Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 343 (2008) 48–55
Carbohydrate RESEARCH

N-Glycosylation in *Chrysosporium lucknowense* enzymes

Alexander V. Gusakov,* Alexey I. Antonov and Boris B. Ustinov

Division of Chemical Enzymology, Department of Chemistry, M. V. Lomonosov Moscow State University, Moscow 119899, Russia

Received 24 September 2007; received in revised form 16 October 2007; accepted 18 October 2007
Available online 26 October 2007

Abstract—Twenty-eight enzymes, encoded by different genes and secreted by different mutant strains of *Chrysosporium lucknowense*, were subjected to MALDI-TOF MS peptide fingerprinting followed by analysis of the MS data using the GlycoMod tool from the ExPASy proteomic site. Various N-linked glycan structures were discriminated in the *C. lucknowense* proteins as a result of the analysis. N-Glycosylated peptides with modifications matching the oligosaccharide compositions contained in the GlycoSuiteDB were found in 12 proteins. The most frequently encountered N-linked glycan, found in 9 peptides from 7 proteins, was $(Man)_3(GlcNAc)_2$, that is, the core pentasaccharide structure forming mammalian-type high-mannose and hybrid/complex glycans in glycoproteins from different organisms. Nine out of 12 enzymes represented variably N-glycosylated proteins carrying common $(Hex)_{0–4}(HexNAc)_{0–6} + (Man)_3(GlcNAc)_2$ structures, most of them being hybrid/complex glycans. Various glycan structures were likely formed as a result of the enzymatic trimming of a 'parent' oligosaccharide with different glycosidases. The N-glycosylation patterns found in *C. lucknowense* proteins differ from those reported for the extensively studied enzymes from *Aspergilli* and *Trichoderma* species, where high-mannose glycans of variable structure have been detected.

Keywords: *Aspergillus* sp.; *Chrysosporium lucknowense*; Glycoproteins; MALDI-TOF mass spectrometry; N-Glycosylation; *Trichoderma reesei*

1. Introduction

Fungal glycoside hydrolases[1,2] and other enzymes belonging to different protein families are often glycosylated, carrying both O-linked and N-linked glycans. The glycosylation is the most frequently encountered post-translational modification in those proteins. The molecules of many glycoside hydrolases have a modular structure, that is, they consist of a catalytic module, flexible peptide linker, and carbohydrate-binding module.[3,4] Linker peptides, which are rich in Ser and Thr residues, are typically O-glycosylated.[4–6] The N-glycosylation seems to be restricted to the catalytic modules, and it is usually absent in other parts of enzyme molecules. The fungal carbohydrate-binding modules are not glycosylated as a rule.

Although a large quantity of information on the glycosylation of yeast proteins has been published, the number of species belonging to filamentous fungi and studied from this point of view is not high. The N-glycosylation has been extensively studied in different enzymes from *Aspergilli*[7–17] as well as in *Trichoderma reesei* cellulases and acetylxylan esterase.[6,18–28] It has been shown that, in most cases, the N-linked oligosaccharides represent mammalian-type high-mannose glycans ($Man_xGlcNAc_2$), which may be phosphorylated (in *T. reesei*). After the secretion of mature proteins, the high-mannose-type glycans may be subjected to enzymatic trimming by an array of hydrolytic enzymes, including α-mannosidases, endoglycosidases F or H, phosphatase, etc. Thus, depending on the fungal strain and fermentation conditions, the observed glycosylation patterns display different variations.[6,22,24,25] Up to 11 mannose residues have been detected in the N-linked oligosaccharide from cellobiohydrolase I of *T. reesei*.[6] A linear $Hex_{7–26}GlcNAc_2$ series of glycans containing up to 24 mannose residues and up to three β-Gal*f* residues has been found in α-galactosidase A from *Aspergillus niger*.[14]

* Corresponding author. Tel.: +7 495 939 5966; fax: +7 495 939 0997; e-mail: avgusakov@enzyme.chem.msu.ru

0008-6215/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2007.10.014

Chrysosporium lucknowense is a filamentous fungus belonging to ascomycetes. It produces a wide range of enzymes that catalyze the biodegradation of cellulose, hemicelluloses, and other polysaccharides. During the last few years, we isolated and characterized about 30 enzymes secreted by different strains of *C. lucknowense*; data on purification and properties of some of the enzymes have been published previously.[29–32] The characterization also included MALDI-TOF mass spectrometry (MS) peptide fingerprinting for the isolated proteins. The full fungal genome of the *C. lucknowense* was sequenced in 2005, and then the genome annotation was carried out (see for instance http://www.dyadic.com/wt/dyad/pr_1143820822). About 200 carbohydrate-active enzymes have already been identified in the *C. lucknowense* genome at this time, so this fungus is very interesting as a source of glycosidases and other enzymes acting on carbohydrates. The amino acid sequences of *C. lucknowense* glycoside hydrolases demonstrate the highest similarity (up to 70–85%) to known enzymes from *Chaetomium* spp., *Humicola* spp., *Neurospora crassa*, while displaying less similarity (up to 60–70%) to the fungal enzymes of genera *Trichoderma (Hypocrea)*, *Aspergillus*, *Phanerochaete*.

