Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1058 (2004) 263–272
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Combining gel and capillary electrophoresis, nano-LC and mass spectrometry for the elucidation of post-translational modifications of *Trichoderma reesei* cellobiohydrolase I

Koen Sandra[a], Ingeborg Stals[b], Pat Sandra[c], Marc Claeyssens[b], Jozef Van Beeumen[a], Bart Devreese[a,*]

[a] *Laboratory of Protein Biochemistry and Protein Engineering, Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium*
[b] *Laboratory of Biochemistry, Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium*
[c] *Laboratory of Separation Science, Ghent University, Krijgslaan 281 S4, B-9000 Gent, Belgium*

Abstract

N-Glycosylation of cellobiohydrolase I from the fungus *Trichoderma reesei* (strain Rut-C30) is studied using a combination of electrophoretic, chromatographic and mass spectrometric techniques. As four potential *N*-glycosylation sites and several uncharged and phosphorylated high-mannose glycans are present, a large number of glycoforms and phospho-isoforms can be expected. Isoelectric focusing both in gel and in capillary format was successfully applied for the separation of the phospho-isoforms. They were extracted in their intact form from the gel and subsequently analysed by nanospray-Q-TOF-MS, thereby making use of a powerful two-dimensional technique. Nano-LC/MS/MS on a Q-Trap MS further allowed the determination of the glycosylation sites. As a novel approach, an oxonium ion was used in precursor ion scanning for selective detection of glycopeptides containing phosphorylated high-mannose glycans.

Keywords: Isoelectric focusing; Nano-liquid chromatography–mass spectrometry; Cellobiohydrolase I; Precursor ion scanning; Q-Trap; Protein glycosylation; Phosphorylation; Glycoforms

1. Introduction

Characterisation of protein glycosylation poses a great analytical challenge, as this type of modification can occur at one or more positions in the polypeptide chain. Additionally, the glycans at a single position are often heterogeneous and may be missing at some positions. As a result, glycoproteins generally exist as populations of glycosylated variants (glycoforms) of a single polypeptide. Moreover, unlike biopolymers such as nucleic acids and proteins, glycans are usually branched structures and the monomers can be linked at various positions. On top of this, sugars often exist as isomers, greatly contributing to their complexity. Complete characterisation of protein glycosylation is therefore a multi-methodological task and demands for state-of-the-art analytical techniques.

Enzymatically or chemically liberated glycans are commonly analysed, as such or after chemical derivatisation, by high performance liquid chromatography (HPLC), capillary electrophoresis (CE) and mass spectrometry (MS) [1–5]. Each technique has its particular advantages and disadvantages and the present trend is towards hyphenation of a separation technique with MS [1,6,7]. When the glycans are released, all information concerning their attachment to the protein is lost. Determination of the glycosylation sites and the oligosaccharide heterogeneity at each site is achieved at the glycopeptide level after enzymatic or chemical hydrolysis of the protein. Commonly, the peptide and glycopeptide mixture is then analysed by HPLC, coupled to electrospray ionisation mass spectrometry (ESI-MS). Glycopeptides can be selectively detected in a digest by the appearance of marker

* Corresponding author. Fax: +32 9 264 5338.
E-mail address: bart.devreese@ugent.be (B. Devreese).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.08.157

oxonium ions like 204 ($HexNac^+$), 292 ($NeuAc^+$) or 366 ($HexHexNac^+$) originating from collisionally excited glycopeptides [8]. Such determinations are usually performed using precursor ion scanning. Once a glycopeptide is detected, MS/MS can provide information on the identity of the peptide, the attached glycan and the attachment site. If only one glycosylation site is present, studies at the glycopeptide level are sufficient to elucidate the glycoforms. However, if multiple sites are occupied, the glycosylated variants should further be investigated at the protein level. Several groups reported on the use of CE to study glycoprotein heterogeneities [1,9,10]. Nearly complete resolution of glycoforms has been demonstrated. Especially capillary isoelectric focusing (CIEF) has proven to be powerful for resolving glycoforms that differ in their isoelectric point due to the presence of sialic acid, phosphate or sulphate groups [11,12]. Study of glycoproteins in their intact form by MS remains challenging. Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight (MALDI-TOF) MS can only resolve glycoforms of small glycoproteins (<20 kDa) [13]. Conventional ESI has only marginally been used in the study of intact glycoproteins. However, when using nanospray in combination with a high-resolution mass spectrometer such as a TOF-analyser, more complicated and higher molecular weight glycoproteins can be analysed successfully [14].

In the present paper, *N*-glycosylation of the cellulase cellobiohydrolase I (CBH I) from *Trichoderma reesei* was studied. CBH I exhibits a multidomain structure. It contains an *N*-glycosylated catalytic domain and a cellulose-binding domain, separated by an *O*-glycosylated linker peptide. As shown recently [15,16], both *N*- and *O*-glycosylation are dependent on the strain and growth medium used. Now we studied the *N*-glycosylation of CBH I from the hypersecreting *T. reesei* strain Rut-C30, grown in minimal medium. Earlier studies on Rut-C30 CBH I [17] have revealed the presence of monoglucosylated high-mannose *N*-glycans and of a remarkable phosphorylated structure with a phosphate in a diester linkage between two mannoses. Recently, CE-MS^n in combination with enzymatic sequencing and methylation studies was applied for a detailed characterisation of the *N*-glycans [18]. Up to now, however, little is known on the repartition of these glycans on the glycosylation sites of the protein. As deduced from the protein sequence, there are four putative *N*-glycosylation sites (with consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except proline). This could give rise to many glycoforms and, in principle, five phospho-isoforms (a phospho-isoform consists of glycosylated variants with the same number of phosphate groups).

Here, we present a powerful approach for studying glycosylation and phosphorylation at the protein level using a combination of isoelectric focusing and MS. Although some papers describe their use [19–22], the on-line combination of CIEF and MS remains to be experimentally difficult. In the present work, an alternative strategy is presented replacing CIEF by gel isoelectric focusing (gel IEF) as first dimension technique. The proteins were liberated from the gel in their intact form by sonication of the excised bands. After dialysis, the proteins were measured as such by nanospray MS on a Q-TOF mass spectrometer.

The *N*-glycosylation sites and the oligosaccharide heterogeneity at each site were determined after tryptic digestion, following reduction and alkylation, using nano-LC/MS/MS on a Q-Trap mass spectrometer, as described recently [23]. Next to the universal oxonium ion for *N*-glycosylation (*m*/*z* 204), we made use of a specific ion for the phosphorylated glycans (*m*/*z* 243) in precursor ion scanning.

