
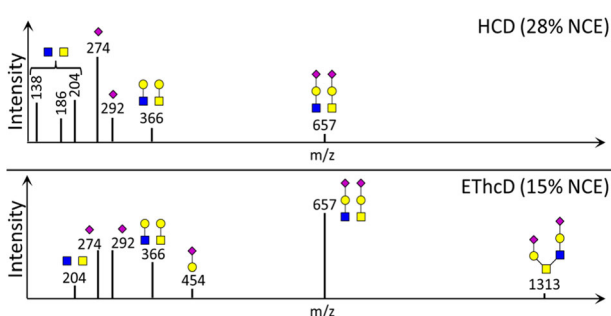


Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures

Adam Pap, Eva Klement, Eva Hunyadi-Gulyas, Zsuzsanna Darula, ,
Katalin F. Medzihradzsky

Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary



Abstract. A very complex mixture of intact, human N- and O-glycopeptides, enriched from the tryptic digest of urinary proteins of three healthy donors using a two-step lectin affinity enrichment, was analyzed by LC-MS/MS, leading to approximately 45,000 glycopeptide EThcD spectra. Two search engines, Byonic and Protein Prospector, were used for the interpretation of the data, and N- and O-linked glycopeptides were assigned from separate searches. The identification rate

was very low in all searches, even when results were combined. Thus, we investigated the reasons why was it so, to help to improve the identification success rate. Focusing on O-linked glycopeptides, we noticed that in EThcD, larger glycan oxonium ions better survive the activation than those in HCD. These fragments, combined with reducing terminal Y ions, provide important information about the glycan(s) present, so we investigated whether filtering the peaklists for glycan oxonium ions indicating the presence of a tetra- or hexasaccharide structure would help to reveal all molecules containing such glycans. Our study showed that intact glycans frequently do not survive even mild supplemental activation, meaning one cannot rely on these oxonium ions exclusively. We found that ETD efficiency is still a limiting factor, and for highly glycosylated peptides, the only information revealed in EThcD was related to the glycan structures. The limited overlap of results delivered by the two search engines draws attention to the fact that automated data interpretation of O-linked glycopeptides is not even close to being solved.

Keywords: EThcD, Fragmentation, Glycan, O-Glycopeptide, MS/MS

Received: 12 December 2017/Revised: 9 March 2018/Accepted: 9 March 2018/Published Online: 5 May 2018

Introduction

Recently, there has been a growing interest in protein glycosylation analysis. The main driving force may be the widespread pharmaceutical use of recombinant proteins as therapeutics, most of which are glycosylated. Since changes in glycosylation may alter the physical properties of these biologicals, as well as their biological activity and

immunogenicity, site-specific, in-depth glycosylation analysis of these proteins is essential [1]. In addition to assisting with proper protein folding, protein processing, and controlling protein survival in the circulation, extracellular glycosylation plays a crucial role in cell adhesion [2–4], influences intracellular processes [3, 5–8], and may be altered by disease [9, 10], and glycosylation defects may cause disease [9, 11]. Site-specific alteration of glycosylation has been implicated in receptor activation [12, 13], and an interplay between mucin-type O-glycosylation of fibroblast growth factor 23 on Thr-178 and its phosphorylation at Ser-180 seems to control phosphate balance [14].

Mass spectrometry, with or without MS/MS analysis, has been used for glycopeptide characterization for decades [15–

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13361-018-1945-7>) contains supplementary material, which is available to authorized users.

Correspondence to: Zsuzsanna Darula; e-mail: darula.zsuzsanna@brc.mta.hu

21]. There are numerous studies where purified glycopeptides, both N-linked [15, 18, 22, 23] and O-linked [16, 19, 24], have been characterized using intact mass measurements, enzymatic digestion, and collision-induced dissociation. Glycosidic bonds are weaker than peptide bonds. Thus, CID spectra are usually dominated by glycan fragments, generally displaying little information about the underlying amino acid sequence. This usually does not represent a significant problem in single-protein analysis when there are only a few potential glycosylation sites, but is a serious problem in mixture analysis, especially in global studies. Radical-based fragmentation techniques ECD [25] and ETD [26] offer an alternative approach where the almost exclusive observation of peptide backbone cleavages provides information for both the peptide sequence and modification site assignments. In these fragmentation modes, the side-chains usually remain intact, although the precursor ion may lose some glycan units, especially sialic acids [27]. Recently, a combination of energy- and radical-based fragmentation, EThcD [28], has been implemented in high-end Orbitrap mass spectrometers (Orbitrap Fusion and Lumos). In EThcD analyses, ETD activation is performed first; then, all fragments as well as the surviving precursor ions are subjected to mild HCD activation, before the resulting products are measured. Based on studies conducted on unmodified [28] and phosphopeptides [29], fragments produced by the ETD process do not fragment further at the HCD energies used (except radical z ions may yield w ions [30]), but the activated precursor ions are fragmented. Although glycopeptides are more prone to fragmentation upon collisional activation than phosphopeptides, secondary fragmentation of c/z ions (for peptide fragment ion nomenclature, see [31]) has not been reported for either N- [32–34] or O-glycopeptides [35]. Hence, this activation method seems to be ideal for glycopeptide characterization as c/z ions generated by ETD enable peptide sequence identification and modification site assignment while supplemental HCD activation yields information on the glycan structure in the form of B/Y ions (for carbohydrate fragment ion nomenclature, see [36]) and may provide additional sequence coverage by generating additional b/y ions.

In practice, ion trap CID using resonance-activation yields information about the glycan structure and the peptide size [27, 37]; beam-type CID (HCD), by producing multiple collisions, may produce a “balanced” spectrum with informative glycan and peptide fragments or may lead to comprehensive peptide fragmentation and break the glycans to smithereens [38–40]; ETD may deliver data that are good enough for both peptide and site identification, provided the modifying oligosaccharide is listed in the queried glycan database [27, 41–45]. In general, each activation method provides some, but not all, of the clues needed to decipher the glycopeptide structure, though good quality EThcD is getting close to the goal (for a recent review of protein O-glycosylation including MS/MS characteristics of O-glycopeptides, see [46]). Mass spectrometry provides limited information about the glycan structure. Normally, LC-MS/MS data do not reveal the ring linkage positions and the linkage stereochemistry cannot be deciphered using mass

spectrometry. Isomeric oligosaccharide building blocks may not be distinguishable, although beam-type CID (HCD) fragmentation has recently been shown to be able to distinguish GlcNAc and GalNAc residues based on their different fragment ion intensity profile [41]. Automated glycopeptide assignments identify the glycan composition and link it to a glycan listed in the database. Glycan assignment is often based on supporting information about the glycans present, such as glycan analysis, knowledge of the potential glycan structures in the sample, or by using glycan structure-specific purification prior to MS.

Data acquired with different activation techniques are currently searched independently by search engines. There have been some attempts to use ETD and CID/HCD data together, but only for N-glycopeptides, and the identification was based on either the HCD data [47–49] or ETD spectra [46], while the other dataset provided confirmation [43–45] or structural information about the glycan [50]. We hope that eventually software that handles all available data in an interactive fashion will be developed, but currently, researchers manually integrate information to produce optimal data interpretation [51–53].

Some large-scale N-glycopeptide studies have used beam-type CID (HCD) data [54–59], whereas others have employed ETD [42, 60–62]. For success in high-throughput O-glycosylation studies, the use of ETD is essential [42–44, 63]. The first large-scale studies using EThcD to study N-glycosylation [32–34] and O-glycopeptides [35] have recently been published.

