HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2016 March 03.

Published in final edited form as:

Anal Chem. 2015 March 3; 87(5): 3064–3071. doi:10.1021/ac504725r.

Characterizing sialic acid variants at the glycopeptide level

Katalin F. Medzihradszky[*], **Krista Kaasik**, and **Robert J. Chalkley**

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, 600 16th Street Genentech Hall, N474A, Box 2240, San Francisco, California 94158-2517, United States

Abstract

Beam-type CID data of intact glycopeptides isolated from mouse liver tissue are presented to illustrate characteristic fragmentation of differentially sialylated glycopeptides. Eight glycoforms of an O-linked glycopeptide from Nucleobindin-1 are distinguished based on the precursor masses and characteristic oxonium ions. We report that all sialic acid variants are prone to neutral loss from the charge reduced species in ETD fragmentation. We show how changes in sialic acid composition affect reverse phase chromatographic retention times: sialic acid addition increases glycopeptide retention times significantly; replacing the N-acetylneuraminic acid with the N-glycolyl variant leads to slightly reduced retention times, while O-acetylated sialic acid-containing glycoforms are retained longer. We then demonstrate how MS-Filter in Protein Prospector can use these diagnostic oxonium ions to find glycopeptides, by showing that a wealth of different glycopeptides can be found in a published phosphopeptide dataset.

Keywords

glycosylation; collision-induced dissociation; glycopeptide; O-acetyl sialic acid; electron-transfer dissociation; NeuAc; NeuGc

Introduction

Glycosylation is one of the most frequent post-translational modifications, being found on essentially all extracellular and secreted proteins [1]. There is a growing understanding of its biological significance: the oligosaccharide structures control the cell adhesion processes [2]; affect cell to cell recognition, including pathogen identification by host cells [3]; direct enzymatic processing [4]; and influence intracellular signaling processes [5]. Understanding glycosylation is of major importance to the pharmaceutical industry, as most protein therapeutics are glycoproteins and the glycosylation affects the protein's activity, specificity, stability and antigenicity [6].

This growing recognition of glycosylation's importance has led to increased interest from researchers and funding agencies to develop new and improved techniques for glycoprotein

[*]Corresponding author Katalin F. Medzihradszky Department of Pharmaceutical Chemistry University of California, San Francisco 600 16th Street Genentech Hall, N474A, Box 2240 San Francisco, CA 94158-2517 Phone: 415-476-5160 Fax: 415-502-1655 folkl@cgl.ucsf.edu.

analysis. However, intact glycopeptide analysis is difficult for multiple reasons. For example, site-specific heterogeneity means a complex set of glycopeptides, at a range of different stoichiometries, will be present after protein digestion, even when analyzing a single protein. Also, mass spectrometry is not effective at characterizing the linkages between different sugar units in the oligosaccharide chain. Many glycomic researchers, particularly those interested in N-linked glycosylation, typically enzymatically release the glycans from the protein, and then analyze the free oligosaccharides. This approach is still the best for in-depth oligosaccharide characterization, but loses all information about the heterogeneity of glycans at a given modification site in a protein (or indeed which protein bears the modification, if there are multiple in the sample under analysis). This type of information is retained in glycopeptide analysis. Both N- and O-linked glycans are frequently terminated (capped) with sialic acids. Sialic acids fulfill a dual role in molecular and cellular interactions: they may mask sugar epitopes, or they themselves may serve as recognition motifs [7]. Sialic acid recognition motifs also can be masked by O-acetylation (esterification in position-9 is the most common). This can have significant biological consequences; for example, masking capping sialic acids has been implicated in “drug-resistant” leukemia cases [8]. N-acetylneuraminic acid (NeuAc) is the most common form of sialic acid in nature, and is found on the surface of all vertebrate cells. Most vertebrates, with the notable exception of humans, regularly utilize another sialic acid in the form of N-glycolylneuraminic acid (NeuGc). NeuGc also may be O-acetylated. The occurrence of NeuGc esterified at positions 7 or 9 has been described for rat liver glycoproteins [9]. Humans do not produce NeuGc endogenously, but can incorporate it from proteins present in their diet. Humans often mount an immune response against NeuGc, and it has been speculated that loss of endogenous production of this sugar was an adaptation against pathogen invasion[10]. However, the incorporation of NeuGc into human glycans was observed a long time ago [11], and this phenomenon has also been linked to cancer [12]. Thus, protein- and site-specific characterization of sialic acids must be an important part of glycosylation studies.

Mass spectrometry has been the most significant tool for glycosylation analysis [13, 14]. The classical approach of analyzing peptides using collision-induced dissociation tandem mass spectrometry extensively fragments glycans attached to the peptide, but provides minimal information about the amino acid sequence [15, 16]. However, more recently developed radical-based fragmentation approaches, most notably electron-transfer dissociation (ETD), produce peptide backbone fragmentation with intact oligosaccharides attached, so provide complementary information to the CID data [15, 16]. Hence, these analytical methods combined can be used to characterize both peptide and glycan components of glycopeptides.