2. Experimental

2.1. Materials and sample preparation

All (bio)chemicals, unless noted, were purchased from Sigma–Aldrich (St. Louis, MO, USA). Both CBH I from the *T. reesei* strain Rut-C30 and standard CBH I from strain QM9414 were produced in house [15,24].

The catalytic domain was prepared by partial proteolytic cleavage of the intact protein with papain at a substrate:enzyme ratio of 80:1 (w/w) in 50 mM sodium acetate (pH 5) for 17 h at room temperature. Purification of the catalytic domain was performed by DEAE anion exchange chromatography, as described in reference [25].

CBH I was reduced and alkylated prior to trypsin proteolysis. The protein (20 μg) was dissolved in 300 μl 50 mM NH_4HCO_3 and 20 μl 45 mM dithiothreitol (DTT) was added. Reduction took place at 60 °C during 30 min. After cooling to room temperature, 20 μl iodoacetamide (100 mM) was added and the mixture was placed in the dark during 30 min. The reduced and alkylated protein was purified from the reaction medium by membrane filtration using an Ultrafree-0.5 centrifugal filter (Millipore, Bedford, MA, USA) with a molecular mass limit of 10 kDa. A volume of 150 μl NH_4HCO_3 (50 mM) and 0.2 μg modified trypsin (Promega, Madison, WI, USA) was added and the protein was incubated overnight at 37 °C. The digest was diluted three times in Milli Q water (Millipore) for LC/MS analysis.

Glycans were released from the glycoproteins by enzymatic deglycosylation with peptide-*N*-glycosidase F (PNGase F–Roche Diagnostics GmbH, Mannheim, Germany), according to the method described by Jackson [26]. Two microliters of a solution consisting of 1% sodium dodecylsulfate (SDS), 0.5% mercaptoethanol and 0.1 M ethylenediamine tetra acetic acid (EDTA) was added to the dried protein. After 30 min incubation at room temperature, 10 μl of 200 mM Na-phosphate buffer (pH 8.6) was added and the mixture was placed in a boiling water bath for 5 min. After cooling to room temperature, 1 μl of 7.5% Triton X100 and 1 unit of PNGase F were added and the mixture was incubated at 37 °C for 20 h. Prior to derivatisation, the proteins were precipitated with ethanol.

The glycans were derivatised by reductive amination. After drying, 2 μl of 100 mM 8-aminopyrene-1,3,6-trisulfonate

(APTS: Beckman–Coulter, Fullerton, CA, USA) in 0.9 M citric acid and 1 μl of 1 M $NaCNBH_3$ (sodium cyanoborohydride) in tetrahydrofuran were added. After incubation at 55 °C during 2 h, 10 μl Milli Q water and 80 μl ice-cold acetone were added and the mixture was placed at −20 °C for 10 min. The reaction tube was centrifuged at 12,500 rpm for 15 min. The supernatant was removed and the precipitated sugars were redissolved in 250 μl Milli Q water.

2.2. CIEF and IEF

CIEF experiments were performed on a BioFocus 3000 capillary electrophoresis system (Bio-Rad, Hercules, CA, USA). Separations were carried out in a 50 μm × 45 cm Beckman eCAP neutral coated capillary. The proteins were dissolved to a concentration of 75 μg/ml in a 3% ampholyte solution consisting of Pharmalyte (pH 2.5–5: Amersham Biosciences, Uppsala, Sweden) and Beckman (pH 3–10) ampholytes (1:3, v/v). A Biomark (Bio-Rad) synthetic p*I* marker (p*I* 5.3) was added (0.5%). To avoid protein precipitation, 7 M ureum was added to the ampholyte solution. This solution was introduced to the capillary by pressure (0.689 MPa) during 60 s. Respectively, 10 mM H_3PO_4 and 20 mM NaOH were used as anolyte and catholyte. Focusing was carried out at 25 kV during 10 min and mobilisation of the focused bands was done by replacing the NaOH solution at the cathode by a mixture of methanol/water/acetic acid (50:49:1). The mobilisation voltage was set at 25 kV and detection was performed at 280 nm.

For gel IEF experiments, a PhastSystem (Amersham Biosciences) was used. A dry precast homogeneous polyacrylamide gel (5% T, 3% C: Amersham Biosciences) was rehydrated with 115 μl Pharmalyte (pH 2.5–5), 25 μl Servalyt (pH 3–7: Serva Electrophoresis, Heidelberg, Germany) and 1860 μl Milli Q water for 2 h. In a prefocusing step the pH gradient was formed, and 4 μl sample (10 mg/ml) was subsequently applied at the cathode position followed by a focusing step as described in the PhastSystem separation technique file No. 100 (Amersham Biosciences). At the end of the run, CBH I activity was revealed by immersing the gel in 2 mM 4-methylumbelliferyl β-lactoside (sodium acetate buffer, pH 5) [27]. The fluorescent spots were excised with a surgical blade and transferred to 0.5 ml tubes (Eppendorf AG, Hamburg, Germany). After adding 200 μl Milli Q water, sonication was performed during 90 min. The supernatant was transferred to an Ultrafree-0.5 centrifugal filter (Millipore) with a molecular mass limit of 10 kDa, and ampholytes and fluorogenic substrate were removed by membrane filtration.

2.3. Mass spectrometry

The intact proteins were measured on a Q-TOF mass spectrometer (Micromass, Manchester, UK) equipped with a nanospray source. The samples were dissolved in 50% acetonitrile (0.1% formic acid) to a final concentration of approximately 5 pmol/μl and measured in the positive ion mode using Protana (Odense, Denmark) needles. The needle and cone voltages were set at 1250 and 35 V, respectively. All data were processed using MassLynx 3.1 and the mass spectra were deconvoluted with the MaxEnt software delivered with MassLynx 3.1.

Tryptic digests were analysed by nano-LC/MS/MS using a Q-Trap mass spectrometer (Applied Biosystems/MDS Sciex, Concord, ON, Canada), as described before [23].

2.4. CE-LIFD

CE-LIFD analyses were performed on a Beckman P/ACE 2100 capillary electrophoresis system equipped with a laser-induced fluorescence detector (3 mW, 488 nm Ar-ion laser). Separations were carried out in a 50 μm × 57 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ, USA) using 25 mM ammonium acetate buffer (pH 4.55). The applied voltage was +25 kV, resulting in a current of 11 μA. Samples were introduced by hydrodynamic injection (3.5 kPa) during 5 s. The temperature was set at 25 °C.

