Identification of the Catalytic Nucleophile of Endoglucanase I from *Fusarium oxysporum* by Mass Spectrometry†

Lloyd F. Mackenzie,‡ Gideon J. Davies,§ Martin Schülein,|| and Stephen G. Withers*,‡

Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada, Department of Chemistry, University of York, Heslington, York Y01 5DD, U.K., and Novo Nordisk A/S, Novo Allé, DK 2880 Bagsvaerd, Denmark

Received December 3, 1996⊗

ABSTRACT: The endoglucanase EG I from *Fusarium oxysporum* catalyzes the hydrolysis of cellulose via a double-displacement mechanism involving the formation and hydrolysis of a glycosyl−enzyme intermediate. Treatment of EG I with 2′,4′-dinitrophenyl-2-deoxy-2-fluoro-β-D-cellobioside results in the time-dependent inactivation of the enzyme ($k_i = 1.36$ min^{-1}, $K_i = 0.88$ mM) via trapping of a covalent 2-deoxy-2-fluorocellobiosyl−enzyme intermediate. This intermediate is, however, catalytically competent undergoing transglycosylation, thus reactivation, in the presence of D-cellobiose. Analysis of a peptic digest of the inactivated enzyme by HPLC/ESMS/MS in the neutral loss mode allowed identification of a 2-fluorocellobiosyl-labeled peptide containing Glu197. This was confirmed by comparative mapping studies and subsequent Edman degradation analysis. This residue is completely conserved in glycosidase family 7, to which EG I belongs, consistent with its key role as the catalytic nucleophile.

Hydrolysis of cellulose by certain fungi and bacteria is achieved by use of a variety of cellulolytic enzymes which in fungi are free species but in some bacteria are organized into a complex dubbed the cellulosome (Bayer et al., 1985; Tomme et al., 1995). Regardless of the source, however, individual cellulases often have a modular structure comprising a catalytic core plus nonhydrolytic domains such as those implicated in cellulose binding. On the basis of sequence alignments, the catalytic core domains of cellulases have been assigned within the classifications of glycosyl hydrolases and, indeed, are found in 12 of the 57 families (Henrissat, 1991; Henrissat & Bairoch, 1993).

The EG I[1] from the thermophilic fungus *Fusarium oxysporum* is found in glycosyl hydrolase family 7 and comprises only a catalytic core. It displays a high degree of sequence similarity to the other catalytic domains of endoglucanases and cellobiohydrolases in this family and particularly (57% identity) to the *Humicola insolens* EG I. This family 7 of glycosyl hydrolases has been shown to cleave β-1,4-glycosidic bonds with net retention of the anomeric configuration (Knowles et al., 1988; Schou et al., 1993). The generally accepted mechanism originally proposed by Koshland (1953) for retaining glycosyl hydrolases involves a double-displacement reaction in which a covalent glycosyl−enzyme intermediate is formed and subsequently hydrolyzed via oxocarbenium ion-like transition states (Figure 1). The mechanism requires the presence of two carboxyl-containing amino acids: one acting as an acid/base catalyst and the other as a nucleophile responsible for formation of the covalent intermediate (McCarter & Withers, 1994; Sinnott, 1990). The acid/base catalyst and catalytic nucleophile of EG I have been proposed by Davies *et al.* (1996) to be Glu202 and Glu197, respectively, on the basis of the recently solved crystal structure (Sulzenbacher et al., 1997) and the similarity with *Trichoderma reesei* CBH1 (Divne et al., 1994). Further support for the proposed identification of these residues comes from the related family 16 endo-1,3(4)-glucanases, which display low sequence homology but share overall structure similarity, including conservation of the catalytic residues equivalent to Glu202 and Glu197 of EG I (Keitel et al., 1993).

Interest in the mechanism of these enzymes has resulted in the generation of probes which are able to selectively label active sites and allow for the identification of key residues (Withers & Aebersold, 1995). Of the many types of compounds available, the 2-deoxy-2-fluoro glycosides have proven to be the most successful in labeling of active site nucleophiles in a number of retaining glycosidases and glycanases (Withers & Aebersold, 1995). These compounds function as mechanism-based inactivators, trapping the enzyme by the accumulation of a relatively stable but catalytically viable 2-deoxy-2-fluoroglycosyl−enzyme intermediate. Accumulation of the intermediate is a consequence of slowing of both the glycosylation and deglycosylation steps by the C-2 fluorine coupled with the presence of a good leaving group (either fluoride or 2,4-dinitrophenol) which accelerates only the first step, glycosylation (Street et al., 1992; Withers et al., 1988). Subsequent analysis of the proteolytic digest of the labeled enzyme by electrospray

† This work was supported by funds from the Protein Engineering Network of Centres of Excellence of Canada and the Natural Sciences and Engineering Research Council of Canada. G.J.D. is a Royal Society University Research Fellow.

* Author to whom correspondence should be addressed. Telephone: 604-822-3402. Fax: 604-822-2847. E-mail: withers@chem.ubc.ca.

‡ University of British Columbia.

§ University of York.

|| Novo Nordisk A/S.

⊗ Abstract published in *Advance ACS Abstracts,* April, 1; 1997.

[1] Abbreviations: BSA, bovine serum albumin; BTC, benzyl thio-β-D-cellobioside; EG I, *Fusarium oxysporum* endoglucanase I; 2ClDNPC, 2′,4′-dinitrophenyl-2-deoxy-2-chloro-β-D-cellobioside; DNP, 2,4-dinitrophenolate; DNPC, 2,4-dinitrophenyl-β-D-cellobioside; 2FDNPC, 2′,4′-dinitrophenyl-2-deoxy-2-fluoro-β-D-cellobioside; LC, liquid chromatography; MS, mass spectrometry; TIC, total ion current; TFA, trifluoroacetic acid.