The formation of glycan-specific reporter ions in the low mass region of CID MS/MS spectra is well documented as a method for recognition of glycopeptide precursors [17], and diagnostic glycan fragment-triggered ETD acquisition has also been introduced recently [18]. Oxonium ions, including NeuAc fragments have been used for decades to decipher peptide-modifying glycan structures [17, 19]. Characteristic oxonium ions for NeuGc also have been described recently [20, 21]. However, only limited use has been made of these ions for glycopeptide assignment from ETD data [22]. If complementary CID and ETD data

are acquired on the same precursor ion, diagnostic oxonium ions formed upon collisional activation could be used to aid ETD spectral interpretation, by eliminating non-glycopeptide spectra or confirming the presence of a particular sugar unit or the combination thereof; steps that would speed up database searches and make the resulting glycopeptide assignments more reliable.

In this work we demonstrate the production of diagnostic oxonium reporter ions in beam-type CID spectra of glycopeptides containing different sialic acid residues. To illustrate this we present data where four different sialic acids were detected in eight different glycoforms of the same O-glycosylation site of the Nucleobindin-1 protein from mouse liver tissue. We demonstrate that the different sialic acids also feature a diagnostic fragment upon ETD activation. We also show an example with these sugar residues detected in different N-linked glycoforms. We present the effect that these different sialylations have on glycopeptide retention in reverse phase chromatography when a weak acid, such as formic acid, is the ion-pairing agent. Finally, we demonstrate that interrogating MS/MS data for these diagnostic fragment ions can reveal the presence of glycopeptides in some `unexpected' places: published raw data from a HeLa cell phosphopeptide-enrichment [23] were scrutinized, and not only a high level of glycopeptide-enrichment was detected, but also the presence of O-acetyl NeuAc and NeuGc. We hope the information supplied here will be useful to researchers interested in the growing field of intact glycopeptide analysis and also that this knowledge eventually could be incorporated into software aimed at the better characterization of glycopeptides.

Methods

Mouse Liver Sample Preparation

All studies were approved by the Institutional Animal Care and Use Committee at UCSF. Liver tissue was dissected from 10 days old mice. Three livers were combined and homogenized by a Dounce homogenizer. The lysis buffer consisted of 10 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES-KOH; pH7.9), 1.5 mM MgCl2, 10 mM KCl with O-GlcNAcase inhibitor PUGNAc (Sigma, St Louis, MI, 50uM), protease inhibitors (Roche, South San Francisco, CA) and phosphatase inhibitor cocktails 1 and 2 (Sigma) added. A crude nuclear preparation was prepared following a previously published two-step differential solubilization and centrifugation protocol [24].

15mg of protein was resuspended in 1 ml buffer containing 50mM ammonium bicarbonate and 6M guanidine hydrochloride. The mixture was incubated at 57 °C with 2mM Tris(2-carboxyethyl)phosphine hydrochloride to reduce cysteine side chains, these side chains were then alkylated with 4.2mM iodoacetamide in the dark for 45 min at room temperature. The mixture was diluted to 0.8M guanidine hydrochloride with 50mM ammonium bicarbonate. Modified porcine trypsin (Promega) was added to 1:20 w/w final concentration. The pH was adjusted to 8.0 and the mixture was digested for 16h at 37 °C. The digests were desalted using a C18 Sep Pak cartridge (Waters) and lyophilized to dryness using a SpeedVac concentrator (Thermo Electron).

Lectin Weak Affinity Chromatography (LWAC)

The LWAC column, consisting of wheat germ agglutinin (Vector Labs, Burlingame, CA) bound to POROS beads (Life Technologies, Grand Island, NY), was prepared as described previously [25] and chromatography was performed using an ÄKTA Purifier (GE Healthcare).

Peptides were resuspended in 100 µl LWAC buffer (100mM Tris-HCl pH 7.7, 150mM NaCl, 2mM MgCl2, 2mM CaCl2, 5% acetonitrile). Sample was split into five aliquots and each sequentially run over the poros-WGA column. The glycopepeptide-enriched fractions from each were collected, pooled, desalted and lyophilized by vacuum centrifugation. This combined fraction was then loaded for a second LWAC separation, and the enriched fraction from this run was loaded for a third round of LWAC purification. Enriched glycopeptides were then separated by high pH reverse phase chromatography using a 1 × 100 mm Gemini 3µ C18 column (Phenomenex, Torrance, CA) as described previously [25].

Mass Spectrometry

LWAC-enriched peptides were analyzed by LC-MS/MS using a NanoAcquity (Waters, Milford, MA) and a LTQ-Orbitrap Velos (Thermo, San Jose, CA). Chromatography was performed using an EASY-Spray Nano-LC source with a 15cm × 75µm ID column packed with 3µm C18 particles. Solvent A was 0.1 % formic acid, 5 % DMSO in water; Solvent B was 0.1% formic acid, 5 % DMSO in acetonitrile. A ninety minute gradient was employed that included 20 minutes at 2% B to load sample, then went from 2 to 18 % solvent B over 55 minutes, then up to 50 % B over a further 4 minutes, before returning to 2 %B. MS scans were acquired from m/z 350–1600, at a resolution of 30,000. MS/MS data was acquired in a data-dependent mode where the top 3 most intense multiply charged precursors were selected for sequential HCD (acquired at a resolution of 7,500) and ETD fragmentation (measured in the ion trap).