3. Results and discussion

3.1. Glycoform and phospho-isoform analysis

CBH I *N*-glycosylation is restricted to the catalytic domain and *O*-glycosylation to the linker peptide. Separation of these domains by papain cleavage thus allows differentiation between both glycosylation types. Since the *N*-glycans have been characterised before [18], ESI-MS of the catalytic domain was expected to give us information on the number of *N*-glycosylation sites that are occupied. The *N*-glycans are presented in Table 1. The phosphorylated glycans (compounds 5–10) are derivatives of the neutral structures (compounds 1–4) to which a $(mannose)_{1,2}$-phosphate residue has been added. The deconvoluted ESI-MS spectrum of the catalytic domain is presented in Fig. 1. The mass

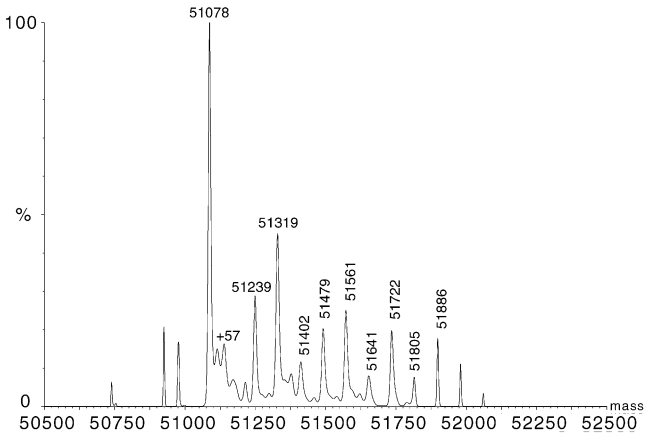
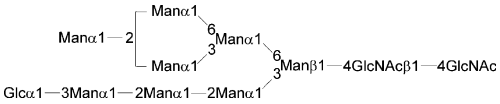
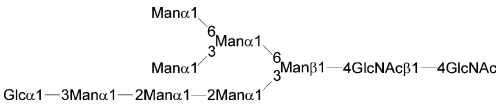
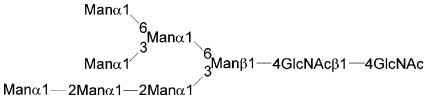
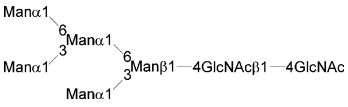
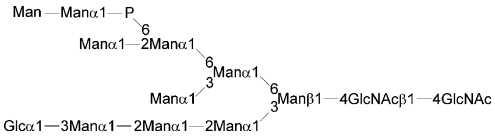
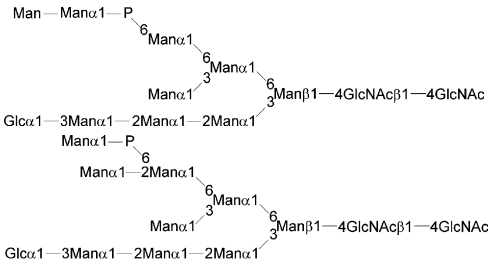
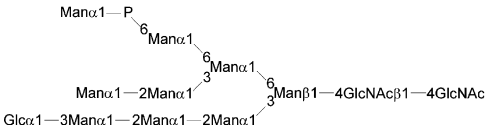
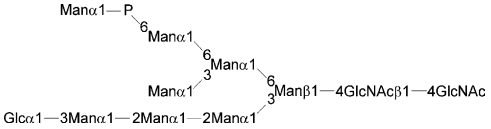
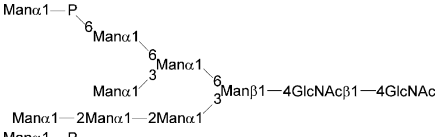
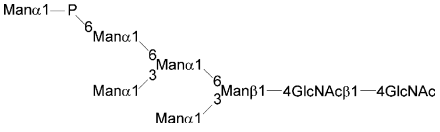


Fig. 1. Deconvoluted ESI-MS spectrum of the catalytic domain of Rut-C30 CBH I. Mass accuracy is typically within 0.01% from the calculated value.

Table 1
Characterised CBH I *N*-glycans from the *T. reesei* strain Rut-C30 [18]

Compound	Composition	Structure	Average MW	Occurrence (%)
1	$GlcMan_8GlcNAc_2$		1865.7	6.0
2	$GlcMan_7GlcNAc_2$		1703.5	44.0
3	$Man_7GlcNAc_2$		1541.4	4.1
4	$Man_5GlcNAc_2$		1217.1	2.3
5	$Man_2PGlcMan_8GlcNAc_2$		2270.0	1.1
6	$Man_2PGlcMan_7GlcNAc_2$		2107.8	2.5
7	$ManPGlcMan_8GlcNAc_2$		2107.8	3.5
8	$ManPGlcMan_7GlcNAc_2$		1945.6	27.9
9	$ManPMan_7GlcNAc_2$		1783.5	1.9
10	$ManPMan_5GlcNAc_2$		1459.2	0.6
Others				6.1

spectrum shows differences of 80 and 162 Da, corresponding to phosphate and hexose, respectively. The calculated average molecular weight of the non-glycosylated catalytic domain is 45,970 Da (major form, see below). When adding three times the mass of the most abundant uncharged *N*-glycan ($GlcMan_7GlcNAc_2$) to the calculated protein mass, a value is obtained corresponding to the most abundant peak in the mass spectrum (51,078 Da). It should be pointed out that papain cleaves the protein at two sites: a major and a minor cleavage site, differing by one glycine residue. This explains the presence of 57 Da spacings in the spectrum. Adding two times the mass of the most abundant uncharged ($GlcMan_7GlcNAc_2$) and once the mass of the most abundant phosphorylated glycan ($ManPGlcMan_7GlcNAc_2$) results in a molecular weight value corresponding to the second most abundant peak in the mass spectrum (51,319 Da).

These simple calculations indicate that three out of four *N*-glycosylation sites are occupied. If two sites are occupied with $GlcMan_8GlcNAc_2$ and one with $GlcMan_7GlcNAc_2$, or if two sites are occupied with $ManPGlcMan_7GlcNAc_2$ and one with $Man_7GlcNac_2$, the molecular weights differ by 2 Da (51,405 Da versus 51,403 Da calculated). Unfortunately, insufficient resolution was achieved to differentiate these species. From this experiment, it is thus impossible to determine the number of phospho-isoforms. In principle, if three glycosylation sites are occupied, four phospho-isoforms are possible, notably one with no phosphorylated glycans at the three sites, one where one glycosylation site bears phosphorylated glycans, one which contains phosphorylated glycans at two sites, and one with only phosphorylated glycans at the three sites. Because the presence of phosphate residues lowers the isoelectric point (p*I*) of the protein, combining a technique that separates proteins on the basis of their p*I* with MS, which could further resolve the glycoforms with the same number of phosphate residues, is thus mandatory.

