ELSEVIER

Carbohydrate Research 304 (1997) 143–154

CARBOHYDRATE RESEARCH

Cellobiohydrolase I from *Trichoderma reesei*: Identification of an active-site nucleophile and additional information on sequence including the glycosylation pattern of the core protein

Klaus Klarskov [a], Kathleen Piens [a], Jerry Ståhlberg [b], Peter B. Høj [c], Jozef Van Beeumen [a], Marc Claeyssens [a,*]

[a] *Department of Biochemistry, Physiology and Microbiology, University of Gent, B-9000 Gent, Belgium*

[b] *Department of Molecular Biology, University of Uppsala, Biomedical Center, S-75124 Uppsala, Sweden*

[c] *Department of Horticulture, Viticulture and Oenology, University of Adelaide, 5064 Glen Osmond, Australia*

Received 7 April 1997; accepted 15 July 1997

Abstract

(*R*,*S*)-3,4-Epoxybutyl β-cellobioside, but not the corresponding propyl and pentyl derivatives, inactivates specifically and irreversibly cellobiohydrolase I from *Trichoderma reesei* by covalent modification of Glu^{212}, the putative active-site nucleophile. The position and identity of the modified amino acid residue were determined using a combination of comparative liquid chromatography coupled on-line to electrospray ionization mass spectrometry, tandem mass spectrometry and microsequencing. It was found that the core protein corresponds to the N-terminal sequence $pyrGlu^{1}$–$Gly^{434}(Gly^{435})$ of intact cellobiohydrolase I. In the particular enzyme samples investigated, the asparagine residues in positions 45, 270 and 384 are each linked to a single 2-acetamido-2-deoxy-D-glucopyranose residue.

Keywords: Cellobiohydrolase I; *Trichoderma reesei*; Epoxybutyl cellobioside; Active-site nucleophile; Electrospray ionization mass spectrometry

1. Introduction

The mechanism of enzymatic cellulose degradation has attracted considerable interest in recent years due to its evident ecological and industrial importance. A fundamental understanding of the mode of action of the enzymes involved is required in order to optimize their role in commercial applications. In nature, certain fungi secrete large amounts of cellulose hydrolyzing enzymes. Among these, the fungus *Trichoderma reesei* produces cellobiohydrolase I (CBH I) in abundance, in addition to several other glycanases.

* Corresponding author. Fax: +32 9 264 53 32; e-mail: Marc.Claeyssens@rug.ac.be

0008-6215/97/$17.00
PII S0008-6215(97)00215-2

CBH I (EC 3.2.1.91) is an exoglucanase that hydrolyzes crystalline cellulose probably by cleavage from the chain ends [1] and has been classified as a family C/7 enzyme [2,3]. This family includes exo- and endo-glucanases that cleave the β-(1 → 4)-glycosidic bond by a double-displacement mechanism, resulting in retention of configuration of the cellobiose product [4,5]. Intact CBH I contains a catalytic domain (core protein) and a C-terminal cellulose binding region which can be separated after limited papain cleavage [6]. The core protein, which contains four potential glycosylation sites, has recently been crystallized and the structure elucidated [7]. Although two putative catalytic amino acid residues were deduced from the three-dimensional structure, direct experimental proof of their roles, as well as detailed information about the extent of glycosylation in the core protein are still lacking.

Epoxide-based glycosyl inhibitors have been used extensively as affinity labels to probe important catalytic residues in a number of glycosidases (briefly reviewed in ref. [8]). These compounds consist of a saccharide moiety, with specificity for subsites in the catalytic domain, and an epoxide group that acts as a target of nucleophile attack thereby forming a covalent ester bond with a catalytic residue.

Here, we demonstrate that CBH I is specifically and irreversibly inhibited by (*R,S*)-3,4-epoxybutyl β-cellobioside. The reaction takes place via covalent modification of a nucleophilic active-site residue. Identification of the residue was achieved by exploiting a combination of liquid chromatography coupled, on-line, to electrospray ionization mass spectrometry (LC–ESMS), tandem mass spectrometry (ESMS/MS) and microsequence analysis.

In addition, detailed information concerning the glycosylation pattern in the core protein was deduced from these results.

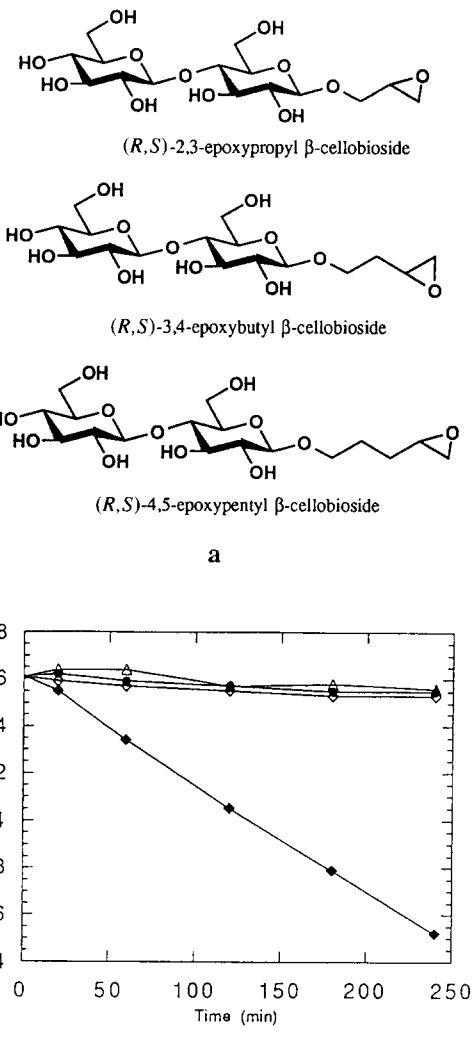

Fig. 1. Structures of affinity inhibitors (a) and inactivation of CBH I core protein (b). The enzyme (2.5 μM), in sodium acetate buffer, pH 5.0, was incubated at 37 °C with, respectively, 0.5 mM (*R,S*)-2,3-epoxypropyl (●), (*R,S*)-3,4-epoxybutyl (◆) and (*R,S*)-4,5-epoxypentyl (◇) β-cellobioside; control (no inhibitor) (△).