In the present study, we applied EThcD for the large-scale analysis of a very complex N- and O-glycopeptide mixture, human urine, with primary focus on O-glycosylation. In order to obtain a comprehensive view of the modifying glycans, tryptic glycopeptides were enriched with lectin weak affinity chromatography using wheat germ agglutinin (WGA), a lectin that has been shown to bind a wide array of glycan structures [42, 64]. The glycopeptide mixtures were analyzed by LC-MS/MS using HCD product-ion-dependent EThcD data acquisition. Since reliable O-glycopeptide assignment, permitting multiple different oligosaccharide structures simultaneously, is still a very hard task to tackle, we used two search engines, Byonic [65] and Protein Prospector (<http://prospector.ucsf.edu>). Both software search for b/y and c/z peptide fragments (for nomenclature, see [31]). Prospector also considers the products of hydrogen migration, ($c - 1$ and $z + 1$ ions) but ignores the glycan fragments. Byonic considers B and additional glycan oxonium ions, as well as Y fragments resulting from glycan fragmentation (for nomenclature, see [36]), and looks for two additional peptide fragments: a- and b-H₂O. Obviously, the search engines also use different scoring systems. Using two different tools should lead to more glycopeptide identifications (as it indeed did) and adds confidence to the shared assignments.

In addition, confidently assigned spectra were inspected manually in order to establish rules about the EThcD fragmentation of glycopeptides. It was obvious at first glance that glycan fragmentation is prominent but somewhat different from the corresponding HCD spectra of the same glycopeptide

precursors. Thus, we investigated how general this phenomenon is and how glycopeptide characterization could benefit from it.

The focus of the present study was not the comprehensive characterization of urinary glycopeptides or comparing the performance of different search engines, but rather to evaluate whether a new analysis approach could help to identify more components, and to draw attention to existing problems that are usually ignored.

Experimental

Sample Preparation

One hundred milliliters of urine from three healthy volunteers (sample A, 46-year-old female; sample B, 46-year-old male; sample C, 26-year-old male) were used for the studies. Samples were collected with appropriate consents approved by the regulatory and ethical authorities (ethics approval number of the Hungarian Scientific and Research Ethics Committee: 1011/16). Protein concentration of the samples ($100 \pm 10 \mu\text{g/ml}$) was determined by the Bradford assay.

Cell debris was removed by centrifugation (5000g, 10 min) and the resulting urine was concentrated on 10-kDa MWCO ultracentrifugation devices (Millipore). The concentrate was supplemented with guanidine hydrochloride (to a final concentration of 6 M), followed by reduction with 20 μl DTT (500 mM in 25 mM ammonium bicarbonate) and alkylation with 40 μl iodoacetamide (500 mM in 25 mM ammonium bicarbonate). The mixtures were washed with guanidine (6 M in 25 mM ammonium bicarbonate) then with ammonium bicarbonate (25 mM) followed by incubation with 100 μg trypsin (37 °C, 12 h). Glycopeptides were enriched by two rounds of affinity chromatography using a homemade column packed with wheat germ agglutinin immobilized on POROS [36] as described in [66]. Two fractions were collected per injection: a “shoulder” fraction (fraction 1) and a “GlcNAc” fraction (fraction 2, eluted with 200 μl GlcNAc (200 mM in 150 mM ammonium bicarbonate)), representing weakly and more strongly bound glycopeptides, respectively. After the first round of enrichment, the GlcNAc fraction was desalted on C18 SepPak cartridges (Waters); the fractions were combined and subjected to a second round of enrichment. After the second round of enrichment, the fraction 2 was desalted, and all samples were dried down.

Mass Spectrometry

The glycopeptide mixtures were analyzed by LC-MS/MS using an Acquity UPLC MClass System (Waters) on-line coupled to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific) operating in positive ion mode. Five percent of the isolated peptide mixtures were injected for each LC-MS/MS analysis. Fractions 1 and 2 were analyzed separately. After trapping at 3% B (Waters Acquity UPLC MClass Symmetry C18 180 $\mu\text{m} \times 20 \text{ mm}$ column, 5- μm particle size,

100-Å pore size; flow rate 10 $\mu\text{l/min}$; solvent A, 0.1% formic acid/water; solvent B, 0.1% formic acid/ACN; flow rate 300 nL/min), peptides were separated using a linear gradient of 10 to 30% B in 60 min (Waters Acquity UPLC MClass BEH C18 75 $\mu\text{m} \times 250 \text{ mm}$ column, 1.7- μm particle size, 130-Å pore size).

Each MS survey scan (m/z 380–1580, $R = 60,000$, acquired in profile mode) was followed by a maximum 3-s cycle collecting MS/MS data of precursors in the order of decreasing charge state ($z = 3-5$) then by increasing m/z (minimum intensity 10^6). Precursor ions were isolated with the quadrupole (isolation window 2 Da). HCD data (AGC target 50000, normalized collision energy (NCE) 28%) were acquired for each precursor, while EThcD data acquisition (AGC target 300000, supplemental activation energy 15% NCE) was triggered by the presence of diagnostic sugar oxonium ion m/z 204.0867 (for N-acetylhexosamine, HexNAc) among the 20 most abundant fragment ions of the HCD spectrum, with a mass tolerance of 10 ppm. All MS/MS spectra were acquired in the Orbitrap ($R = 15,000$, centroid mode). Dynamic exclusion was enabled (maximum 1 HCD and EThcD spectra/precursor in 30 s).

Data Interpretation

Proteome Discoverer (Thermo Scientific, v2.2.0.388) was used to generate separate HCD and EThcD peaklists from the raw data in mgf format. A minimum peak count of 10 was required to retain the MS/MS spectrum. EThcD peaklists were filtered using the MS-Filter program of Protein Prospector [67] for the presence of sialic acid oxonium ion $m/z = 292.1027$ (mass tolerance 10 ppm) within the 80 most abundant fragment ions (since most spectra feature less fragment ions), then searched using the Protein Prospector Batch Tag Web (v5.16.0.) and Byonic (v2.13.17) search engines. Protein Prospector used the 40 most abundant ions from each half of the spectral mass range in the database searches. Byonic uses practically all the observed peaks. N- and O-glycopeptides were searched separately with the following parameters: database human subset of the Swissprot database (2017.9.19.version, 20,219 sequences) concatenated with a randomized sequence for each protein entry; enzyme: semitrypsin with maximum 1 missed cleavage site; mass accuracy: 5 ppm for precursor ions and 10 ppm for fragment ions specified as monoisotopic values; fixed modification: carbamidomethylation (Cys); variable modifications: acetylation (protein N-terminus), cyclization (peptide N-terminal Gln), and oxidation (Met); and maximum number of variable modifications per peptide: 2. For O-glycopeptide identifications, the HexNAcHex, HexNAcHexNeuAc, HexNAcHexNeuAc₂, and HexNAc₂Hex₂NeuAc₂ glycan structures were also considered as “common” variable modifications. For N-glycopeptide searches, the “57 human N-glycans” database was considered as additional “rare” variable modification (1 N-glycan per peptide). Acceptance criteria for identifications are as follows: Protein Prospector searches: maximum FDR values: fraction 1: 5 and 1%, fraction 2: 10 and 5% for protein and peptide identifications, respectively,

and SLIP score ≥ 6 for estimation of site assignment reliability [68]; Byonic searches: maximum protein FDR value: 1%, Pep2D score < 0.1 [69].

MS-Filter was further used to screen for the presence of specific carbohydrate oxonium ions as described in the “Results and Discussion” section. All ions were searched with ± 10 -ppm mass tolerance using the instrument type specification as ESI-EThcD-high-res except for m/z 1313.4625 that was searched with instrument type ESI-Q-high-res.