Capillary isoelectric focusing (CIEF) is a powerful technique for p*I*-based separations and has extensively been reviewed [28–30]. It combines the high-resolving power of conventional gel isoelectric focusing (gel IEF) with well-known advantages of CE. Despite the strength of the technique, there are some drawbacks, such as the variable migration times and protein precipitation. Addition of internal markers can correct for the former. Protein precipitation at the p*I* is a well-known phenomenon; in CIEF this can result in irreproducible patterns, appearance of spikes, fluctuating current, variable mobilisation times, and loss of resolution. Protein solubility in CIEF can be improved by adding, for example, urea at high concentrations to the ampholyte solution. The CIEF electropherogram of the catalytic domain of CBH I isolated from the Rut-C30 strain in the presence of 7 M urea is presented in Fig. 2, as well as the electropherogram of the catalytic domain bearing only three GlcNAc residues (obtained from the *T. reesei* strain QM9414 and named standard CBH I in what follows). CIEF analysis of Rut-C30 CBH I clearly shows more charge-heterogeneity than standard CBH I, due to the presence of the phosphate residues. Four peaks are present, indicating the existence of four phospho-isoforms. The migration time of the first peak corresponds with the migration time of standard non-phosphorylated CBH I. Apparently, the presence of higher molecular weight glycans in the former does not affect the p*I*. In the absence of urea, broad marginally resolved peaks were observed and it was impossible to determine the number of phospho-isoforms. The major and minor papain cleavage sites are not distinguished by CIEF (their p*I* difference is 0.01 pH units).

As stated before, on-line combination of CIEF and MS is not straightforward. Therefore, we performed gel IEF and attempted to isolate the proteins from the gel. Very thin gels with relative large pore sizes are normally used in gel isoelectric focusing. Thus, in principle, the intact proteins can easily be recovered from the gel. The electropherograms of the catalytic domain of Rut-C30 CBH I and of standard CBH I are presented in Fig. 3. Conform to our results in CIEF-analysis, Rut-C30 CBH I displays four bands. Visualisation of the proteins was realised with the fluorogenic substrate 4-methylumbelliferyl β-lactoside that could easily be removed from the protein. The Rut-C30 CBH I bands were excised and the proteins liberated by sonication and separated from ampholytes and fluorogenic substrate by membrane filtration prior to further analysis. The deconvoluted ESI-MS spectra of the four bands are presented in Fig. 4. Within one band, mass differences of 162 Da corresponding to a hexose unit remain, but the spacings of 80 Da (corresponding to phosphate) are no longer present (compare with Fig. 1). The different spectra show similar patterns, although

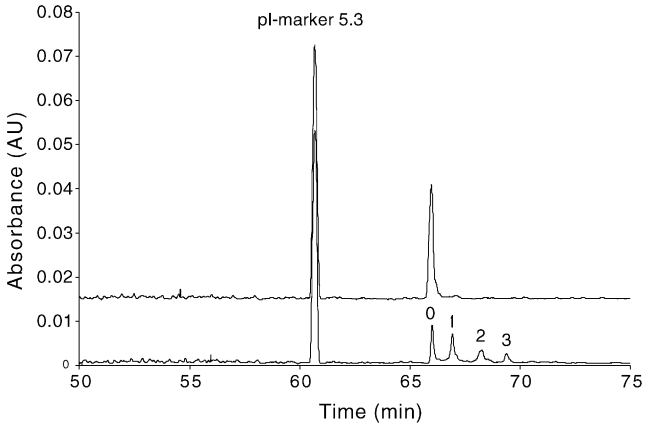


Fig. 2. CIEF electropherograms of the catalytic domain of standard CBH I containing only three GlcNAc residues (upper trace) and of CBH I isolated from the Rut-C30 strain (lower trace). The numbers 0–3 correspond with the number of phosphate residues and are thus phospho-isoforms. The region right of the p*I* marker corresponds to the more acidic region.

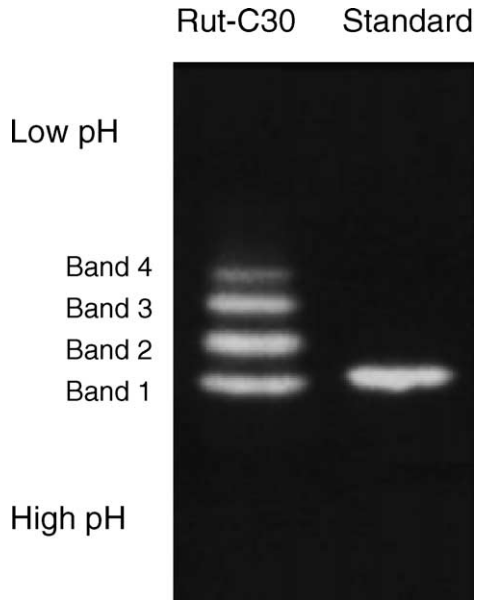


Fig. 3. Gel IEF electropherogram of the catalytic domain of standard CBH I (right) and of CBH I isolated from the Rut-C30 strain (left).

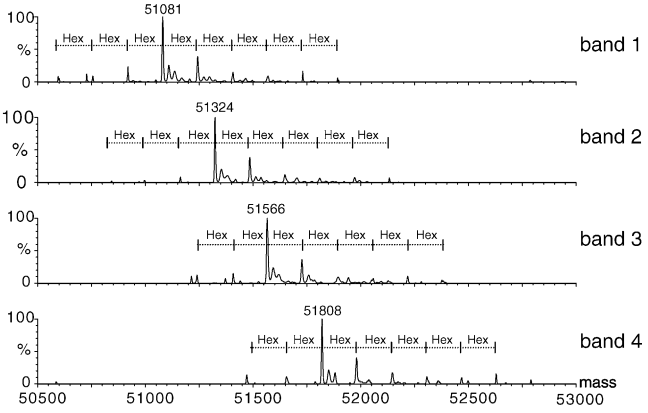


Fig. 4. Deconvoluted ESI-MS spectra of the proteins eluted from the different IEF bands (see Fig. 3 left lane). Hex corresponds to hexose, with a molecular weight of 162 Da. Band 1 corresponds to the lower band, band 4 to the upper band in Fig. 3.

a shift of 242 Da, corresponding to mannose-phosphate, is noticed for subsequent bands. The most abundant peak in the spectrum of bands 1–4 corresponds to the catalytic domain that bears respectively three $GlcMan_7GlcNAc_2$ residues, two $GlcMan_7GlcNAc_2$ and one $ManPGlcMan_7GlcNAc_2$, one $GlcMan_7GlcNAc_2$ and two $ManPGlcMan_7GlcNAc_2$, and three $ManPGlcMan_7GlcNAc_2$. Subsequent bands thus differ in the number of phosphate residues. It is now possible to assign all the peaks and to determine all existing glycoforms. In comparison to the mass spectrum of the unfractionated protein, some new peaks are detected in the spectra of the different phospho-isoforms.