2. Results

Specificity of epoxyalkyl–cellobioside inactivation of CBH I core protein.—Incubation of CBH I core protein with low concentrations of (*R,S*)-3,4-epoxybutyl β-cellobioside results in rapid and effective inactivation of the enzyme while no inactivation was observed using the corresponding (*R,S*)-2,3-epoxypropyl or (*R,S*)-4,5-epoxypentyl β-cellobioside derivatives (Fig. 1). Double reciprocal replotting of the results [9] allowed determination of the inhibition constant ($K_i \cong 1$ mM). This value is considerably lower than those reported for other cellulases irreversibly inhibited with the same compound, but compares favorably with the K_i value for cellobiose (0.3 mM) [10]. This compound also effectively protects the enzyme against inactivation by (*R,S*)-3,4-epoxybutyl β-cellobioside (results not shown). In contrast to most endoglucanases investigated earlier, e.g. ref. [11], inactivation seems limited to the epoxybutyl derivative suggesting sterically restricted binding and reaction with CBH I. This was further underscored when the effect of aglycon chain length was determined for the corresponding D-glucosides and *C*-cellobiosides. While (*R,S*)-2,3-epoxybutyl β-D-

glucopyranoside is required at much higher concentrations for effective inhibition ($K_i \cong 10$ mM), the aglycon length dependency is just as strict and this undoubtedly reflects the higher affinity of CBH I for the cellobiosyl glycon. The dependence on the stereochemistry of the aglycon was also investigated but, in contrast to the 1,3-1,4-β-glucanases of bacterial and plant origin [11,12], the potencies of the (*R*)- and (*S*)-diastereoisomers of 3,4-epoxybutyl β-cellobioside were indistinguishable (data not shown). Since hydrolysis of glucosides containing an *O*-glycosyl bond can be a potential drawback when epoxy-based inhibitors are used, several *C*-glycosides were tested. In this case the epoxypentyl cellobioside was optimally inactivating (data not shown). Indeed, its aglycon chain length corresponds to this of the epoxybutyl *O*-glycosides. The pH dependence for the inactivation of CBH I with 3,4-epoxybutyl β-cellobioside

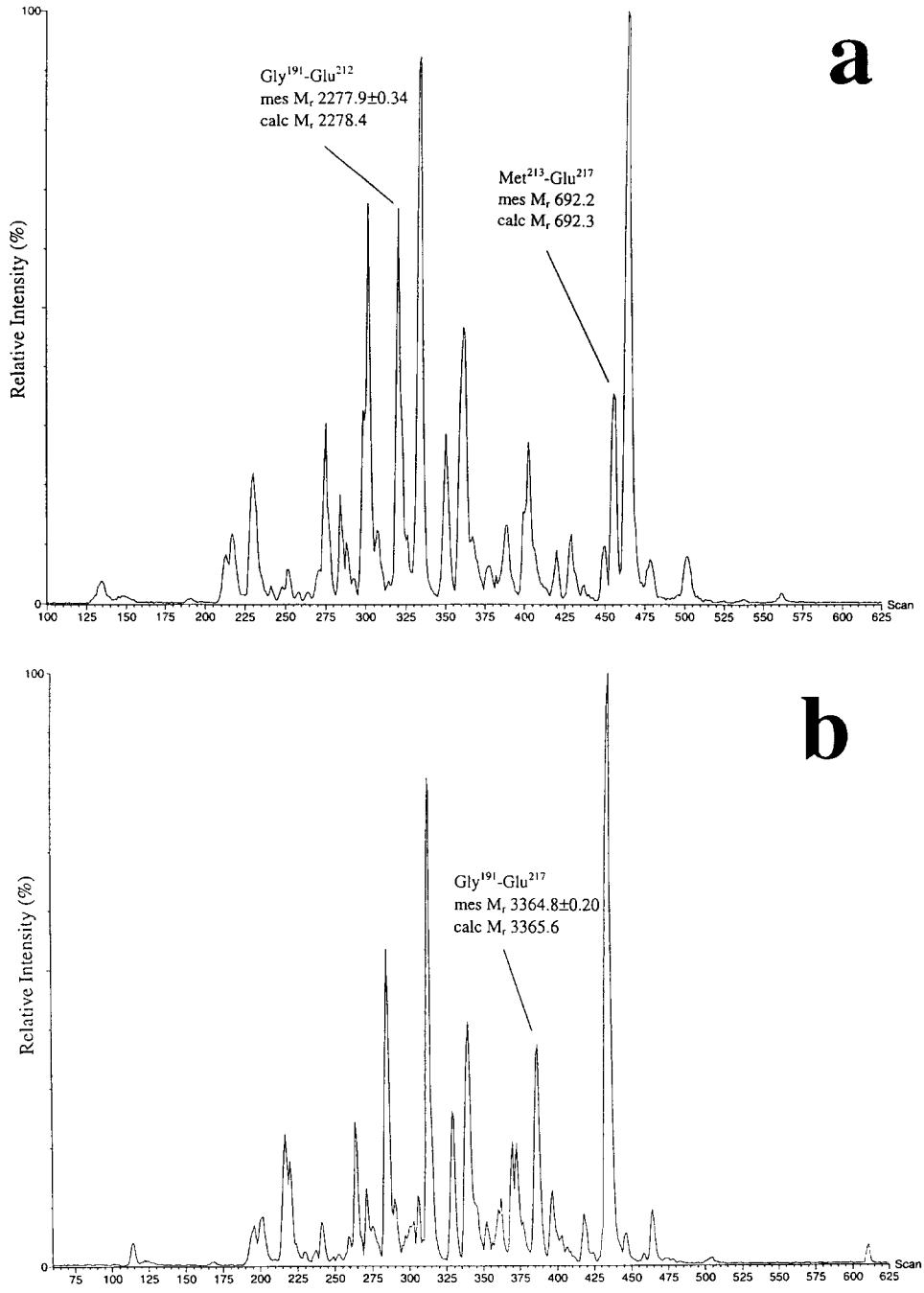

Fig. 2. Capillary LC–ESMS BPI chromatogram of peptides derived from V8 proteolysis of CBH I core protein before (a) and after (b) inactivation with (*R,S*)-3,4-epoxybutyl β-cellobioside. Approximately 10 pmol of sample were injected.

(optimum around 5, not shown) closely reflects that for the enzyme activity on 2-chloro-4-nitrophenyl β-lactoside [13].

Labeling stoichiometry.—Liquid scintillation counting of CBH I, completely inactivated with ^{14}C-labeled (*R*,*S*)-3,4-epoxybutyl β-cellobioside, revealed the incorporation of 0.8 mol of inhibitor per mol of enzyme. The stoichiometry of inactivation was also determined by ESMS (Table 1). A mass increase of 413.7 Da indicates that, under the condi-

