Glycobiology vol. 12 no. 12 pp. 837–849, 2002

Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from *Trichoderma reesei*

Joseph P.M. Hui[1,3], Theresa C. White[4], and Pierre Thibault[2,3]

[3]Institute for Biological Sciences, 100 Sussex Drive, Ottawa, Ontario, Canada, K1A 0R6, and [4]Iogen Corporation, 400 Hunt Club Road, Ottawa, Ontario, Canada K1V 1C1

Received on May 22, 2002; revised on July 22, 2002; accepted on July 24, 2002

Mass spectrometric techniques combined with enzymatic digestions were applied to determine the glycosylation profiles of cellobiohydrolase (CBH II) and endoglucanases (EG I, II) purified from filamentous fungus *Trichoderma reesei*. Electrospray mass spectrometry (ESMS) analyses of the intact cellulases revealed the microheterogeneity in glycosylation where glycoforms were spaced by hexose units. These analyses indicated that glycosylation accounted for 12–24% of the molecular mass and that microheterogeneity in both N- and O-linked glycans was observed for each glycoprotein. The identification of N-linked attachment sites was carried out by MALDI-TOF and capillary liquid chromatography–ESMS analyses of tryptic digests from each purified cellulase component with and without PNGase F incubation. Potential tryptic glycopeptide candidates were first detected by stepped orifice-voltage scanning and the glycan structure and attachment site were confirmed by tandem mass spectrometry. For purified CBH II, 74% of glycans found on Asn310 were high mannose, predominantly $Hex_{7-9}GlcNAc_2$, whereas the remaining amount was single GlcNAc; Asn289 had 18% single GlcNAc occupancy, and Asn14 remained unoccupied. EG I presented N-linked glycans at two out of the six potential sites. The Asn56 contained a single GlcNAc residue, and Asn182 showed primarily a high-mannose glycan $Hex_8GlcNAc_2$ with only 8% being occupied with a single GlcNAc. Finally, EG II presented a single GlcNAc residue at Asn103. It is noteworthy that the presence of a single GlcNAc in all cellulase enzymes investigated and the variability in site occupancy suggest the secretion of an endogenous endo H enzyme in cultures of *T. reesei*.

Key words: cellulases/glycoforms/nanoelectrospray/tandem mass spectrometry/*Trichoderma reesei*

Introduction

The role of cellulase is very important in the biosphere to recycle the rich carbon source (cellulose) available in the plant kingdom to maintain the ecological balance. The filamentous fungus *Trichoderma reesei* is the microorganism conventionally chosen as a model to study the cellulase enzymatic system (as well as for commercial applications) because of its remarkable secretory ability. It is capable of secreting in excess of 30 g/L of protein into the extracellular medium (Conesa *et al.*, 2001).

The cellulolytic enzyme system of *T. reesei* comprises two major cellobiohydrolases (CBHs) and several endoglucanases (EGs), together with small amounts of β-glucosidase and several hemicellulases (Kubicek, 1992). Degradation of cellulose involves an interplay between different cellulolytic enzymes. Synergy exists between the cellulolytic enzymes and the most widely accepted hypothesis (International Union of Biochemistry and Molecular Biology, 1992) suggests that hydrolyses start with EGs that randomly hydrolyze the β-(1→4) glycosidic linkages in the amorphous region of the cellulose. This significantly reduces the degree of polymerization of the substrate and creates new chain ends that facilitate the accessibility for the CBHs to split off cellobiose units. Subsequently, β-glucosidases hydrolyze cellobiose into glucose units (Kleywegt *et al.*, 1997).

Although this synergistic enzyme activity has been reported previously (Nidetzky *et al.*, 1994; Henrissat *et al.*, 1985), studies from three-dimensional crystal structure have further reinforced the functional relationships between cellulase and its substrate. Both CBH I and II have an enclosed tunnel structure formed by parallel β strands containing the active site to mediate the hydrolysis of cellulose from the chain ends (Divine *et al.*, 1994; Rouvinen *et al.*, 1990). The presence of this tunnel structure allows for close interaction between the catalytic domain and the loose chain ends appearing on the crystalline cellulose surface. In contrast to CBHs, the structure of EGs, in particular the EG I catalytic domain, resembles an open substrate-binding cleft rather than a tunnel (Kleywegt *et al.*, 1997). This enables them to interact with the amorphous or disordered crystalline cellulose more effectively.

Both CBHs and EGs are glycoproteins containing significant microheterogeneities in N- and/or O-linked glycan attachment (Kubicek, 1992). With the exception of EG III, these cellulases are structurally similar and are composed of a catalytic core domain and a cellulose-binding domain (CBD) connected via a linker peptide showing extensive O-linked glycosylation. The CBD serves as anchor on the cellulose surface and facilitates the catalytic action of the cellulolytic enzymes. The role of glycosylation is related to the secretion of these cellulolytic enzymes, to provide sufficient spatial separation between the core and CBD domains, and to protect the linker peptide against proteolysis (Srisodsuk *et al.*, 1993; Clarke, 1997).

Mammalian high-mannose type N-linked glycans have been observed in CBH I from *T. reesei* RUT-C30 strain (Hui *et al.*,

[1]Present address: Caprion Pharmaceuticals Inc., 7150 Alexander Fleming, St-Laurent, Quebec, H4S 2C8, Canada
[2]To whom correspondence should be addressed; E-mail: pierre.thibault@nrc.ca

2001; Maras *et al.*, 1997). The presence of single GlcNAc in CBH I catalytic domain have been found independently from strains QM9414 (Klarskov *et al.*, 1997) and ALKO2877 (Harrison *et al.*, 1998). However, the nature of glycosylation in CBH II, EG I, and EG II is not as well documented in the literature as is CBH I glycosylation, possibly due to the difficulties in their purification. The putative N-linked sites, Asn56 and Asn182 in EG I core from VTT mutant strains, have been reported to bear a single GlcNAc residue (Kleywegt *et al.*, 1997). The Asn289 and Asn310 in CBH II have also been reported to be glycosylated, although the nature of the attached oligosaccharide was not determined (Rouvinen *et al.*, 1990). There are no reports on the characterization of the putative N-linked glycan at Asn103 in EG II, although its glycan structure and catalytic properties have been studied to some extent (Macarron *et al.*, 1993; Saloheimo *et al.*, 1998).

A previous investigation from this laboratory reported that CBH I from RUT-C30 contained a single phosphorylated disaccharide in the linker domain (Hui *et al.*, 2001). This is a first report in the literature of phosphorylated O-linked glycans in CBH I from *T. reesei*, although previous investigation on mutant strains of *Saccharomyces cerevisiae* have described a mannosylphosphorylated mannotriose (Nakayama *et al.*, 1998). Such a finding prompted us to investigate the occurrence of mannosylphosphorylation of O-linked oligosaccharide in other cellulases secreted from *T. reesei* to address the question as to whether CBH I is a specific substrate for this mannosylphosphotransferase or if protein phosphorylation is dependent on the cell cycle or environmental changes. Furthermore, the observation of both single GlcNAc and high mannose at Asn270 of CBH I from RUT-C30 has suggested the expression of endogenous endo H enzymatic activity. To address these questions and their influences on other cellulases, we undertook the structural characterization of N-linked glycans in other cellulase components.

This article focuses on CBH II, EG I and EG II cellulase components produced by *T. reesei* strain RUT-C30. This allows for further study of the substrate specificity of the mannosylphosphotransferase and the extent of enzymatic activity of the suspected endogenous endo H enzymes. To our knowledge, this is the first detailed report describing the nature and extent of N-linked glycosylation of these cellulase components. A combination of ion-exchange chromatography and chromatofocusing was used to purify the different cellulases. Each intact cellulase was first characterized by electrospray mass spectrometry (ESMS) to determine its molecular mass. Proteolytic fragments from trypsin and papain digestion were analyzed by capillary liquid chromatography–ESMS (LC-ESMS) and matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) to identify the nature and sites of N-linked glycosylation and to monitor the extent of heterogeneity in O-linked glycans on the linker domain.

Results

MS analysis of intact CBH II, EG I, and EG II from RUT-C30

The amino acid sequence information for the three investigated cellulases were obtained from SWISS-PROT database (Figure 1). All identified N-linked glycans in this study were annotated. When occupied, the N-linked glycosylation site was made up of either a single GlcNAc or a high-mannose glycan. High-performance anion exchange chromatography/pulsed amperometric