In contrast to on-line methods, further investigations can be performed on the fractionated phospho-isoforms using the above-described method. By way of example, Fig. 5 represents the CE-LIFD electropherograms of the APTS-derivatised *N*-glycans isolated from each phospho-isoform. Indeed, an increasing intensity of the phosphorylated sugars is noticed for subsequent bands (see Fig. 3 left lane). The CE-LIFD method has been described elsewhere [18]. Because the EOF drives the sugars towards the detector, which is located at the cathode, the phosphorylated species migrate more slowly than their neutral counterparts. Further analysis is not restricted to structural analysis but also kinetic experiments, amongst others, can be performed on the different phospho-isoforms.

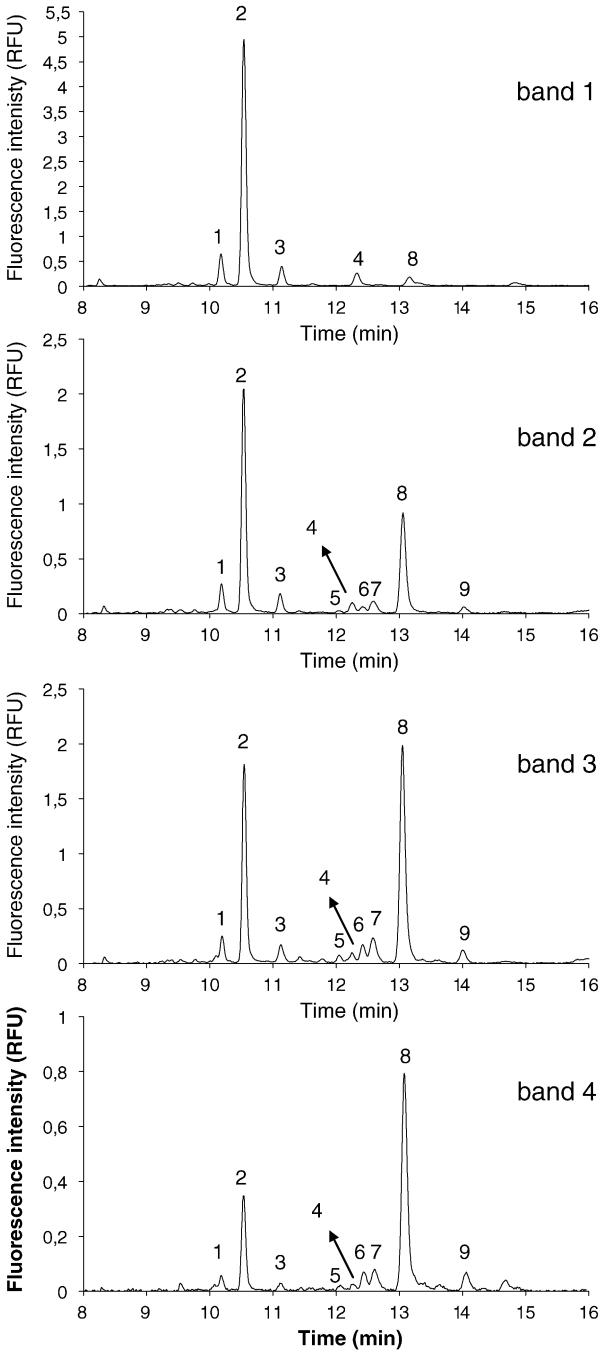


Fig. 5. CE-LIFD electropherograms of the APTS derivatised Rut-C30 CBH I *N*-glycans enzymatically liberated from the different phospho-isoforms by PNGase F. The numbers on top of the peak correspond with those in Table 1.

3.2. Glycopeptide analysis

The above-described measurements do not indicate which asparagine residues are glycosylated. Four *N*-glycosylation sites are present, although only three sites are actually glycosylated. Information on the glycosylation sites can be obtained from the study of glycopeptides. Carr and co-workers [8] demonstrated that glycopeptides can easily be detected in tryptic digests by precursor ion scanning using triple quadrupole instruments. The Q-Trap system, with its capabilities to perform typical triple quadrupole scans, next to improved ion trap scan functions, can also be used for that purpose. The precursor ion scan mode can be used as survey scan in an information dependent acquisition (IDA) experiment. This is typically useful in combination with LC/MS/MS, and glycopeptides can be identified ‘on the fly’. The IDA procedure consisted of the following experiments: a precursor ion scan as survey scan, followed by an enhanced resolution

Fig. 6. Amino acid sequence of the catalytic domain of *T. reesei* CBH I. The four possible *N*-glycosylation sites are encircled. The tryptic cleavage sites and the generated tryptic peptides are indicated. The minor form contains an additional glycine residue at the C-terminus.

(ER) scan of the most intense ion, and an enhanced product ion (EPI) scan of this ion if the charge state is higher than two (total cycle time with two ER and EPI scans summed ∼5 s). The collision energy is automatically adjusted based upon the ion's charge state and mass. The enhanced resolution scan is not only useful to obtain the peptide's charge state but also to improve mass accuracy for proper precursor selection. If an ion is detected twice in an enhanced resolution scan it is automatically excluded during 60 s, so as to obtain as much information as possible. For a more detailed description of the use of the Q-Trap mass spectrometer in the study of protein glycosylation we refer to [23].