MS/MS Data Analysis

For mouse liver data, raw files were converted into peak list files using in-house software (PAVA) based on the Raw_Extract script from Xcalibur v2.4 (Thermo Scientific), which creates separate peak lists for HCD and ETD data [26]. Each fragmentation type was then searched separately using Batch-Tag in Protein Prospector [27]. An iterative searching strategy using mass modification searching was employed to identify glycopeptides, similar to as previously published [28]. Briefly, data was initially searched considering mass modifications within a mass range of 200–3000 on N, S or T residues. The results from this initial search were used to create an accession number list, and a second search was performed only with these proteins permitted, which allowed detection of more glycoforms of glycopeptides already identified.

For HeLa cell data, six raw files (20120126_EXQ5_KiSh_SA_LabelFree_HeLa_Phospho_Control_rep1_Fr1 through Fr6) were downloaded from the PRIDE repository via the ProteomeXchange submission PXD000612 [23]. Peak lists were created using PAVA [26]. The six peak list files were compressed into a single .zip file, then submitted to MS-Filter in Protein Prospector.

Initially, peak lists were filtered based on the presence of a peak at m/z 204.087 ± 15 ppm in the sixty most intense peaks in the spectrum. A peptide fragment at m/z 204.1343, a y_2 ion derived from Gly-Lys C-termini could be misidentified as a HexNAc oxonium ion in low resolution, low mass accuracy measurements. However, the oxonium ion and this fragment differ by 232 ppm, and even if both were present in a spectrum they would be resolved at the resolution of the acquired HCD data. The resulting peak lists were then analyzed/filtered for the presence of m/z 274.092 ± 15 ppm (for NeuAc) or m/z 290.087 ± 15ppm (for NeuGc).

Results

Glycosylation in mouse liver

All peptides enriched after lectin weak affinity chromatography were analyzed sequentially by beam-type CID (referred to as HCD on Thermo instruments) and ETD. Hundreds of unique glycopeptides were identified using Protein Prospector, but this article will focus only on results from two of the more heterogeneous glycosylation sites to show diagnostic fragments produced by CID as a result of the presence of different sialic acid residues in different locations in the glycan structure; and the resulting effect these different sialylation states have on glycopeptide retention by reverse phase chromatography. We will also illustrate sialic acid fragmentation in ETD MS.

Eight different glycoforms of the peptide ^{31}AAPPQEDSQAgTETPDTGLYYHR^{52} of Nucleobindin-1 were identified (Thr-41, labeled with `g' was identified as the glycosylation site). These glycoforms are listed in Table 1.

The low mass region of the CID spectra of these eight glycopeptides are presented in Figure 1. These eight glycoforms feature differently sialylated versions of a mucin core 1 type O-linked glycan structure, i.e. GalNAc-Gal, which produces a disaccharide fragment ion at m/z 366.1. All of these spectra also contained an ion at m/z 204.087, corresponding to the GalNAc oxonium ion, and the relative intensity ratios of other low mass fragments make this assignment unambiguous [30], but the intention of Figure 1 is to draw attention to the diagnostic differences between spectra, so only the mass range from m/z 270 to m/z 600 is displayed.

The spectrum in the first panel contains fragment ions at m/z 292.1 and m/z 274.1, which indicate the presence of a NeuAc (the latter ion is water loss from the former). The second spectrum instead contains ions at m/z 334.1 and m/z 316.1, which are 42 Da higher in mass, indicating O-acetylation on the NeuAc (probably in position-9, i.e. Neu5,9Ac2). The precursor of the spectrum that is panel three was 291 Da higher in mass than that in panel one (see Table 1), indicating the presence of a second NeuAc. An additional fragment ion observed in this spectrum is at m/z 495.2. This corresponds to a HexNAcNeuAc disaccharide oxonium ion, showing the presence of NeuAc attached to the GalNAc. Panel four contains m/z 292.1 and m/z 274.1, showing the presence of a NeuAc, but there are also new fragment ions at m/z 308.1, m/z 290.1 and m/z 511.2. The first two of these are diagnostic for the presence of a NeuGc. The peak at m/z 511.2 corresponds to GalNAcNeuGc, so this fragment ion determines the locations of the NeuGc, and by inference the NeuAc, on the glycan structure. The spectrum in panel five contains diagnostic

ions for the presence of NeuAc and previously unreported ions for acetylated NeuAc. The ion at m/z 537.2, 42 Da higher in mass than the m/z 495.2 ion observed in panel three, determines that the Neu5Ac,xAc ($NeuAc_2$, since the site of O-acetylation cannot be determined from these data) is attached to the GalNAc. Low mass ions in panel six show this glycan contained NeuGc and acetylated NeuAc. The ion at m/z 511.2 establishes that the GalNAc has a NeuGc attached to it, meaning the $NeuAc_2$ is attached to the Gal. The low mass region of panel seven contains ions indicating the presence of NeuAc, $Neu(Ac)_2$, NeuGc and NeuGc,Ac. There are also ions at m/z 511.2 and 553.2, indicating that glycoforms with both NeuGc or NeuGc,Ac attached to the GalNAc are present. This spectrum is of a mixture of two precursors: that bearing a glycan with NeuGc,Ac and NeuAc as illustrated in the panel, and the other glycan is that presented in panel six, where the NeuAc is O-acetylated instead of the NeuGc. The final panel (eighth) is of the glycopeptide containing NeuGc and NeuAc both O-acetylated. The peak at m/z 553.2 indicates the NeuGc,Ac is attached to the GalNAc.