S0006-2960(96)02962-5 CCC: $14.00

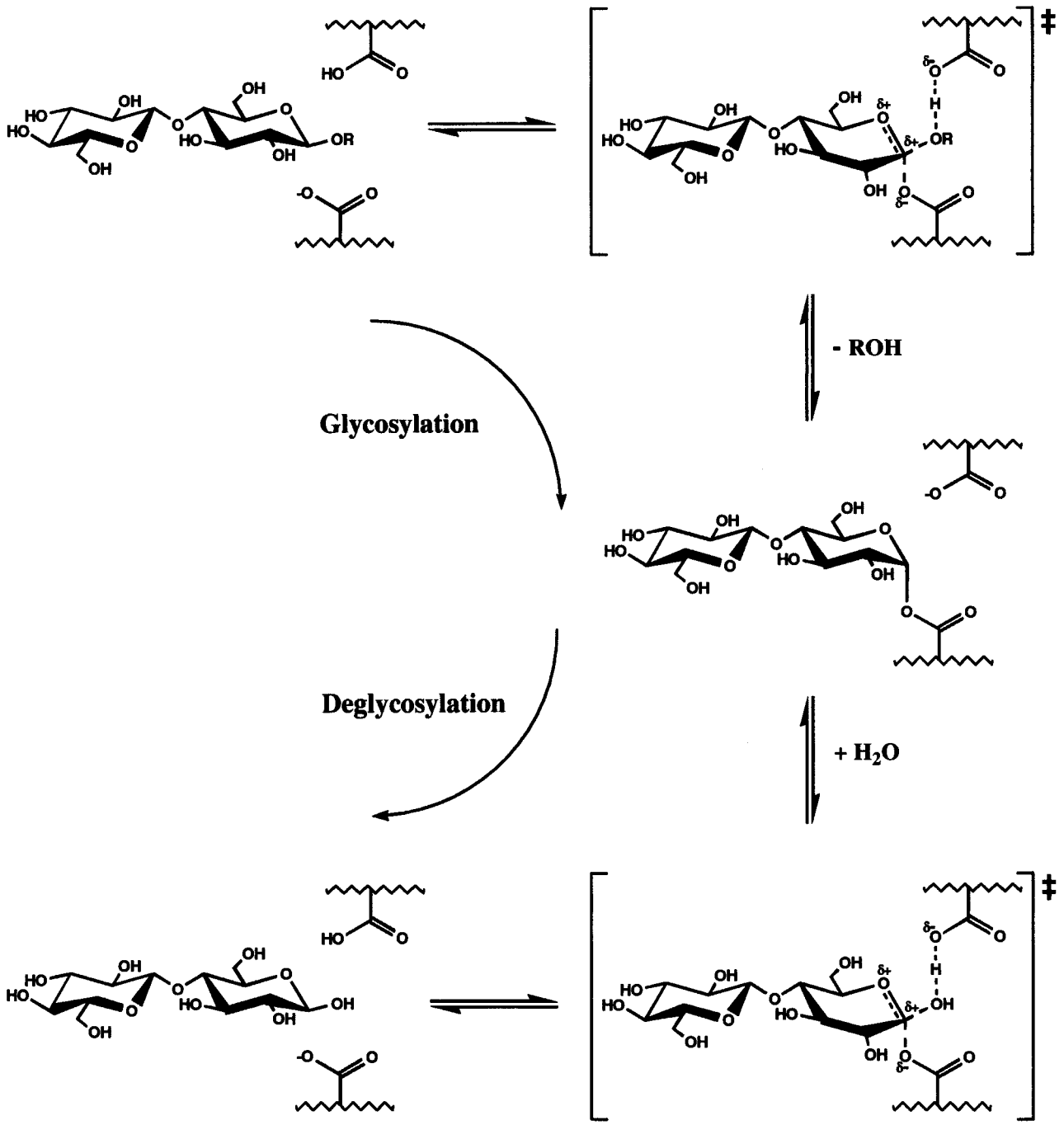

FIGURE 1: Proposed mechanism for a retaining β-endoglucanase.

mass spectrometry has permitted the identification of active site nucleophiles in several glycosidases (Mackenzie et al., 1997; McCarter & Withers, 1996; Miao et al., 1994a,b; Tull et al., 1995).

In this paper, we report the use of 2′,4′-dinitrophenyl-2-deoxy-2-fluoro-β-D-cellobioside and 2′,4′-dinitrophenyl-2-deoxy-2-chloro-β-D-cellobioside to inactivate endoglucanase 1 from *F. oxysporum*. In addition, we describe the use of these reagents in conjunction with electrospray mass spectrometry for the identification of the key catalytic nucleophile.

MATERIALS AND METHODS

General. Buffer chemicals and other reagents were obtained from Sigma Chemical Co. unless otherwise noted. *F. oxysporum* EG I was expressed in a recombinant *Aspergillus* system, essentially as outlined (Christensen et al., 1988), and purified as described in the following paper by Sulzenbacher *et al*. (1997). The substrate 2,4-dinitrophenyl-β-D-cellobioside (DNPC) was synthesized as previously described (Tull & Withers, 1994). The inactivators 2′,4′-dinitrophenyl-2-deoxy-2-fluoro-β-D-cellobioside (2FDNPC) and 2′,4′-dinitrophenyl-2-deoxy-2-chloro-β-D-cellobioside (2ClDNPC) and the competitive inhibitor benzyl thio-β-D-cellobioside (BTC) were prepared from the corresponding monosaccharides, and their synthesis will be described elsewhere (L. Mackenzie, Q. P. Wang, and S. G. Withers, unpublished result). Pepsin (from porcine gastric mucosa) was obtained from Boehringer Mannheim.

Kinetics. 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 hy-

drolysis of DNPC was used to monitor enzyme activity by measurement of the rate of 2,4-dinitrophenolate release (λ = 400 nm, ϵ = 10 910 M^{-1} cm^{-1}), using a Unicam UV4 UV-Visible spectrophotometer equipped with a circulating water bath. Michaelis−Menten parameters for the substrate, previously untested with this enzyme, were determined using a range of DNPC concentrations from 0.1 to 2.4 mM.

Inactivation Kinetics. The inactivation of EG I by either 2FDNPC or 2ClDNPC was monitored by incubation of the enzyme (0.010 mg/mL) under the above conditions in the presence of various concentrations of the inactivators (0−0.5 mM). Residual enzyme activity was determined at appropriate time intervals by addition of an aliquot (10 μL) of the inactivation mixture to a solution of DNPC (0.29 mM, 800 μL) 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 the equation (I = 2FDNPC or 2ClDNPC)