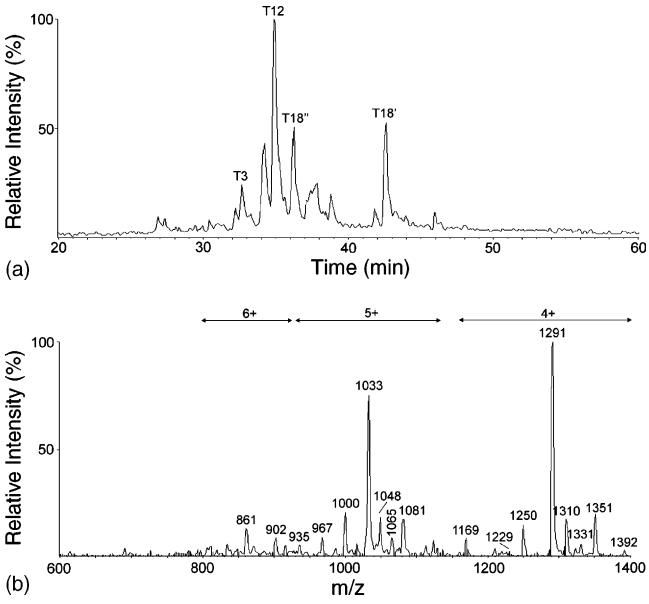


Fig. 7. (a) LC/MS/MS base peak chromatogram of the precursor ion scan at *m/z* 204. (b) Combined precursor ion spectra of 42.2–43.3 min. The charge states of the peptides are indicated.

The amino acid sequence of the catalytic domain of CBH I is presented in Fig. 6. The potential glycosylation sites are marked. The catalytic domain contains 10 disulfide bonds and reduction and alkylation prior to proteolysis is essential for

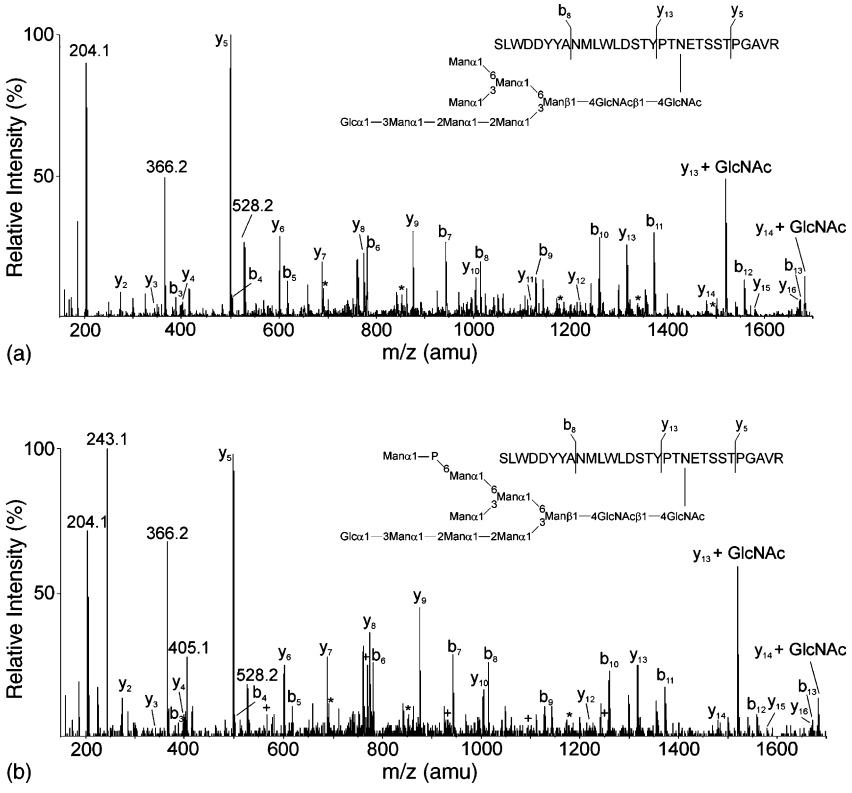


Fig. 8. MS/MS spectra and structures of the four fold charged glycopeptides at *m/z* 1291 (a) and 1351 (b). The asterisk (*) corresponds to sugar oxonium ions without the phosphate moiety; the plus (+) corresponds to sugar oxonium ions with the phosphate group. Collision energy was relatively high, respectively 58 and 60 eV in (a) and (b).

Table 2
Glycopeptides identified from the LC/MS/MS experiment using *m/z* 204 in precursor ion scanning

Tryptic peptide	Sequence[a]	Retention time of glycopeptides with uncharged glycans	Retention time of glycopeptides with phosphorylated glycans[b]
T3	WTHATNSSTNCYDGNTWSSTLCPDNETCAK	32.7	n.d.
T12	LGNTSFYGPGSSF	34.9	35.2
T18′	SLWDDYYANMLWLDSTYPTNETSSTPGAVR	42.6	42.9
T18′′	MLWLDSTYPTNETSSTPGAVR	36.2	36.5

[a] The glycosylated Asn residues are underscored.
[b] The phosphorylated T3 glycopeptide could not be unambiguously determined using precursor ion scanning at *m/z* 204.

identification of all glycosylation sites. The nano-LC/MS/MS base peak chromatogram of the precursor ion scan of *m/z* 204 is presented in Fig. 7a and, as an example, the combined precursor ion spectra of 42.2–43.3 min in Fig. 7b.

The MS/MS spectrum of the most abundant glycopeptide (*m/z* 1291) is presented in Fig. 8a. Peptide fragments are accompanied with glycan fragments. From the b- and y-ions, one can easily elucidate the sequence corresponding to the T18 peptide that has been aspecifically cleaved by trypsin between methionine and serine at position 364 (T18′). The intact T18 peptide is not detected in the LC/MS run. By combining the sequence information gathered from the b- and y-ions, nearly complete sequence coverage is obtained (except for one amino acid). Since only one Asn-X-Ser/Thr consensus sequence is present in the peptide there is no doubt about the occupied *N*-glycosylation site. The peptide mass is known (3455 Da), and combining this with the mass of the $GlcMan_7GlcNAc_2$ glycan (1703.5 Da) explains the quadruply charged ion at *m/z* 1291. Adding the other uncharged glycans (compounds 1, 3 and 4, see Table 1) to the peptide explains the quadruply charged ions at *m/z* 1169.0, 1250.0 and 1331.0 seen in Fig. 7b. Similar abundance is noticed as for the released uncharged glycans. Also five and six times charged glycopeptides are detected. For these more energetic species, minor in-source fragmentation is observed.