Table 2 lists all of the diagnostic oxonium mono- and disaccharide ions described here for sugar composition and `linkage' analysis. For all versions of sialic acid diagnostic ions were observed, and in each case the water loss peak was more prominent than the native oxonium ion, although this ratio may be influenced by CID collision energy (this data was probably acquired with slightly higher collision energy than optimal for only monitoring glycan oxonium ions, as the hope was to also produce some peptide fragments). It should be noted that the data here is neither able to determine the linkage types between sugars, nor is able to tell sites of acetylation on the sialic acid residues.

The HCD spectra produced sufficient information to assign all of the glycan compositions, which sialic acids were O-acetylated, and to determine which sialic acid was linked to the core GalNAc (Figure 1). However, the identification of the peptide sequence and the site of O-glycosylation were both provided by ETD data. In ETD fragmentation the conventional wisdom is that labile modifications are not lost, so only modified fragment ions are expected. However, in a previous study we already demonstrated that NeuAc is prone to fragmentation under ETD conditions [31]. The same fragmentation was observed for all sialic acid variants detected in this study. To illustrate this, the ETD spectrum of the fifth glycoform in Table 1 and Figure 1, which features both NeuAc and NeuGc is presented in Figure 2. In the ETD spectrum the most intense fragment ions correspond to losses of each individual sialic acid and the loss of both from the charge-reduced precursor ion. Such fragmentation is typical for sialic acid containing glycopeptides and these neutral losses may occur even without supplemental activation. However, it should be noted that no losses of sugar moieties from any peptide backbone cleavage products (c or $z^{\bullet}$ ions) are usually observed.

It has previously been described that glycan structures containing different compositions of neutral sugar residues co-elute in reverse phase chromatography, but that the presence of charged residues such as sialic acids alter the chromatographic retention [32]. Hence, we investigated the chromatographic effect the different sialic acid compositions present on the Nucleobindin-1 glycopeptide described in Figure 1 and Table 1 had on their retention times.

Figure 3 shows composite extracted ion chromatogram profiles for the seven different precursor m/z values listed in Table 1 (glycoforms 6 and 7 are isomers, differing only by the location of the acetyl group, so have identical masses). Two consistent effects can be observed from these results. The presence of an additional acetylation shifts the glycopeptide retention time significantly later (e.g. compare black and dark blue profiles in Figure 3). The second repeatable effect is that a glycan with a NeuGc elutes slightly earlier than one with a NeuAc (compare dark green and bright red chromatograms in Figure 3).

To confirm the consistency of these chromatographic effects we investigated other glycopeptides: another Nucleobindin-1 O-linked peptide ^{399}KQQLQEQgSAPPSKPDGQLQFR^{419}, which displayed similar chromatographic behavior (data not shown); and a heterogeneous N-linked glycopeptide ^{177}VVLHPgNHSVVDIGLIK^{192} from mouse haptoglobin. This second peptide was identified with seven different complex glycan structures, including bi- and tri-antennary structures containing three types of sialic acid residues: NeuAc, NeuGc and NeuGc,Ac (see Table 3). The chromatographic profiles of these glycoforms are presented in Figure 4. These N-linked glycopeptides eluted in two groups. The first of these contained glycoforms featuring two sialic acids; the second cluster consists of trisialo glycoforms. Although the chromatographic shifts in these larger glycans were less dramatic, it was still observed that acetylation of NeuGc to NeuGc,Ac makes the glycopeptide more hydrophobic, extending its retention time (compare black and brown profiles). The difference caused by NeuGc compared to NeuAc is again small, but the glycoform containing NeuGc eluted earlier (red peak is earlier than green). The addition of neutral sugar structures, for example, the fucosylation of the disialo biantennary oligosaccharide or triantennary versus biantennary structure bearing the three sialic acids lead to shorter retention times (compare the yellow profile to the black one; and the red peak to the turquoise one). A monosialo biantennary glycoform of low abundance was detected featuring the same retention time as the disialo structure (Supplementary Figure 1). We assume that this was a product of in-source fragmentation, and indeed it was detected at a significantly lower intensity than the fully capped counterpart (Supplementary Figure 2). The neutral, oligomannose glycoforms of this glycopeptide eluted more than 10 minutes earlier than the sialo glycopeptides (Supplementary Figure 1).

Finding glycopeptides in published phosphopeptide data

To demonstrate how effective these diagnostic sugar oxonium ions are at finding glycopeptide spectra we analyzed some data from a large HeLa cell phosphopeptide dataset acquired by HCD fragmentation and recently published [23], to see if glycopeptides were also enriched, since it has previously been shown that the titanium dioxide chromatography, used by these authors for phosphopeptide enrichment, also purifies sialylated glycopeptides [33]. Peak lists were created from the six runs that made up control replicate 1 in their analysis. These peak lists were then analyzed for diagnostic oxonium fragment ions using MS-Filter in Protein Prospector (www.prospector.ucsf.edu). This software, when supplied with a peak list file/s (e.g. mgf files) can be used to find MS/MS spectra that contain a particular fragment (or neutral loss), within the specified mass accuracy, among the

specified number of most abundant fragment ions. The output is a new peak list file containing only spectra containing the ion of interest.