Results and Discussion

Affinity-based glycopeptide enrichment was performed from tryptic digests of three individual human urine samples (labeled A, B and C), using wheat germ agglutinin, which binds a wide array of glycopeptides [40]. Using human serum tryptic digests, we have previously observed that singly glycosylated O-glycopeptides tend to elute at the end of the flow-through fraction, and the background of non-glycosylated peptides is high. Multiple modified O-glycopeptides are predominantly present in the fraction eluted with GlcNAc (unpublished data). In order to maximize peptide spectrum matches (PSMs), we collected and analyzed these fractions separately (referred to as “fraction 1” and “fraction 2”). During LC-MS/MS analysis, the presence of the diagnostic HexNAc oxonium ion, m/z 204, in the HCD spectrum triggered EThcD data acquisition. Since glycopeptides are usually larger than unmodified peptides, not only singly but also doubly charged ions were excluded from precursor ion selection. In the EThcD experiments, the default value of supplemental activation energy (15%) was used. Approximately 45,000 EThcD spectra were acquired (Online Resource 1).

Since the enrichment method has been reported as non-discriminative [40, 59, 60, 63], the glycan structures present could not be predicted. Initial screening of the MS/MS data for the diagnostic monosaccharide oxonium ion m/z 292.103 of N-acetylneuraminic acid indicated a predominance of sialylated glycopeptides (Online Resource 2). Thus, we focused on such structures, and database searches were performed with the m/z 292-filtered peaklists. For O-glycopeptide identification, two search engines were used, Protein Prospector and Byonic. Glycan structures representing the di-, mono-, and nonsialylated core-1 and the disialylated core-2 O-glycans were considered as potential modifications, presuming that urine has a similar O-glycan distribution to plasma [70]. A previous study using sialic acid-based enrichment of urinary glycoproteins also indicated the dominance of core-1 and core-2 O-glycans [49]. For N-glycopeptide identification, only Byonic was used, and a larger N-glycan database representing the major plasma N-glycans was considered. This glycan database contains all the N-glycans identified on urinary glycoproteins in an earlier study [49]. All O-glycopeptide identifications meeting the acceptance criteria are listed in Online Resource 3, while N-glycopeptide identifications are presented in Online Resource 4. More O-glycopeptides were identified than

N-glycopeptides. O-glycopeptide identifications are in good agreement with previous results: 67% of the identified O-glycosylated sequences reported by Halim [49] were identified in the present study except that we also report on the number of sialic acids present. The overlap of N-glycopeptides is much less impressive (20%). The overall spectral identification rate was rather low. Fraction 1 yielded more assignments. For the O-glycopeptides, the search engines performed quite similarly: Byonic yielded 556 PSMs in comparison to 552 delivered by Protein Prospector, with 343 shared identifications. Combined with the 328 N-glycopeptide assignments, $\sim 5\%$ of the fragmentation spectra are accounted for. Fraction 2 yielded less, 326 and 304 PSMs by Byonic and Prospector, respectively, with 173 shared identifications. The O-glycopeptide identifications combined with the 109 assignments from the N-glycosylation searches cover $\sim 2.5\%$ of the data. Since the later eluting species most likely feature larger and/or more oligosaccharide structures, these results are not entirely unexpected. However, the overall success rate is disheartening. There are numerous analysis approaches one could try to increase identifications. For example, only two O-glycans/peptide were permitted, even though sequences containing up to 12 GalNAc modifications have been reported from Simple Cell experiments [41]. We have previously reported triply and quadruply O-glycosylated peptides from human serum, albeit after removing the sialic acids [71]. One could also search for additional glycoforms in glycoproteins already confidently assigned in the mixture. Searches could be performed with relaxed enzyme specificity, since the presence of other proteolytic activity was amply detected. In addition, deamidation of Asn and Gln residues most likely occurred during the sample preparation; thus, additional variable modifications could be introduced. However, opening up the search space introduces additional issues, and as we present below, the reliability of glycopeptide assignments, at least for O-glycopeptides, has not been solved even on this “conservative” analysis level.

We scrutinized the data to find some explanations for our low success rate and in the hope to increase the number of reliable assignments. Firstly, we found that the majority of the EThcD precursors displayed low charge density: over 90% of the precursor ions were $m/z > 900$. Identification rate was $\sim 13\%$ for precursors $m/z < 900$; higher m/z precursors yielded identification rates of $\sim 2\%$. This suggests that ETD efficiency is a limiting factor and the collisional activation in this particular setting does not deliver sufficient information about the amino acid sequence. In many EThcD spectra, radical-based fragmentation was limited or non-existent; ions representing glycan fragmentation as a result of the supplemental HCD activation dominated. This was especially true for glycopeptides bearing larger glycan structures, or for multiply glycosylated peptides. Figure 1 depicts the EThcD spectrum of an unidentified O-glycopeptide. From the data, one can decipher that the peptide bears a hexasaccharide HexNAc₂Hex₂NeuAc₂, probably representing a disialylated core-2 O-glycan. However, the low number of potential sequence ions prevents the identification of the underlying amino acid sequence. Another