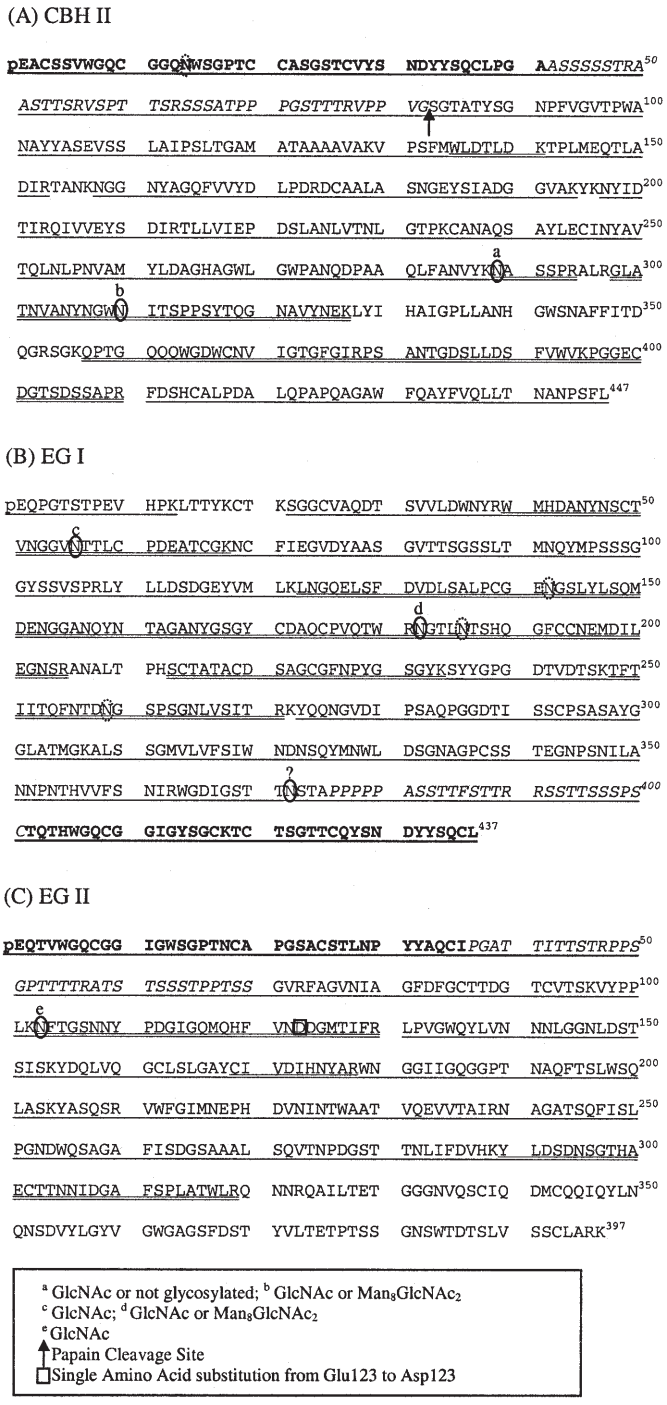

Fig. 1. Amino acid sequence of cellulolytic enzymes from *T. reesei*. (**A**) CBH II; (**B**) EG I; (**C**) EG II. Sequence information is obtained from SWISS-PROT database (*http://www.expasy.ch/sprot*). The CBD domain is highlighted in bold, the linker domain in italics, and the remaining sequence constitutes the catalytic core. The putative N-glycosylation sites are circled. pE represents pyroglutamic acid; the known papain cleavage site reported in the literature is indicated by arrow. Insets denote the glycan determined in this study. Tryptic peptides observed with both MALDI-TOF and capillary LC/ESMS analyses are underscored. Sequence coverage for CBH II, EG I, and EG II are 91.5%, 96.3%, and 80.4%, respectively. Amino acid sequences identified by nano-ESMS/MS analyses are indicated by a double underscore.

detection (HPAEC/PAD) analysis of oligosaccharides released from the hydrazinolysis of CBH II, EG I, and EG II (data not shown) were similar to those obtained previously for CBH I (Hui *et al.*, 2001). These analyses revealed a high proportion of di- and trisaccharides presumably located on the linker region as reported previously (Harrison *et al.*, 1998; Hui *et al.*, 2001). In contrast to previous investigations on CBH I (Hui *et al.*, 2001), no phosphorylated trisaccharide was observed on any of the cellulases reported in the present study. In addition, high-mannose glycans, predominantly $Man_8GlcNAc_2$ or $Man_7Glc_1GlcNAc_2$, were also observed for CBH II and EG I (see details to follow). Carbohydrate composition analyses (data not shown) revealed predominantly Man, Glc, and glucosamine (hydrolyzed GlcNAc) and were consistent with the suspected glycan structures.

The papain cleavage site of CBH II from strain QM9414 has been reported by partial sequencing at the N terminus of CBH II core (Tomme *et al.*, 1988). This position corresponds to the amide linkage between Gly82 and Ser83 (Figure 1A). No papain cleavage sites for EG I or EG II have been reported in the literature.

To profile the glycan microheterogeneity, purified cellulases were subjected to flow injection analysis by ESMS (Figure 2). The multiply charged ion envelopes were displayed in the range *m/z* 2000–3000 (Figure 2A) and a reconstructed molecular mass profile could be obtained for each cellulase (Figures 2B–D). The minor glycoform distribution in the mass profile of RUT-C30 CBH II from 55,500 – 56,500 Da was later shown to be the series with single GlcNAc at Asn310 (Figure 2B). In the reconstructed mass spectra for all three components, the glycoforms are regularly spaced by hexose units (162 Da) and the corresponding glycosylation percentages were estimated using the difference between the observed mass and the theoretical average mass. The relative proportion of glycans in CBH II, EG I, and EG II with respect to the protein mass was calculated to be 17.7–24.0%, 13.0–16.5%, and 11.9–16.7%, respectively. Unlike the CBH I from RUT-C30 (Hui *et al.*, 2001), no 80-Da series of peaks embedded between the glycoforms was observed in the ESMS profiles of these three cellulases, suggesting the absence of mannosylphosphorylation in the linker regions.

Papain proteolysis

Papain proteolysis was previously used to separate the core domains from the linker/CBD glycopeptides (Tomme *et al.*, 1988; Woodward *et al.*, 1994). Purified cellulases from RUT-C30 were digested with papain and subsequently analyzed by MALDI-TOF MS (Figure 3). In each case, the results for the released linker/CBD glycopeptides enlarged in the inset indicated that the glycans are highly heterogeneous. Consistent with the ESMS mass spectra of the intact CBH II, EG I, and EG II, there were no additional 80-Da peak series from phosphate modifications embedded between the hexose spacing.

Interestingly, the released CBH II core from RUT-C30 resulted in two peaks spaced by 1564 Da (Figure 3A). Similar spacing was observed previously for CBH I core and was attributed to variability in N-linked site occupancy. Incubation of CBH II core with endo H enzymes effectively removed the peak with high mannose, in this case resulting in a single protein mass with single GlcNAc attachment (data not shown). This experiment also enabled the identification of a single

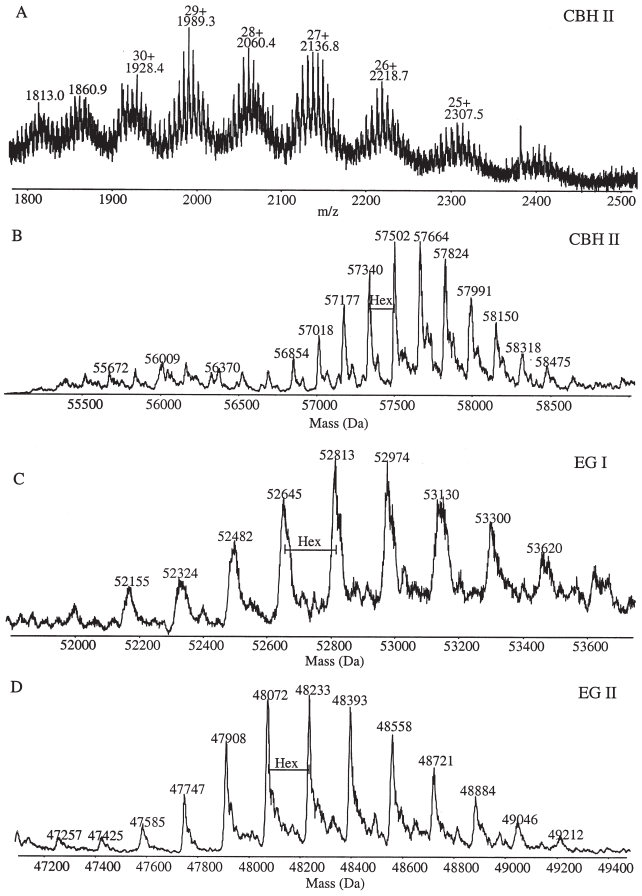

Fig. 2. ESMS mass spectra of (**A**) CBH II and reconstructed molecular mass profiles of (**B**) CBH II; (**C**) EG I; and (**D**) EG II. All cellulases were purified from Iogen strain RUT-C30. Conditions: An aliquot of 0.5 µg of purified cellulases dissolved in 50% aqueous acetonitrile (1% acetic acid) was introduced at 10 µl/min to the mass spectrometer.

high-mannose glycan from one of the other three putative N-linked sites. The high-mannose oligosaccharide was later located at Asn310 using capillary LC-ES tandem mass spectrometry (MS/MS) analyses described in a later section.

Papain cleaves within the CBH II linker between Gly82 and Ser83 and therefore, in the absence of glycosylations, the predicted linker/CBD peptide mass would be 8155.8 Da. From this information, the number of O-linked hexoses in CBH II can be calculated from the MALDI-TOF results. The heterogeneous 15,000-Da linker/CBD shown in the inset of Figure 3a correspond to the O-linked glycans and possibly a N-linked glycan at Asn14 appended to the native peptide backbone of a molecular mass of 8155.8 Da (Figure 1A). Results from chymotryptic digest (described later) showed that the putative N-linked position at Asn14 is not glycosylated. It was concluded that there are 39–46 O-linked hexoses appended to the CBH II linker peptide. When compared to CBH I (Hui *et al.*, 2001), this more extensive glycosylation of the linker region can be rationalized by the fact that this cellulase has a longer linker domain (Figure 1A).