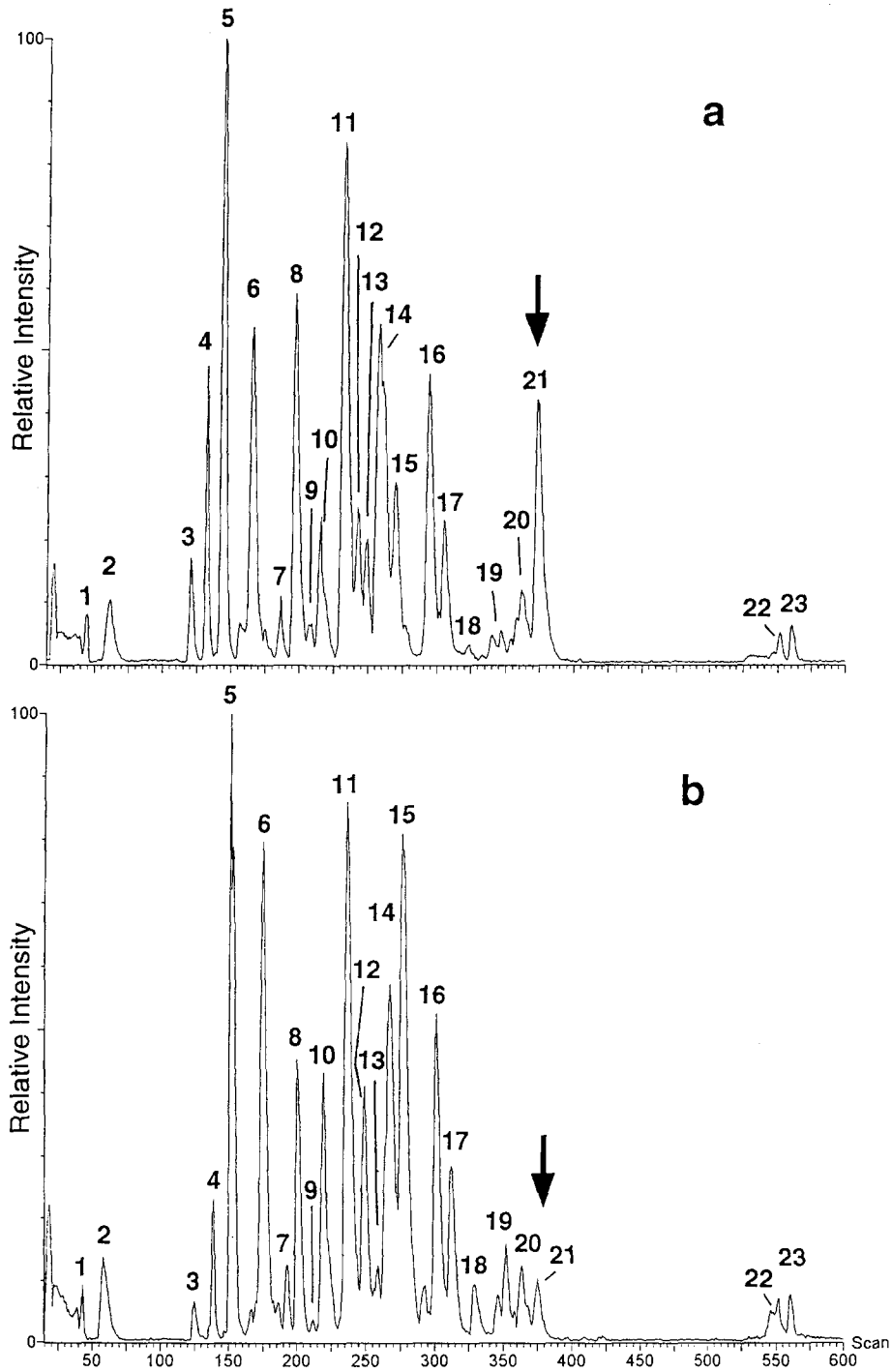

Fig. 3. Preparative LC–ESMS BPI chromatogram of peptides obtained after tryptic digestion of CBH I core protein before (a) and after (b) inactivation with (*R*,*S*)-3,4-epoxybutyl β-cellobioside. The arrows in (a) and (b) indicate the fractions containing the unlabeled peptide Cys^{210}–Cys^{230} as well as the labeled peptide Ser^{211}–Cys^{230}, respectively. Approximately 560 pmol of sample were loaded onto the column.

Table 1
Molecular weight determination of CBH I core protein by ESMS prior to and after inactivation with (*R,S*)-3,4-epoxybutyl β-cellobioside (calculated mass 412.4 Da). The two values in each column represent a major and a minor component

Observed mass		
Before	After	ΔM_r
46,576.7 ± 4.8	46,990.4 ± 3.2	413.7
46,636.1 ± 6.4	47,045.0 ± 4.5	408.9

tions used, one mol of inhibitor is incorporated per mol of CBH I core protein. These results suggest that inactivation most likely occurs at the active site and that non-specific reactions with non-catalytic nucleophilic amino acid residues can be excluded.

Identification of the active-site nucleophilic amino acid residue.—To identify the amino acid residue modified by (*R,S*)-3,4-epoxybutyl β-cellobioside, we adapted a recently proposed mass spectrometric strategy [14]. Labeled and unlabeled core protein were

Table 2
Summarized results from LC–ESMS analysis of trypsin digested CBH I core protein

HPLC fraction	Measured mass[a]	Calculated mass[b]	ΔM_r	Peptide position[c]
1	447.3	447.2	0.1	pyrGlu^{1}–Cys^{4}
2	515.4	515.3	0.1	Asn^{103}–Arg^{107}
3	420.3			not found
4	922.0	922.0	0.0	Tyr^{158}–Lys^{166}
5	1463.6 ± 0.33	1463.7	0.1	Tyr^{167}–Arg^{178}
6	2078.0 ± 0.23	2078.2	0.2	Gly^{395}–Lys^{415}
7	877.9 ± 0.13	878.3	0.4	Gly^{347}–Lys^{354}
8	2276.3 ± 0.33	2276.4	0.1	Thr^{231}–Arg^{251}
9	749.5	749.4	0.1	Gly^{347}–Lys^{353}
10	807.8 ± 0.27	807.9	0.1	Val^{416}–Lys^{422}
	1089.9	1090.2	0.3	Phe^{423}–Gly^{434}
	1146.6	1147.2	0.6	Phe^{423}–Gly^{435}
	4251.1 ± 0.45	4250.8	0.3	Gly^{139}–Arg^{178}
11	1933.2 ± 0.41	1933.1	0.1	Tyr^{252}–Arg^{267}
12	3682.8 ± 0.48	3479.8	203.0	Trp^{40}–Lys^{69}
13	3483.3 ± 0.21	3280.5	202.8	Trp^{40}–Cys^{67}
14	1665.9 ± 0.24	1665.9	0.0	Thr^{5}–Lys^{18}
	2284.3 ± 0.36	2284.6	0.3	Cys^{19}–Arg^{39}
15	2095.1 ± 0.30	2095.4	0.3	pyrGlu^{1}–Lys^{18}
16	1764.0 ± 0.19	1764.0	0.0	Lys^{287}–Arg^{302}
	2324.3 ± 0.23	2121.3	203.0	Leu^{268}–Lys^{287}
17	1635.6	1635.8	0.2	Leu^{288}–Arg^{302}
18	3429.5 ± 0.10	3429.8	0.7	Asn^{70}–Lys^{102}
19	2463.6 ± 0.10	2447.8	15.8	Cys^{210}–Cys^{230}
	4889.3 ± 0.19	4890.1	0.8	Tyr^{303}–Lys^{346}
	5041.8 ± 0.31	4837.2	204.6	Thr^{26}–Cys^{67}
	5749.1 ± 0.35	5750.1	1.0	Tyr^{303}–Lys^{354}
20	2804.7 ± 0.24	2805.1	0.4	Gly^{139}–Lys^{166}
	4705.8 ± 0.68	4706.2	0.4	Cys^{210}–Arg^{251}
	5621.1 ± 0.17	5622.0	0.9	Tyr^{303}–Lys^{353}
	7562.4 ± 0.99	7563.3	0.9	Phe^{182}–Arg^{251}
21	2447.5 ± 0.33	2447.8	0.3	Cys^{210}–Cys^{230}
	5304.5 ± 0.22	5304.8	0.3	Phe^{182}–Cys^{230}
22	3604.5 ± 0.01	3605.1	0.6	Leu^{108}–Cys^{138}
	4620.3 ± 0.02	4401.9	218.4	Ala^{355}–Arg^{394}
	4732.4 ± 0.19	4530.1	202.3	Lys^{354}–Arg^{394}
	4748.2 ± 0.09	4530.1	218.1	Lys^{354}–Arg^{394}
	6391.3 ± 0.58	6392.2	0.9	Leu^{108}–Lys^{166}
23	4604.3 ± 0.15	4401.9	202.4	Ala^{355}–Arg^{394}