The MS/MS spectrum of the ion at *m/z* 1351 is presented in Fig. 8b. The same b- and y-ions are present as for the ion at *m/z* 1291, however, additional glycan fragments (oxonium ions) are noticed (at *m/z* 243.1, 405.1, 567.2, 770.3, 932.3, 1094.3, 1256.4 and 1418.4) corresponding to fragments containing the phosphate group. The unusual mannose-phosphate oxonium ion (*m/z* 243.1) is very intense and can thus also be used in precursor ion scanning for selective detection of glycopeptides containing the phosphate moiety. It is unlikely that

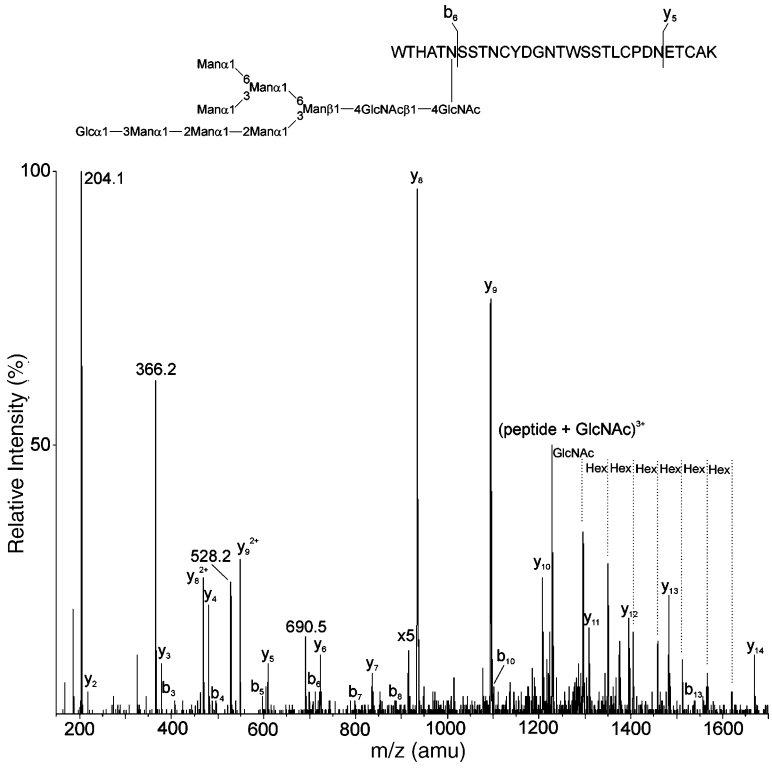


Fig. 9. MS/MS spectrum and structure of the four-fold charged T3 glycopeptide at *m/z* 1296. 'x5' means that the intensity of the peak is multiplied with a factor five. This ion corresponds to the b_6 ion with the GlcNAc attached, indicating that Asn in the consensus sequence NSS is actually glycosylated. Collision energy was at 58 eV.

the phosphate is bound to the outer mannose unit since this linkage is very labile. The ion at *m/z* 1351 corresponds to a quadruply charged species with the $ManPGlcMan_7GlcNAc_2$ attached to the T18′ fragment. The ion at *m/z* 1310 in Fig. 7b corresponds to a form where the labile mannosyl-phosphate bond of the $ManPGlcMan_7GlcNAc_2$ is broken in the source, although it may also correspond to a species where the $ManPMan_7GlcNAc_2$ is attached to the T18′ peptide. The ions at *m/z* 1229.0 and 1392.0 further correspond to glycopeptides where the $ManPMan_5GlcNAc_2$ and $ManPGlcMan_8GlcNAc_2$ or $(Man)_2PGlcMan_7GlcNAc_2$ are attached to T18′. Again, five and six times charged phosphorylated glycopeptides are present accompanied by in-source fragmentation. It should further be noted that the phosphorylated glycopeptides elute later than their corresponding glycopeptides containing uncharged glycans, although they are not completely resolved. This phenomenon is not restricted to the T18′ glycopeptide, but occurs with all glycopeptides.

All identified CBH I glycopeptides are presented in Table 2. The T3 glycopeptide contains two consensus sequences. To determine the actual glycosylation site, fragments have to be present in the MS/MS spectrum where a portion of the glycan moiety (usually GlcNAc) is still attached to the peptide backbone. No such y-ions are present in the spectrum. On the other hand, b-ions are noticed with GlcNAc bound at low intensity, indicating that Asn 45 is glycosylated. This correlates with previous findings [31]. The MS/MS spectrum of the T3 glycopeptide with the $GlcMan_7GlcNAc_2$ moiety attached is presented in Fig. 9.

Some peaks in the chromatogram (Fig. 7a) correspond to non-glycosylated peptide material, e.g. at time 34.2 min. Carr and co-workers, and Roepstorff and co-workers [8,32] reported a lack of specificity in precursor ion scanning using low-resolution quadrupole instrumentation, since non-glycosylated peptides can give rise to fragment ions with similar mass as the sugar oxonium ions. The LC/MS/MS base peak chromatogram of the precursor ion scan monitoring the phosphorylated oxonium ion at *m/z* 243 and the combined spectra in the peak at time 41.4 min are presented in Fig. 10. The spectrum is similar to that in Fig. 7b but the glycopeptides containing uncharged glycans are no longer detected. Only the peptides containing phosphorylated glycans give rise to the specific oxonium ion, thus presenting an elegant method for its detection in proteolytic digests. The peak at time 32.9 corresponds to a non-glycosylated peptide. This peptide was identified as the N-terminal peptide T1 from the MS/MS data. The reason why this peptide was detected is that an a_3 ion can be formed with an *m/z* value of 242.1. Q3 possessed insufficient resolution to resolve this ion from the phosphorylated oxonium ion at *m/z* 243.1. The phosphorylated T3 peptide could not be detected when monitoring the HexNAc oxonium ion. Monitoring the ion at *m/z* 243, however, allowed this peptide to be detected.