In contrast to the high coring rate of CBHs, papain proteolysis of EGs was not effective and showed less conclusive

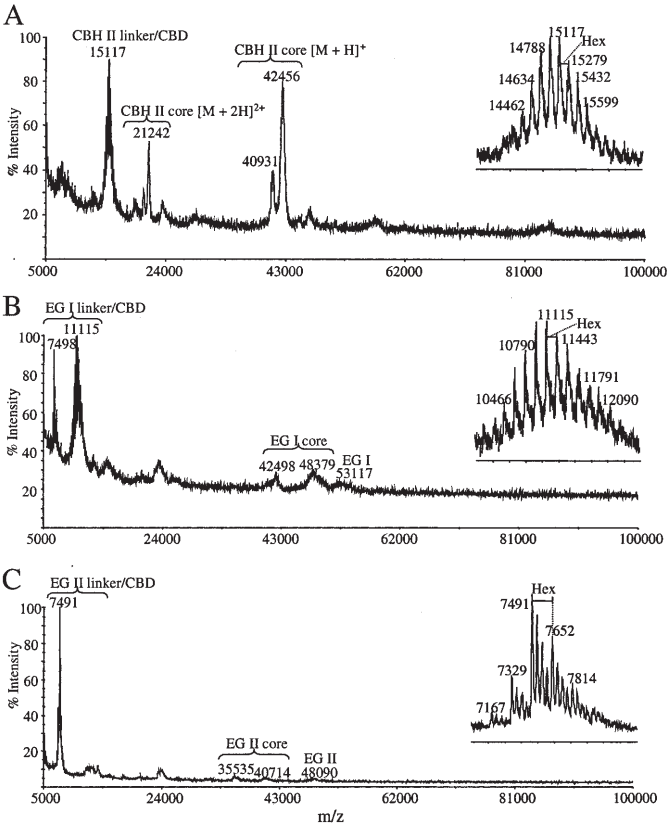

Fig. 3. MALDI-TOF analysis of papain proteolysis of purified cellulases from RUT-C30. (**A**) CBH II; (**B**) EG I; and (**C**) EG II. The released linker and CBD domains were enlarged and shown in the insets.

results (Figures 3B–C). MALDI-TOF analyses of the EG I and II papain digestion products revealed more than one cleavage site. In spite of the use of high enzyme dosage (enzyme-to-substrate ratio of 5:1 for EG I and EG II), relatively large amounts of intact cellulase compared to the core were still visible after 24 h of digestion, indicating the poor coring of the EGs by papain. The large EG I linker/CBD glycopeptide of about 11,000 Da displayed an incremental number of O-linked hexose residues (inset of Figure 3B), whereas the smaller of the two EG II linker/CBD glycopeptides of about 7500 Da showed a relatively small O-linked heterogeneity with an extra three series of potassium adduct ions (39 Da) embedded between the hexose spacings, possibly arising from the papain buffer (inset of Figure 3C). The large EG II linker/CBD glycopeptide of about 13,000 Da was broad and poorly resolved, thus preventing accurate peak definition of Hex spacing. Despite the poor coring rate of EGs by papain and the lack of knowledge of cleavage site in the literature, the number of O-linked glycans attached to both EG I and II was calculated using the difference between the observed mass of the intact protein from ESMS analysis and the theoretical average mass of their respective sequences (EG I: 46,011.4, EG II: 42,150.5 Da). The number of O-linked hexoses appended to the CBD/linker domain of EG I was determined to be 24–34 following subtraction of the amino acid sequence and the two N-linked glycans (discussed later) from the observed mass (Figure 2C). For EG II, which comprised a single GlcNAc at Asn_{103} (see following), the corresponding number of O-linked glycans was higher and ranged from 32 to 42 residues.

MALDI-TOF and capillary LC-ESMS analyses of tryptic digestion

Tryptic digestion was used to characterize the attachment sites and the nature of N-linked oligosaccharides in CBH II, EG I, and EG II from RUT-C30. The digests were analyzed primarily by MALDI-TOF using 2,5-dihydroxybenzoic acid (DHB) as a matrix for small peptides, and sinapinic acid was used for relatively large peptide. The observed tryptic peptides from MALDI-TOF analyses of CBH II, EG I, and EG II are listed in Tables I to III, respectively. They were matched against the theoretical digest. It is noteworthy that isotopic resolution was not typically achieved above *m/z* 4000, and the isotope average molecular masses are thus reported.

The assignment of N-linked glycopeptides was first made in tryptic peptides before and after PNGase F treatment. For example, MALDI-TOF analysis showed that peaks 14–16 from EG I (Table II) were shifted to *m/z* 2755.1 after incubation with PNGase F corresponding to the native sequence Asn182–Arg205 (Figure 4). However, the use of MALDI-TOF does not provide an accurate measurement of the glycan distribution, and the three cellulase tryptic digests were further analyzed by online capillary LC-ESMS.

To facilitate the identification of tryptic glycopeptides on the Q-TOF instrument, a mixed scan function was used to promote the in-source formation of characteristic oxonium ions under high-orifice voltage at 120 V while detecting the multiply protonated peptides under normal operating orifice voltage at 30 V (Hui *et al.*, 2001; Bateman *et al.*, 1998; Carr *et al.*, 1993). Similar to the previous study on CBH I (Hui *et al.*, 2001), the oxonium disaccharide ion Hex-HexNAc corresponding to *m/z* 366 was selected as a diagnostic ion for N-linked glycopeptide. The extracted ion chromatogram for *m/z* 366 together with the total ion chromatogram for *m/z* 400–2000 are shown in Figure 5. The observed peptide masses from capillary LC/ESMS analysis together with the expected masses and their assignments are summarized in Tables I–III. Peptides indicated by asterisks on Tables I–III were identified using nano-ESMS/MS following fractionation on an analytical C_8 column.

The extracted ion chromatogram for *m/z* 366 showed distinct peak at 18.1 min and 18.6 min in Figures 5A and 5B respectively. These peaks correspond to tryptic peptides 16–18 of CBH II (Figure 5A, Table I) and 14–16 of EG I (Figure 5B, Table II), both sets containing a high-mannose attachment confirmed subsequently using nano-ESMS/MS analysis described in a later section. The predominant glycoform is $Hex_8GlcNAc_2$ with minor contributions of $Hex_7GlcNAc_2$ and $Hex_9GlcNAc_2$. It is noteworthy that Hex is used throughout this article to identify either Man or Glc residues as recent investigations indicated that the major N-linked glycan on CBH I from the same RUT C30 sample is $Man_7Glc_1GlcNAc_2$ (Stals and Claessens, personal communication). No peak was observed in the reconstructed ion chromatogram for *m/z* 366 of EG II, suggesting the absence of high mannose glycans for this cellulase (Figure 5C). Indeed, subsequent experiments indicated only a single GlcNAc residue at Asn103. Mass accuracy for all observed fragment ions below *m/z* 3000 was generally less than 0.01%. All the suspected tryptic glycopeptides comprising hexose and N-acetyl-hexosamine residues were

Table I. Assignment of tryptic masses of CBH II from RUT-C30

Peak	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed mass (Da)	
				LC-MS	MALDI-TOF
1	130–134	VPSFM	579.3	579.3	—
2*	289–294	NASSPR	630.3	630.3	—
3*	289–294	NASSPR + GlcNAc	833.4	833.4	833.4
4*	135–141	WLDTLDK	889.5	889.6	889.5
5	197–203	NYIDTIR	893.5	893.5	893.7
6	204–213	QIVVEYSDIR	1220.6	1220.6	1220.7
7	142–153	TPLMEQTLADIR	1386.7	1386.7	1386.7
8	130–141	VPSFMWLDTLDK	1450.7	1450.7	1451.0
9	158–174	NGGNYAGQFVVYDLPDR	1883.9	1883.9	1883.7
10	214–234	TLLVIEPDSLANLVTNLGTPK	2207.2	2207.2	2206.8
11*	357–378	QPTGQQQWGDWCNVIGTGFGIR	2504.2	2504.2	2504.6
12*	379–410	PSANTGDSLLDSFVWVKPGGECDGTSDSSAPR	3308.5	3308.5	3309.2
13*	298–327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + GlcNAc	3445.6	3446.5	3446.3
14	158–194	NGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAK	3833.8	—	3834.1
15	411–447	FDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	4105.0	—	4105.2
16	298–327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_7GlcNAc_2$	4783.1	4783.8	4780.4[b]
17*	298–327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_8GlcNAc_2$	4945.1	4946.1	4944.4[b]
18	298–327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_9GlcNAc_2$	5108.2	5108.5	5104.1[b]
19	235–288	CANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGW-PANQDPAAQLFANVYK	5981.8	—	5978.1[b]

*Peptides subjected to MS/MS analyses.
[a]Refers to ^{12}C monoisotopic component.
[b]Average masses observed in MALDI-TOF analyses using sinapinic acid as matrix.
Cysteine is carbamidomethylated.

selected and later confirmed by MS/MS. The percentage sequence coverage for CBH II, EG I, and EG II, including the large heterogeneous linker peptides observed in MALDI-TOF was calculated to be 91.5%, 96.3%, and 80.4%, respectively. Large hydrophobic peptides above *m/z* 4000 were not generally eluted from the C_{18} reversed-phase column.

Among the three putative N-glycan sites of CBH II (Asn14, 289, 310), only Asn310 showed a high mannose. This site is predominantly occupied by $Hex_8GlcNAc_2$ with additional heterogeneity conferred by a single GlcNAc (peaks 13, 16–18 of Table I) as indicated in Figure 5A. The distribution of high-mannose glycoform at Asn310 as deduced from the integrated peak areas of the extracted ion chromatograms is in the ratio of 2:6:5 for Hex_7:Hex_8:Hex_9. The putative site at Asn289 remained primarily unglycosylated with only 18% of this site being occupied by a single GlcNAc (peaks 2–3 in Table I). The Asn14 was not glycosylated as indicated from MALDI analysis of chymotryptic fragments described in the following section. Among all the listed tryptic fragments, peaks 14, 15, and 19 were detected by MALDI-TOF alone. Interestingly, peaks 1 and 4 shown in Table I originate from peak 8 as a result of cleavage between the Met134 and Trp135 amide linkage.

There are six putative N-glycan sites in EG I core (Asn56, 142, 182, 186, 259, 372). A high-mannose glycan, predominantly $Hex_8GlcNAc_2$, was found at either Asn182 or Asn186, with a minor glycoform corresponding to a single GlcNAc in the ratio of 12:1 (peaks 9, 14–16 of Table II). This high-mannose N-glycan is also heterogeneous and the ratio of Hex_7:Hex_8:Hex_9 was calculated to be 1:10:3. The unambiguous assignment of the glycosylation sites was difficult in the present experiment due to the close proximity of the two Asn residues. Further experiments to elucidate the precise glycosylation site are described later. A single GlcNAc was first assigned at Asn56 from the precursor ion mass alone (peak 10 of Table II) and later confirmed by nano-ESMS/MS. There was no mass modification found on the tryptic peptides containing Asn142 and Asn259 indicating the absence of any posttranslational modifications in these putative sites (peaks 8, 18 of Table II). The large hydrophobic tryptic peak 18 was not eluted from the PepMap C18 column but was collected from the Zorbax C8 column and subjected to nano-ESMS/MS analysis to rule out the presence of a glycan at Asn142. No information of glycosylation at Asn372 was successfully obtained because of its proximity to the linker region (Figure 1B).