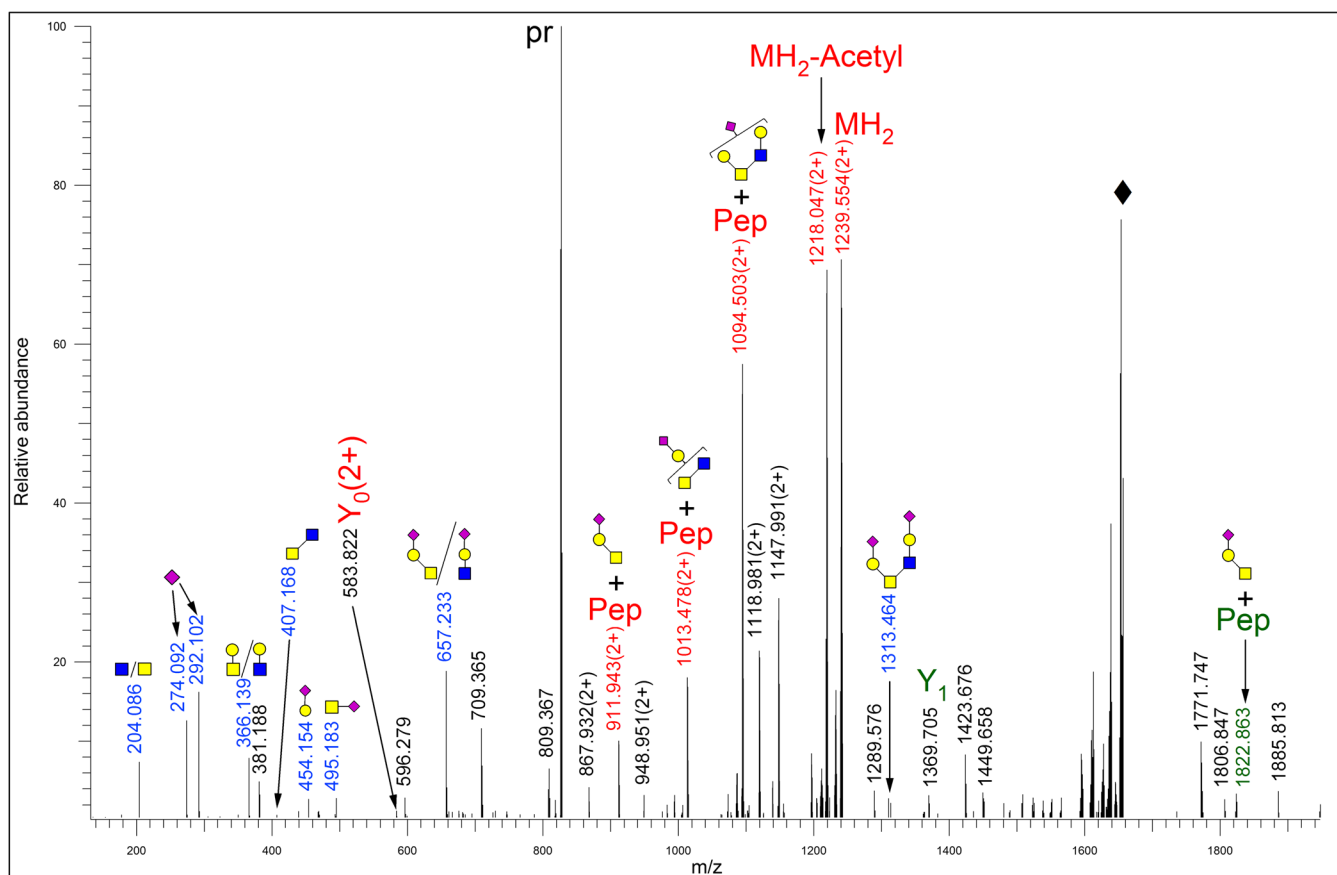


Figure 1. EThcD spectrum of an unidentified glycopeptide m/z 827.037(3+) bearing a disialo mucin-type core-2 structure featuring very informative glycan fragments, which help to assign the modifying sugar, but the spectrum does not contain sufficient data for peptide identification. Oxonium ions are in blue, while sugar losses from charge-reduced precursors are marked in red (2+) or green (1+). Y_0 represents the unmodified peptide, while Y_1 indicates the peptide modified with the core GalNAc only. The diamond symbol (◆) indicates the charge-reduced ion of a co-eluting doubly charged precursor, indicating multiple components are being co-fragmented. One glycan fragment, m/z 495, does not represent the normal sialylated core-2 structure. Its presence may indicate (i) precursor ion interference (as mentioned above); (ii) an unexpected structure; (iii) a co-eluting isomeric glycoform; or (iv) sialic acid migration, i.e., a rearrangement reaction. The peaklist of this spectrum is included as Online Resource 5

group, performing N-glycopeptide analysis, reported similar observations [34].

We also found that precursor ion interference is the rule rather than the exception when using high-sensitivity MS in such complex mixtures. Ion clusters corresponding to the charge-reduced forms of a co-eluting, different-charge state precursor ion was detected in the majority of the EThcD spectra (e.g., see Fig. 1). Automated data interpretation also indicated precursor ion interference: in spite of the m/z 204 HCD product ion trigger and the presence of the neuraminic acid-specific oxonium ion at m/z 292, unmodified peptides were confidently assigned from the 292-filtered EThcD peaklists (Online Resources 3 and 4; slides 30–32 of Online Resource 5). In addition, from the same peaklists, quite a large number of N-glycopeptides featuring neutral structures were identified in both fractions (Online Resource 4).



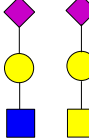
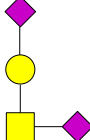
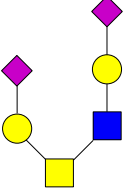

The abundance of glycan oxonium ions and the detection of even the intact hexasaccharide (Fig. 1) seemed to be unusual in comparison to HCD spectra. In order to confirm such differences and to assess the prevalence of ions representing larger

structures (Table 1), we performed a statistical analysis of the dataset using the MS-Filter tool in Protein Prospector [61]. Since it is likely that common O-glycans may yield diagnostic fragment ions, filtering for such fragments is an obvious path to investigate. In addition, ascertaining the presence of certain structures on multiply modified O-glycopeptides would be a great improvement.

The oxonium ions of HexNAc or sialic acid are usually abundant upon collisional activation, while the m/z 657 fragment representing a HexNAcHexNeuAc structure is much weaker. At the same time, our personal observation suggested that both m/z 292 and m/z 657 are more intense in EThcD. Thus, we tested how many spectra featured these masses within the 20 most abundant ions. Filtering with the sialic acid, oxonium ion delivered about the same number of matching spectra in both activation methods (Online Resource 2). However, the trisaccharide ion was present in about 50% less HCD spectra than EThcD spectra (Table 2).

The increased frequency of the trisaccharide (B_3) ion in EThcD spectra confirmed our hunch about the improved

Table 1. Structure and monoisotopic m/z values of the sugar oxonium ions used in the filtering process

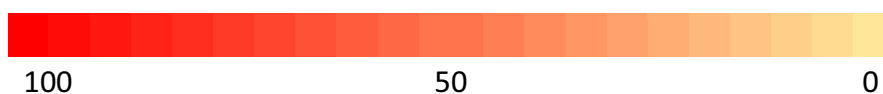
Oxonium ion	Structure	m/z	Glycan
$[\text{NeuAc}+\text{H}]^+$		292.1027	N, O
$[\text{HexNAc}_2+\text{H}]^+$		407.1661	O
$[\text{HexNAcHexNeuAc}+\text{H}]^+$		657.2349	N, O
$[\text{HexNAcHexNeuAc}_2+\text{H}]^+$		948.3303	O
$[\text{HexNAc}_2\text{Hex}_2\text{NeuAc}_2+\text{H}]^+$		1313.4625	O
			

survival rate of larger glycan fragments. The detection of intact O-glycans should improve the assignment of O-glycopeptides, since the differentiation between sequences modified by a hexasaccharide or by two smaller glycans (such as two trisaccharides or a di- and tetrasaccharide) often proves impossible. The disialylated core-1 O-glycan, GalNAc(NeuAc)GalNeuAc can produce a B ion at m/z 948, while the disialylated core-2 O-glycan GalNAc(GalNeuAc)GlcNAcGalNeuAc would yield a B ion at m/z 1313. These ions were detected in a subset of the spectra (Table 2), but their intensities were lower than that of the trisaccharide oxonium ion (albeit much higher compared to HCD). The frequency of m/z 948 seemed to plateau only when considering the top 80 most intense peaks. The B fragment of the intact hexasaccharide was the least common; its frequency did not reach 1% (obviously, this number reflects not only the fragility of the structure but also the lower occurrence of this glycoform). We searched for other core-2-specific candidate ions. A fragment at m/z 407 representing HexNAc₂ was detected in a reasonable number of spectra (Table 2). The disialylated core-2 hexasaccharide could also yield HexNAc₂ Hex, HexNAc₂Hex₂, HexNAc₂HexNeuAc, and HexNAc₂Hex₂NeuAc at m/z 569, 731, 860, and 1022, respectively. However, none of these were observed at a significant level.

Hoping that the characteristic oxonium ions would help to identify most tetra- and hexasaccharide-bearing glycoforms, database searches were performed with shorter, pre-screened peaklists. For tetrasaccharides, the presence of m/z 292 and the intact B ion at m/z 948 were required. To attempt to find hexasaccharide-bearing glycopeptides, we also required the detection of m/z 657 along with the diagnostic internal fragment (m/z 407) or the B ion representing the intact glycan (m/z 1313). This additional confirmation was needed, as at high masses, there is an increased chance of interference from peptide fragments, while at lower masses, precursor ion interference seems to be higher. Since O-glycopeptides frequently feature multiple glycosylations [71], probably also with different glycan structures, the smaller, potentially underlying glycans were also listed as variable modifications in these searches. The search with the m/z 407 pre-filtered list yielded glycopeptides with many kinds of glycans (data not shown), several of which cannot produce an m/z 407 ion, confirming that low-mass oxonium ions are frequently non-specifically observed due to precursor ion interference. Using the intact glycan masses as specific markers for the presence of glycans of such sugar compositions proved to be of limited benefit. More than 90% of the confident assignments featured the

Table 2. Heatmap of percentage of the MS/MS spectra featuring the specified fragment ions within 10 ppm among the specified number of most abundant ions. A, B, and C represent different samples, while 1 and 2 represent fraction 1 and 2 of affinity enrichment, respectively