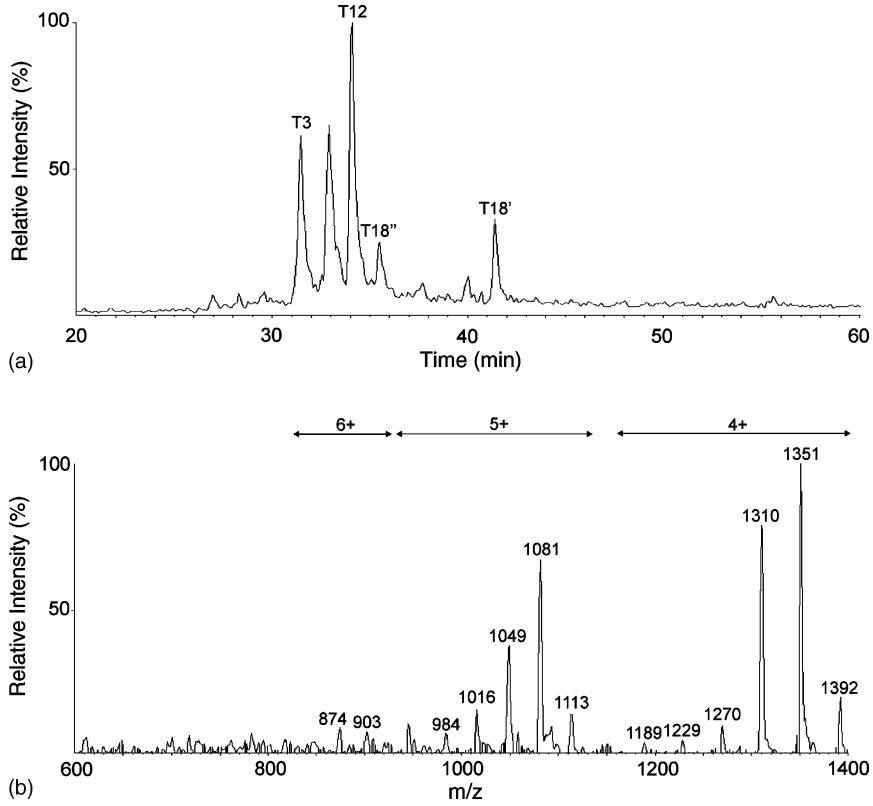


Fig. 10. (a) LC/MS/MS base peak chromatogram of the precursor ion scan at *m/z* 243. (b) Combined precursor ion spectra over the peak at retention time 41.4 min.

From the above-described measurements it is assumed that all glycans (see Table 1) can be present at the three characterised glycosylation sites.

As a general conclusion it can be stated that IEF combined with MS allowed to distinguish the different glyco- and phospho-isoforms of CBH I. Precursor ion scanning, monitoring both a glycan and phospho-sugar specific oxonium ion allowed unambiguous identification of the glycosylation site and to determine the heterogeneity at each site. The method can now be applied for complete analysis of cellulase glycosylation from different strains, different organisms, mutated organisms, cellulases expressed in other organisms, etc. The described methodology, however, is not restricted to cellulase, and can also be applied to study glycosylation of other phosphorylated or sulphated glycoproteins if chromo- or fluorogenic substrates are present.

Acknowledgements

This work was supported by a research grant to K.S. from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT-Vlaanderen) and by the Fund for Scientific Research-Flanders (FWO-Vlaanderen) through Project G.0312.02 to J.V.B. and B.D.

References

[1] Y. Mechref, M. Novotny, Chem. Rev. 102 (2002) 321.
[2] P.M. Rudd, C. Colominas, L. Royle, N. Murphy, E. Hart, A.H. Merry, H.F. Hebestreit, R.A. Dwek, Proteomics 1 (2001) 285.
[3] A. Paulus, A. Klockow-Beck, in: K.D. Altria (Ed.), Analysis of Carbohydrates by Capillary Electrophoresis, Vieweg, Wiesbaden, Germany, 1999.
[4] D.J. Harvey, Proteomics 1 (2001) 311.
[5] J. Zaia, Mass Spec. Rev. 23 (2004) 161.
[6] A.H. Que, M.V. Novotny, Anal. Chem. 74 (2002) 5184.
[7] L.A. Gennaro, D.J. Harvey, P. Vouros, Rapid Commun. Mass Spectrom. 17 (2003) 1528.
[8] M.J. Huddleston, M.F. Bean, S.A. Carr, Anal. Chem. 65 (1993) 877.
[9] J.P. Landers, R.P. Oda, B.J. Madden, T.C. Spelsberg, Anal. Biochem. 205 (1992) 115.
[10] P.M. Rudd, I.G. Scragg, E. Coghil, R.A. Dwek, Glycoconjugate J. 9 (1992) 86.
[11] G. Hunt, T. Hotaling, A.B. Chen, J. Chromatogr. A 800 (1998) 355.
[12] A. Cifuentes, M.V. Moreno-Arribas, M. de Frutos, J.C. Diez-Masa, J. Chromatogr. A 830 (1999) 453.
[13] D.J. Harvey, Int. J. Mass Spectrom. 226 (2003) 1.
[14] M. Karas, U. Bahr, T. Dülcks, Fresenius J. Anal. Chem. 366 (2000) 669.
[15] I. Stals, K. Sandra, S. Geysens, R. Contreras, J. Van Beeumen, M. Claeyssens, Glycobiology 14 (8) (2004) 713.
[16] I. Stals, K. Sandra, B. Devreese, J. Van Beeumen, M. Claeyssens, Glycobiology 14 (8) (2004) 725.
[17] M. Maras, A. De Bruyn, J. Schraml, P. Herdewijn, M. Claeyssens, W. Fiers, R. Contreras, Eur. J. Biochem. 245 (1997) 617.
[18] K. Sandra, J. Van Beeumen, I. Stals, P. Sandra, M. Claeyssens, B. Devreese, Anal. Chem (2004), in press.
[19] J. Wei, L. Yang, A.K. Harrata, C.S. Lee, Electrophoresis 19 (1998) 2356.
[20] Q. Tang, A.K. Harrata, C.S. Lee, Anal. Chem. 68 (1996) 2482.
[21] C.X. Zhang, F. Xiang, L. Pas-Tolic, G.A. Anderson, T.D. Veenstra, R.D. Smith, Anal. Chem. 72 (2000) 1462.
[22] A. Chartogne, U.R. Tjaden, J. van der Greef, Rapid Commun. Mass Spectrom. 14 (2000) 1269.
[23] K. Sandra, B. Devreese, I. Stals, M. Claeyssens, J. Van Beeumen, J. Am. Soc. Mass Spectrom. 15 (2004) 413.
[24] K. Klarskov, K. Piens, J. Stahlberg, P.B. Hoj, J. Van Beeumen, M. Claeyssens, Carbohydr. Res. 304 (1997) 143.
[25] P. Tomme, S. McRae, T.M. Wood, M. Claeyssens, Methods Enzymol. 160 (1988) 187.
[26] P. Jackson, Methods Enzymol. 230 (1994) 253.
[27] H. van Tilbeurgh, F.G. Loontiens, C.K. De Bruyne, M. Claeyssens, Methods Enzymol. 160 (1988) 45.
[28] X. Liu, Z. Sosic, I.S. Krull, J. Chromatogr. A 735 (1996) 165.
[29] P.G. Righetti, C. Gelfi, M. Conti, J. Chromatogr. B 699 (1997) 91.
[30] T. Wehr, R. Rodriguez-Diaz, M. Zhu, Chromatographia 53 (2001) S-45.
[31] J.P.M. Hui, P. Lanthier, T.C. White, S.G. McHugh, M. Yaguchi, R. Roy, P. Thibault, J. Chromatogr. B 752 (2001) 349.
[32] J. Jebanathirajah, H. Steen, P. Roepstorff, J. Am. Soc. Mass Spectrom. 14 (2003) 777.