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

Reactivation Kinetics. Reactivation of the 2-fluorocellobiosyl- or 2-chlorocellobiosyl-inactivated enzyme was studied as follows. Enzyme (250 μL, 0.041 mg/mL inactivated as above, but in the absence of BSA) was concentrated using 10 kDa nominal cutoff centrifugal concentrators (Amicon Corp., Danvers, MD) to a volume of approximately 50 μL and then diluted with 250 μL of buffer without BSA. This was repeated twice, and the retentate was diluted to a final volume of 250 μL 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 cellobiose (19.5 mM) or BTC at various concentrations (0−34 mM). Reactivation was monitored by removal of aliquots (10 μL) 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, $k_{obs,react}$, for each reactivator concentration was determined from a direct fit of the activity (full rate minus observed rate) *vs* time data to a first-order equation. The reactivation rate constant (k_{react}) and dissociation constant (K_{react}) for the reactivation were determined by using Grafit (Leatherbarrow, 1990) to fit values of $k_{obs,react}$ to the following equation:

$$k_{obs,react} = \frac{k_{react}[BTC]}{K_{react} + [BTC]} \quad (2)$$

Mass Spectrometry. Mass spectra were recorded using a PE-Sciex API 300 (LC/MS experiments) triple-quadrupole mass spectrometer (Sciex, Thornhill, ON) equipped with an Ionspray ion source. Peptides were separated by reverse phase HPLC on an Ultrafast Microprotein Analyzer (Michrom BioResources Inc., Pleasanton, CA) directly interfaced with the mass spectrometer. In each of the MS experiments, the proteolytic digest was loaded onto a C18 column (Reliasil, 1 × 150 mm) and then eluted with a gradient of 0 to 60% solvent B over 30 min (for neutral loss experiments) or 60 min (for LC/MS experiments) followed by 100% B over 10 min at a flow rate of 50 μL/min (solvent A, 0.05% trifluoroacetic acid and 2% acetonitrile in water; solvent B, 0.045% trifluoroacetic acid and 80% acetonitrile in water). A post-column splitter was present in all experiments unless stated otherwise, splitting off 70% of the sample into a fraction collector and sending 30% into the mass spectrometer. Spectra were obtained in either the single-quadrupole scan mode (LC/MS) or tandem MS neutral loss mode.

Labeling of EG I. EG I (20 μL, 4.1 mg/mL, in 100 mM sodium phosphate buffer at pH 7.5) was incubated with 2FDNPC or 2ClDNPC (2 μL, 10 mM) at 40 °C for 60 min. Complete inactivation, >99%, was confirmed by assaying with DNPC as described earlier. This mixture was immediately digested with pepsin as described below.

Pepsin Digest Conditions. EG I (20 μL, native or labeled, ∼4 mg/mL) was mixed with 100 mM phosphate buffer at pH 2 (40 μL) and pepsin (20 μL, 0.5 mg/mL in 100 mM phosphate buffer at pH 2). Mixtures were incubated at room temperature for 120 min and then immediately analyzed by ESMS or frozen until required.

ESMS Analysis of the Proteolytic Digest. For LC/MS experiments, the quadrupole mass analyzer of an API 300 triple-quadrupole mass spectrometer was scanned over a *m/z* range of 300−2500 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. In the neutral loss scanning mode, MS/MS spectra were obtained by searching for the mass loss of *m/z* 327, 163.5, or 109, corresponding to the loss of the 2-fluorocellobiosyl label from a peptide ion in the singly, doubly, or triply charged states. Thus, the conditions were as follows: scan range, *m/z* 400−1800; 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; and CAD = 1 (collision gas is N_2).

Reduction/Oxidation of Disulfide Bonds. Cleavage of the disulfide bonds present in the digested protein [EG I contains 18 cysteine residues (Sulzenbacher et al., 1997), all of which form disulfide bonds] was achieved by either reduction with DTT or oxidation with performic acid as follows. A sample (25 μL) of the 2-fluorocellobiosyl-labeled crude digest or purified peptide was treated either with 10 mM DTT and 200 mM sodium phosphate at pH 7.5 (100 μL) for 15 min at room temperature and then acidified to pH ∼2 with 50% TFA or with a freshly prepared solution of performic acid (Hirs, 1967) (50 μL) for 15 min at 0 °C. The samples were then immediately analyzed by ESMS as described earlier, except that no splitter was used so that sensitivity was maximized.

Aminolysis of the 2-Fluorocellobiosyl-Labeled Peptide. To a sample (20 μL, 1.0 mg/mL) of the purified 2-fluorocellobiosyl-labeled peptide was added concentrated ammonium hydroxide (5 μL). The mixture was incubated for 15 min at 50 °C, acidified with 50% TFA, and analyzed by ESMS as described above.

Chemical Sequencing. Purification of the 2-fluorocellobiosyl-labeled EG I peptide was achieved by HPLC separation of the peptic digest as described above and by collecting the appropriate fractions containing the labeled peptide via a postcolumn splitter. The amino acid sequence of the labeled peptide was determined by S. Perry of NAPS, UBC, using standard pulsed liquid phase protocols and instrumen-

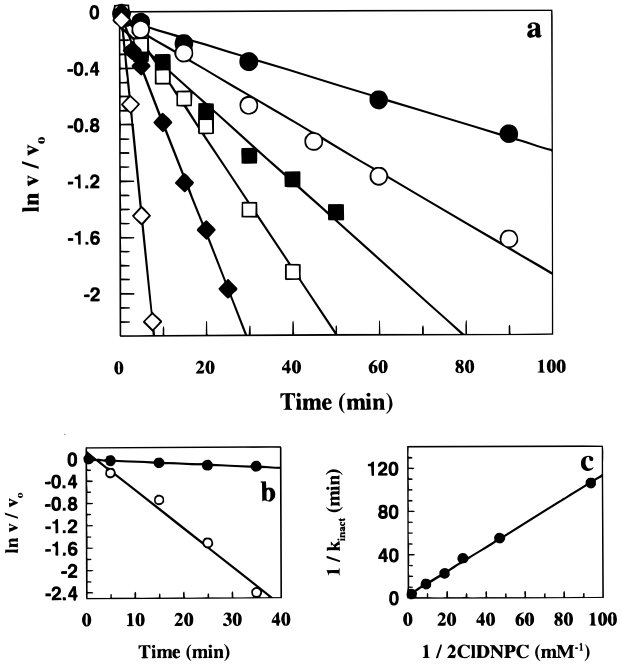

FIGURE 2: Inactivation of *F. oxysporum* EG I by 2ClDNPC. (a) Semilogarithmic plot of residual activity *vs* time at the indicated inactivator concentrations: (●) 0.011 mM, (○) 0.021 mM, (■) 0.036 mM, (□) 0.053 mM, (◆) 0.106 mM, and (◇) 0.53 mM (b) Inactivation with 0.106 mM 2ClDNPC in the (○) absence and (●) presence of 0.51 mM BTC. (c) Replot of the first-order rate constants from panel a.

tation on a Perkin-Elmer model 476A sequencer and model 120A PTH analyzer (Applied Biosystems, Foster City, CA) as described previously (Wang et al., 1993).