In contrast, only one putative N-glycan site is expected in EG II core (Asn103) and the corresponding tryptic peptide showed a single GlcNAc addition (peak 13 of Table III). This assignment was further confirmed by nano-ESMS/MS and further indicated that Glu123 is substituted to Asp, described later. Such finding is in agreement with the earlier results from

Table II. Assignment of tryptic masses of EG I from RUT-C30

Peak	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed mass (Da)	
				LC-MS	MALDI-TOF
1	1–13	pEQPGTSTPEVHPK	1387.7	1387.6	1387.5
2	235–247	SYYGPGDTVDTSK	1388.6	1388.6	—
3*[c]	165–177	IPDNLAPGNYVLR	1440.8	1440.8	1440.8
4	109–122	LYLLDSDGEYVMLK	1657.8	1657.8	1657.7
5	22–39	SGGCVAQDTSVVLDWNYR	2025.9	2025.9	2026.2
6	419–437	TCTSGTTCQYSNDYYSQCL	2307.9	2307.9	2308.0
7*	213–234	SCTATACDSAGCGFNPYGSGYK	2329.9	2329.9	2330.0
8*	248–271	TFTIITQFNTDNGSPSGNLVSITR	2582.3	2582.3	2582.6
9*	182–205	NGTLNTSHQGFCCNEMDILEGNSR + GlcNAc	2956.2	2956.2	—
10*	40–68	WMHDANYNSCTVNGGVNTTLCPDEATCGK + GlcNAc	3474.4	3475.2	3475.8
11	273–307	YQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGK	3484.6	3484.7	—
12	272–307	KYQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGK	3612.7	—	3614.0[b]
13	69–108	NCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVSPR	4150.8	4150.8	4153.2[b]
14	182–205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_7GlcNAc_2$	4293.7	4293.8	4297.2[b]
15*	182–205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_8GlcNAc_2$	4455.7	4455.6	4458.4[b]
16	182–205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_9GlcNAc_2$	4617.8	4617.8	4621.2[b]
17	308–363	ALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSN-ILANNPNTHVVFSNIR	6095.8	6097.0	—
18*	123–181	LNGQELSFDVDLSALPCGENGSLYLSQMDENGGAN-QYNTAGANYGSGYCDAQCPVQTWR	6461.8	—	6467.7[b]

*Peptides subjected to MS/MS analyses.
[a]Refers to ^{12}C monoisotopic component.
[b]Average masses observed in MALDI-TOF analyses using sinapinic acid as a matrix.
[c]Unexpected tryptic peptides from EG IV of *T. reesei*.
Cysteine is carbamidomethylated.

stepped orifice-voltage scanning experiment, which predicted the absence of high mannose in EG II. It is also noteworthy that peaks 3 and 6 listed in Table III originate from the same tryptic peptide unexpectedly cleaved at the Tyr168–Cys169 bond. Likewise, peaks 5 and 7 are also from the same tryptic peptide produced by cleavage between two asparagine residues (Asn305–Asn306).

The capillary LC/ESMS/MS analyses of tryptic peptides from CBH II, EG I, and EG II also confirmed the occurrence of unexpected fragments arising from nonspecific cleavage sites. A list of all proteolytic peptides observed is presented in Tables I–III. Interestingly, the sequence IPDNLAPGNYLR listed in peak 3 of EG I and peak 4 of EG II (Tables II and III) was identified as a EG IV tryptic peptide by nano-ESMS/MS sequencing followed by protein database search and will be discussed later.

Chymotryptic digestion

Because no further information on the assignment of N-glycans to the remaining sites in CBH II and EG I was possible from analyses of tryptic peptides, subsequent experiments used chymotrypsin to obtain glycopeptide fragments of convenient size surrounding Asn14 from CBH II and Asn372 of EG I. Chymotrypsin digestion typically results in numerous small peptides that complicate the mass assignments. The observed chymotryptic peptides of CBH II from MALDI-TOF analysis were matched with the expected masses and these results are presented in Table IV. The chymotryptic fragment from peak 8 has two possible matches and no definite assignment could be made from these mass measurements alone. Table IV also shows a chymotryptic fragment at 905.3 (peak 9) consistent with the peptide GQCG-GQNW with an unmodified Asn14. Similarly no modification was observed on Asn289 as evidenced by the chymotryptic peptide at 942.5 Da (peak 10 of Table IV).

Similar mass spectral analyses were conducted on EG I to identify the putative N-linked site at Asn372. Unfortunately, the expected chymotryptic fragment contains a truncated part of the linker region Gly365–Phe386 modified with short O-linked di- and trisaccharides, thus preventing the unambiguous identification of the N-linked glycan chain.

Nano–ESMS-MS analysis of tryptic peptides

To complete the assignment of glycosylation sites on CBH II, EG I, and EG II, the separation of large hydrophobic peptides >4000 Da was conducted on a larger-bore C8 column (2.1 × 150 mm) to fractionate potential glycopeptides for subsequent nanoelectrospray tandem mass spectrometry (nano–ESMS-MS) analyses.

The variability in site occupancy at Asn310 of CBH II was confirmed by conducting product ion scan on peaks 13 and 17 indicated in Table I (Figure 6). By using relatively low collision

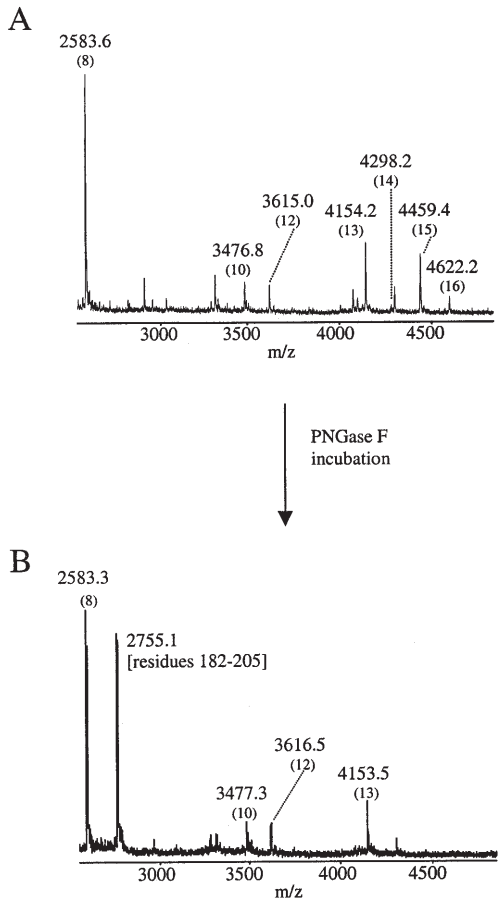

Fig. 4. MALDI-TOF analysis of EG I tryptic digest (**A**) before and (**B**) after PNGase F incubations. Numbers in parentheses underneath the indicated masses correspond to peptides listed in Table II.

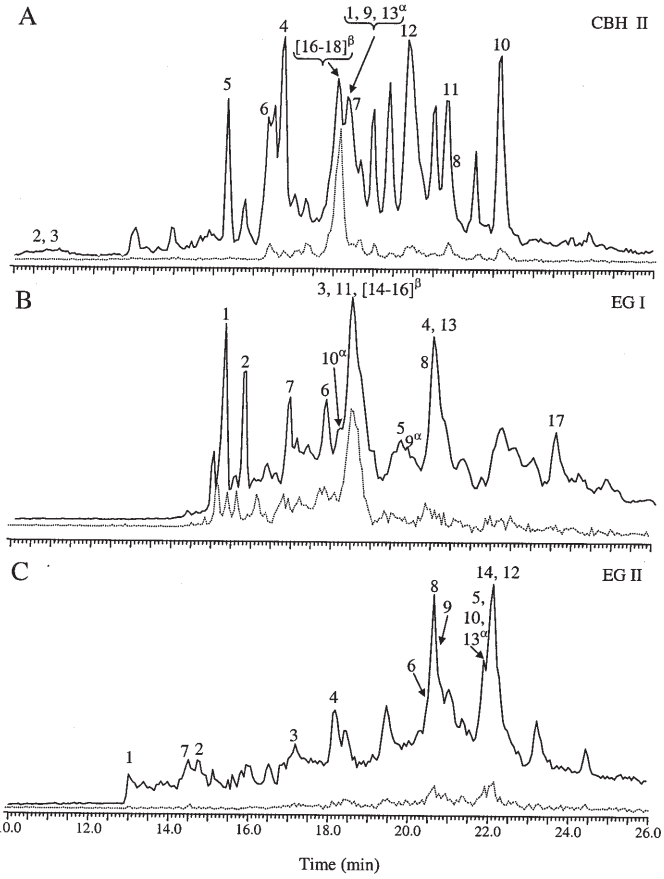

Fig. 5. Capillary LC/ESMS analyses of tryptic digests of purified cellulases from RUT-C30. (**A**) CBH II; (**B**) EG I; and (**C**) EG II. The numbers with the corresponding peptide assignment are shown in Tables I–III, respectively (numbers with α superscript represent peptides with single GlcNAc attachment, whereas β are those with high mannose). Solid lines represent the high mass acquisition chromatograms (*m/z* 400–2000) at 30 V sampling orifice voltage, and dotted lines are extracted mass chromatograms for *m/z* 366 at 120 V orifice voltage. Separation conditions: Pepmap C_{18} reversed-phase column (15 cm × 350 μm ID), linear gradient of 5–95% acetonitrile (0.2% formic acid) in 35 min delivering at a flow rate of 3.5 μl/min.

energy (30 eV), both the quadruply charged ion series (*m/z* 1196.9, 1156.4, 1115.9, 1075.4, 1034.9, 994.4) and the singly charged series of peaks (*m/z* 1176.3, 1014.3, 852.2, 690.2, 528.1, 366.1) corresponding to the loss of hexose and N-acetylhexosamine units were clearly observed (Figure 6A). This characteristic fragmentation pattern together with the diagnostic oxonium ions at *m/z* 163.1, 204.1, and 366.1 and the mass of the precursor ion further support the proposal of a high-mannose oligosaccharide $Hex_8GlcNAc_2$ is attached to Asn310. Though the Hex_8 glycoform is the most abundant oligosaccharide chain attached at this site, some microheterogeneity in the glycan distribution ranging from Hex_7 to Hex_9 is observed. The occurrence of $Hex_8GlcNAc_2$ as the major glycoform is consistent with the earlier results from CBH I (Hui *et al.*, 2001).