The availability of the protein amino acid sequences from the proprietary annotated *C. lucknowense* genome combined with the MS data for secreted proteins allowed one to discriminate possible modifications in peptides derived from *C. lucknowense* proteins. This paper describes N-glycosylation patterns observed in different *C. lucknowense* enzymes, which were found by the analysis of MS data using the GlycoMod tool from the ExPASy proteomic site.[33]

2. Results

Twenty-eight enzymes in total, encoded by different genes and secreted by different strains of *C. lucknowense*, were subjected to MALDI-TOF MS peptide fingerprinting, followed by analysis of the MS data with the MASCOT program for identification of enzymes in the proprietary *C. lucknowense* translated protein database as well as proteomic tools (MotifScan, PeptideMass, FindPept, FindMod, GlycoMod) from the ExPASy site created and maintained by the Swiss Institute of Bioinformatics (http://cn.expasy.org/tools/). The proteins analyzed belonged to different classes of enzymes (hydrolases and oxidoreductases) and included nine cellulases (*endo*-(1→4)-β-glucanases, EC 3.2.1.4, and *exo*-cellobiohydrolases, EC 3.2.1.91), six *endo*-(1→4)-β-xylanases (EC 3.2.1.8), two cellobiose dehydrogenases (EC 1.1.99.18), β-glucosidase (EC 3.2.1.21), β-xylosidase (xylan (1→4)-β-xylosidase, EC 3.2.1.37), *exo*-β-D-glucosaminidase (EC 3.2.1.-), glucoamylase (EC 3.2.1.3), chitinase (EC 3.2.1.14), arabinogalactan *endo*-(1→4)-β-galactosidase (EC 3.2.1.89), laminarinase (glucan (1→3)-β-glucosidase, EC 3.2.1.58), xyloglucanase (xyloglucan-specific *exo*-β-(1→4)-glucanase, EC 3.2.1.155), non-classified β-(1→3)-glucanase belonging to the GH17 family (EC 3.2.1.-), acetylxylan esterase (EC 3.1.1.72), and alkaline serine protease (EC 3.4.21.-).

Although the MotifScan tool found potential N-glycosylation sites (the consensus sequence Asn-Xaa-Ser/Thr, where Xaa cannot be Pro) in most of the proteins analyzed (in 23 enzymes out of 28 in total), the analysis of the MS data with the GlycoMod tool revealed the possible N-glycosylated peptides with modifications matching the oligosaccharide compositions contained in the GlycoSuiteDB[34,35] only in 12 proteins. They are listed in Table 1.

To demonstrate how the analysis was carried out, we will examine the *C. lucknowense* xylanase (Xyl IV) that represents a good example of a variably glycosylated protein. The amino acid sequence of the Xyl IV is shown in Figure 1. The enzyme consists of 375 amino acid residues (including a signal peptide). The MALDI-TOF mass spectrum of peptides obtained after a tryptic digestion of the protein band from the SDS-PAGE gel is shown in Figure 2. Using the PeptideMass and FindPept tools, 22 specific tryptic peptides matching the sequence of Xyl IV were found. They are shown in bold in Figure 1. The GlycoMod tool found additional nine peaks in the mass spectrum of the Xyl IV digest that corresponded to two peptides (shaded in Fig. 1) carrying variable modifications of Asn-98 and Asn-285 with oligosaccharides (Table 2). Any N-linked glycans bound to the third potential N-glycosylation site (Asn-356) were not discriminated.

The peak with m/z 2645.2 in Figure 2 corresponds to the first N-glycosylated peptide in the Xyl IV sequence (Fig. 1), where the $(Hex)_3(GlcNAc)_2$ oligosaccharide is bound to the Asn-98 residue. This oligosaccharide seems to represent the well-known conserved $(Man)_3(GlcNAc)_2$ core structure that forms mammalian-type high-mannose and hybrid/complex glycans in glycoproteins from different organisms.[37] Two other peaks (with m/z 2807.3 and 2969.3), differing cumulatively from the first mentioned peak by 162.1 and 162.0 Da, that is, by the mass of an anhydrohexose residue, correspond to the same peptide modified with glycans containing the same core pentasaccharide that is longer by one and two Hex residues, respectively (Table 2, Nos. 2 and 3). Two additional peaks (with m/z 3010.4 and 3172.4), differing cumulatively from the peak 2645.2 by 365.1 Da (HexNAc + Hex) and 162.0 Da (Hex), correspond to the same first peptide modified with $(Hex)_1(HexNAc)_1 + (Man)_3(GlcNAc)_2$ and $(Hex)_2(HexNAc)_1 + (Man)_3(GlcNAc)_2$ oligosaccharides, respectively. These oligosaccharides represent hybrid/complex glycan structures according to the nomenclature accepted in the GlycoSuiteDB.

Table 1. Potential and occupied N-glycosylation sites in *C. lucknowense* enzymes found by the MotifScan and GlycoMod tools from the ExPASy site (http://cn.expasy.org/tools/)

Enzyme[a]	GH family designation[b]	N-Glycosylation sites potential	N-Gycosylation sites occupied	Glycosylated peptides
EG I	Cel7C	N77, N254, N264	N264	VNVTDYYGEGAEFR
EG II	Cel5A	N93, N156, N260	N156	NLTEVVNFVTNAGK
EG VI	Cel6C	N44, N232, N246, N312	N246, N312	ENFTEWSNSWDESH…, VNNTVVDALVWVK…
Xyl IV	Xyn10C	N98, N285, N356	N98, N285	See Table 2
Xyl V	Xyn11C	N59, N90	N90	NITFSGSTQHTSGTV…
EGA	Gls2A	N96, N142, N194, N231, N243, N347, N584, N704, N739, N753, N834	N584	RPNVSLFHMSR
BXL II	Bxl3A	N233, N388, N602	N233	INGTWACENDR
GA	Gla15A	N97, N199, N272	N97	DAGLVLTGIVDALSQNYS…
Lam	Lam55A	N232, N254	N232	NMKFNNCK
XGL	Xgl74A	N236, N433, N456, N589	N589	NNTVFYAASGASFYR
CDH I	—	N140, N421, N458, N537, N692, N699	N458, N692 or N699	LWLNTSVK, NVANLTWLFPNSTITPR
CDH II	—	N164, N288, N419, N627	N164, N288, N419, N627	GNVSTSNGNLVLGR, GPPSTGLWNGTMK…, EISANEAPSEKNHT…, TNSTHAMTLSQYLGR