RESULTS AND DISCUSSION

Inactivation Kinetics. Incubation of endoglucanase 1 with either 2FDNPC or 2ClDNPC resulted in the inactivation of the enzyme in a time-dependent manner according to pseudo-first-order kinetics. The data for the inactivation of EG I with 2ClDNPC are shown in Figure 2a. The observed rates of inactivation for 2FDNPC and 2ClDNPC were shown to be dependent on the inactivator concentrations in a saturable manner, permitting determination of the inactivation rate constant ($k_i = 1.36 \pm 0.01$ min^{-1}; $k_i = 0.70 \pm 0.01$ min^{-1}) and the dissociation constant ($K_i = 0.88 \pm 0.04$ mM; $K_i = 1.19 \pm 0.06$ mM) for the 2-fluoro and 2-chloro analogs, respectively. Thus, inactivation is somewhat faster with 2FDNPC, but the two compounds bind with similar affinity. Incubation with 2ClDNPC (0.106 mM) in the presence of a competitive inhibitor BTC (0.51 mM; $K_i = 0.04$ mM) reduced the apparent rate constant for inactivation from 0.059 $\pm$ 0.004 to 0.0042 $\pm$ 0.0004 min^{-1} (Figure 2b). A similar reduction was observed for inactivation by 2FDNPC. 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* β-glucosidase (Withers & Street, 1988; Withers et al., 1990), that the inactivation is a consequence of the accumulation of a relatively stable 2-deoxy-2-halo-α-D-cellobiosyl− enzyme intermediate.

Further evidence that this inactivation occurs via stabilization and trapping of the normal intermediate in catalysis was obtained by demonstration of its catalytic competence.

Table 1: Kinetic Parameters

	inactivation			reactivation		
	k_i (min^{-1})	K_i (mM)	k_i/K_i (mM^{-1} min^{-1})	k_{spon} (min^{-1})[a]	k_{trans} (min^{-1})[b]	K_{trans} (mM)[b]
2FDNPC	1.36	0.88	1.55	0.000 11	−	−
2ClDNPC	0.70	1.19	0.588	0.000 17	0.016	7.5

[a] Spontaneous reactivation in 100 mM sodium phosphate at pH 7.5. [b] Reactivation carried out in the presence of BTC.

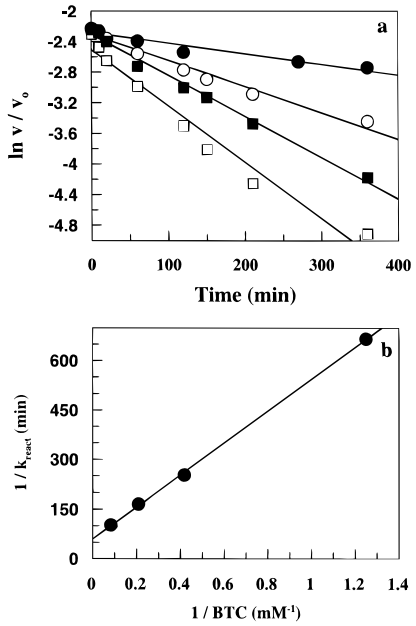

FIGURE 3: Reactivation of 2-deoxy-2-chlorocellobiosyl EG I by BTC. (a) Semilogarithmic plot of activity *vs* time for (●) 0.80 mM, (○) 2.4 mM, (■) 4.8 mM, and (□) 12.0 mM BTC. (b) Double-reciprocal replot of first-order rate constants from panel a *vs* BTC concentration.

Following removal of excess inactivator from the labeled EG I, 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 0.000 11 min^{-1} ($t_{1/2} = 6300$ min) and 0.000 17 min^{-1} ($t_{1/2} = 4100$ min) for the 2-fluorocellobiosyl- and 2-chlorocellobiosyl− enzymes, respectively (Table 1). It is of interest to compare these values, which reflect deglycosylation, with those for the substrates DNPC and D-cellotriose ($k_{cat} = 81 \pm 4$ s^{-1} and $K_m = 0.123 \pm 0.005$ mM, and $k_{cat} = 12 \pm 1$ s^{-1} and $K_m = 0.4 \pm 0.04$ mM, respectively). Although the rate-determining step for DNPC is not known, it is likely to be the deglycosylation step, which in any case must have a rate constant of at least 81 s^{-1}. On this basis, replacement of the 2-hydroxyl by fluorine or chlorine slows the deglycosylation step by at least (2×10^7)-fold. Such large reductions in rates by fluorine substitution have been seen in several other glycosyl hydrolases (McCarter et al., 1992; Miao et al., 1994b; Wang et al., 1993). Addition of D-cellobiose (19.5 mM) to the reactivation mixtures resulted in a greatly increased rate of regeneration of free enzyme, with pseudo-first-order rate constants of 0.014 $\pm$ 0.002 min^{-1} ($t_{1/2} = 50$ min) and 0.0048 min^{-1} ($t_{1/2} = 140$ min) being determined for the 2-chlorocellobiosyl− and 2-fluorocellobiosyl− enzymes, respectively. This increase in the rate of reactivation in the presence of D-cellobiose (~100-fold) suggests, as seen with other enzymes (McCarter & Withers, 1996;