In the six files analyzed a total of 165654 MS/MS spectra were acquired. 20631 MS/MS spectra (12.5 % of those acquired) contained a HexNAc oxonium ion at m/z 204.087, showing there were a considerable number of glycopeptides enriched in the sample. Only 2220 of these contained the 274.092 ion, indicating presence of a NeuAc and 22 spectra contained the 316.103 indicating O-acetylated NeuAc. Human glycoproteins normally should not contain NeuGc. However the NeuGc diagnostic fragment ions: m/z 290.087 and m/z 308.098, were found in 64 spectra, an example of which is shown in Figure 5. This spectrum is from an O-linked glycopeptide, and is a mixture spectrum of two glycoforms, both with a sugar composition of NeuAcNeuGcGalGalNAc. The identity of the peptide itself could not be determined.

Discussion

Glycosylation is the most complex post-translational modification to study, since it involves a wide array of different modifications: different amino acid side-chains can be modified with single sugar units that can be extended to form short or large linear or branched oligosaccharides. Site-specific heterogeneity is manifested in both site-occupancy and in the number of different structures modifying the same residue. In order to understand the biological function of glycosylation, site-specific heterogeneity has to be studied, which means glycopeptide analysis is essential. The two different MS/MS activation methods of CID and ETD analysis provide complementary information about glycopeptides. In simplified terms, CID delivers information about the sugar structure *via* glycosidic bond cleavages, and provides molecular mass information for the modified peptide in the form of Y_0 and Y_1 ions, for O-linked and N-linked glycopeptides, respectively [13, 14]. ETD mostly leads to peptide backbone fragments, and thus, can identify the modified sequence as well as determine the site, but rarely yields information about the glycan mass or structure. Hence, for ETD spectrum identification by database searching the glycan mass has to be guessed. Searching a protein and glycan database simultaneously is the approach the commercially available Byonic software takes, whereas an iterative searching approach using Protein Prospector was employed here, similar to an earlier study [28]. Each approach can be quite successful, but also each may deliver false glycan assignments. Methionine or tryptophan oxidation (very common fortuitous modifications), when occurring close to a glycosylation site, may make the distinction between glycan structures differing by 16 Da (hexose vs fucose; NeuAc vs NeuGc) practically impossible.

Similarly, misidentification of the monoisotopic precursor ion may lead to the assignment of two fucose residues in the glycan instead of NeuAc (m/z 292.116 vs m/z 291.095). It has been pointed out by Wu et al. [34], that some of the ETD-based mis-assignments could be corrected if CID data were available and used. However, presently no information from the CID data is utilized in the ETD searches by any search engines, as far as we know. However, there is a computational framework that uses CID, HCD and ETD data, searched separately, in concert for N-linked glycopeptide identification[35].

In the present study we report diagnostic fragment ions for the four most common sialic acids in mammalian glycoproteins, and also describe the disaccharide fragments indicating HexNAc modification. The high proportion of oxygens makes the sugar oxonium ions mass-deficient in comparison to peptide fragments of nominally the same mass. This means that high mass accuracy fragment measurements achieved by Orbitrap or time-of-flight detectors allow exquisite separation of sugar fragment ions from any peptide products, and thus, these oxonium ions are highly specific. We show that using these diagnostic glycan fragment ions one can find different sialic acid variants in large scale or single protein glycosylation studies. O-acetylation of sialic acids, normally not probed for, can be detected this way, and NeuGc incorporation into human samples can be revealed. For higher specificity one can make use of disaccharide fragment ions when trying to find glycopeptide spectra with glycans bearing specific motifs, such as sialylated HexNAc residues, or antennae fucosylated structures. In addition, observing a sialylated HexNAc fragment in N-linked glycopeptide spectra indicates a Gal beta 1–3 GlcNAc linkage instead of the more common Gal beta 1–4 linkage. Only these so called type-1 structures may be sialylated on the subterminal GlcNAc. However, truncated, sialyl GlcNAc-containing glycans have been reported in some glycosylation studies without further confirmation.

The number of researchers performing glycopeptide analysis is currently dwarfed by the number of researchers studying phosphorylation. However, phosphopeptide enrichment by TiO_2 has been reported as suitable for the isolation of sialylated glycopeptides [33]. As shown, even under not optimized conditions [23] ~12% of the acquired MS/MS data corresponded to glycopeptides. Interestingly, barely a tenth of these glycopeptides contained sialic acid, indicating that the presence of the acidic capping is not a requirement in the enrichment process. We demonstrated the usefulness of the diagnostic fragment ions by identifying O-acetyl NeuAc-bearing glycopeptides and also a not insignificant occurrence of NeuGc. Human cells do not have the enzymatic machinery to create this residue, however, they can get it from dietary sources *in vivo*, or in this particular case from bovine glycoproteins present in the cell culture media. Some of these spectra may represent tryptic peptides from contaminating bovine proteins themselves (the CID spectra did not contain sufficient information for the identification of the modified sequences). However, a significant level of NeuGc incorporation in human cancer cell lines has been reported [36] and the HeLa cells analyzed are derived from cervical cancer cells.