[a] Abbreviations: BXL, β-xylosidase; CDH, cellobiose dehydrogenase; EG, *endo*-(1→4)-β-glucanase; EGA, *exo*-β-D-glucosaminidase; GA, glucoamylase; Lam, laminarinase; Xyl, xylanase; XGL, xyloglucanase.

[b] Enzyme names according to the standardized nomenclature based on glycoside hydrolase (GH) families, the classification introduced by Henrissat et al.[1,2,36] (see also http://www.cazy.org/).

```
  1 MHLSSSLLLLAALPLGIAGKGKGHGHGPHTGLHTLAKQAGLKYFGSATDS 50
 51 PGQRERAGYEDKYAQYDQIMWKSGEFGLTTPTNGQKWLFTEPERGVFNFT 100
101 EGDIVTNLARKHGFMQRCHALVWHSQLAPWVESTEWTPEELRQVIVNHIT 150
151 HVAGYYKGKCYAWDVVNEALNEDGTYRESVFYKVLGEDYIKLAFETAAKV 200
201 DPHAKLYYNDYNLESPSAKTEGAKRIVKMLKDAGIRIDGVGLQAHLVAES 250
251 HPTLDEHIDAIKGFTELGVEVALTELDIRLSIPANATNLAQQREAYKNVV 300
301 GACVQVRGCIGVEIWDFYDPFSWVPATFPGQGAPLLWFEDFSKHPAYDGV 350
351 VEALTNRTTGGCKGKGKGKGKVWKA                         375
```

Figure 1. Amino acid sequence of Xyl IV from *C. lucknowense*. Putative signal peptide is shown in italic; peptides identified by MALDI-TOF MS are shown in bold; potential sites of trypsin cleavage are underlined; potential N-glycosylation sites are shown in boxes; two identified glycopeptides are shaded.

A very similar situation was observed for the second N-glycosylated peptide in the Xyl IV (second shaded region in Fig. 1). This peptide also had variable modifications with glycans including the core $(Man)_3(GlcNAc)_2$ oligosaccharide (the peak with m/z 2389.2 in Fig. 2) as well as two longer oligosaccharides differing from the 'parent' peak by 162.0 Da (Hex) and 365.1 Da (HexNAc + Hex), respectively. The structures of those oligosaccharides are shown in Table 2 (Nos. 7 and 8). The GlycoMod tool also found one additional possible N-linked glycan structure (No. 9) for the second N-glycosylated peptide. It should be noted that in all cases the difference between the observed and theoretical peptide masses (m/z) in Table 2 did not exceed 0.1 Da.

Similar analyses of the MS data were carried out for other *C. lucknowense* enzymes. The N-glycosylated peptides (either complete or their fragments) are shown in Table 1, while glycan structures are presented in Table 3. The most often encountered N-linked glycan was found to be $(Man)_3(GlcNAc)_2$ (No. 2 in Table 3), that is, the same conserved core pentasaccharide already observed in the Xyl IV above. This structure was found in nine peptides from seven different proteins. The next predominant oligosaccharide was $(Hex)_1(HexNAc)_1 + (Man)_3(GlcNAc)_2$ (No. 6) that was found in seven peptides from six proteins. Then $(Hex)_1 + (Man)_3(GlcNAc)_2$ (No. 3) oligosaccharide follows (six peptides from five proteins). Glycan structures 1, 4, 5, 7, and 10 were also observed in two or more *C. lucknowense* enzymes. The common feature of all the N-linked glycans mentioned (except for the No. 1 representing a single *N*-acetylglucosamine residue) is the presence of the $(Man)_3$-$(GlcNAc)_2$ core structure. Oligosaccharides 8, 9, and 11 (each of them was observed as a singular case—either in the CDH I or Lam) also contain the same core pentasaccharide as a common structural feature.