[a]The measured masses were derived from the LC–ESMS analysis shown in Fig. 3a.
[b]The masses were calculated using Biolynx software.
[c]Refers to the position in the sequence shown in Fig. 5.

analyzed by comparative LC–ESMS, following derivatization of cysteine residues with 3-bromopropylamine and subsequent proteolytic degradation. The derivatization of cysteine residues was found advantageous because additional tryptic cleavage sites were introduced after aminopropylcysteines (see later). Capillary LC–ESMS analysis of peptides obtained after limited V8 protease digestion of labeled and unlabeled core CBH I, resulted in the base peak intensity (BPI) chromatograms shown in Fig. 2a and b. Although minor differences in the retention times can be observed, the two BPI chromatograms are comparable except for three fractions (indicated in the figure). The peptides with the measured mass 2277.9 and 692.2 Da, assigned to the sequences Gly^{191}–Glu^{212} and Met^{213}–Glu^{217}, respectively, were absent in the analysis of labeled core CBH I. Instead, a peptide of mass 3364.8 Da was observed which corresponds to the calculated molecular weight of the peptide Gly^{191}–Glu^{217} including the mass of one conjugated epoxide (calculated mass 3365.6 Da). Absence of cleavage after the glutamate in position 212 indicates that this amino acid residue is modified. However, labeling of the nearby Asp^{214} could not completely be ruled out. To further identify the exact position of the modified amino acid residue, labeled and unlabeled core CBH I were submitted to digestion with trypsin and the peptides were analyzed by preparative LC–ESMS (Fig. 3 and Table 2). In this procedure, 10% of the sample was directed to the mass spectrometer for molecular weight determination and the remaining material was collected for further analysis. LC–ESMS analysis of tryptic digests of labeled and unlabeled core CBH I were compared to identify the peptide modified by the epoxide. The peptide (measured mass 2447.5 Da) corresponding to Cys^{210}–Cys^{230} (calcd mass 2447.8 Da) was absent in the analysis of modified core CBH I while a new peptide appeared with a mass of 2699.9 Da and eluting at a slightly lower acetonitrile concentration. The mass difference (252.1 Da) of these peptides is 160.3 Da below the expected value for one epoxide (calcd mass 412.4 Da) incorporated. Since this difference corresponds well to the calculated mass of one aminopropylated cysteine residue (calcd mass 160.2 Da), the modified peptide could be

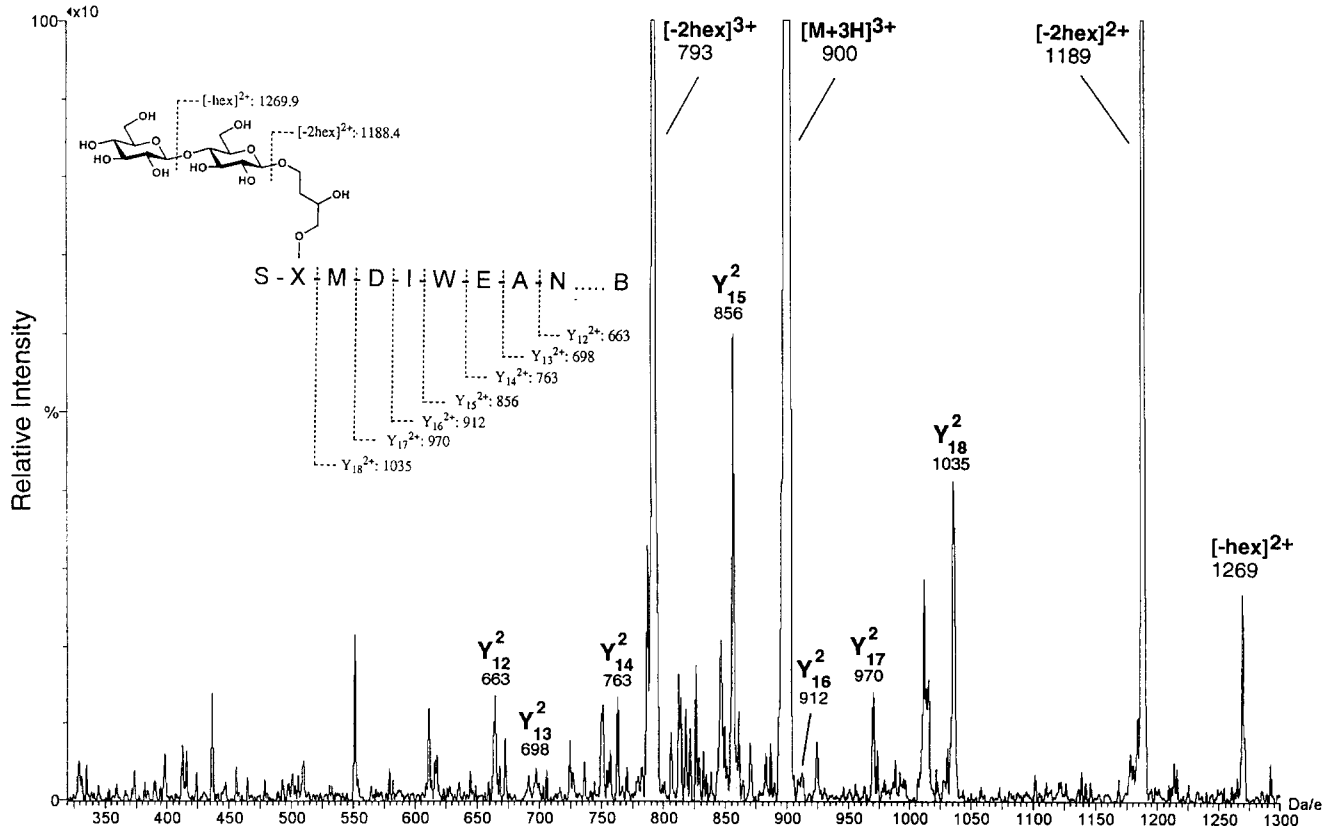

Fig. 4. Collision induced ESMS/MS of the triply charged parent ion at *m/z* 900 from the labeled peptide S^{211}–B^{230} (B indicates aminopropylcysteine). Only doubly charged Y″ fragment ions, used to confirm the sequence, are indicated. The peptide was obtained from the LC–ESMS analysis shown in Fig. 3b (Fraction 21).