Sample	657		948		1313		407	
	Top20		Top80		Top80		Top40	
	HCD	EThcD	HCD	EThcD	HCD	EThcD	HCD	EThcD
A1	44	87	0.41	23	0	0.16	3	2
B1	45	89	0.49	17	0	0.21	9	4
C1	53	91	2	36	0	0.59	17	11
A2	47	92	1.4	33	0	0.59	30	19
B2	50	95	1.4	30	0	0.30	35	17
C2	66	96	14	57	0	0.12	48	39



targeted glycan (Online Resources 3 and 5). However, the assignment rate was no higher compared to the results from the peaklists that were only filtered for m/z 292 (see the Summary sheet of Online Resource 3). Moreover, depending on the dataset and the search engine used, 33–50% of the PSMs linked to tetrasaccharide-modified sequences were lost when the pre-screened peaklists were used. This means that these spectra did not feature the diagnostic B fragment. However, we do not know how reliable these assignments are, despite the probability-based measures used by the search engines (see some examples, from the hexasaccharide dataset, Online Resource 5). Protein Prospector benefited more from using a filtered dataset: 83 and 28 novel PSMs were assigned to tetrasaccharide-modified glycopeptides from the pre-screened peaklists, when analyzing fractions 1 and 2, respectively (Online Resource 3). Byonic practically did not gain by using the 948-filtered peaklist: it added 3 and 11 novel PSMs to the assignments, from the fractions 1 and 2, respectively. This could be explained by the fact that Byonic also scores the glycan fragments. The inferior results for the stronger binding glycopeptide mixture (fractions 2) may be explained by the more extensive glycosylation.

Since the core-2 structure is probably the rarest O-glycan structure considered in the present study and thus, should represent the smallest spectrum pool, we illustrate the difficulties encountered in data interpretation with these data. Querying the peaklists filtered for the presence of m/z 292 allowed assignment of 110 individual PSMs to hexasaccharide-bearing O-glycopeptides in fraction 1 (Online Resource 3). A significant portion of these data indicated multiple glycosylations, including modification by two hexasaccharides. In numerous cases, such as when there were adjacent potential modification sites, there was not sufficient information to decide between the

different potential glycoforms (a single hexasaccharide as opposed to, for example, two trisaccharides). Additionally, since only two modifications per peptide were permitted, sequences reported modified with two hexasaccharides could actually be glycopeptides more extensively decorated with smaller glycans, an ambiguity produced by the analysis approach. From the peaklist filtered for m/z 1313 (95 spectra), Protein Prospector delivered 22 PSMs (23%) representing 16 unique sequences (see list in Online Resource 5). All but one of these was modified by the hexasaccharide, although for a few sequences the software could not make a confident site assignment because of adjacent potential modifications sites (for the best spectrum, see Fig. 2). From the same data, Byonic delivered 17 PSMs representing 12 unique glycopeptides. The overlapping with Prospector identifications was only 12 PSMs related to 7 unique glycopeptides (Online Resource 5). Both search engines identified an unmodified peptide from a mixture spectrum. Interestingly, for this data, the glycan scoring did not work to Byonic's advantage, as the search engine assigned doubly modified peptides and also listed a trisaccharide-modified sequence. The relevant EThcD spectra, their assignments, and also the corresponding HCD data are presented in Online Resource 5. These data illustrate that glycopeptide assignments may benefit from using the corresponding HCD data. In our data, some of the HCD spectra provide unambiguous peptide sequence assignment, some spectra do not offer any supporting information besides the accurate Y_0 mass, some indicate that multiple components were present, and some indicate that the EThcD assignment is incorrect.

Finally, we have also investigated the presence of the intact tetra- and hexasaccharide glycan oxonium ions in the EThcD spectra of confidently identified (score ≥ 200 , Delta Mod score ≥ 20 for Byonic identifications; score ≥ 15 , SLIP score ≥ 6 for

Protein Prospector identifications) glycopeptides. Only singly glycosylated peptides were considered in order to minimize interference from ambiguous site assignments that frequently also translates into ambiguous glycan structure identification. Fifty-nine percent of the tetrasaccharide-related spectra displayed m/z 948, while 13% of the hexasaccharide-related glycopeptide data contained m/z 1313 within the top 80 most abundant fragment ions. These findings indicate that larger glycan structures are unstable even under mild collisional activation.

In summary, the non-reducing end glycan fragments may aid database searches and may strengthen the reliability of glycopeptide assignments. Primarily, they improve the characterization of O-glycosylation where the same mass addition to a peptide sequence may correspond to a single larger or multiple smaller glycans. Such a contribution is probably less significant in N-glycosylation analysis, although it may help to identify antenna fucosylation or less common structural features. Thus, scoring glycan fragments is beneficial in the evaluation of EThcD spectra for determining glycosylation state, albeit that

complementary information on the peptide part is also necessary. Hence, glycopeptide data should be interpreted similarly to cross-linked peptides—only those identifications that contain sufficient evidence to identify both parts of the structure should be considered reliable. The Bionic search engine does predict and score B/Y ions. However, several EThcD spectra meeting the acceptance criteria displayed exclusively glycan fragmentation. On the other hand, Protein Prospector currently does not score glycan fragmentation, meaning both the score and the reliability of the glycopeptide identifications might be underestimated.

Reducing end Y fragments might contribute tremendously to the proper assignment of glycopeptides. Unfortunately, there is no current software implementation for their de novo recognition in O-glycosylation, especially when multiple, even different, glycans may be present. There is software where N-glycopeptide assignment heavily relies on the identification of Y_1 from ion trap CID [72] or HCD data [47]. Both software also incorporated glycan fragmentation scoring in order to decipher structures. According to two recent publications, N-