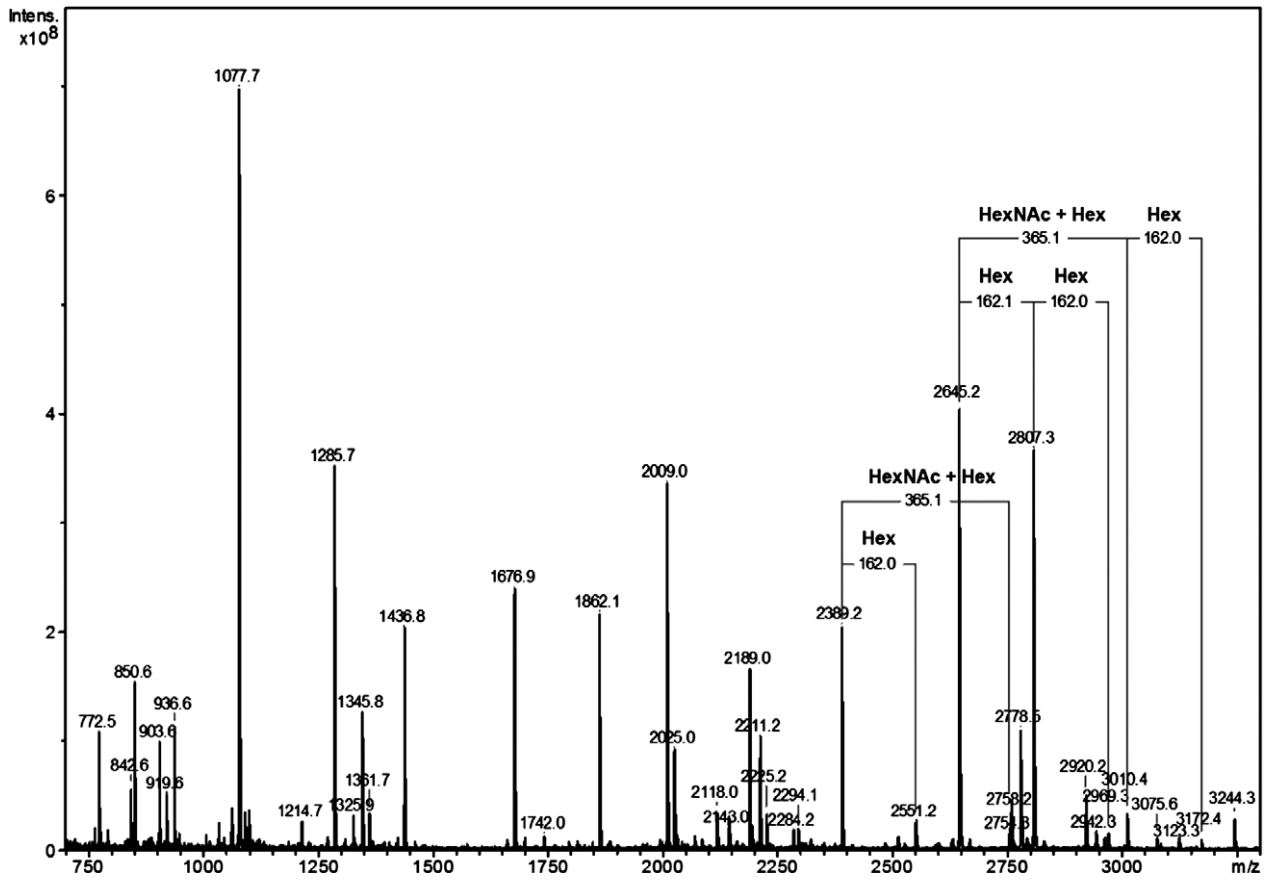

Figure 2. MALDI-TOF mass spectrum of peptides derived from the in-gel tryptic digest of *C. lucknowense* Xyl IV.

Table 2. N-Glycosylated peptides and glycan structures found by the GlycoMod tool in the Xyl IV of *C. lucknowense* using MS data shown in Figure 2

No.	*m/z* observed	*m/z* theoretical	Peptide	Modification
1	2645.2	2645.21	GVFNFTEGDIVTNLAR	$(Man)_3(GlcNAc)_2$
2	2807.3	2807.26	GVFNFTEGDIVTNLAR	$(Hex)_1 + (Man)_3(GlcNAc)_2$
3	2969.3	2969.31	GVFNFTEGDIVTNLAR	$(Hex)_2 + (Man)_3(GlcNAc)_2$
4	3010.4	3010.34	GVFNFTEGDIVTNLAR	$(Hex)_1(HexNAc)_1 + (Man)_3(GlcNAc)_2$
5	3172.4	3172.39	GVFNFTEGDIVTNLAR	$(Hex)_2(HexNAc)_1 + (Man)_3(GlcNAc)_2$
6	2389.2	2389.14	LSIPANATNLAQQR	$(Man)_3(GlcNAc)_2$
7	2551.2	2551.19	LSIPANATNLAQQR	$(Hex)_1 + (Man)_3(GlcNAc)_2$
8	2754.3	2754.27	LSIPANATNLAQQR	$(Hex)_1(HexNAc)_1 + (Man)_3(GlcNAc)_2$
9	2211.1	2211.09	LSIPANATNLAQQR	$(Hex)_1(HexNAc)_2(Deoxyhexose)_1$

It should be noted that one of the tryptic peptides in CDH I (689 NVANLTWLFPNSTITPR 705) contained two potential N-glycosylation sites (Asn-692 and Asn-699), so in this case the actual position of the modification could not be discriminated. One cannot exclude the possibility that both N-glycosylation sites in this peptide were modified. In such case, the $(Man)_3(GlcNAc)_2$ core structure could be linked to both the Asn-692 and Asn-699 residues instead of the structure No.8 linked to only one of those residues. A similar situation could take place in other cases where the glycosylation of the same CDH I peptide was detected (Nos. 5, 9, and 10 in Table 3). However, the possibility that both Asn residues were modified is not very high since the glycosylation of only one of those residues was clearly detected in case No. 2.