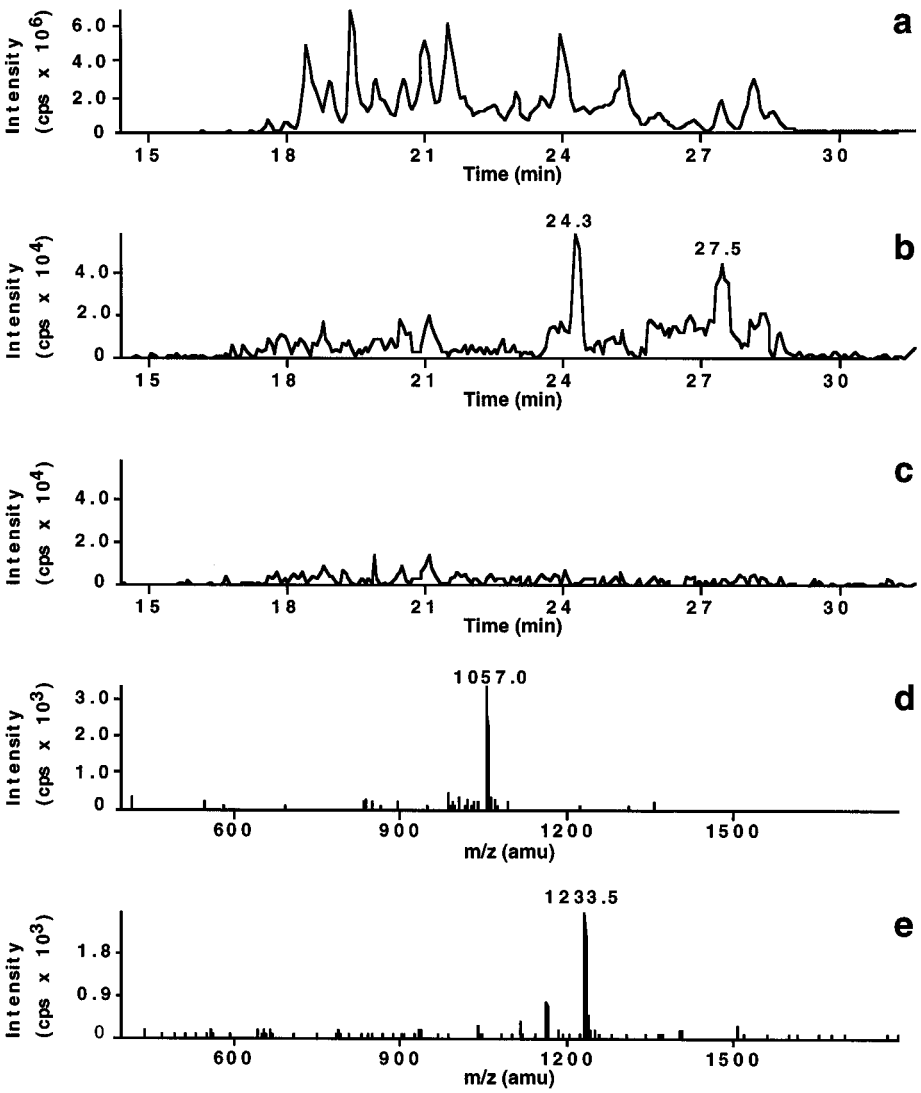

FIGURE 4: ESMS experiments on peptic digest of EG I (a) labeled with 2FDNPC, TIC in normal MS mode, (b) labeled with 2FDNPC, TIC in neutral loss mode, and (c) unlabeled, in neutral loss mode. (d) Mass spectrum of peptide at 24.3 min. (e) Mass spectrum of peptide at 27.5 min.

Miao et al., 1994b; Staedtler et al., 1995; Wang et al., 1993), that reactivation is significantly accelerated by transglycosylation to an acceptor sugar. The reactivation process was investigated in more detail by determining rate constants for reactivation of the 2-chlorocellobiosyl−enzyme in the presence of a range of concentrations of benzyl thio-β-D-cellobioside (BTC), a nonhydrolyzable glycosyl acceptor. The reactivation process followed pseudo-first-order kinetics at each BTC concentration, with the rate being dependent on the acceptor (BTC) concentration in a saturable manner (Figure 3a). A direct fit of k_{obs} vs BTC concentration, as described in Materials and Methods, yielded kinetic parameters for the reactivation process of $k_{react} = 0.016 \pm 0.004$ min^{-1} and $K_{react} = 12.5 \pm 0.9$ mM. This is illustrated in Figure 3b, as a double-reciprocal replot. Therefore, reactivation by transglycosylation occurs some 100-fold faster than by spontaneous hydrolysis, which is consistent with the ability of the intermediate to be intercepted by glycosyl acceptors, as is seen in the normal reaction catalyzed by retaining glycosyl hydrolases (Boons, 1996; Singh et al., 1995; Toone et al., 1989).

Identification of the Labeled Active Site Peptide by ESMS. Peptic digestion of the EG I inactivated with either 2FDNPC or 2ClDNPC 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 (TIC) displayed a large number of peaks which arise from every peptide in the digest (Figure 4a). Identification of the peptide(s) bearing the 2-halocellobiosyl 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 inactivator 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 sugar of known mass. Therefore, if 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. When the spectrometer was scanned in the neutral loss tandem MS/MS mode, a search for the mass loss *m/z* 327 (corresponding to the loss of the 2-fluorocellobiosyl label) from a singly charged peptide in the labeled digest mixture revealed no peaks that were not also present in a control digest (unlabeled). The spectrometer was then set up to detect a neutral loss of mass *m/z* 163.5, corresponding to the loss of the label from a doubly charged species, and again no peaks were detected. However, when the mass spec-