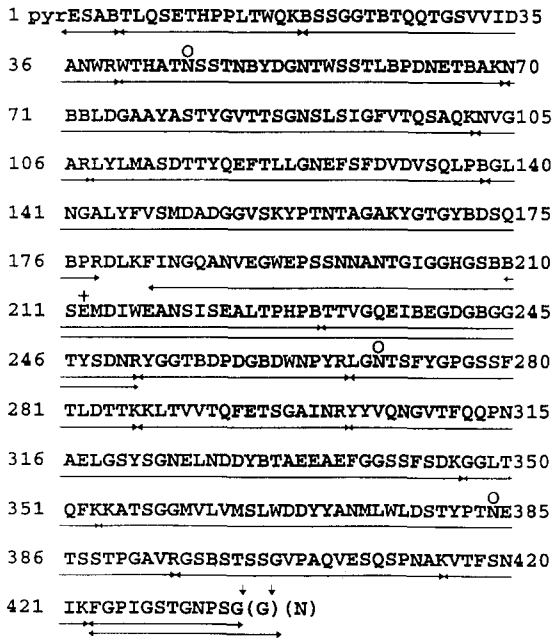

Fig. 5. Amino acid sequence of the catalytic core of CBH I [29–31]. B represents cysteine residues modified with 3-bromopropylamine. Only tryptic peptides, used to confirm the primary sequence, are shown; ↓ represents papain cleavage site(s). Glycosylation sites and the labeled Glu are indicated by ○ and +, respectively.

assigned to Ser^{211}–Cys^{230} (calcd mass 2699.9 Da). Partial sequence determination of the peptide was obtained by collision induced dissociation (CID) ESMS/MS (Fig. 4). Loss of one and two glucose units, respectively, confirms that the peptide is labeled. In addition, the complete absence of the affinity label on the C-terminal Y″ fragment ions observed suggests that the label is located on either the N-terminal Ser^{211} or Glu^{212}. Edman degradation of the first four amino acid residues revealed the sequence SXMD with serine in position 211 clearly present and glutamic acid in position 212 absent. Hence, it can be concluded unambiguously that Glu^{212} had been modified.

Peptide mapping and characterization of glycosylation sites.—Peptide mapping, using the molecular weight information obtained from LC–ESMS analysis of tryptic peptides, allowed verification of 99% of the published primary structure of core CBH I (Table 2, Fig. 5). The tripeptide Asp^{179}–Lys^{181} was not detected because its mass is outside the scanned mass range. Based on the measured masses, two peptides could be tentatively assigned to the C-terminal sequence Phe^{423}–Gly^{434}/Gly^{435} (Table 2, fraction 10). ESMS/MS of the peptide Phe^{423}–Gly^{434} allowed unambiguous identification of the four C-terminal residues –Asn–Pro–Ser–Gly confirming the sequence assignment (data not shown). ESMS analysis of unmodified core CBH I showed the presence of two compounds with the masses of the most abundant and a minor species being 46,576.7 ± 4.8 and 46,636.1 ± 6.8 Da, respectively. This corresponds well with the calculated molecular weight of $PyrGlu^{1}$–Gly^{434}/Gly^{435} (46,576.6/46,633.6) when the mass of three *N*-acetylglucosamine residues is included in the calculation. The identity of the sugar units was established from carbohydrate analysis of the intact CBH I core protein. Several fractions contain peptides with a measured mass of 16, 203 or 219 Da above the calculated M_r weight (Table 2). These increments can be explained by oxidation of methionine or/and the presence of a single *N*-acetylglucosamine. The glycopeptide containing fractions were identified in a separate capillary LC–ESMS analysis using collisional excitation and single ion recording [15] (data not shown). The observed mass of glycopeptides Leu^{268}–Lys^{287} and Ala^{355}–Arg^{394} unambiguously confirms the presence of one *N*-acetylglucosamine at positions 270 and 384, respectively. The location of the third *N*-acetylglucosamine was established by LC–ESMS analysis of peptide Trp^{40}–Lys^{69} (contains two potential glycosylation sites at positions 45 and 64, respectively) after chymotryptic subdigestion. It was recently demonstrated that electrospray ionization mass spectrometers can be used as selective detectors to identify fractions containing glycopeptides during a LC–ESMS analysis [16,17]. This is achieved by single ion recording (SIR) of diagnostic carbohydrate oxonium fragment ions that can be produced from glycopeptides by increasing the collision excitation potential in the electrospray ionization source. Thus, performing this procedure in combination with full-scan LC–ESMS analysis al-

Table 3
Summarized results from LC–ESMS analysis of tryptic peptide Trp^{40}–Lys^{69} after subdigestion with chymotrypsin

HPLC fraction	Measured mass[a]	Calculated mass[b]	ΔM_r	Peptide position[c]
1	1321.0 ± 0.29	1118.1	202.9	Trp^{40}–Asn^{49}
2	1482.7 ± 0.30	1482.7	0.0	Ser^{57}–Lys^{69}
3	1644.7 ± 0.23	1441.5	203.2	Trp^{40}–Tyr^{51}
4	591.3	591.2	0.1	Asp^{52}–Trp^{56}
5	2218.1 ± 0.26	2015.1	203.0	Trp^{40}–Trp^{56}

[a] The measured masses were derived from the LC–ESMS analysis shown in Fig. 6.
[b] The masses were calculated using Biolynx software.
[c] Refers to the position in the sequence shown in Fig. 5.