If one were to look at the distribution of the N-glycan structures found amongst different enzymes, then nine out of 12 enzymes listed in Table 3 (except for the Xyl

Table 3. Structures of N-linked glycans found by the GlycoMod tool in the *C. lucknowense* enzymes

No.	N-linked glycan	Enzyme[a]
1	$(HexNAc)_1$	EG VI (N246), XGL
2	$(Man)_3(GlcNAc)_2$	EG I, Xyl IV (N98, N285), Xyl V, BXL II, GA, CDH I (N692 or N699), CDH II (N419, N627)
3	$(Hex)_1 + (Man)_3(GlcNAc)_2$	EG II, Xyl IV (N98, N285), EGA, BXL II, CDH II (N419)
4	$(Hex)_2 + (Man)_3(GlcNAc)_2$	EG II, EG VI (N312), Xyl IV (N98)
5	$(HexNAc)_1 + (Man)_3(GlcNAc)_2$	EG I, CDH I (N458, N692 or N699)
6	$(Hex)_1(HexNAc)_1 + (Man)_3(GlcNAc)_2$	EG I, EG II, EG VI (N246), Xyl IV (N98, N285), XGL, CDH II (N627)
7	$(Hex)_2(HexNAc)_1 + (Man)_3(GlcNAc)_2$	Xyl IV (N98), CDH II (N288)
8	$(Hex)_3(HexNAc)_2 + (Man)_3(GlcNAc)_2$	CDH I (N692 or N699)
9	$(Hex)_4(HexNAc)_2 + (Man)_3(GlcNAc)_2$	CDH I (N692 or N699)
10	$(Hex)_3(HexNAc)_3 + (Man)_3(GlcNAc)_2$	Lam, CDH I (N692 or N699)
11	$(Hex)_2(HexNAc)_6 + (Man)_3(GlcNAc)_2$	Lam
12	$(HexNAc)_2(Deoxyhexose)_1$	CDH II (N164)
13	$(Hex)_1(HexNAc)_2(Deoxyhexose)_1$	Xyl IV (N285)
14	$(Hex)_3(HexNAc)_2(Deoxyhexose)_2$	Lam
15	$(Hex)_1(HexNAc)_1(Deoxyhexose)_1(Pent)_1 + (Man)_3(GlcNAc)_2$	Lam
16	$(Hex)_4(Phos)_1 + (Man)_3(GlcNAc)_2$	Lam

[a] For those enzymes in which only one N-glycosylation site was found to be modified, the glycosylated residues are given in Table 1; otherwise, they are shown in parenthesis. Enzyme abbreviations are given in the footnote in Table 1.

V, EGA, and GA) represented variably glycosylated proteins, that is, when the same N-glycosylation site is occupied by oligosaccharides differing in length (the number of Hex and HexNAc residues). As in the case of Xyl IV described in detail above, the most pronounced examples of variably glycosylated proteins demonstrated endoglucanases I and II, laminarinase, and cellobiose dehydrogenases I and II. The highest number of total Hex residues was observed in the oligosaccharide structure No. 9 from CDH I ($Hex_4 + Man_3$, seven Hex residues in total), while the structure No. 11 found in laminarinase was most rich in HexNAc (eight residues in total). So, the structures 2–11 were most likely formed as a result of the enzymatic trimming of a hybrid/complex glycan, containing at least seven Hex residues and eight HexNAc residues, with α-mannosidases, β-hexosaminidases, and/or other glycosidases. The minimal structure of that hypothetical glycan may be $(Hex)_4(HexNAc)_6 + (Man)_3(GlcNAc)_2$.

The N-linked glycan structures 12–16 fall out of the series of homologous glycosylation patterns 2–11 having different numbers of Hex and HexNAc residues linked to the $(Man)_3(GlcNAc)_2$ core (Table 3). Four of them contain deoxyhexose residue(s). The last two structures contain either a pentose residue or a phosphorylated hexose residue while sharing the core pentasaccharide as a common structural feature. Since each of the structures 12–16 was observed as the only one of its kind, further studies are needed to confirm whether they really exist in *C. lucknowense* enzymes or whether they are false-positive hits, which are often encountered in MS analysis where the computer-based search and identification are based on the comparison of experimental masses of molecules with theoretical ones contained in different databases.

3. Discussion

Various N-linked glycan structures were found in 12 different enzymes from *C. lucknowense* belonging to different enzyme classes and protein families (Tables 1 and 3). Nine of these enzymes represented variably glycosylated proteins, that is, when the same Asn residue was modified with oligosaccharides having different structure. Only one Asn residue out of two or more potential N-glycosylation sites was found to be occupied in the molecules of endoglucanases I and II, xylanase V, *exo*-β-D-glucosaminidase, β-xylosidase II, glucoamylase, laminarinase, and xyloglucanase (Table 1). Endoglucanase VI, xylanase IV, and cellobiose dehydrogenase I presented N-linked glycans at two out of 3–6 potential sites, while all four potential N-glycosylation sites were modified in the cellobiose dehydrogenase II.

The most frequently met N-linked glycan was $(Man)_3(GlcNAc)_2$. This pentasaccharide represents a well-known conserved core structure that forms mammalian-type high-mannose and hybrid/complex glycans in glycoproteins from different organisms.[37] Two terminal mannose residues in the core pentasaccharide are connected by α-(1→3)- and α-(1→6)-linkages to the Man-β-(1→4)-GlcNAc-β-(1→4)-GlcNAc trisaccharide residue. Most of other oligosaccharides detected in *C. lucknowense* glycoproteins had a common structure $(Hex)_x(HexNAc)_y + (Man)_3(GlcNAc)_2$, where x and y were varied in the range of 0–4 and 0–6 residues, respectively (Table 3). Such a structure is typical either for hybrid or complex N-linked glycans according to the nomenclature accepted in the GlycoSuiteDB.