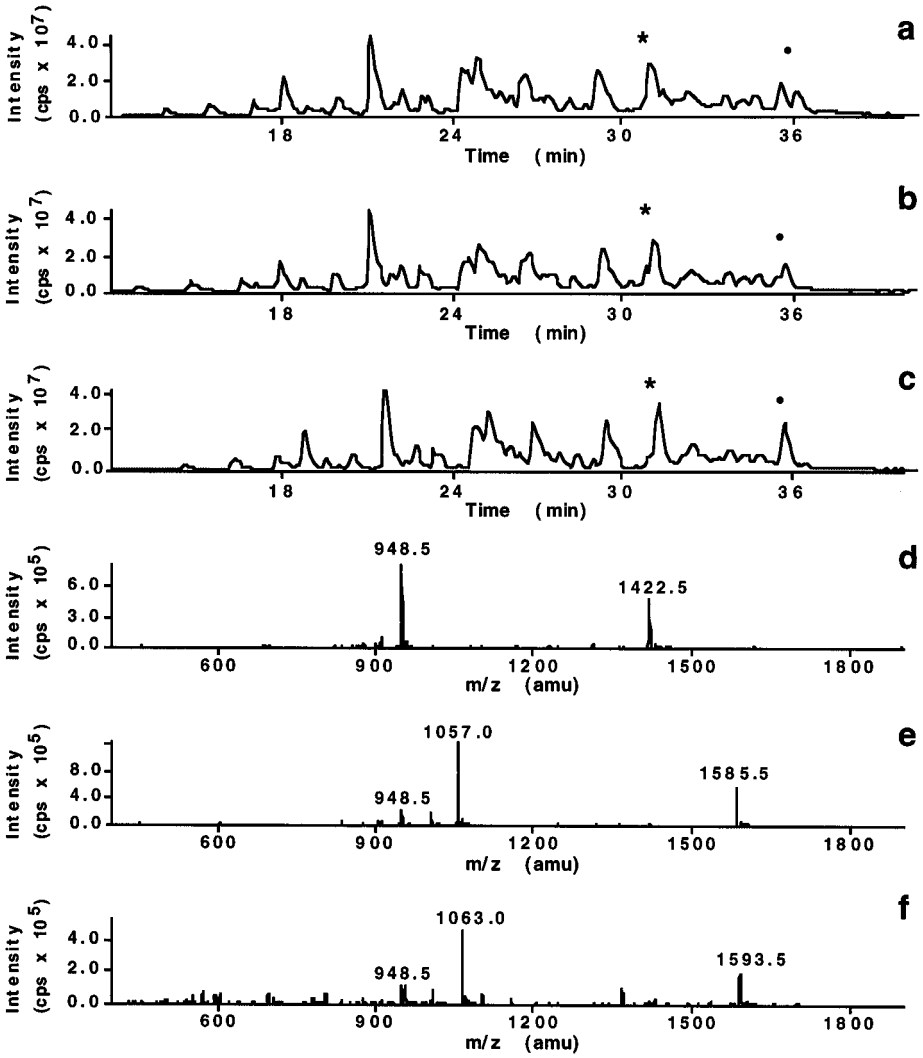

FIGURE 5: Comparative mapping ESMS experiments on peptic digest of EG I (TIC in normal MS mode) (a) unlabeled, (b) labeled with 2FDNPC, and (c) labeled with 2ClDNPC. (d−f) Mass spectra of peptides at 31.1 min for the unlabeled, 2-fluorocellobiosyl-labeled, and 2-chlorocellobiosyl-labeled peaks, respectively.

trometer was scanned using *m/z* 109 (neutral loss corresponding to a triply charged species), two peaks were detected at 24.3 and 27.5 min (Figure 4b) that were absent in the control experiment (Figure 4c). The triply charged peptides were measured at *m/z* 1057 and 1233.5 (Figure 4d,e), thus corresponding to two labeled peptides with masses of 3168 Da [(3 × 1057) − 3H] and 3697.5 Da [(3 × 1233.5) − 3H]. Since the mass of the 2-fluorocellobiosyl label is 327, the unlabeled peptides must therefore have molecular masses of 2842 Da (peptide A) (3168 − 327 + 1 Da) and 3371.5 Da (peptide B) (3697.5 − 327 + 1 Da).

Comparative mapping of the LC/MS profiles provided further evidence for the identity of the labeled peptides. Peptic digests of each of the following—unlabeled, 2FDNPC-labeled, and 2ClDNPC-labeled EG I—were separated by HPLC using a long gradient (60 min) and were scanned in LC/MS mode. The three digests showed very similar profiles with the exception of differences seen at 31.1 and 35.8 min (Figure 5a−c). The peak at 31.1 min in the unlabeled EG I digest contained a series of masses arising from a doubly charged peptide at *m/z* 948.5 Da and a triply charged peptide at *m/z* 1422.5 Da (Figure 5d), corresponding to a peptide with a molecular mass of 2842 Da [(948.5 × 3) − 3H and (1422 × 2) − 2H]. A similar series of doubly and triply charged peptides was seen in digests of the 2-fluorocellobiosyl- and 2-chlorocellobiosyl-labeled proteins, corresponding to neutral labeled peptides of 3168.5 and 3185.5 Da, respectively (see Figure 5e,f). Similar results obtained for the peptides seen at 36 min (data not shown) are summarized in Table 2 along with data from 31.1 min. The mass difference between the labeled peptides and that of the unlabeled peptides was determined to be 326 ± 2 and 344 ± 2 Da for the 2-fluorocellobiosyl- and 2-chlorocellobiosyl-labeled EG I, respectively. These correspond within error to the mass of the particular label used (326 and 343 Da), thereby identifying these as the peptides of interest, as identified independently by neutral loss.

These peptides are surprisingly large, considering that this was a peptic digest. However, analysis of the amino acid sequence of the endoglucanase (Sheppard et al., 1994) revealed that there are 18 cysteine residues. All of these form disulfide bonds (Sulzenbacher et al., 1997), suggesting that the peptide in question might be cross-linked via a disulfide to another peptide. A sample of the purified 2-fluorocellobiosyl-labeled peptide A with a mass of 3168 Da was therefore reduced using dithiothreitol as described in Materials and Methods and analyzed by LC/MS. Two new peaks were detected: a singly charged species (*m/z* 1414) and a doubly charged species (*m/z* 718) which combine to give a mass of 2847 ± 2 Da. After correction for the

Table 2: Comparative Mapping ESIMS Results of 2FDNPC- and 2ClDNPC-Labeled EG I

	m/z of peaks (doubly, triply charged)		corresponding peptide mass[a] (Da)		mass difference from control (Da)	
EG I digests	at 31 min	at 36 min	at 31min	at 36 min	at 31min	at 36 min
2FDNPC	1057, 1585	1233.5, 1849.5	3168	3697	326[b]	326[b]
2ClDNPC	1063, 1593.5	1239, 1858	3186	3714	344[c]	343[c]
control	948.5, 1422	1124.5, 1686.5	2842	3371	–	–

[a] Peptide mass calculated as follows: doubly charged ($2 \times m/z - 2H$) and triply charged ($3 \times m/z - 3H$). [b] Mass of 2FC = 326 Da. [c] Mass of 2ClC = 343 Da.

Family 7	**Acc #**	* ‡
Fusarium oxysporum EG I	P46237	192-G V C C N **E** L **D** I W **E** A N S R A-207
Humicola insolens EG I	[a]	192-G S C C N **E** M **D** I W **E** V N S R A-207
Trichoderma reesei EG I	P07981	191-G F C C N **E** M **D** I L **E** G N S R A-208
Fusarium oxysporum CBH1	P46238	207-G T C C P **E** M **D** I W **E** A N S I S-222
Humicola grisea CBH1	D63516	192-G T C C S **E** M **D** I W **E** A N N M A-207
Trichoderma reesei CBH1	P00725	207-G S C C S **E** M **D** I W **E** A N S I S-222
Family 16		
Bacillus circulans Laminarinase	P23903	545-G T - S G **E** I **D** V M **E** A A R G L-561
Bacillus macerans Lichenase	P23904	98-G T Q W D **E** I **D** I - **E** F - L G K-111

[a] US Patent. 5,457,046

FIGURE 6: Alignment of the region containing the proposed catalytic nucleophile glutamate, labeled by 2FDNPC for *F. oxysporum* EG I, in selected family 7 and family 16 glycosyl hydrolases. Accession numbers correspond to those in the SWISS-PROT or GenBank data bases. Absolutely conserved residues of family 7 and family 16 enzymes are shown in bold, with the proposed active site nucleophile shown with an asterisk and the acid/base catalyst with a double dagger.