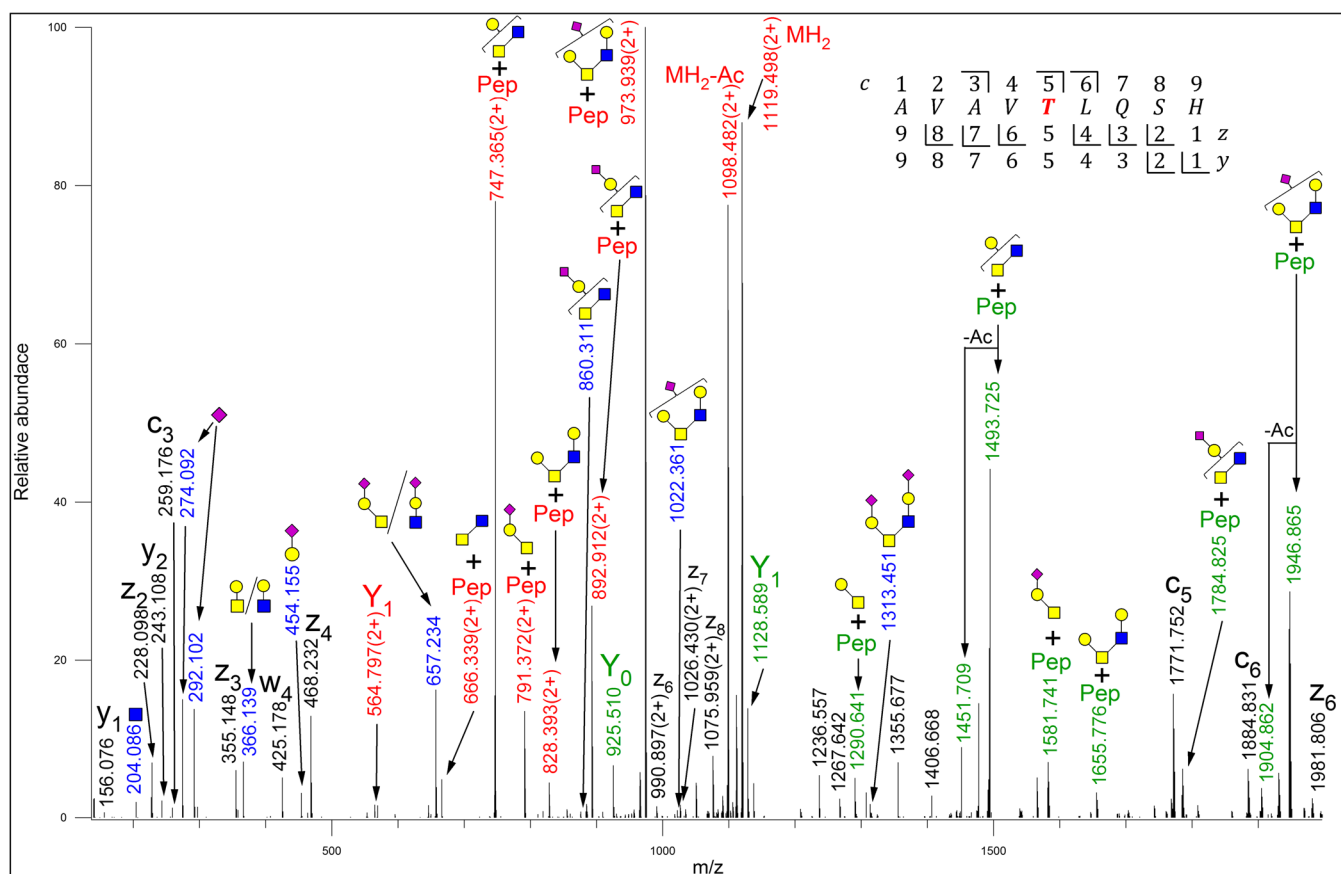


Figure 2. EThcD spectrum of m/z 746.660 (3+) identified as 342 AVAVTLQSH 350 from Protein YIPF3 (Q9GZM5) bearing a disialylated mucin-type core-2 glycan at Thr-5 (Thr-346 in the protein sequence). The oxonium ions (in blue) and reducing end, sugar-loss fragments (2+ in red; 1+ in green) are labeled with the appropriate CFG symbols, while the peptide fragments are assigned according to the nomenclature [31], except the radical z ions were originally called z+1 fragments. The m/z of the precursor ion is within 5 ppm of the calculated value, while all the fragments were measured within 10 ppm. “Pep” indicates the bare peptide backbone. Y_0 refers to the unmodified peptide, while Y_1 indicates the peptide modified with one N-acetyl-galactosamine. “-Ac” denotes the neutral loss of an acetyl group. The peaklist of this spectrum is included as Online Resource 7

glycopeptide glycans can even be sequenced de novo from HCD data [48, 49]. Unfortunately, the X; X + 203; X + 365 pattern that helps to identify the Y₁ fragment of N-glycopeptides may also represent Y₀, Y₁, and Y₂ in mucin-type O-glycopeptides, or partially gas-phase deglycosylated peptide fragments in these molecules. Thus, sophisticated informatic tools are required to utilize reducing end fragments even in the structure confirmation of the most promising candidates.

In comparison to ETciD, EThcD data provided additional information on the modifying glycans. On the other hand, the efficiency of the ETD fragmentation is still lower than that of collision-based fragmentation methods. In a recent study, the EThcD settings were altered for the more efficient characterization of N-glycopeptides [34]. The ETD activation time was shortened and the supplemental activation energy was increased. This improved the peptide fragmentation data (more *y* and *b* ions were detected) at the expense of glycan characterization. This approach may be counterproductive for O-glycosylation analysis, as the larger oxonium ions that allow differentiation between single and multiple glycosylations may not survive, and gas-phase elimination of the glycan(s) may also occur. Fragment ions generated by ETD activation are lower charge state than the precursor; therefore, following supplemental activation normally does not induce secondary fragmentation [28]. However, glycosidic bonds are highly labile upon collisional activation; therefore, we wanted to rule out the possibility that secondary fragmentation biases automated data interpretation. We have selected high-scoring EThcD spectra assigned to O-glycopeptides with only one potential modification site (to exclude the possibility that the spectrum is acquired on a mixture of co-eluting glycoforms) then checked if there are any unassigned fragment ions that might be interpreted as loss of sialic acid(s) (−291 and −582 Da) or the complete O-glycan (−947 Da for the tetra-, and −1312 Da for the hexasaccharide) from peptide sequence ions. We did not observe secondary fragmentation with the low supplemental activation (NCE 15%) applied in the present study (data not shown).

Conclusions

Glycopeptides represent two or more covalently linked biomolecules, and not only does each part have to be reliably identified, but also their linkage position has to be determined. This is especially problematic in O-glycosylation analysis, where multiple different oligosaccharides may be present on the peptide studied, and the mass addition can be translated into a series of different glycan combinations (although isomeric N-glycans also exist). This elevated complexity, compared to an unmodified peptide, demands the use of multiple activation methods to provide comprehensive analysis. Glycan fragmentation (observed in CID, HCD, and EThcD MS/MS spectra) should be scored, perhaps independently in the database searches, just like the cross-linked peptides are in some search engines [73].

Moreover, if software were available that reliably assessed accuracy of precursor charge state and monoisotopic mass determination from combined MS1 scans, and then combined information from HCD and EThcD data, i.e., oxonium ions, peptide fragment ions, and B and Y ions—as we currently do manually—then the rate of confident glycopeptide identification could be significantly higher.

We have observed higher *m/z* glycan oxonium ions in EThcD spectra of glycopeptides acquired with mild supplemental activation (NCE 15%) compared to HCD spectra employing a higher collision energy. These ions are specific for some glycan structures, so it could be used for confirming correct glycan identification in automated data interpretation. However, as these ions tend to be of low intensity, their presence cannot be considered as a pre-requisite for targeted LC-MS/MS data acquisition (such as the *m/z* 204 HCD product-ion-dependent ETD approach for general glycopeptide data acquisition [74]). On the other hand, the detection of the intact glycan oxonium ions such as *m/z* 948 for the disialylated core-1 O-glycan, or *m/z* 1313 for the disialylated core-2 O-glycan can be used as an orthogonal post-identification validator.

Acknowledgements

We appreciate the help of Robert Chalkley in editing our manuscript.

Funding Information

This work was supported by the following grants: Hungarian Scientific Research Fund 105611 (to Z. Darula) and the Economic Development and Innovation Operative Programmes GINOP-2.3.2-15-2016-00001 and GINOP-2.3.2-15-2016-00020 from the Ministry for National Economy.

Compliance with Ethical Standards

The study was approved by the regulatory and ethical authorities (ethics approval number of the Hungarian Scientific and Research Ethics Committee: 1011/16).

References

1. Albrecht, S., Hilliard, M., Rudd, P.: Therapeutic proteins: facing the challenges of glycobiology. *JHPOR*. **1**, 12–17 (2014)
2. Lasky, L.A., Singer, M.S., Dowbenko, D., Imai, Y., Henzel, W., Fennie, C., Watson, S., Rosen, S.D.: Glycosylation-dependent cell adhesion molecule 1: a novel mucin-like adhesion ligand for L-selectin. *Cold Spring Harb. Symp. Quant. Biol.* **57**, 259–269 (1992)
3. Pan, Y., Yago, T., Fu, J., Herzog, B., McDaniel, J.M., Mehta-D'Souza, P., Cai, X., Ruan, C., McEver, R.P., West, C., Dai, K., Chen, H., Xia, L.: Podoplanin requires sialylated O-glycans for stable expression on lymphatic endothelial cells and for interaction with platelets. *Blood*. **124**, 3656–3665 (2014)
4. Kariya, Y., Kanno, M., Matsumoto-Morita, K., Konno, M., Yamaguchi, Y., Hashimoto, Y.: Osteopontin O-glycosylation contributes to its phosphorylation and cell-adhesion properties. *Biochem. J.* **463**, 93–102 (2014)
5. Leuenberger, B., Hahn, D., Pischitzis, A., Hansen, M.K., Sterchi, E.E.: Human meprin beta: O-linked glycans in the intervening region of the type I membrane protein protect the C-terminal region from proteolytic cleavage and diminish its secretion. *Biochem. J.* **369**, 659–665 (2003)