Various glycans, whose structure may be described by the above-mentioned formula were most likely formed as a result of the enzymatic trimming of some 'parent'

oligosaccharide with different glycosidases: α-mannosidases (EC 3.2.1.24; EC 3.2.1.113; EC 3.2.1.114), β-hexosaminidases (EC 3.2.1.52), and/or other glycosidases (if Hex and HexNAc are not Man and GlcNAc but represent other monosaccharides). At least five different β-hexosaminidases (*N*-acetylhexosaminidases) belonging to the GH families 2, 3, and 20 were found in the *C. lucknowense* genome. One enzyme with such specificity (*exo*-β-D-glucosaminidase) was studied in this paper. At least one GH38 and a few GH47 family α-mannosidases were also detected in the fungal genome. However, further studies should be done to find out which glycosidases of the above mentioned are indeed secreted and may participate in oligosaccharide trimming.

It should be noted that the N-linked glycans $(Man)_3(GlcNAc)_2$, $(Man)_4(GlcNAc)_2$ as well as the oligosaccharides with variable mannose content—$(Man)_{3-7}(GlcNAc)_3$—have been previously detected in the endoglucanase III (Cel12A) of *C. lucknowense* using enzyme deglycosylation followed by HPLC analysis of the released oligosaccharides (unpublished data). The first core pentasaccharide made up ~90% of the total glycans released from the EG III. The structure of oligosaccharides from EG III displayed the same general pattern as the structure Nos. 2, 3, 5, 6, and 7 in Table 3, thus confirming the results of the MALDI-TOF MS analysis described in this paper. It is also noteworthy that the EG III was amongst 28 enzymes in the present studies, whose peptides were analyzed by MS for possible glycosylation. The EG III indeed contains one potential N-glycosylation site (Asn-241); however, a molecular mass (*m*/*z*) of the theoretical non-modified tryptic peptide containing this residue is 4894 Da. This means that the N-modified glycopeptide would have a molecular mass of more than 5000 Da. Such heavy peptides are hardly visible in the MALDI-TOF mass spectra, and this explains why the EG III glycosylation was not detected by the MS. Similar situation could take place in other cases of *C. lucknowense* enzymes containing potential N-glycosylation sites, which were not found to be modified with glycans according to the MS data.

In most cases of reported glycoproteins from filamentous fungi belonging to the genera *Aspergillus* and *Trichoderma*, the N-linked glycans represented mammalian-type high-mannose oligosaccharides $(Man)_x(GlcNAc)_2$, where *x* typically varied in the range between 5 and more than 20 residues.[6–28] High-mannose glycans from the CBH I, EG I, and acetylxylan esterase of *T. reesei* have been found to be phosphorylated, one or two phosphate residues linked to mannose residues by the phosphodiester bond being detected.[6,19,24–26,28] A rather unusual oligosaccharide $(Glc)_1(Man)_7(GlcNAc)_2$ containing one terminal glucose residue has been observed in the CBH I from *T. reesei* RUTC 30 strain.[20] A single *N*-acetylglucosamine residue, linked to the Asn residues representing protein N-glycosylation sites, has been found in *T. reesei* cellulases.[6,21,24,26,27] We detected the N-linked single *N*-acetylglucosamine residue also in xyloglucanase (Cel74A) of *T. reesei*.[30] Its presence is typically attributed to the secretion of an endogenous Endo-H- or Endo-F-like activity in *T. reesei* cultures.[6,24,27] The N-linked high-mannose-type oligosaccharides containing α-D-galactofuranosyl residues have been found in α-glucosidase and α-galactosidase A of *A. niger*.[11,14]

Unlike in glycoproteins from *Aspergilli* and *T. reesei*, the N-linked glycans detected in the enzymes from *C. lucknowense* represented in most cases not high-mannose but either hybrid or complex oligosaccharides. Whereas a single *N*-acetylglucosamine residue has been often observed in *T. reesei* enzymes,[6,21,24,26,27,30] it was found only in two peptides from *C. lucknowense* (peptides from the EG VI and xyloglucanase, Table 3). This means that the Endo-H or Endo-F pathway of N-linked oligosaccharide trimming is not a common case in *C. lucknowense*. Only in one case (No. 16 in Table 3) a potential phosphorylation of the N-linked glycan in *C. lucknowense* was detected, thus demonstrating another difference of this fungus from *T. reesei*. Further studies are needed to determine the precise structure of the hybrid/complex oligosaccharides found in the *C. lucknowense* glycoproteins and to find out if deoxyhexose and pentose residues (Nos. 12–15 in Table 3) are indeed a part of N-linked glycans in this filamentous fungus.

4. Experimental

4.1. Enzymes

Enzymes were isolated from crude multienzyme preparations (from Dyadic International, Inc., USA) produced by various mutant strains of *C. lucknowense*. The enzyme purification was carried out using different types of chromatography and chromatofocusing on a Pharmacia FPLC system (Sweden).[29–32,38,39]

The enzyme purity was characterized by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectrofocusing. SDS-PAGE was carried out in 12% gel using a Mini Protean II equipment (Bio-Rad Laboratories, USA). Isoelectrofocusing was performed on a Model 111 Mini IEF Cell (Bio-Rad Laboratories, USA). Staining of protein bands was carried out with Coomassie Blue R-250 (Ferak, Germany).

Enzymes for MALDI-TOF MS peptide fingerprinting were taken in the form of small pieces of gel (~1 mm^3) cut from the SDS-PAGE protein bands. The enzymes were either homogeneous samples according to the data of SDS-PAGE and isoelectrofocusing (in most cases) or they represented protein bands from chromatographic

fractions containing the target protein with purity not less than 90–95%.