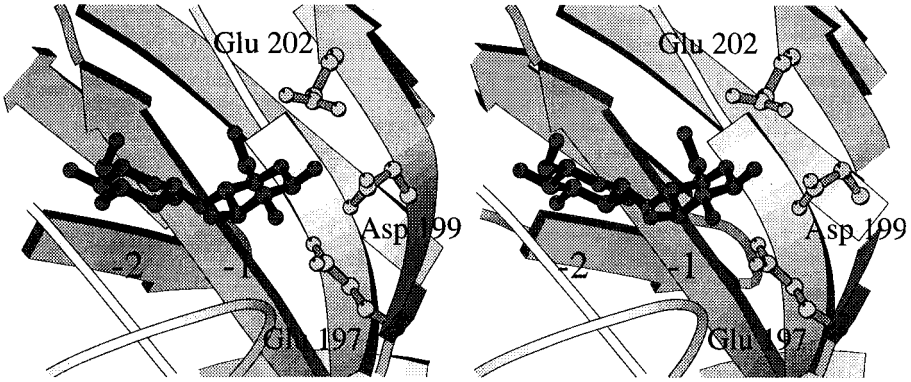

FIGURE 7: Stereodiagram of the *F. oxysporum* EG I active site showimg the location of bound D-cellobiose and the residues Glu197 (nucleophile), Glu202 (acid/base), and Asp199. For a full description of the interactions between D-cellobiose and EG I, see Sulzenbacher et al. (1997).

four hydrogens removed to reduce the disulfide bonds, a mass of 2843 Da is obtained which agrees well with that of the unlabeled peptide A of 2842 Da, indicating that the reduction with DTT at pH 7.5 resulted in the cleavage not only of the disulfide bonds but also of the labile ester linkage between the peptide and the 2-fluorocellobiosyl moiety. In order to preserve the ester linkage and cleave the disulfide bonds, an acidic oxidation was used in which the cysteine residues are fully oxidized to the corresponding cysteic acid (Hirs, 1967). Oxidation of cysteine to cysteic acid should result in a mass increase of 48 Da per cysteine residue. Analysis of the peptide after treatment with performic acid resulted in the detection of two new peaks, a singly charged species (*m/z* 1510) and a doubly charged species (*m/z* 928.5). Similar treatment of a control sample (purified unlabeled peptide with a mass of 2842 Da) also gave two new peptide fragments, this time of *m/z* 1510 and a doubly charged peptide of *m/z* 765. Comparison of the masses of the two peptide fragments from DTT treatment, 1413 and 1434 Da, with those from the control performic acid treament, 1509 and 1530 Da, indicates that each of the dipeptide halves contains two cysteine residues, as seen by a 96 Da mass increase for both. The mass difference between the control and the 2-fluorocellobiosyl-labeled performic acid-treated peptides allowed the identification of the peptide with a mass of 1855 ± 1 Da ($1855 - 1530 = 325 \pm 2$ Da) as that which contained the 2-fluorocellobiosyl label (326 Da). This in turn corresponds to a reduced, unlabeled peptide with a mass of 1433 ± 3 Da.

Computer analysis of the amino acid sequence of EG I (Sheppard et al., 1994) revealed seven candidate peptides for the mass of 1433 Da. However, since the peptide sequence should contain either an aspartate or glutamate residue [known nucleophiles of retaining glycosyl hydrolases (McCarter & Withers, 1994)] and two cysteine residues as determined from the performic acid experiments, only one

Table 3: Edman Degradation of 2F-Cell-Labeled Active Site Peptide

cycle	PTH derivative(s)		yield (pmol)		cDNA sequences	
1	Y	V	54	61	Y	V
2	G	G	99		G	G
3	A	N	49	36	A	N
4	G	I	59	50	G	I
5	Y	K	49	50	Y	K
6	C	G	−	54	C	G
7	D	Q	35	40	D	Q
8	A	G	39	40	A	G
9	Q	V	38	42	Q	V
10	C	C	−		C	C
11	Y	C	34	−	Y	C
12	V	N	20	18	V	N
13	T	E	11	3	T	E
14	−	L	−	3	−	L

candidate peptide remained, 185-VGNIKGQGVCCNEL-198. Similar analysis for the second portion of the peptide A, mass of 1413 Da, resulted in nine candidate peptides, all of which could be eliminated but two, 167-YGAGYCDAQ-CYVT-179 and 163-AGAYYGAGYCDAQC-176, from the requirement for two cysteine residues. Interestingly, these two peptides contain the same two cysteine residues (C172 and C176). Results from the crystal structure of this enzyme (Sulzenbacher et al., 1997) indicate that the cysteine pairs C172·C195 and C176·C194 form disulfides. However, since the two fragments above both contain cysteines C172 and C176, both are still possible candidates, though this does not affect the identification of the nucleophile.

The sequence of peptide A was confirmed using standard chemical sequencing. Edman degradation of the 2-fluoro-cellobiosyl-labeled peptide yielded the sequence information shown in Table 3, confirming the earlier identification of the 2-fluorocellobiosyl peptide fragment and allowing the identification of the second peptide fragment as 167-YGAGYCDAQCYVT-179. These results demonstrate that Glu197 is the essential active site nucleophile and give the amino acid sequence of peptide A, which was initially observed in the neutral loss experiments as

$$_{167}\text{-YGAGYCDAQCYVT-}_{179}$$
$$_{185}\text{-VGNIKGQGVCCNEL}_{198}$$

Similar treatment of peptide B from the neutral loss experiment (data not shown) produced two unlabeled peptide fragments with masses of 1963 ± 1 and 1413 ± 1 Da. A computer search of the amino acid sequence for a peptide with a mass of 1963 Da produced four candidate peptides, with only one containing the same two cysteine residues and the same glutamic acid residue present in peptide A, 180-PFINGVGNIKGQGVCCNEL-198. The second portion of the peptide was found to be identical to that observed for peptide A, namely 167-YGAGYCDAQCYVT-179. Therefore, the structure of peptide B is

$$_{167}\text{-YGAGYCDAQCYVT-}_{179}$$
$$_{180}\text{-PFINGVGNIKGQGVCCNEL}_{198}$$

Covalent attachment of the 2-fluorocellobiosyl label to Glu197 through an ester linkage was confirmed by treatment of the labeled peptide digest with ammonium hydroxide. After treatment, the labeled peptide with a molecular mass of 3168.5 Da (doubly charged *m/z* 1585.5 and triply charged *m/z* 1057) was replaced by one new peptide having a molecular mass of 2841 Da (doubly charged *m/z* 1421.5 and triply charged *m/z* 948). This would indicate that the labeled glutamate reacted by aminolysis, generating Gln197, thereby implicating the catalytic nucleophile as Glu197.