In addition to using the diagnostic fragment ions, considering the influence of sialic acids on chromatographic retention also can be beneficial in both small scale and large scale glycopeptide analyses. Relative quantification of the different glycoforms when characterizing site-specific glycosylation is important. However, in-source fragmentation may occur even in electrospray ionization, leading to the detection of less sialylated glycoforms “coeluting“. On the other hand, in large scale glycosylation studies when an orthogonal separation such as high pH reverse phase chromatography is used for off-line fractionation, glycoforms containing different degree of acidity may end up in different fractions [28].

In conclusion, we have presented the characterization of differently sialylated O-linked and N-linked glycopeptides, pointing out the highly specific, diagnostic fragment ions of the

sialic acid variants (some of which have not been reported before), as well as the influence of these sugar units on the reverse phase chromatographic fractionation. Both of these information will be useful to people working in the growing field of glycopeptide analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by NIH grant NIGMS 8P41GM103481, and by the Howard Hughes Medical Institute (to the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF, Director: A.L. Burlingame).

Supporting information is available. Supplementary Figures 1 and 2 were uploaded in a single file. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1). Apweiler R, Hermjakob H, Sharon N. Biochim Biophys Acta. 1999; 1473:4–8. [PubMed: 10580125]

(2). Jacob GS, Welply JK, Scudder PR, Kirmaier C, Abbas SZ, Howard SC, Keene JL, Schmuke JJ, Broschat K, Steininger C. Adv Exp Med Biol. 1995; 376:283–290. [PubMed: 8597260]

(3). Baum LG, Garner OB, Schaefer K, Lee B. Front Immunol. 2014; 5:284. doi:10.3389/fimmu.2014.00284. [PubMed: 24995007]

(4). Schjoldager KT, Vester-Christensen MB, Goth CK, Petersen TN, Brunak S, Bennett EP, Levery SB, Clausen H. J Biol Chem. 2011; 286:40122–40132. [PubMed: 21937429]

(5). Woronowicz A, Amith SR, De Vusser K, Laroy W, Contreras R, Basta S, Szewczuk MR. Glycobiology. 2007; 17:10–24. [PubMed: 16971381]

(6). Lingg N, Zhang P, Song Z, Bardor M. Biotechnol J. 2012; 7:1462–1472. [PubMed: 22829536]

(7). Schauer R. Curr Opin Struct Biol. 2009; 19:507–14. [PubMed: 19699080]

(8). Parameswaran R, Lim M, Arutyunyan A, Abdel-Azim H, Hurtz C, Lau K, Müschen M, Yu RK, von Itzstein M, Heisterkamp N, Groffen J. J Exp Med. 2013; 210:805–819. [PubMed: 23478187]

(9). Butor C, Diaz S, Varki A. J Biol Chem. 1993; 268:10197–10206. [PubMed: 8486687]

(10). Varki A. Glycoconj J. 2009; 26:231–245. [PubMed: 18777136]

(11). Higashi H, Naiki M, Matuo S, Okouchi K. Biochem Biophys Res Commun. 1977; 79:388–395. [PubMed: 412499]

(12). Samraj AN, Läubli H, Varki N, Varki A. Front Oncol. 2014; 4:33. doi:10.3389/fonc.2014.00033. [PubMed: 24600589]

(13). Medzihradszky KF. Methods Enzymol. 2005; 405:116–138. [PubMed: 16413313]

(14). Peter-Katalinić J. Methods Enzymol. 2005; 405:139–171. [PubMed: 16413314]

(15). Thaysen-Andersen M, Packer NH. Biochim Biophys Acta. 2014; 1844:1437–1452. [PubMed: 24830338]

(16). Zauner G, Kozak RP, Gardner RA, Fernandes DL, Deelder AM, Wuhrer M. Biol Chem. 2012; 393:687–708. [PubMed: 22944673]

(17). Huddleston MJ, Bean MF, Carr SA. Anal Chem. 1993; 65:877–884. [PubMed: 8470819]

(18). Zhao P, Viner R, Teo CF, Boons GJ, Horn D, Wells L. J. Proteome Res. 2011; 10:4088–104. [PubMed: 21740066]

(19). Medzihradszky KF, Gillece-Castro BL, Townsend RR, Burlingame AL, Hardy MR. J. Am. Soc. Mass Spectrom. 1996; 7:319–328. [PubMed: 24203358]

(20). Moody AM, North SJ, Reinhold B, Van Dyken SJ, Rogers ME, Panico M, Dell A, Morris HR, Marth JD, Reinherz EL. J Biol Chem. 2003; 278:7240–7246. [PubMed: 12459555]

(21). Mechref Y, Muzikar J, Novotny MV. Electrophoresis. 2005; 26:2034–2046. [PubMed: 15841499]

(22). Darula Z, Sherman J, Medzihradszky KF. Mol Cell Proteomics. 2012; 11 O111.016774. doi: 10.1074/mcp.

