- Wang, Q., Trimbur, D., Graham, R., Warren, R. A. J. and Withers, S. G. (1995) *Biochemistry* **34**, 14554–14562
- Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J. and Withers, S. G. (1994) *J. Am. Chem. Soc.* **116**, 11594–11595