4.2. MALDI-TOF mass spectrometry of peptides

The in-gel tryptic digestion of the protein bands after the SDS-PAGE was carried out essentially as described by Smith.[40] Trypsin (Promega, modified, 5 μg/mL) in 50 mM NH_4HCO_3 was used for protein digestion. The resulting peptides were extracted from a gel with 20% aq acetonitrile containing 0.1% trifluoroacetic acid and subjected to MALDI-TOF MS using either α-cyano-4-hydroxycinnamic or 2,5-dihydroxybenzoic acid as a matrix.[41] Ultraflex and Autoflex II mass spectrometers (Bruker Daltonik Gmbh, Germany) were used in the MS experiments.

4.3. Analysis of the MS data

Mass spectra of peptides obtained after a digestion of proteins from SDS-PAGE bands with trypsin were analyzed using Bruker Data Analysis for TOF 1.6g software. The list of monoisotopic peptide masses (as $[M+H]^+$ ions) was generated for each protein, and then the peptide masses were analyzed with the FindPept, PeptideMass, FindMod, GlycoMod MS tools from the ExPASy WWW site: http://cn.expasy.org/tools/ (for those enzymes that have been identified in *C. lucknowense* earlier[29–32] and for which the amino acid sequences have been already available). In the case of a few new enzymes isolated recently, the analysis included a preliminary search for matching peptides in the proprietary *C. lucknowense* translated protein database (from annotated fungal genome) using the in-house version of the MASCOT program[42] to identify the enzyme, and then the above-mentioned MS tools from the ExPASy site were used. The MotifScan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan) was used for identification of potential N-glycosylation sites in proteins.

Analysis for possible N-linked glycans in the enzymes was carried out using the GlycoMod tool (http://cn.expasy.org/tools/glycomod/).[33] The possibility of cysteine modification with acrylamide (CYS_PAM) and methionine oxidation (MSO), which are typical variable modifications in mass spectrometry of proteins, was included in the analysis as well as the possibility for one missed cleavage with trypsin. Only those peptide masses from a mass spectrum were taken into account during the GlycoMod search that did not match the theoretical non-modified tryptic peptides from a concrete protein. Since a search with the GlycoMod tool is based on a comparison of experimental masses of glycopeptides with theoretical ones, its results sometimes contain false-positive hits (displaying rather exotic glycan structures). So, finally only those N-linked glycan structures were taken for consideration and further manual analysis that matched the oligosaccharide compositions contained in the GlycoSuiteDB (http://glycosuite.proteomesystems.com/glycosuite/glycodb).[34,35] As a rule, the difference between the observed and theoretical glycopeptide masses (m/z) did not exceed 0.2–0.3 Da.

Acknowledgments

The authors thank Peter Punt from TNO Nutrition and Food Research, The Netherlands, for providing amino acid sequences of the enzymes and for the useful discussions. The authors also express their gratitude to Mark Gosink from the Scripps Research Institute, Florida, USA, for help in identifying a few new proteins in the *C. lucknowense* translated protein database.

References

1. Henrissat, B. *Biochem. J.* **1991**, *280*, 309–316.
2. Henrissat, B.; Bairoch, A. *Biochem. J.* **1996**, *316*, 695–696.
3. Van Tilbeurgh, H.; Tomme, P.; Claeyssens, M.; Bhikhabhai, R.; Pettersson, G. *FEBS Lett.* **1986**, *204*, 223–227.
4. Gilkes, N. R.; Henrissat, B.; Kilburn, D. G.; Miller, R. C.; Warren, R. A. J. *Microbiol. Rev.* **1991**, *55*, 303–315.
5. Langsford, M. L.; Singh, G. B.; Moser, B.; Miller, R. C., Jr.; Warren, R. A. J.; Kilburn, D. G. *FEBS Lett.* **1987**, *225*, 163–167.
6. Harrison, M. J.; Nouwens, A. S.; Jardine, D. R.; Zachara, N. E.; Gooley, A. A.; Nevalainen, H.; Packer, N. H. *Eur. J. Biochem.* **1998**, *256*, 119–127.
7. Yamaguchi, H.; Ikenaka, T.; Matsushima, Y. *J. Biochem.* **1971**, *70*, 587–594.
8. Minobe, S.; Nakajima, H.; Itoh, N.; Funakoshi, I.; Yamashina, I. *J. Biochem.* **1979**, *86*, 1851–1854.
9. Takegawa, K.; Fujiwara, K.; Iwahara, S.; Yamamoto, K.; Tochikura, T. *Biochem. Cell Biol.* **1989**, *67*, 460–464.
10. Chiba, Y.; Yamagata, Y.; Iijima, S.; Nakajima, T.; Ichishima, E. *Curr. Microbiol.* **1993**, *27*, 281–288.
11. Takayanagi, T.; Kimura, A.; Chiba, S.; Ajisaka, K. *Carbohydr. Res.* **1994**, *256*, 149–158.
12. Colangelo, J.; Licon, V.; Benen, J.; Visser, J.; Bergmann, C.; Orlando, R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1448–1453.
13. Wallis, G. L. F.; Swift, R. J.; Hemming, F. W.; Trinci, A. P. J.; Peberdy, J. F. *Biochim. Biophys. Acta* **1999**, *1472*, 576–586.
14. Wallis, G. L. F.; Easton, R. L.; Jolly, K.; Hemming, F. W.; Peberdy, J. F. *Eur. J. Biochem.* **2001**, *268*, 4134–4143.
15. Warren, M. E.; Kester, H.; Benen, J.; Colangelo, J.; Visser, J.; Bergmann, C.; Orlando, R. *Carbohydr. Res.* **2002**, *337*, 803–812.
16. Woosley, B. D.; Kim, Y. H.; Kolli, V. S. K.; Wells, L.; King, D.; Poe, R.; Orlando, R.; Bergmann, C. *Carbohydr. Res.* **2006**, *341*, 2370–2378.
17. Kasajima, Y.; Yamaguchi, M.; Hirai, N.; Ohmachi, T.; Yoshida, T. *Biosci. Biotechnol. Biochem.* **2006**, *70*, 2662–2668.
18. Salovuori, I.; Makarow, M.; Rauvala, H.; Knowles, J.; Kääriäinen, L. *Biotechnology (NY)* **1987**, *5*, 152–156.
19. Maras, M.; De Bruyn, A.; Schraml, J.; Herdewijn, P.; Claeyssens, M.; Fiers, W.; Contreras, R. *Eur. J. Biochem.* **1997**, *245*, 617–625.