A later-eluting peak with a molecular mass corresponding to peptide Gly298–Lys327 of CBH II with a single GlcNAc was assigned as peak 13 in Table I. The product ion spectrum of the corresponding triply charged ion at *m/z* 1149.8 shows a series of y-type fragment ions together with an abundant oxonium ion for GlcNAc at *m/z* 204.1 (Figure 6B). The appearance of y-type ions modified by a single GlcNAc starting from y_{18} was consistent with a modified Asn310 residue. Similarly, the nature of the N-linked glycans at putative site Asn289 was confirmed by nano-ESMS/MS analysis. Both the native peptide Asn289–Arg294 (peak 2 of Table I) and the suspected glycopeptide with single GlcNAc attachment (peak 3 of Table I) were selected as precursor ions for product ion scan experiments. The two spectra show very similar fragmentation patterns from y_1 to y_6 ions, with the exception that an abundant oxonium at *m/z* 204.1 was observed in the glycosylated peptide (data not shown). The ratio of tryptic peptide Asn289–Arg294 with and without GlcNAc was 2:9 as calculated from extracted chromatogram of the corresponding ions. In the first publication of CBH II 3D structure, Rouvinen and co-workers reported the presence of glycosylation at Asn289 and Asn310 in CBH II, though no precise structural assignment on the N-linked oligosaccharides was made (Rouvinen *et al.*, 1990).

The assignment of high mannose in EG I is more challenging due to the presence of two proximal N-linked sites (Asn182, 186). The product ion spectrum of the quadruply charged ion of the suspected glycopeptide at *m/z* 1114.9 (Figure 7A), comprised a series of triply charged fragment ions between *m/z* 986.7 and 1432.2 corresponding to consecutive losses of seven mannose residues together with the cleavage of a GlcNAc at *m/z* 1054.5 (Figure 7A). The precise location of the N-linked site was achieved by using Edman sequencing of the

Table III. Assignment of tryptic masses of EG II from RUT-C30

Peak	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed mass (Da)	
				LC-MS	MALDI-TOF
1	205–210	YASQSR	710.3	710.3	710.2
2	97–102	VYPPLK	715.4	715.4	715.3
3*	169–178	CIVDIHNYAR	1259.6	1259.6	1259.5
4* [c]	165–177	IPDNLAPGNYVLR	1440.8	1440.8	1440.7
5*	306–319	NIDGAFSPLATWLR	1559.8	1559.8	1559.8
6	155–168	YDQLVQGCLSLGAY	1585.7	1585.8	—
7*	290–305	YLDSDNSGTHAECTTN	1783.7	1783.6	1783.7
8	74–96	FAGVNIAGFDFGCTTDGTCVTSK	2424.1	2424.0	2424.1
9	131–154	LPVGWQYLVNNNLGGNLDSTSISK	2588.3	2588.3	2587.7
10	179–204	WNGGIIGQGGPTNAQFTSLWSQLASK	2717.4	2717.3	2717.4
11	211–239	VWFGIMNEPHDVNINTWAATVQEVVTAIR	3309.7	—	3309.7
12	290–319	YLDSDNSGTHAECTTNNIDGAFSPLATWLR	3325.5	3325.4	3326.6
13*	103–130	NFTGSNNYPDGIGQMQHFVNDDGMTIFR + GlcNAc	3377.5	3377.4	3377.5
14	240–289	NAGATSQFISLPGNDWQSAGAFISDGSAAA-LSQVTNPDGSTTNLIFDVHK	5077.4	5077.0	5079.9[b]

*Peptides subjected to MS/MS analyses.
[a]Refers to ^{12}C monoisotopic component.
[b]Average mass observed in MALDI-TOF analysis.
[c]Unexpected tryptic peptides from EG IV of *T. reesei*.
Cysteine is carbamidomethylated.

purified glycopeptide peak because the second-generation fragment ions of tryptic peptides with a single GlcNAc failed to provide unambiguous assignment. This analysis revealed that Asn182 and not Asn186 was modified with a N-linked high-mannose glycan chain.

The putative N-linked site at Asn56 of EG I was suspected to contain a single GlcNAc, and this proposal was confirmed by selecting the corresponding $[M+3H]^{3+}$ at *m/z* 1159.4 as a precursor ion (Figure 7B). The initial y_1–y_{12} ions were matched, and the GlcNAc attachment was observed starting from the fragment ion y_{13} onward consistent with a modifed Asn56. Kleywegt and co-workers previously reported the occupancy of single GlcNAc at position Asn56 and Asn182 by electron-density mapping (Kleywegt *et al.*, 1997). Our results also support that the same sites are occupied except that in EG I from RUT-C30, Asn182 comprised a high-mannose glycan. This difference could be accounted for by variability in endogenous endo H activity of *T. reesei*.

The only putative N-glycan site at Asn103 in EG II was suspected to be modified with a single GlcNAc residue (Figure 8A). Interestingly, the mass of the precursor ion is 14 Da less than the calculated molecular mass of the peptide Asn103–Arg130 (peak 13 of Table III). The product ion spectrum of the $[M+3H]^{3+}$ at *m/z* 1126.8 showed a prominent GlcNAc oxonium ion at *m/z* 204.1 and a series of y-type ions from y_1 to y_7. Starting from y_8 ion onward, the observed masses were 14 Da less than that expected from the peptide sequence. The y_8 ion corresponds to a glutamic acid residue from the C-terminus and the 14-Da shift suggests a substitution of the glutamic acid by an aspartic acid. Therefore, MS/MS of glycopeptide (*m/z* 1126.8) from EG II confirmed both a single GlcNAc at Asn103 as well as a single amino acid substitution from Glu123 to Asp123. This was confirmed by DNA sequence analysis showing GAC codon (Asp) for the amino acid 123 instead of a GAG (Glu).

A number of unexpected tryptic peptides were also identified by MS/MS sequencing. Interestingly, an unique sequence that was later found to be EG IV tryptic peptides by performing protein data search via the European Molecular Biology Laboratory (EMBL) peptideSearch program (http://www.narrador.embl-heidelberg.de/GroupPages/homepage.html), was observed in both EG I and EG II (Figure 8B). It is not clear why only one EG IV tryptic peptide was observed, but trace amounts of unstable EG IV could have been copurified with this preparation.

Peaks 11 and 12 in Table I originated from the same CBH II tryptic peptide (Gln357–Arg410) as a result of an unexpected cleavage between Arg378 and Pro379 (data not shown). Cleavage of the Arg–Pro bond is very unusual in tryptic digest, although examples have been reported in the literature, such as in the sequence –Trp–Arg–Pro–Ala– in bovine erythrocyte carbonic anhydrase (Nyman *et al.*, 1966). The current sequence is –Ile–Arg–Pro–Ser– and does not display any consensus domain with the previously mentioned example. Another unexpected cleavage occurred between the Asn305 and Asn306 of EG II tryptic peptide Tyr290–Arg319 (peaks 5 and 7 in Table III). This cleavage may be attributed to the enzymatic activities of small amount of asparagine protease present in the original crude enzymes.

Discussion

This is a first report on the assignment and characterization of N-linked glycans in CBH II, EG I, and EG II purified from a

Table IV. Assignment of chymotryptic masses of CBH II (RUT-C30) observed by MALDI-TOF analysis

Peak	Residues	Assignment	$Mass_{mono}$[a] (Da)	Observed mass (Da)
1	432–435	QAYF	527.2	527.2
2	343–347	SNAFF	584.3	584.3
3	94–99	VGVTPW	657.4	657.4
4	325–329	NEKLY	665.3	665.4
5	337–342	LANHGW	696.3	696.4
6	330–336	IHAIGPL	719.4	719.4
7	243–248	LECINY	810.4	810.4
8	263–270	DAGHAGWL	825.4	825.4
	262–269	LDAGHAGW	825.4	
9	8–15	GQCGGQNW	905.4	905.3
10	288–296	KNASSPRAL	942.5	942.5
11	186–195	SIADGGVAKY	979.5	979.5
12	297–306	RGLATNVANY	1077.6	1077.6
13	170–179	DLPDRDCAAL	1144.5	1144.5
14	199–209	IDTIRQIVVEY	1347.7	1347.8
15	348–364	ITDQGRSGKQPTGQQQW	1913.9	1913.9
16	347–364	FITDQGRSGKQPTGQQQW	2061.0	2060.9
17	348–367	ITDQGRSGKQPTGQQQWGDW	2272.1	2271.9

[a]Refers to ^{12}C monoisotopic component.
Cysteine is carbamidomethylated.