(23). Sharma K, D'Souza RC, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M. Cell Rep. 2014; 8:1583–1594. [PubMed: 25159151]

(24). Akashi M, Okamoto A, Tsuchiya Y, Todo T, Nishida E, Node K. Cell Rep. 2014; 7:1056–1064. [PubMed: 24794436]

(25). Trinidad JC, Barkan DT, Gulledge BF, Thalhammer A, Sali A, Schoepfer R, Burlingame AL. Mol Cell Proteomics. 2012; 11:215–229. [PubMed: 22645316]

(26). Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL. Mol Cell Proteomics. 2011; 10 M111.010728 doi: 10.1074/mcp.M111.010728.

(27). Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL. Mol Cell Proteomics. 2008; 7:2386–2398. [PubMed: 18653769]

(28). Trinidad JC, Schoepfer R, Burlingame AL, Medzihradszky KF. Mol Cell Proteomics. 2013; 12:3474–3488. [PubMed: 23816992]

(29). Domon B, Costello CE. Glycoconjugate J. 1988; 5:397–409.

(30). Halim A, Westerlind U, Pett C, Schorlemer M, Rüetschi U, Brinkmalm G, Sihlbom C, Lengqvist J, Larson G, Nilsson J. J Proteome Res. 2014; 13:6024–6032. [PubMed: 25358049]

(31). Darula Z, Medzihradszky KF. Mol Cell Proteomics. 2009; 8:2515–2526. [PubMed: 19674964]

(32). Medzihradszky KF, Maltby DA, Hall SC, Settineri CA, Burlingame AL. J. Am. Soc. Mass Spectrom. 1994; 5:350–358. [PubMed: 24222589]

(33). Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH. Mol Cell Proteomics. 2007; 6:1778–1787. [PubMed: 17623646]

(34). Wu SW, Pu TH, Viner R, Khoo KH. Anal Chem. 2014; 86:5478–5486. [PubMed: 24796651]

(35). Mayampurath A, Yu CY, Song E, Balan J, Mechref Y, Tang H. Anal Chem. 2014; 86:453–63. [PubMed: 24279413]

(36). Inoue S, Sato C, Kitajima K. Glycobiology. 2010; 20:752–762. [PubMed: 20197272]

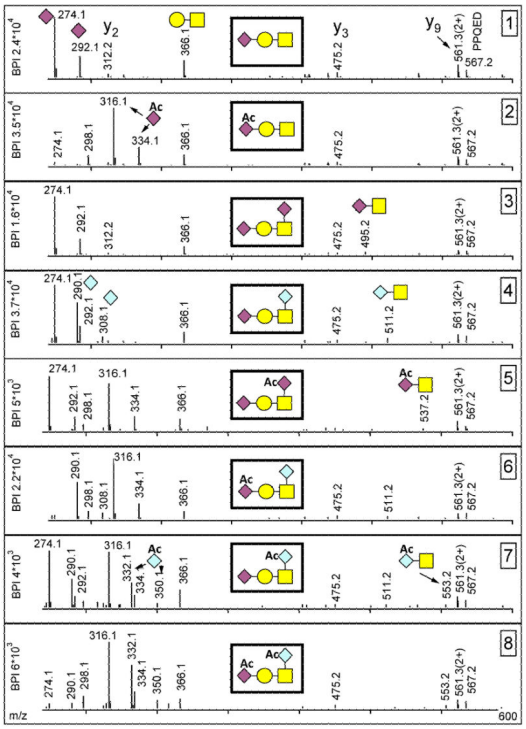

Figure 1.

Low mass region of the beam-type CID (HCD) spectra of eight different glycoforms of a Nucleobindin-1 O-linked glycopeptide, ^{31}AAPPQEDSQAgTETPDTGLYYHR^{52}. Thr-41 indicated with a 'g' in the sequence was identified as the glycosylation site from ETD data. The structures in the middle of the panels indicate the glycan. The sugar-unit symbols are: GalNAc = yellow square; Gal = yellow circle; NeuAc = purple diamond; NeuGc = light blue diamond; Ac indicates O-acetylation of the sialic acid. The identity of the Hex modifying the GalNAc cannot be determined from the CID data: it was assumed that the most common, mucin-type core 1 structure, Gal-GalNAc modified this protein. The different sialic acids produced at least 2 diagnostic ions: a B1 fragment and a more abundant fragment formed *via* water loss from the B1 oxonium ion [Nomenclature: 29], as indicated in the appropriate panels. Diagnostic fragments that show the sialic acid attached to the core GalNAc are also labeled. For simplicity, fragments are only labeled in the panel where they are first observed. Panel 7 shows HCD data from a mixture spectrum: the tetrasaccharide featuring both sialic acid variants was detected in two differently acetylated forms.

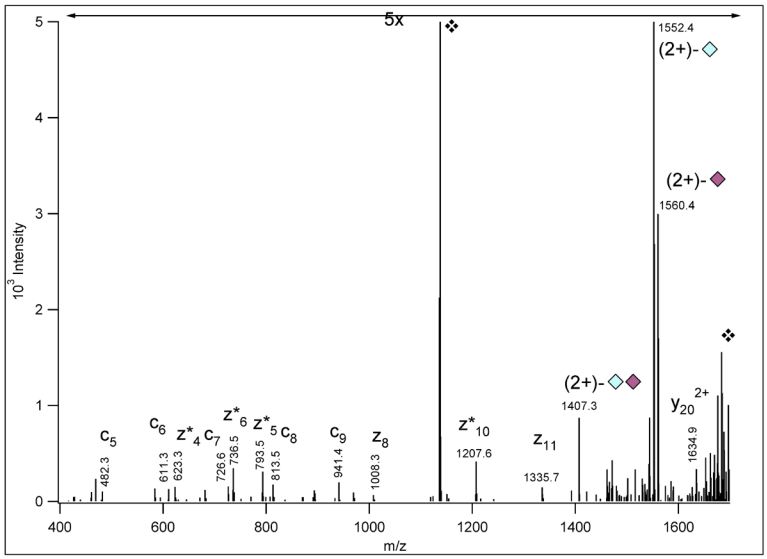

Figure 2.