20. De Bruyn, A.; Maras, M.; Schraml, J.; Herdewijn, P.; Contreras, R. *FEBS Lett.* **1997**, *405*, 111–113.
21. Klarskov, K.; Piens, K.; Ståhlberg, J.; Hoj, P. B.; van Beeumen, J.; Claeyssens, M. *Carbohydr. Res.* **1997**, *304*, 143–154.
22. Hui, J. P. M.; Lanthier, P.; White, T. C.; McHugh, S. G.; Yaguchi, M.; Roy, R.; Thibault, P. *J. Chromatogr., B* **2001**, *752*, 349–368.
23. Garcia, R.; Cremata, J. A.; Quintero, O.; Montesino, R.; Benkestock, K.; Ståhlberg, J. *Biotechnol. Appl. Biochem.* **2001**, *33*, 141–152.
24. Stals, I.; Sandra, K.; Geysens, S.; Contreras, R.; van Beeumen, J.; Claeyssens, M. *Glycobiology* **2004**, *14*, 713–724.
25. Stals, I.; Sandra, K.; Devreese, B.; van Beeumen, J.; Claeyssens, M. *Glycobiology* **2004**, *14*, 725–737.
26. Eriksson, T.; Stals, I.; Collén, A.; Tjerneld, F.; Claeyssens, M.; Stålbrand, H.; Brumer, H. *Eur. J. Biochem.* **2004**, *271*, 1266–1276.
27. Hui, J. P. M.; White, T. C.; Thibault, P. *Glycobiology* **2002**, *12*, 837–849.
28. Harrison, M. J.; Wathugala, I. M.; Tenkanen, M.; Packer, N. H.; Nevalainen, K. M. H. *Glycobiology* **2002**, *12*, 291–298.
29. Bukhtojarov, F. E.; Ustinov, B. B.; Salanovich, T. N.; Antonov, A. I.; Gusakov, A. V.; Okunev, O. N.; Sinitsyn, A. P. *Biochemistry (Moscow)* **2004**, *69*, 542–551.
30. Grishutin, S. G.; Gusakov, A. V.; Markov, A. V.; Ustinov, B. B.; Semenova, M. V.; Sinitsyn, A. P. *Biochim. Biophys. Acta* **2004**, *1674*, 268–281.
31. Gusakov, A. V.; Sinitsyn, A. P.; Salanovich, T. N.; Bukhtojarov, F. E.; Markov, A. V.; Ustinov, B. B.; van Zeijl, C.; Punt, P.; Burlingame, R. *Enzyme Microb. Technol.* **2005**, *36*, 57–69.
32. Gusakov, A. V.; Salanovich, T. N.; Antonov, A. I.; Ustinov, B. B.; Okunev, O. N.; Burlingame, R.; Emalfarb, M.; Baez, M.; Sinitsyn, A. P. *Biotechnol. Bioeng.* **2007**, *97*, 1028–1038.
33. Cooper, C. A.; Gasteiger, E.; Packer, N. *Proteomics* **2001**, *1*, 340–349.
34. Cooper, C. A.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. *Nucleic Acids Res.* **2001**, *29*, 332–335.
35. Cooper, C. A.; Joshi, H. J.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. *Nucleic Acids Res.* **2003**, *31*, 511–513.
36. Henrissat, B.; Teeri, T. T.; Warren, R. A. J. *FEBS Lett.* **1998**, *425*, 352–354.
37. Dwek, R. A.; Edge, C. J.; Harvey, D. J.; Wormald, M. R.; Parekh, R. B. *Annu. Rev. Biochem.* **1993**, *62*, 65–100.
38. Markov, A. V.; Gusakov, A. V.; Kondratyeva, E. G.; Okunev, O. N.; Bekkarevich, A. O.; Sinitsyn, A. P. *Biochemistry (Moscow)* **2005**, *70*, 657–663.
39. Grishutin, S. G.; Gusakov, A. V.; Dzedzyulya, E. I.; Sinitsyn, A. P. *Carbohydr. Res.* **2006**, *341*, 218–229.
40. Smith, B. E. *Protein Sequencing Protocols*; Humana Press: Totowa, NJ, 1997.
41. *Proteome Research: Mass Spectrometry*; James, P., Ed.; Springer: Berlin, 2001.
42. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.