6. Peng, J., Jiang, J., Wang, W., Qi, X., Sun, X.L., Wu, Q.: Glycosylation and processing of pro-B-type natriuretic peptide in cardiomyocytes. *Biochem. Biophys. Res. Commun.* **411**, 593–598 (2011)
7. Goth, C.K., Halim, A., Khetarpal, S.A., Rader, D.J., Clausen, H., Schjoldager, K.T.: A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14623–14628 (2015)
8. Goth, C.K., Tuhkanen, H.E., Khan, H., Lackman, J.J., Wang, S., Narimatsu, Y., Hansen, L.H., Overall, C.M., Clausen, H., Schjoldager, K.T., Petäjä-Repo, U.E.: Site-specific O-glycosylation by polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-transferase T2) co-regulates β 1-adrenergic receptor N-terminal cleavage. *J. Biol. Chem.* **292**, 4714–4726 (2017)
9. Ungar, D.: Golgi linked protein glycosylation and associated diseases. *Semin. Cell Dev. Biol.* **20**, 762–769 (2009)
10. An, H.J., Kronewitter, S.R., de Leoz, M.L., Lebrilla, C.B.: Glycomics and disease markers. *Curr. Opin. Chem. Biol.* **13**, 601–607 (2009)
11. Wada, Y.: Mass spectrometry in the detection and diagnosis of congenital disorders of glycosylation. *Eur J Mass Spectrom (Chichester)*. **13**, 101–103 (2007)
12. Fu, H.L., Valiathan, R.R., Payne, L., Kumarasiri, M., Mahasenan, K.V., Mobashery, S., Huang, P., Fridman, R.: Glycosylation at Asn211 regulates the activation state of the discoidin domain receptor 1 (DDR1). *J. Biol. Chem.* **289**, 9275–9287 (2014)
13. Soto, A.G., Trejo, J.: N-linked glycosylation of protease-activated receptor-1 second extracellular loop: a critical determinant for ligand-induced receptor activation and internalization. *J. Biol. Chem.* **285**, 18781–18793 (2010)
14. Tagliabracchi, V.S., Engel, J.L., Wiley, S.E., Xiao, J., Gonzalez, D.J., Nidumanda Appaiah, H., Koller, A., Nizet, V., White, K.E., Dixon, J.E.: Dynamic regulation of FGF23 by Fam20C phosphorylation, GalNAc-T3 glycosylation, and furin proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5520–5525 (2014)
15. Carr, S.A., Roberts, G.D.: Carbohydrate mapping by mass spectrometry: a novel method for identifying attachment sites of Asn-linked sugars in glycoproteins. *Anal. Biochem.* **157**, 396–406 (1986)
16. Settineri, C.A., Medzihradzky, K.F., Masiarz, F.R., Burlingame, A.L., Chu, C., George-Nascimento, C.: Characterization of O-glycosylation sites in recombinant B-chain of platelet-derived growth factor expressed in yeast using liquid secondary ion mass spectrometry, tandem mass spectrometry and Edman sequence analysis. *Biomed Environ Mass Spectrom.* **19**, 665–676 (1990)
17. Hemling, M.E., Roberts, G.D., Johnson, W., Carr, S.A., Covey, T.R.: Analysis of proteins and glycoproteins at the picomole level by on-line coupling of microbore high-performance liquid chromatography with flow fast atom bombardment and electrospray mass spectrometry: a comparative evaluation. *Biomed. Environ. Mass Spectrom.* **19**, 677–691 (1990)
18. Barr, J.R., Anumula, K.R., Vettese, M.B., Taylor, P.B., Carr, S.A.: Structural classification of carbohydrates in glycoproteins by mass spectrometry and high-performance anion-exchange chromatography. *Anal. Biochem.* **192**, 181–192 (1991)
19. Buko, A.M., Kentzer, E.J., Petros, A., Menon, G., Zuiderweg, E.R., Sarin, V.K.: Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3992–3996 (1991)
20. Carr, S.A., Huddleston, M.J., Bean, M.F.: Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci.* **2**, 183–196 (1993)
21. Huddleston, M.J., Bean, M.F., Carr, S.A.: Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **65**, 877–884 (1993)
22. Medzihradzky, K.F., Maltby, D.A., Hall, S.C., Settineri, C.A., Burlingame, A.L.: Characterization of protein N-glycosylation by reversed-phase microbore liquid chromatography / electrospray mass spectrometry, complementary mobile phases, and sequential exoglycosidase digestion. *J. Am. Soc. Mass Spectrom.* **5**, 350–358 (1994)
23. Medzihradzky, K.F., Besman, M.J., Burlingame, A.L.: Structural characterization of site-specific N-glycosylation of recombinant human factor VIII by reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **69**, 3986–3994 (1997)
24. Medzihradzky, K.F., Gillece-Castro, B.L., Hardy, M.R., Townsend, R.R., Burlingame, A.L.: Structural elucidation of O-linked glycopeptides by high energy collision-induced dissociation. *J. Am. Soc. Mass Spectrom.* **7**, 319–328 (1996)
25. Zubarev, R.A., Hom, D.M., Fridriksson, E.K., Kelleher, N.L., Kruger, N.A., Lewis, M.A., Carpenter, B.K., McLafferty, F.W.: Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72**, 563–573 (2000)
26. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F.: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528–9533 (2004)
27. Darula, Z., Medzihradzky, K.F.: Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell. Proteomics.* **8**, 2515–2526 (2009)
28. Frese, C.K., Altmeyer, A.F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J., Mohammed, S.: Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668–9673 (2012)
29. Frese, C.K., Zhou, H., Taus, T., Altmeyer, A.F., Mechtler, K., Heck, A.J., Mohammed, S.: Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **12**, 1520–1525 (2013)
30. Zhokhov, S.S., Kovalyov, S.V., Samgina, T.Y., Lebedev, A.T.: An EThcD-based method for discrimination of leucine and isoleucine residues in tryptic peptides. *J. Am. Soc. Mass Spectrom.* **28**, 1600–1611 (2017)
31. Biemann, K.: Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* **193**, 886–887 (1990)
32. Parker, B.L., Thaysen-Andersen, M., Fazakerley, D.J., Holliday, M., Packer, N.H., James, D.E.: Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes. *Mol. Cell. Proteomics.* **15**, 141–153 (2016)
33. Glover, M. S., Yu, Q., Chen, Z., Shi, X., Kent, K. C., Li L. Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass. Spectrom.* **427**, 35–42 (2018)
34. Yu, Q., Wang, B., Chen, Z., Urabe, G., Glover, M.S., Shi, X., Guo, L.W., Kent, K.C., Li, L.: Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide / glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **28**, 1751–1764 (2017)
35. Zhang, Y., Xie, X., Zhao, X., Tian, F., Lv, J., Ying, W., Qian, X.: Systems analysis of singly and multiply O-glycosylated peptides in the human serum glycoproteome via EThcD and HCD mass spectrometry. *J. Proteomics* **170**, 14–27 (2018)
36. Doman, B., Costello, C.E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **5**, 397–409 (1988)
37. Wu, S.W., Pu, T.H., Viner, R., Khoo, K.H.: Novel LC-MS² product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **86**, 5478–5486 (2014)
38. Medzihradzky, K.F.: Characterization of protein N-glycosylation. *Methods Enzymol.* **405**, 116–138 (2005)
39. Peter-Katalinic, J.: Methods in enzymology: O-glycosylation of proteins. *Methods Enzymol.* **405**, 139–171 (2005)
40. Darula, Z., Chalkley, R.J., Lynn, A., Baker, P.R., Medzihradzky, K.F.: Improved identification of O-linked glycopeptides from ETD data with optimized scoring for different charge states and cleavage specificities. *Amino Acids.* **41**, 321–328 (2011)
41. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., Nilsson, J.: Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* **13**, 6024–6032 (2014)
42. Trinidad, J.C., Schoepfer, R., Burlingame, A.L., Medzihradzky, K.F.: N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics.* **12**, 3474–3488 (2013)
43. Steentoft, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., Schjoldager, K.T., Lavrsen, K., Dabelsteen, S., Pedersen, N.B., Marcos-Silva, L., Gupta, R., Bennett, E.P., Mandel, U., Brunak, S., Wandall, H.H., Levery, S.B., Clausen, H.: Precision mapping