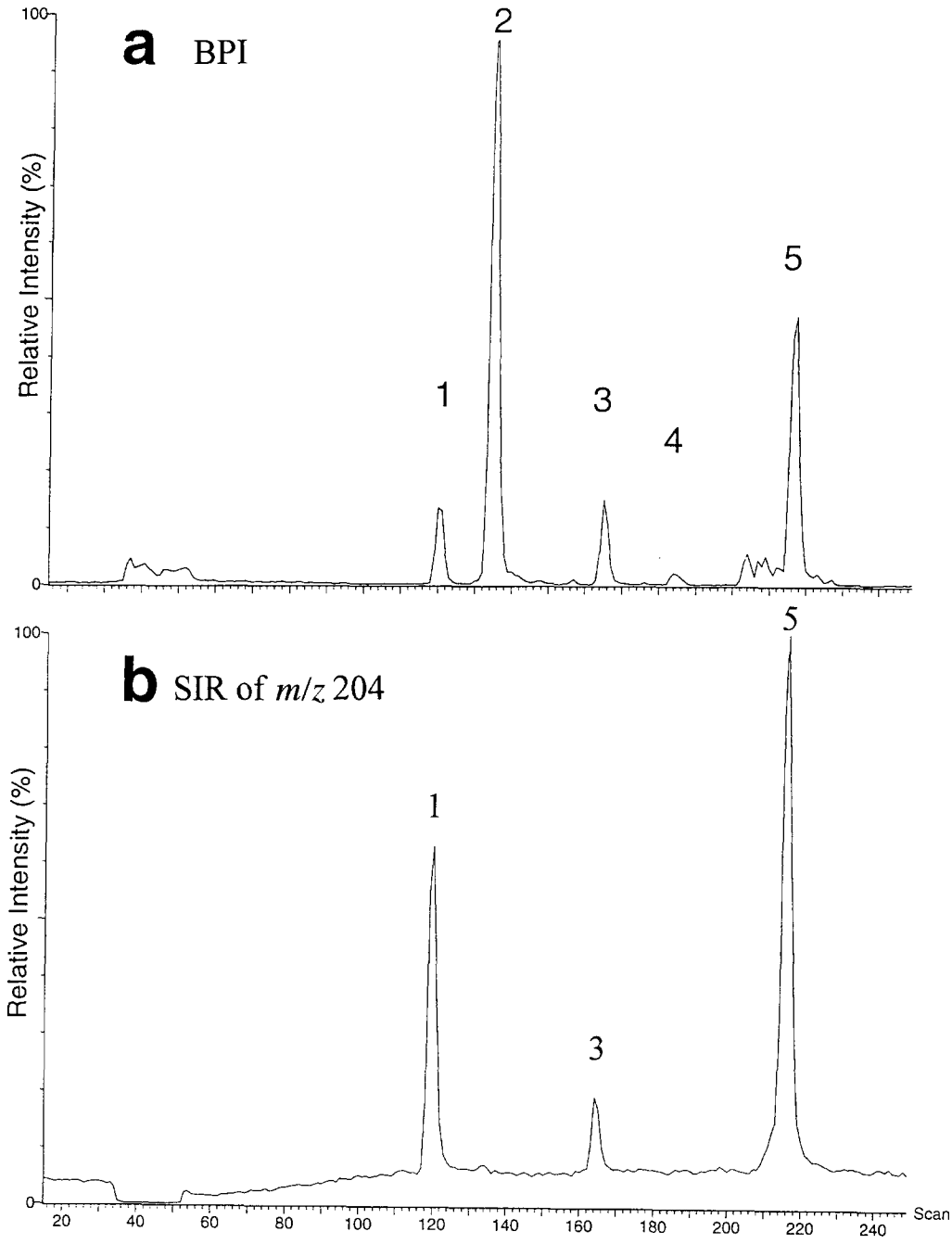

Fig. 6. Combined full-scan/collision excitation LC–ESMS analysis of peptide Trp^{40}–Lys^{69} after subdigestion with chymotrypsin. (a) Base peak intensity chromatogram obtained by scanning the *m/z* range 375–1800 in 5.8 s/scan. The extraction cone voltage was 42 V. (b) Fractions containing glycopeptides were identified during the same separation by fragmentation in the electrospray ionization source and single ion recording of the diagnostic *N*-acetylhexosamine oxonium fragment ion at *m/z* 204. The fragmentation was achieved by adjusting the extraction cone voltage to 115 V for 0.2 s/scan.

lows to simultaneously determine the molecular weight and to identify fractions containing glycopeptides [15]. Applying this approach to the LC–ESMS analysis of peptide Trp^{40}–Lys^{69}, after digestion with chymotrypsin, indicated that the peptides in Fractions 1, 3, and 5 are glycosylated (Fig. 6). Based on the observed mass and taking into account the mass of one linked GlcNAc, the identified glycopeptides could all be assigned to cover the N-terminal region of peptide Trp^{40}–Lys^{69}, thus confirming the presence of one *N*-acetylglucosamine residues at Asn in position 45 (Table 3).

3. Discussion

(*R*,*S*)-3,4-Epoxybutyl β-cellobioside specifically and irreversibly inactivates CBH I. Taking into account the chemical nature of the inhibitor, covalent modification of a nucleophilic residue within the active center of the catalytic core protein is expected [18]. Using liquid scintillation counting and ESMS, it was shown that one mole of inhibitor is incorporated per mole of enzyme. This makes reaction with nucleophiles outside the active site unlikely.

The present study provides two lines of evidence that glutamate in position 212 is modified by the inhibitor. Firstly, hydrolysis after Glu^{212} could not be observed when inactivated core CBH I protein was digested with V8 endoprotease. Similar results have been reported for affinity-labeled endoglucanase II (EC 3.2.1.4) from *Trichoderma reesei* [19] and *B. amyloliquefaciens* endohydrolase [11]. Secondly, further characterization of the conjugated tryptic peptide by tandem mass spectrometry showed that either the N-terminal serine or the glutamic acid residue could be modified. Edman degradation revealed the sequence Ser^{211}–Xxx–Met–Asp–, where Xxx represents a blank cycle. Since the corresponding sequence of the unlabeled peptide is Ser^{211}–Glu–Met–Asp–, we conclude that Glu^{212} is the target of the epoxide.

CBH I cleaves β-(1 → 4)-glycosidic bonds via a double-displacement mechanism [4,5]. Hence, two amino acid residues are postulated to act either as an acid-catalyst (proton donor) or as a nucleophile, respectively. Both residues have been tentatively identified in the three-dimensional structure of CBH I core in complex with small ligands, and Glu^{217} was proposed to be the proton donor [7]. Furthermore, it was observed that the glutamic acid residues at positions 212 and 217 are situated in the catalytic tunnel on opposite sides of the scissile β-(1 → 4)-glycosidic bond, and that the spatial location of Glu^{212} makes it the most likely nucleophile. The catalytic-site sequence of CBH I (E^{212}MDIWE^{217}) shows striking similarity with the sequence E^{105}IDIE^{109} found in 1,3-1,4-β-D-glucan 4-glucanohydrolase (EC 3.2.1.73) from *B. amyloliquefaciens* [11]. Affinity labeling and site-directed mutagenesis have previously demonstrated the role of Glu^{105} as a potential nucleophile in this endohydrolase [11,20]. Interestingly, even with an additional Trp in the sequence of CBH I, superposition of the three-dimensional structures shows that the spatial location of Glu^{105} in endohydrolase corresponds to Glu^{212} in CBH I [7]. The catalytic nucleophile of another member of the C/7 glycosyl hydrolase family, to which CBH I belongs, has recently been identified [21]. Inactivation of endoglucanase EG I (EC 3.2.1. 4) from *Fusarium oxysporum* with 2′,4′-dinitrophenyl-2-deoxy-2-fluoro β-cellobioside resulted in covalent modification of Glu^{197} corresponding to Glu^{212} in CBH I. Like in the present study, labeling of the assumed catalytic nucleophile in EG I is supported by results obtained by X-ray crystallography [22]. However, although X-ray crystallographic data and our results using epoxide-type affinity labeling support the role of Glu^{212} in CBH I as the catalytically active nucleophile, the possibility remains that the residue can adapt, functionally, as nucleophile or acid-base catalyst depending on the substrate or affinity label used [5]. Studies with endocellulases from *T. reesei*, e.g. for EG II, have shown that the inactivation with ω-epoxyalkyl cellobiosides is much less aglycon chain length dependent [19]. The rates of inactivation for EG I, also belonging to family C/7, at 2 mM inhibitor concentration are 0/0.05/0.02 min^{-1} for, respectively, the propyl, butyl and pentyl derivatives (K. Piens, unpublished data). Structural factors are likely to be implicated. For the 'exocellulases' a tunnel-shaped active site has been shown, while for EG I a cleft-like substrate binding site without the interconnecting loops present in CBH I has been proposed [7]. Thus, in the later case, less strict sterical constraints would be imposed upon inhibitor binding.