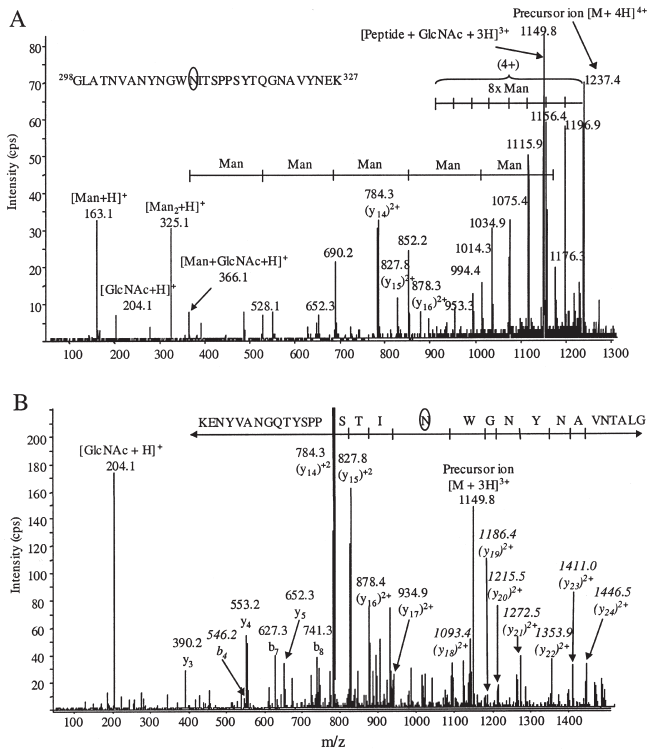

Fig. 6. Nano-ESMS/MS spectra of N-linked glycopeptides of CBH II tryptic digest with oligosaccharide attached at Asn310. (**A**) Product ion scan of *m/z* 1237.4 (peak 17 of Table I) to confirm the presence of high mannose at Asn310; (**B**) production ion scan of *m/z* 1149.8. (peak 13 of Table I) to confirm single GlcNAc at Asn310. The y-type ions in italics indicate addition of a GlcNAc. A collision energy (laboratory frame of reference) of 140 eV was used in (**A**) and 150 eV in (**B**).

fermentation broth of *T. reesei* strain RUT-C30. ESMS was used initially to probe the complexity of glycosylation. A number of enzymatic digests of the cellulase proteins, including papain, trypsin, chymotrypsin, endo H, and PNGase F, were used to facilitate the elucidation of the structures of the N-linked oligosaccharides. These provided an exhaustive assignment of the majority of the putative N-linked glycan sites (see Tables I–III).

Only two of the three potential N-glycosylation sites in the CBH II sequence were found to be glycosylated. Asn310 contained mainly a high-mannose-type glycan ($Hex_{7–9}GlcNAc_2$) in a 3:1 ratio with a single GlcNAc. Asn289 was primarily unglycosylated with a small fraction (18%) of the CBH2 proteins bearing a single GlcNAc at this site. Asn14, located within the CBD, was not glycosylated. The presence of glycans at Asn289 and Asn310 is consistent with the previously reported 3D structure analysis of the CBH II catalytic domain (Rouvinen *et al.*, 1990), though only a single GlcNAc was observed at Asn310. This may simply be a result of the high-mannose-containing glycoforms not crystallizing. The crystal structure also identified O-linked mannose residues at Thr87 and 97 and at Ser106, 109, 110, and 115. However, the absence of basic amino residues between these O-linked sites prevented the determination of the nature of glycans at each residue. Information from papain cleavage sites and from N-linked glycan mass allow us to calculate indirectly the number of O-linked hexoses appended to the linker peptide to be ranged from 39 to 46. However, analyses of the chymotryptic digests indicated that Thr97 is primarily unglycosylated as the native peptide for amino acids 94–99 was observed by MALDI-TOF analysis (Table IV).

Three of the six potential N-glycosylation sites in EG I were unglycosylated: Asn142, 186, and 259. Also, a single GlcNAc was found at Asn56. The N-glycan at Asn182 was primarily a high-mannose structure $Hex_{7–9}$ $GlcNAc_2$ with a small percentage of the molecules displaying a single GlcNAc at this site (8%). The 3D structure of EG I also identified single GlcNAc modifications at Asn56 and Asn182 (Kleywegt *et al.*, 1997). Assignment of a specific glycan structure to Asn372 was not possible due to its proximity to the heavily O-glycosylated linker region. Finally, the single N-glycosylation site in the EG II sequence, Asn103, was found to contain a single GlcNAc.

The heterogeneity of the glycan structures found on these three *Trichoderma* cellulases, as well as for CBH I from the same strain (Hui *et al.*, 2001), is indicative of the presence of several glycan-trimming enzyme activities. The detection of $Hex_{7–9}GlcNAc_2$ structures is indicative of mannosidase trimming. The gene encoding the α-1,2-mannosidase has been cloned from *T. reesei* and the recombinant enzyme produced from *S. cerevisiae* characterized (Maras *et al.*, 2000). This purified recombinant enzyme was characterized as a class I mannosidase based on its ability to trim $Man_{6–9}GlcNAc_2$ structures to $Man_5GlcNAc_2$. This *in vitro* activity is consistent with the release of $Man_5GlcNAc_2$ structures from purified CBH I and EG I catalytic domain observed by other groups (Maras *et al.*,

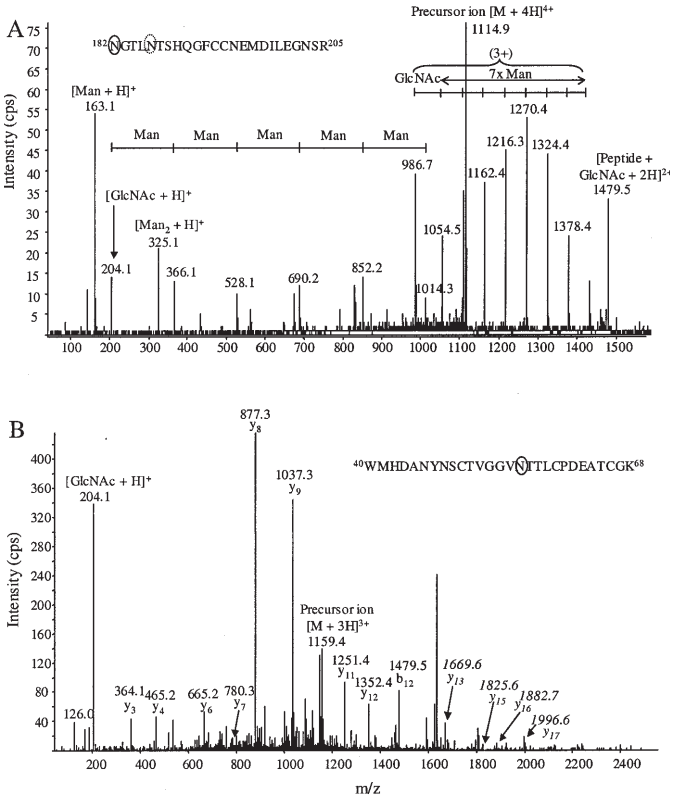

Fig. 7. Nano-ESMS/MS spectra of N-linked glycopeptides of EG I tryptic digest. (**A**) Product ion scan of *m/z* 1114.9 (peak 15 of Table II) to confirm the presence of high mannose at either Asn182 or Asn186; (**B**) production ion scan of *m/z* 1159.4 (peak 10 of Table II) to confirm the single GlcNAc at Asn56. The y-type ions in italics indicate addition of a GlcNAc. A collision energy (laboratory frame of reference) of 140 eV was used in (**A**) and 150 eV in (**B**).

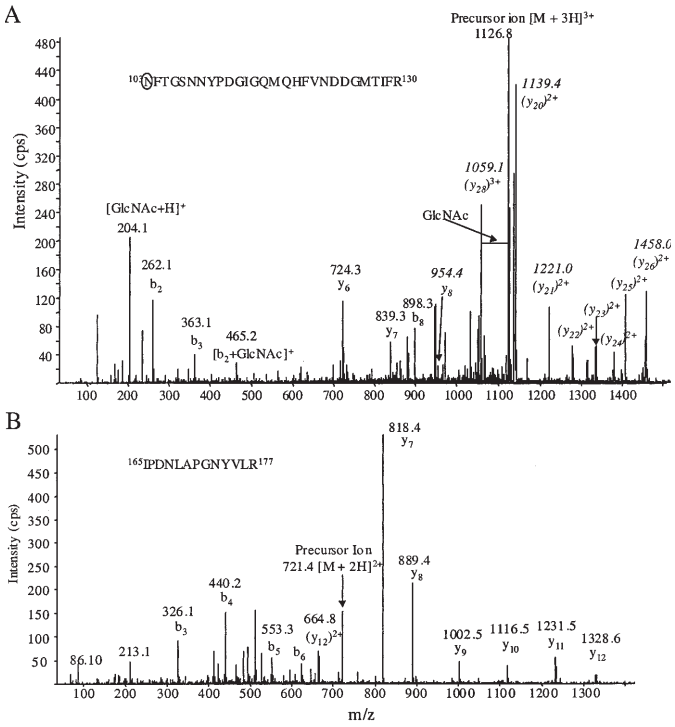

Fig. 8. Nano-ESMS/MS spectra of N-linked glycopeptides of EG II tryptic digest. (**A**) Product ion scan of *m/z* 1126.8 (peak 13 of Table III) confirming the single GlcNAc at Asn103. The y-type ions in italics indicate a decrease in 14 Da from Glu123 to Asp123. (**B**) Product ion scan of *m/z* 721.4 (peak 3 of Table II and peak 4 of Table III). A collision energy (laboratory frame of reference) of 150 eV was used in for (**A**) and 80 eV in (**B**).

1997; Salovouri *et al.*, 1987; Garcia *et al.*, 2001) By contrast, $Man_5GlcNAc_2$ was not identified as a major N-linked glycan on CBH II, EG I, or EG II from the RUT-C30 fermentation broth used in the present study. In addition, other groups have identified glucosylated and phosphorylated high-mannose structures on CBH I and EG I produced by various *Trichoderma* strains (Maras *et al.*, 1997; Garcia *et al.*, 2001).

The differences in the glycan structures identified here compared to previous investigations probably result from the difference in enzyme production methods used. The enzymes analyzed in the previous reports were produced in small-scale flask cultures in defined media. The enzymes analyzed in the present study were produced in a highly aerated fed-batch fermentation using a rich medium; as a result, the productivity of that system is much higher. The significance of this is as follows. It is possible that the residence time of the cellulases in the endoplasmic reticulum and Golgi is too short in a high-productivity fermentation to allow for extensive trimming or modification by resident mannosidases or mannosylphospho-transferases. By contrast, it is possible that the cellulases reside in the endoplasmic reticulum and Golgi for longer times in a lower-productivity shake flask culture on defined medium, and this could result in more extensive glycan processing. This would be consistent with the observation that cleavage of the first mannose from $Man_9GlcNAc_2$ by the *Trichoderma* mannosidase occurs at a faster rate than cleavage of the next three mannoses (Van Petegem *et al.*, 2001).