- of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013)
44. Vakhrushev, S.Y., Steentoft, C., Vester-Christensen, M.B., Bennett, E.P., Clausen, H., Lavery, S.B.: Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol. Cell. Proteomics.* **12**, 932–944 (2013)
 45. Zauner, G., Kozak, R.P., Gardner, R.A., Fernandes, D.L., Deelder, A.M., Wührer, M.: Protein O-glycosylation analysis. *Biol. Chem.* **393**, 687–708 (2012)
 46. Darula, Z., Medzihradzky, K. F., Analysis of mammalian O-glycopeptides—we have made a good start, but there is a long way to go. *Mol. Cell. Proteomics.* **17**, 2–17 (2018)
 47. He, L., Xin, L., Shan, B., Lajoie, G. A., Ma, B., GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res.* **13**, 3881–3895 (2014)
 48. Sun, W., Kuljanin, M., Pittock, P., Ma, B., Zhang, K., Lajoie, G. A., An effective approach for glycan structure de novo sequencing from HCD spectra. *IEEE Trans Nanobioscience.* **15**, 177–184 (2016)
 49. Sun, W., Liu, Y., Lajoie, G., Ma, B., Zhang, K., An improved approach for N-linked glycan structure identification from HCD MS/MS spectra. *IEEE/ACM Trans Comput Biol Bioinform.* <https://doi.org/10.1109/TCBB.2017.2701819> (2017) [Epub ahead of print]
 50. Mayampurath, A., Yu, C.Y., Song, E., Balan, J., Mechref, Y., Tang, H.: Computational framework for identification of intact glycopeptides in complex samples. *Anal. Chem.* **86**, 453–463 (2014)
 51. Halim, A., Nilsson, J., Rüetschi, U., Hesse, C., Larson, G.: Human urinary glycoproteomics; attachment site-specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell. Proteomics.* <https://doi.org/10.1074/mcp.M111.013649> (2012)
 52. Halim, A., Rüetschi, U., Larson, G., Nilsson, J.: LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **12**, 573–584 (2013)
 53. Hoffmann, M., Marx, K., Reichl, U., Wührer, M., Rapp, E.: Site-specific O-glycosylation analysis of human blood plasma proteins. *Mol. Cell. Proteomics.* **15**, 624–641 (2016)
 54. Thaysen-Andersen, M., Venkatakrishnan, V., Loke, I., Laurini, C., Diestel, S., Parker, B.L., Packer, N.H.: Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *J. Biol. Chem.* **290**, 8789–87802 (2015)
 55. Cheng, K., Chen, R., Seebun, D., Ye, M., Figeys, D., Zou, H.: Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J. Proteomics* **110**, 145–154 (2014)
 56. Parker, B.L., Thaysen-Andersen, M., Solis, N., Scott, N.E., Larsen, M.R., Graham, M.E., Packer, N.H., Cordwell, S.J.: Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity. *J. Proteome Res.* **12**, 5791–5800 (2013)
 57. Scott, N.E., Marzook, N.B., Cain, J.A., Solis, N., Thaysen-Andersen, M., Djordjevic, S.P., Packer, N.H., Larsen, M.R., Cordwell, S.J.: Comparative proteomics and glycoproteomics reveal increased N-linked glycosylation and relaxed sequon specificity in *Campylobacter jejuni* NCTC11168 O. *J. Proteome Res.* **13**, 5136–5150 (2014)
 58. Liu, M., Zhang, Y., Chen, Y., Yan, G., Shen, C., Cao, J., Zhou, X., Liu, X., Zhang, L., Shen, H., Lu, H., He, F., Yang, P.: Efficient and accurate glycopeptide identification pipeline for high-throughput site-specific N-glycosylation analysis. *J. Proteome Res.* **13**, 3121–3129 (2014)
 59. Kawahara, R., Saad, J., Angeli, C.B., Palmisano, G.: Site-specific characterization of N-linked glycosylation in human urinary glycoproteins and endogenous glycopeptides. *Glycoconj. J.* **33**, 937–951 (2016)
 60. Snovida, S.I., Bodnar, E.D., Viner, R., Saba, J., Perreault, H.: A simple cellulose column procedure for selective enrichment of glycopeptides and characterization by nano LC coupled with electron-transfer and high-energy collisional-dissociation tandem mass spectrometry. *Carbohydr. Res.* **345**, 792–801 (2010)
 61. Medzihradzky, K.F., Kaasik, K., Chalkley, R.J.: Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell. Proteomics.* **14**, 2103–2110 (2015)
 62. Xu, S.L., Medzihradzky, K.F., Wang, Z.Y., Burlingame, A.L., Chalkley, R.J.: N-Glycopeptide profiling in arabidopsis inflorescence. *Mol. Cell. Proteomics.* **15**, 2048–2054 (2016)
 63. King, S.L., Joshi, H.J., Schjoldager, K.T., Halim, A., Madsen, T.D., Dziegiel, M.H., Woetmann, Vakhrushev, S.Y., Wandall, H.H.: Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Advances.* **1**, 429–442 (2017)
 64. Chang, C.F., Pan, J.F., Lin, C.N., Wu, I.L., Wong, C.H., Lin, C.H.: Rapid characterization of sugar-binding specificity by in-solution proximity binding with photosensitizers. *Glycobiology.* **21**, 895–902 (2011)
 65. Bern, M., Kil, Y. J., Becker, C.: Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics.* <https://doi.org/10.1002/0471250953.bi1320s40> (2012)
 66. Darula, Z., Samyai, F., Medzihradzky, K.F.: O-glycosylation sites identified from mucin core-1 type glycopeptides from human serum. *Glycoconj. J.* **33**, 435–445 (2016)
 67. Medzihradzky, K.F., Kaasik, K., Chalkley, R.J.: Characterizing sialic acid variants at the glycopeptide level. *Anal. Chem.* **87**, 3064–3071 (2015)
 68. Baker, P.R., Trinidad, J.C., Chalkley, R.J.: Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics.* (2011). <https://doi.org/10.1074/mcp.M111.008078>
 69. Bern, M. W., Kil, Y. J. Two-dimensional target decoy strategy for shotgun proteomics. *Journal of Proteome Research.* <https://doi.org/10.1021/pr200780j> (2011)
 70. Yabu, M., Korekane, H., Miyamoto, Y.: Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology.* **24**, 542–553 (2014)
 71. Pap, A., Medzihradzky, K.F., Darula, Z.: Using “spectral families” to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited. *Anal. Bioanal. Chem.* **409**, 539–550 (2017)
 72. Wu, S.W., Liang, S.Y., Pu, T.H., Chang, F.Y., Khoo, K.H.: Sweet-Heart—an integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *J. Proteomics* **84**, 1–16 (2013)
 73. Tmka, M.J., Baker, P.R., Robinson, P.J., Burlingame, A.L., Chalkley, R.J.: Matching cross-linked peptide spectra: only as good as the worse identification. *Mol. Cell. Proteomics.* **13**, 420–434 (2014)
 74. Saba, J., Dutta, S., Hemenway, E., Viner, R.: Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int. J. Proteomics* 560391 (2012). <https://doi.org/10.1155/2012/560391>. (2012)