The catalytic core of CBH I contains four potential glycosylation sites (Asn–X–Ser/Thr) of which the Asn in positions 45, 270, and 384 each were found to be linked to one *N*-acetylglucosamine. Glycosylation of this fungal enzyme is expected to be heterogeneous and complex [23]. The presence of a single GlcNAc at these sites suggests that the oligosaccharides have been modified by cleavage by an endoglycosidase (with a specificity similar to that of endoglycosidase F or H) directed at the chitobiose core of the N-glycan. Single GlcNAc units have also been found in the catalytic domain of *T. reesei* endoglucanase I (Jose Cremata, personal communication). It is thus possible that *Trichoderma* produces such an endoglucanase as part of the glycosylation trimming machinery in the Golgi organelle.

4. Experimental

Inhibitors.—(*R*,*S*)-3,4-Epoxyalkyl β-cellobiosides (*n*-propyl, *n*-butyl and *n*-pentyl) were synthesized as described [24]. The corresponding *C*-glyco-

sides were obtained from Dr. Stick (University of Western Australia). Optically pure (*R*)- and (*S*)-*O*-epoxybutyl β-cellobiosides and (^{14}C) (*R,S*)-*O*-epoxybutyl β-cellobioside were generously provided by Dr B.E. Rodriguez (University of Western Australia).

Preparation of enzyme and activity measurements.—Cultivation of *Trichoderma reesei* wild-type strain QM9414 was carried out using a continuous feed of a lactose–cellobiose solution (ratio 9:1) as carbon source and intact CBH I (EC 3.2.1.91) was fractionated from the culture filtrate in several steps by anion-exchange chromatography [25]. The core protein was prepared by partial proteolytic cleavage [26] and purified by ion-exchange chromatography [25]. The protein concentration was determined spectrophotometrically at 280 nm and CBH I was assayed with the chromogenic substrates 2-chloro-4-nitrophenyl β-lactoside or the corresponding cellobioside as described [5].

Inactivation experiments.—CBH I core protein (2.5 μM) was incubated with different concentrations (0–10 mM) of the inhibitor in 50 mM sodium acetate buffer, pH 5.0, at 37 °C. At different intervals, aliquots were withdrawn to measure activity and the results were plotted and analyzed conventionally [9,10]. Larger amounts of CBH I core protein (0.12 mM) were inactivated with (*R,S*)-3,4-epoxybutyl β-cellobioside (2.4 mM) in distilled water for 18 h at 4 °C. Excess of inhibitor was removed by gel filtration on a Sephadex G-25 column (Pharmacia, Uppsala, Sweden).

Peptide analysis.—Reduction and derivatization were performed on 90 μg labeled or unlabeled CBH I core protein as described previously [27]. The protein was incubated in 30 μL 0.5 M Tris–HCl buffer, pH 8.2, containing 6 M guanidine–HCl, 50 mM EDTA and 20 mM DTT (previously flushed with argon) for 1.5 h at 37 °C. Alkylation was carried out in the dark by adding 15 μL buffer without DTT but containing 250 mM 3-bromopropylamine, at room temperature for 1 h. Reagents were removed by gel filtration on a Bio-Gel P6 column (0.8 × 10 cm) (Bio-Rad, Belgium) in 50 mM ammonium acetate, pH 4. Later during this work, we found that rapid and effective desalting could also be performed by precipitating protein by adding 500 μL acetone–water (9:1, vol) to 20 μL solution containing 0.5–1.2 nmol alkylated CBH I core protein. After centrifugation, the supernatant was removed and the procedure repeated twice followed by washing the precipitate with 1 mL water. The precipitated protein was redissolved in 2 μL saturated guanidine–HCl solution in water, prior to enzymatic digestion.

Protease V8 (EC 3.4.21.19) digestion of the protein after desalting was carried out for 20 h at 37 °C (enzyme–substrate ratio: 1:20, w/w). Trypsin (EC 3.4.21.4) digestion was performed in a 200 mM Tris–HCl buffer, pH 8.2, for 4 h at room temperature (enzyme–substrate ratio: 1:20, w/w). The total amount of protein used was estimated to be 800 pmol per digest. Subdigestion of peptide Trp^{40}–Lys^{69} with chymotrypsin (enzyme–substrate ratio: 1:40, w/w) was carried out in a 50 mM Tris–HCl buffer, pH 8.3, for 30 min at room temperature.

The LC–ESMS interface used is described in detail elsewhere [28]. Peptides derived from enzymatic hydrolysis of labeled and unlabeled core CBH I were separated on a capillary (0.32 × 200 mm) or a narrowbore column (2.1 × 100 mm) packed with YMC-AQ C18 (5 μm particle size, 120 Å pore size) reverse phase (YMC Inc, Wilmington, NC). The HPLC solvents consisted of 0.05% aq TFA containing 5% acetonitrile (solvent A) and 0.05% TFA in 80% acetonitrile–20% water (solvent B). Proteolytic peptides were eluted in a linear gradient from 2 to 45% solvent B in 50 min, followed by a linear increase to 100% solvent B in 20 min at a flow-rate of 150 μL/min (narrowbore column) or 4 μL/min (capillary column).

N-terminal sequencing was performed using an 476A pulsed liquid-phase sequenator equipped with an on-line 120A PTH amino acid analyzer (Perkin Elmer, Applied Biosystems Division, Foster City, CA).

Peptides were assigned by searching the sequence of core CBH I with the respective measured masses as input data (BioLynx, Micromass Ltd., Altrincham, UK).