Further evidence for the identification of Glu197 as the active site nucleophile comes from analysis of sequence alignments. The glutamic acid equivalent to Glu197 is completely conserved in all members of family 7 and, interestingly, among the lichenases and laminarinases of family 16 (Figure 6), as would be expected if this residue has such a crucial role. Indeed, conversion of the equivalent residue in the *Bacillus macerans* lichenase to Gln by site-directed mutagenesis (Glu103Gln) resulted in a complete loss of activity (Hahn et al., 1995a). Interestingly, as has been seen in a number of other glycosidases (McCarter & Withers, 1994), there are two other highly conserved carboxylic acids, Glu202 and Asp199, in *F. oxysporum* EG I numbering which are seen in the active site (Figure 7). Glu202 has been suggested as the acid/base catalyst, largely on the basis of crystal structures of two other members of family 7, *T. reesei* CBH1 (Divne et al., 1994) and *H. insolens* EG I (Davies et al., 1996), and the structurally related family 16 lichenase from *Bacillus sp*. (Keitel et al., 1993). Consistent with this, mutation of the Glu202 equivalent residue in the *B. macerans* enzyme (Glu107Gln) resulted in complete activity loss. The other conserved residue, Asp199, appears to be less crucial since mutation of it to Asn (Asp105Asn in *B. macerans*) reduced the activity some 300-fold, but did not obliterate it (Hahn et al., 1995b).

These data, therefore, clearly confirm the role of Glu197 as the catalytic nucleophile in the *F. oxysporum* enzyme and therefore, by analogy, with the members of this family. Further insight is provided in the following paper describing the 3-dimensional structure of *F. oxysporum* EG I both free and complexed with substrate and a covalent inactivator.

ACKNOWLEDGMENT

The authors thank Karen Rupitz for assistance with enzyme kinetics and David Chow and Shouming He for technical assistance with the ESMS experiments.

REFERENCES

Bayer, E. A., Setter, E., & Lamed, R. (1985) *J. Bacteriol. 163*, 552−559.
Boons, G. J. (1996) *Tetrahedron 52*, 1095−1121.
Christensen, T. W., Boel, H. E., Mortensen, S. B., Hjortshoj, K., Thim, L., & Hansen, M. T. (1988) *Bio/Technology 6*, 1419−1422.
Davies, G. J., Tolley, S. P., Divne, C., Jones, T. A., Woldike, H., & Schülein, M. (1996) *J. Mol. Biol.* (submitted for publication).
Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Petterson, G., Knowles, J. K. C., Teeri, T., & Jones, T. A. (1994) *Science 265*, 524−528.
Hahn, M., Keitel, T., & Heinemann, U. (1995a) *Eur. J. Biochem. 232*, 849−858.
Hahn, M., Olsen, O., Politz, O., Borriss, R., & Heinemann, U. (1995b) *J. Biol. Chem. 270*, 3081−3088.
Henrissat, B. (1991) *Biochem. J. 280*, 309−316.
Henrissat, B., & Bairoch, A. (1993) *Biochem. J. 293*, 781−788.
Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 197−199.
Keitel, T., Simon, O., Borriss, R., & Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5287−5291.

Knowles, J. K. C., Lentovaara, P., Murray, M., & Sinnott, M. L. (1988) *J. Chem. Soc., Chem. Commun.*, 1401−1402.
Koshland, D. E. (1953) *Biol. Rev. 28*, 416−436.
Leatherbarrow, R. J. (1990) *GraFit*, version 3.0, Erithacus Software Ltd., Staines, UK.
Mackenzie, L. F., Brooke, G. S., Cutfield, J. F., Sullivan, P. A., & Withers, S. G. (1997) *J. Biol. Chem. 272*, 3161−3167.
McCarter, J. D., & Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
McCarter, J. D., & Withers, S. G. (1996) *J. Biol. Chem. 271*, 6889−6894.
McCarter, J. D., Adam, M., & Withers, S. G. (1992) *Biochem. J. 286*, 721−727.
Miao, S., McCarter, J. D., Grace, M., Grabowski, G., Aebersold, R., & Withers, S. G. (1994a) *J. Biol. Chem. 269*, 10975−10978.
Miao, S., Ziser, L., Aebersold, R., & Withers, S. G. (1994b) *Biochemistry 33*, 7027−7032.
Schou, C., Rasmussen, G., Kaltoft, M. B., Henrissat, B., & Schülein, M. (1993) *Eur. J. Biochem. 217*, 947−953.
Sheppard, P. O., Grant, F. J., Oort, P. J., Sprecher, C. A., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L., & O'Hara, P. J. (1994) *Gene 150*, 163−167.
Singh, S., Packwood, J., Samuel, C. J., Critchley, P., & Crout, D. (1995) *Carbohydr. Res. 279*, 293−305.
Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171−1202.
Staedtler, P., Hoenig, S., Frank, R., Withers, S. G., & Hengstenberg, W. (1995) *Eur. J. Biochem. 232*, 658−663.
Street, I. P., Kempton, J. B., & Withers, S. G. (1992) *Biochemistry 31*, 9970−9978.
Sulzenbacher, G., Schulein, M., & Davies, G. J. (1997) *Biochemistry 36*, 5902−5911.
Tomme, P., Warren, R. A. J., & Gilkes, N. R. (1995) *Adv. Microbiol. Physiol. 37*, 1−81.
Toone, E. J., Simon, E. S., Bednarski, M. D., & Whitesides, G. M. (1989) *Tetrahedron 45*, 5365.
Tull, D., & Withers, S. G. (1994) *Biochemistry 33*, 6363−6370.
Tull, D., Miao, S., Withers, S. G., & Aebersold, R. (1995) *Anal. Biochem. 224*, 509−514.
Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aebersold, R., & Withers, S. G. (1993) *J. Biol. Chem. 268*, 14096−14103.
Withers, S. G., & Street, I. P. (1988) *J. Am. Chem. Soc. 110*, 8551−8553.
Withers, S. G., & Aebersold, R. (1995) *Protein Sci. 4*, 361−372.
Withers, S. G., Rupitz, K., & Street, I. P. (1988) *J. Biol. Chem. 263*, 7929−7932.
Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B., & Aebersold, R. (1990) *J. Am. Chem. Soc. 112*, 5887−5889.

BI962962H