Conversely, the lower rate of protein production in the shake-flasks will facilitate transferase activity. Alternatively, other reports have suggested that mannosylphosphate transfer occurs in yeast under conditions of oxidative or nutritional stress (Jigami and Odani, 1999). The higher aeration rate and richer medium used for the fed-batch fermentation would tend to minimize this transferase activity. However, we have previously detected a mannosylphosphate on the O-linked glycans of the linker peptide of CBH 1 (Hui *et al.*, 2001). The fact that we did not observe this modification on the CBH II, EG I, or EG II proteins purified from the same fermentation broth is not necessarily surprising considering that CBH I can constitute over 80% of the extracellular protein produced by RUT-C30 under fed-batch fermentation conditions. CBH II, EG I, and EG II constitue a much smaller percentage of the total cellulase and the amount of phosphorylated CBH II, EG I, or EG II would be very minor and difficult to detect and purify.

The identification of single GlcNAc structures linked to various sites on all three enzymes is strongly suggestive of an endo H–type activity present within the *Trichoderma* secretory pathway or secreted into the extracellular medium. Similarly, the observation that a single site can either have a single GlcNAc or no glycan is indicative of N-glycosidase activity or, possibly, partial lack of glycosylation within the endoplasmic reticulum. The identification and characterization of the endo H and N-glycosidase activities along with the localization of these enzymes and the mannosidase within *Trichoderma* cells

will lead to a better understanding of glycan synthesis and trimming by filamentous fungi.

Materials and methods

Culture and growth conditions

T. reesei crude enzyme from a RUT-C30 fermentation was provided by Iogen (Ottawa, ON, Canada). It was grown on potato dextrose agar at 28–30°C until a confluent lawn of spores was obtained. Spores were collected and used to inoculate 750 ml Berkeley media (10 g/L glucose, 1.4 g/L $(NH_4)_2SO_4$, 2.0 g/L KH_2PO_4, 0.31 g/L $MgSO_4 \cdot 7H_2O$, 0.53 g/L $CaCl_2$, 11.6 g/L wet corn steep liquor, 5.1 g/L dry corn steep, 5 mg/L $FeSO_4 \cdot 7H_2O$, 0.8 mg/L $MnSO_4 \cdot H_2O$, 0.7 mg/L $ZnSO_4 \cdot 7H_2O$) in a 2-L baffled flask. After 3 days of growth at 28°C and 150 rpm, this culture was used to inoculate 10 L of fermentation medium with the following initial composition: 10 g/L glucose, 4.4 g/L $(NH_4)_2SO_4$, 2.77 g/L KH_2PO_4, 1.4 g/L $MgSO_4 \cdot 7H_2O$, 0.37 g/L $CaCl_2$, 6 g/L dry corn steep, 3.75 mg/L $FeSO_4 \cdot 7H_2O$, 1.2 mg/L $MnSO_4 \cdot H_2O$, 1.05 g/L $ZnSO_4 \cdot 7H_2O$. A fed-batch aerobic fermentation using a proprietary soluble inducing carbohydrate source was run for 6 days at pH 4.0 and 28–30°C in a 14-L New Brunswick Microferm fermentor. After 6 days, the culture was filtered over Harborlite and the culture filtrate adjusted to pH 4.5 and preserved with 0.5% benzoate to prevent microbial growth. The total secreted protein was 24 g/L.

CBH II purification

Purification of CBH II from RUT-C30 was carried out using two cation exchangers: methyl sulfonate sepharose followed by carboxymethyl sepharose. Both exchangers were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). CBH I was initially removed according to a procedure described previously (Hui *et al.*, 2001), with the exception that the starting material was increased to approximately 50 mg of protein and a 5-ml DEAE sepharose volume bed was used. The unretained fraction enriched in CBH II and EGs was collected, concentrated, and desalted with Biomax-5 filter unit from Millipore (Bedford, MA). This fraction was loaded onto a poly-prep column from Bio-Rad (Hercules, CA) packed with 5-ml volume bed of sulfonate sepharose previously equilibrated with 3 mM acetate buffer, pH 4.0. Detection of cellulolytic components was monitored online using an UV detector set at 280 nm wavelength. After elution of the unretained components, the buffer was changed to a high-salt acetate buffer (50 mM sodium acetate, pH 4.0, 1 M NaCl) to elute the CBH II–enriched fraction. This fraction was again concentrated and desalted using the Biomax-5 filter unit and was further purified using carboxymethyl sepharose (2 ml volume bed) equilibrated with 6 mM sodium acetate, pH 4. The unretained components were first removed, followed by elution of CBH II using 6 mM sodium acetate buffer with 40 mM NaCl, pH 4. The collected CBH II fraction was again concentrated and desalted.

Chromatofocusing purification of EG I and EG II

Chromatofocusing was performed using the Pharmacia Polybuffer exchanger PBE 94 and Polybuffer 74 (pH 4–7) adjusted to pH 4 as elution buffer. The PBE 94 was mixed with small amount of starting buffer (0.025 imidazole-HCl, pH 7.4), degassed, and slowly poured into a column of 20 cm in length and 4 ml volume capacity. Prior to the purification, the PBE 94 was equilibrated with 10–15 columns of the starting buffer flowing at 0.2 ml/min, followed by 1 column volume of the elution buffer. The unretained fraction from sulfonate sepharose (enriched in EGs) was slowly introduced into the column, and the Polybuffer was maintained at a flow rate of 0.2 ml/min. Detection of cellulase elution was monitored online using a UV detector set at 280 nm wavelength. Collected fractions were concentrated using the Biomax-10 filter (Millipore) to remove the Polybuffer ampholytes. EG II enriched fraction was collected in about 15 min of Polybuffer introduction and the EG I enriched fraction in about 90 min.

Enzymatic digestion of cellulases

The tryptic digestion, papain, and PNGase F incubations of cellulases were conducted as described previously (Hui *et al.*, 2001). Typically 200–500 µg of purified cellulase was used as the starting material for tryptic digestion. In the case of papain digestion, the enzyme-to-substrate ratio of CBH II was increased to 1:10 (w/w), whereas a ratio of 5:1 (w/w) was used for EG I and EG II.

The endo H digestion was performed on CBH II papain digest from RUT-C30. Approximately 50 µg of starting material was freeze-dried and redissolved in 45 µl 50 mM ammonium acetate at pH 5.5 and 5 µl 10× G5 buffer supplied by New England BioLabs (Beverly, MA). An aliquot of 5 µl endo H (New England BioLabs), equivalent to 2500 units, was added, and the mixture was incubated at 37°C for 5 h.

Chymotryptic digestions were performed on reduced and alkylated cellulases. Approximately 20 µg cellulase component was freeze-dried and redissolved in 20 µl 100 mM ammonium bicarbonate, 0.1 mM $CaCl_2$, pH 7.8. An aliquot of 1 µl chymotrypsin prepared as 1 mg/ml in 1 mM HCl was introduced, resulting in an enzyme-to-substrate ratio of 1:20. The mixture was incubated at 37°C overnight. Lyophilized sequencing-grade chymotrypsin was purchased from Roche Diagnostics (Laval, Quebec, Canada).

Mass spectrometry

All MALDI-TOF mass spectra were acquired on a Voyager Elite STR Biospectrometry Workstation (PerSeptive Biosystems, Framingham, MA) equipped with delayed extraction. Sinapinic acid was used as a MALDI matrix for protein analysis and was prepared at a concentration of 10 mg/ml in a mixture of methanol-acetonitrile–deionized water (1:1:1, v/v/v). For peptide detection, 2,5-DHB was used as a matrix and was prepared as 0.2 M solution in 50:50 methanol:deionized water. A nitrogen laser (337 nm) operating at a frequency of 10 Hz was used for ionization.

For the intact cellulase and papain digests analyses, a 1-µl aliquot of a 1–5 mg/ml solution was mixed with 10 µl sinapinic acid matrix solution, and an aliquot of 0.5 µl was deposited on the MALDI plate. These sample spots were analyzed in linear positive ion mode with typical conditions of 25 kV accelerating voltage, 91% grid voltage, 0.3% guide wire voltage, and a delay time varying from 200 to 300 ns. External mass calibration was carried out using the singly protonated ions of insulin (5734.59 Da) and apo-myoglobin (16,952.56 Da) for papain digest, and the singly and doubly protonated ions of bovine serum albumin with average masses of 66,431 Da and 33,216 Da, respectively, for the analyses of intact cellulase components. For peptide digests' analyses, a matrix-to-sample (1 mg/ml)

ratio of 2:1 was prepared and an aliquot of 0.5 µl was applied on the MALDI plate. These sample spots were analyzed in reflectron positive ion mode with 20 kV accelerating voltage, 72% grid voltage, 200 ns delay time, and the guide wire voltage varying from 0.02% to 0.05%. External calibration was carried out using the protonated molecular ions of des-Arg_1-bradykinin (904.4681 Da) and PTH 1–31 Amide (3716.9845 Da).

ESMS

ESMS experiments were acquired using a PE/Sciex Q-Star (Thornhill, Ontario, Canada) hybrid quadrupole/time-of-flight instrument. Flow injection analyses were obtained by dissolving the purified cellulase to a concentration of 500 µg/ml in 50% acetonitrile and 1% acetic acid. A 1-µl injection volume was injected into a stream of 50% acetonitrile (0.2% acetic acid) introduced at a flow rate of 10 µl/min to the mass spectrometer. The orifice voltage was varied between 90 and 100 V, and the ring voltage was adjusted between 300 and 350 V. Mass spectra were typically collected over *m/z* 1600–3000.

Capillary LC-ESMS analysis

A capillary high-performance LC system coupled to a Q-TOF2 hybrid quadrupole/time-of-flight instrument (Micromass, Manchester, UK) was used for online capillary LC/ESMS experiments. Chromatographic separations were achieved on a 15 cm × 0.32 mm PepMap C_{18} capillary column (LC Packings, San Francisco, CA) using a linear gradient elution of 5–95% acetonitrile (0.2% formic acid) in 35 min. A constant flow rate of 3.5 µl/min was used. Conventional mass spectra were obtained by operating the quadrupole in a RF-only mode while a pusher electrode was pulsed to transfer all ions to the time-of-flight analyzer. Mass spectra were acquired using stepped orifice-voltage scanning similar to that described previously (Hui *et al.*, 2001; Bateman *et al.*, 1998; Carr *et al.*, 1993). In the present work, the sampling orifice voltage was maintained at 120 V during low mass acquisition (*m/z* 150–400) with a collection time of 1 s and 30 V during high mass scanning (*m/z* 400–2000) with a collection time of 2 s. A delay of 0.1 s was set between acquisition modes. Data were acquired and processed in the MassLynx Window NT based data system version 3.4.