Mass spectrometry.—All analyses were performed on a triple quadrupole mass spectrometer (Bio-Q upgraded to the Quattro II version, Micromass Ltd., UK) equipped with a pneumatically assisted electrospray source. Optimization and calibration have been described elsewhere [28]. Labeled and unlabeled CBH I core, dissolved in water, were diluted to a final concentration of 5–10 pmol/μL in 50% (vol) aq MeCN containing 0.1% formic acid and introduced into the source at a flow-rate of 5 μL/min. Spectra were acquired from *m/z* 1300 to 2300 in 15 s. The extraction cone voltage was 42 V and the temperature 70 °C. Calibration was performed using the multiple charged ions from a separate acquisition of bovine pancreatic trypsinogen at a resolution so

that the peak at m/z 1714 was 1 amu wide 50% above base line. Background subtracted spectra were processed to convergence using a maximum entropy algorithm (Maxent, part of the Masslynx software) and a resolution of 0.5 Da/channel. For LC–ESMS analysis the mass spectrometer was scanned from m/z 390 to 1800 in 6 s and the total ion current (TIC) chromatogram reconstructed using the base peak intensity (BPI) ion mode. The resolution was adjusted to resolve the monoisotopic peaks from polyethylene glycol (average mass 2000) at m/z 1840, 20% above base line. LC–ESMS analysis using collisional excitation and single ion monitoring of the diagnostic oxonium fragment ion from *N*-acetylglucosamine (m/z 204) was performed by increasing the extraction cone voltage to 115 V for 0.2 s/scan followed by scanning the m/z range 375–1800 at 42 V as described above for 5.8 s/scan. ESMS/MS was performed by collision induced dissociation of the doubly or triply charged molecular ion using argon and a collision energy offset of −38 V.

Carbohydrate analysis.—CBH I core protein (200 mg) was hydrolyzed in 4 N TFA for 4 h at 100 °C. After freeze drying, the sample was redissolved in 200 μL deionized water and analyzed on a Carbopac PA10 anion-exchange column equipped with an amino trap precolumn (Dionex, Sunnyvale, CA) at 40 °C. The HPLC system used for separations and detection of the monosaccharides consisted of a DX500 equipped with a GP40 gradient pump and an ED 40 electrochemical detector (Dionex, Sunnyvale, USA). Separations were performed isocratically using a 16 mM NaOH soln, and data were processed with the Dionex PeakNet software package (release 4.11).

Acknowledgements

K.P., M.C. and J.S. thank the EU for BIO2-CT94-3030 and the Department of Industry, Technology and Commerce, Australia for support. R.V. Stick and E.B. Rodriguez are acknowledged for providing the *C*-epoxypropyl and pentyl derived inhibitors as well as the diastereoisomers of 3,4-epoxybutyl β-cellobioside. J.V.B. is indebted to the Flemish government for a concerted Research Action (GOA, 12052293).

References

[1] T.M. Wood, *Enzyme Systems for Lignocellulose Degradation*, M.P. Coughlan (Ed.), Elsevier, London and New York, 1989, pp 19–35.
[2] B. Henrissat, *Biochem. J.*, 280 (1991) 309–316.
[3] B. Henrissat and A. Bairoch, *Biochem. J.*, 293 (1993) 781–788.
[4] J.K. Knowles, P. Lehtovaara, M. Murray, and M.L. Sinnott, *J. Chem. Soc., Chem. Commun.*, (1988) 1401–1402.
[5] M. Claeyssens, P. Tomme, C.E. Brewer, and E.J. Hehre, *FEBS Lett.*, 263 (1990) 89–92.
[6] H. Van Tilbeurgh, P. Tomme, M. Claeyssens, R. Bhikhabhai, and G. Pettersson, *FEBS Lett.*, 204 (1986) 223–227.
[7] C. Divne, J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J.K.C. Knowles, T.T. Teeri, and T.A. Jones, *Science*, 265 (1994) 524–528.
[8] S.G. Withers and R. Aebersold, *Prot. Sci.*, 4 (1995) 361–372.
[9] R. Kitz and I.B. Wilson, *J. Biol. Chem.*, 237 (1962) 3245–3249.
[10] B. Nidetzky and M. Claeyssens, *Biotechnol. Bioeng.*, 44 (1994) 961–966.
[11] P.B. Høj, R. Condron, J.C. Traeger, J.C. McAuliffe, and B.A. Stone, *J. Biol. Chem.*, 267 (1992) 25059–25066.
[12] P.B. Høj, E.B. Rodriguez, J.R. Iser, R.V. Stick, and B.A. Stone, *J. Biol. Chem.*, 266 (1991) 11628–11631.
[13] P. Tomme and M. Claeyssens, *FEBS Lett.*, 243 (1989) 239–243.
[14] D. Tull, D.L. Burgoyne, D.T. Chow, S.G. Withers, and R. Aebersold, *Anal. Biochem.*, 234 (1996) 119–125.
[15] M. Allen, J. Anacleto, R. Bonner, L. Nuwaysir, and B. Shushan, *Characterization of protein digests using novel scanning (funky ramps)*, in *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA, USA, 1995, p. 646.
[16] K.L. Duffin, J.K. Welply, E. Huang, and J.D. Henion, *Anal. Chem.*, 64 (1992) 1440–1448.
[17] S.A. Carr, M.J. Huddleston, and M.F. Bean, *Prot. Sci.*, 2 (1993).
[18] G. Legler, *Adv. Carbohydr. Chem. Biochem.*, 48 (1990) 319–385.
[19] R. Macarron, J. Van Beeumen, B. Henrissat, I. Mata, and M. Claeyssens, *FEBS Lett.*, 316 (1993) 137–140.
[20] A. Planas, M. Juncosa, J. Lloberas, and E. Querol, *FEBS Lett.*, 308 (1992) 141–145.
[21] L.F. Mckenzie, G.J. Davies, M. Schülein, and S.G. Withers, *Biochem.*, 36 (1997) 5893–5901.
[22] G. Sulzenbacher, M. Schülein, and G.J. Davies, *Biochem.*, 36 (1997) 5902–5911.
[23] I. Salovuori, M. Makarow, H. Rauvala, J. Knowles, and L. Kääriäinen, *Bio / Technology*, 5 (1987) 152–156.
[24] P.B. Høj, E.B. Rodriguez, R.V. Stick, and B.A. Stone, *J. Biol. Chem.*, 264 (1989) 4939–4947.
[25] J. Ståhlberg, C. Divne, A. Koivula, K. Piens, M. Claeyssens, T.T. Teeri, and T.A. Jones, *J. Mol. Biol.*, 264 (1996) 337–349.
[26] P. Tomme, H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens, *Eur. J. Biochem.*, 170 (1988) 575–581.

[27] R.A. Jue and J.E. Hale, *Anal. Biochem.*, 210 (1993) 39–44.
[28] K. Klarskov, D. Roecklin, B. Bouchon, J. Sabatié, A. Van Dorsselaer, and R. Bischoff, *Anal. Biochem.*, 216 (1994) 127–134.
[29] S. Shoemaker, V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo, and M. Innis, *Bio/Technology*, 1 (1983) 691–695.
[30] T. Teeri, I. Salovuori, and J. Knowles, *Bio/Technology*, 1 (1983) 696–699.
[31] L.G. Fägerstam, L.G. Pettersson, and J.Å. Engström, *FEBS Lett.*, 167 (1984) 309–315.