Nano–ESMS-MS

Type A nanoelectrospray glass tips were purchased from Micromass. Typically, the tips were carefully loaded with 1–2 µl tryptic peptides prepared in 50% methanol, 0.2% formic acid; a nanoelectrospray voltage of 1000–1200 V was applied for ionization. In MS/MS experiments, glycopeptide precursor ions were selected by the first quadrupole while a pusher electrode was pulsed at a rate of 62 µs to transfer fragment ions formed in the RF-only hexapole cell to the time-of-flight analyzer. Collisional activation was performed using argon collision gas with a typical offset voltage of 25–50 V between the DC voltage of the entrance quadrupole and the RF-only hexapole cell. Data were acquired and processed in the MassLynx Window NT–based data system version 3.4.

Fraction collection

A Zorbax C8 from Agilent Technologies (Palo Alto, CA) reversed-phase column (2.1 mm × 15 cm) was used for fraction collection with a flow rate of 0.2 ml/min and a linear gradient elution of 5–95% acetonitrile (0.1% trifluroacetic acid) in 35 min. In this case, the cellulase tryptic digests (1 mg/ml) was manually injected via a Rheodyne six-port valve (Rohnert Park, CA) fitted with a 5-µl sample loop into an HPLC HP1100 system (Hewlett-Packard, Palo Alto, CA). The elution of peptides was monitored by an online UV detector. Each fraction was then analyzed by MALDI-TOF, and all suspected glycopeptides were subjected to nano-ESMS/MS analysis.

Edman degradation

Sequencing of the EG I tryptic peptide Asn182–Arg205 was achieved using the Applied Biosystems model 491 Procise sequencing system (Foster City, CA).

Database searching

All the amino acid sequence information of *T. reesei* was obtained from Swiss-Prot database (http://www.expasy.ch/sprot). The accession numbers for CBH I, CBH II, EG I, and EG II are P00725, P07987, P07981, and P07982, respectively. Unidentified sequences from MS/MS analyses were further examined using peptideSearch developed by EMBL (http://www.narrador.embl-heidelberg.de/GroupPages/homepage.html) or MS-Seq search engine from ProteinProspector by UCSF (http://prospector.ucsf.edu). Typically, the peptide sequence tag approach was used where a partial amino acid sequence together with the molecular mass of the peptide were entered to match for potential proteins available in the database. The searching parameters were limited to tryptic peptides with all cysteine residues modified to carbamidomethylated cysteine. A protein mass range from 0 to 300 kDa was used; the peptide mass accuracy was less than 0.5 Da, and all origins of species were included in the search.

Acknowledgments

Joseph P.M. Hui acknowledges the Fonds pour la Formation de Chercheurs et l'Aide á la Recherche (FCAR) for a graduate student scholarship. The authors would like to thank Iogen for their financial support, Pat Lanthier (IBS, NRC) for performing monosaccharide and oligosaccharide analyses, Dave Watson (IBS, NRC) for Edman degradation amino acid sequencing, Genny Giroux (Iogen) for technical assistance, and Roger MacKenzie and Tomoko Hirama (IBS, NRC) for their assistance in cellulase purification. This article is contribution number 42463 from NRC.

Abbreviations

CBD, cellulose binding domain; CBH, cellobiohydrolase; DHB, dihydroxybenzoic acid; EG, endoglucanase; HPAEC/PAD, high-performance anion exchange chromatography/pulsed amperometric detection; LC-ESMS, liquid chromatography electrospray mass spectrometry; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MS/MS, tandem mass spectrometry; nano–ESMS-MS, nanoelectrospray tandem mass spectrometry.

References

Bateman, K.P., White, R.L., Yaguchi, M., and Thibault, P. (1998) Characterization of protein glycoforms by capillary-zone electrophoresis—nanoelectrospray mass spectrometry. *J. Chromatogr. A*, **794**, 327–344.

Carr, S.A., Huddleston, M.J., and Bean, M.F. (1993) Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci.*, **2**, 183–196.

Clarke, A.J. (1997) Biodegradation of cellulose. In *Enzymology and biotechnology*. Technomic Publishing, Pennsylvania, p. 55.

Conesa, A., Punt, P.J., van Luijk, N., and van den Hondel, C.A.M.J.J. (2001) The secretion pathway in filamentous fungi: a biotechnological view. *Fung. Genet. Biol.*, **33**, 155–171.

Divine, C., Ståhlberg, J., Reinikaninen, T., Rouhonen, L., Petterson, G., Knowles, K.J.C., Teeri, T.T., and Jones, T.A. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science*, **265**, 524–528.

Garcia, R., Cremata, J.A., Quintero, O., Montesino, R., Benkestock, K., and Ståhlberg, J. (2001) Characterization of protein glycoforms with N-linked neutral and phosphorylated oligosaccharides: studies on the glycosylation of endoglucanase I (Cel7B) from *Trichoderma reesei*. *Biotech. Appl. Biochem.*, **33**, 141–152.

Harrison, M.J., Nouwens, A.S., Jardine, D.R., Zachara, N.E., Gooley, A.A., Nevalainen, H., and Packer, N.H. (1998) Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *Eur. J. Biochem.*, **256**, 119–127.

Henrissat, B., Driguez, H., Viet, C., and Schülein, M. (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Bio/Technology*, **3**, 722–726.

Hui, J.P.M., Lanthier, P., White, T.C., McHugh, S.G., Yaguchi, M. Roy, R., and Thibault, P. (2001) Characterization of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. *J. Chromatogr. B*, **752**, 349–368.

International Union of Biochemistry and Molecular Biology. (1992) *Enzyme Nomenclature*. Academic Press, New York.

Jigami, Y. and Odani, T. (1999) Mannosylphosphate transfer to yeast mannan. *Biochim. Biophys. Acta*, **1426**, 335–345.

Klarskov, K., Piens, K., Ståhlberg, J., HØj, P.B., Van Beeumen, J., and Claeyssens, M. (1997) Cellobiohydrolase I from *Trichoderma reesei*: identification of an active-site nucleophile and additional information on sequence including the glycosylation pattern of the core protein. *Carbohydr. Res.*, **304**, 143–154.

Kleywegt, G.J., Zou, J.Y., Divne, C., Davies, G.J., Sinning, I., Ståhlberg, J, Reinikainen, T., Srisodsuk, M., Teeri, T.T., and Jones, T.A. (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.*, **272**, 383–397.

Kubicek, C.P. (1992) The cellulase proteins of *Trichoderma reesei*: structure, multiplicity, mode of action and regulation of formation. *Adv. Biochem. Eng. Biotechnol.*, **45**, 1–25.

Macarron, R., van Beeumen, J., Henrissat, B., de la Mata, I., and Claeyssens, M. (1993) Identification of an essential glutamate residue in the active site of endoglucanase III from *Trichoderma reesei*. *FEBS Lett.*, **316**, 137–140.

Maras, M., Callewaert, N., Piens, K., Claeyssens, M., Martinet, W., Deawele, S., Contreras, H., Dewerte, I., Penttilä, M., and Contreras, R. (2000) Molecular cloning and enzymatic characterization of a *Trichoderma reesei* 1, 2-α-D-mannosidase. *J. Biotechnol.*, **77**, 255–263.

Maras, M., De Bruyn, A., Schraml, J., Herdewijn, P., Claeyssens, M., Fiers, W., and Contreras, R. (1997) Structural characterization of N-linked oligosaccharides from cellobiohydrolase I secreted by the filamentous fungus *Trichoderma reesei* RUTC 30. *Eur. J. Biochem.*, **245**, 617–625.

Nakayama, K., Feng, Y., Tanaka, A., and Jigami, Y. (1998) The involvement of *mnn4* and *mnn6* mutations in mannosylphosphorylation of O-linked oligosaccharide in yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta*, **1425**, 255–262.

Nidetzky, B., Steiner, W., Hayn, M., and Claeyssens, M. (1994) Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: a new model for synergistic interaction. *Biochem. J.*, **298**, 705–710.

Nyman, P.O., Strid, L., and Westermark, G. (1966) Carboxyl-terminal amino acid sequences of human and bovine erythrocyte carbonic anhydrase. *Biochim. Biophys. Acta*, **122**, 554–556.

Rouvinen, J., Bergfors, T., Teeri, T., and Knowles, J.K.C. (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*, **249**, 380–386.

Saloheimo, M., Lehtovaara, P., Penttila, M. Teeri, T.T., Ståhlberg, J., Johansson, G., Petterson, G., Claeyssens, M., Tomme, P., and Knowles, J.K.C. (1998) EG III, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene*, **63**, 11–21.

Salovouri, I., Makarow, M., Rauvala, H., Knowles, J., and Kääriäinen L. (1987) Low molecular weight high-mannose type glycans in a secreted protein of the filamentous fungus *Trichoderma reesei*. *Bio/Technology*, **5**, 152–156.

Srisodsuk, M., Reinikainen, T., Penttilä, M., and Teeri, T.T. (1993) Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *J. Biol. Chem.*, **268**, 20756–20761.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeyssens M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414—analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.*, **170**, 575–581.

Van Petegem, F., Contreras, H., Contreras, R., and Van Beeumen, J. (2001) *Trichoderma reesei* α-1, 2-mannosidase: structural basis for the cleavage of four consecutive mannose residues. *J. Mol. Biol.*, **312**, 157–165.

Woodward, J., Brown, J.P., Evans, B.R., and Affholter, K.A. (1994) Papain digestion of crude *Trichoderma reesei* cellulase: purification and properties of cellobiohydrolase I and II core proteins. *Biotechnol. Appl. Biochem.*, **19**, 141–153.