ETD spectrum of ^{31}AAPPQEDSQAgTETPDTGLYYHR^{52} peptide from Nucleobindin-1 bearing a NeuAcGal(NeuGc)GalNAc modification on Thr-41, i.e. glycoform 5 in Table 1 and Figure 1. ❖ labels the original precursor, m/z 1137.4800(3+), and its charge-reduced form. The three most intense fragment ions were formed *via* neutral losses of the different individual sialic acid residues and both sialic acids from the charge-reduced species. NeuAc = purple diamond; NeuGc = light blue diamond. Asterisks indicate z+1 fragments.

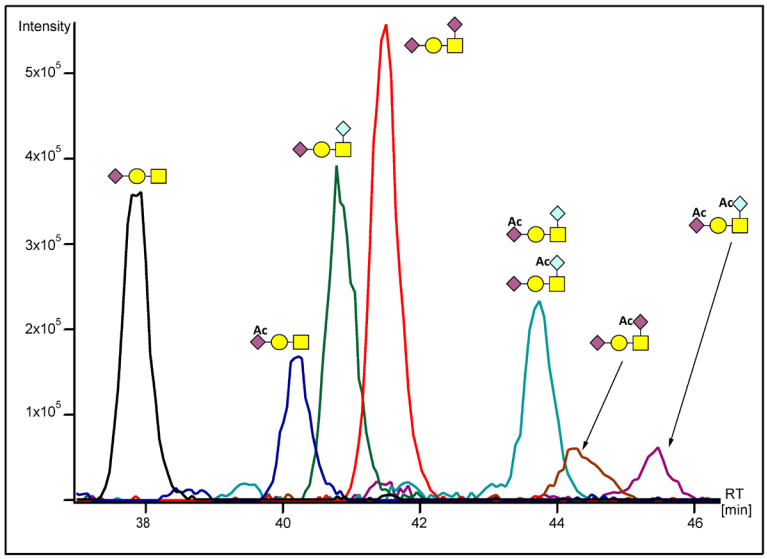

Figure 3.

Superimposed extracted ion chromatograms of eight different glycoforms of ^{31}AAPPQEDSQAgTETPDTGLYYHR^{52} from Nucleobindin-1. The glycosylation site has been assigned from ETD data and indicated with a 'g'. Monoisotopic masses listed in Table 1 were extracted with a mass accuracy of 20 ppm. The sugar-unit symbols are: GalNAc = yellow square; Gal = yellow circle; NeuAc = purple diamond; NeuGc = light blue diamond; Ac indicates O-acetylation of the sialic acid.

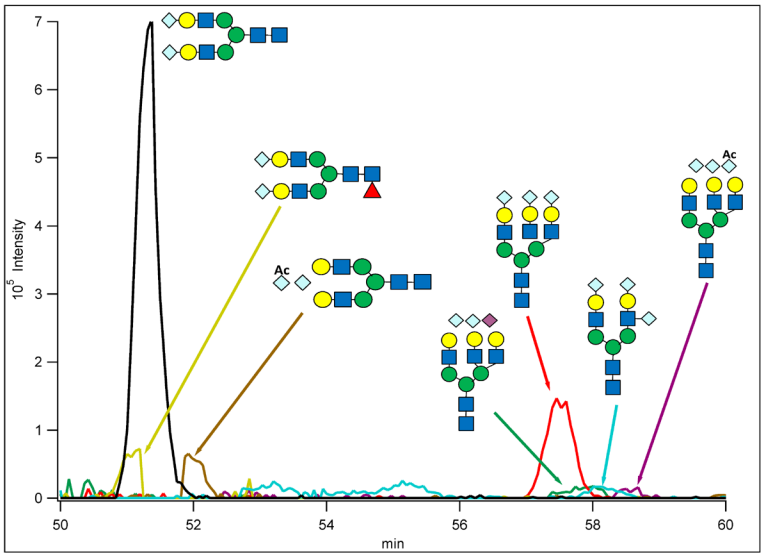

Figure 4.

Superimposed extracted ion chromatograms of seven different glycoforms of the N-linked glycopeptide ^{177}VVLHPNHSVVDIGLIK^{192} from mouse Haptoglobin (see Table 3). The sugar-unit symbols are: GlcNAc = dark blue square; Man = green circle; Gal = yellow circle; Fuc = red triangle; NeuAc = purple diamond; NeuGc = light blue diamond; Ac indicates O-acetylation of the sialic acid. When it was not possible to determine at which branch the different sialic acids were attached they are shown 'disconnected'. Monoisotopic (4+) masses listed in Table 3 were extracted with a 20 ppm mass tolerance.

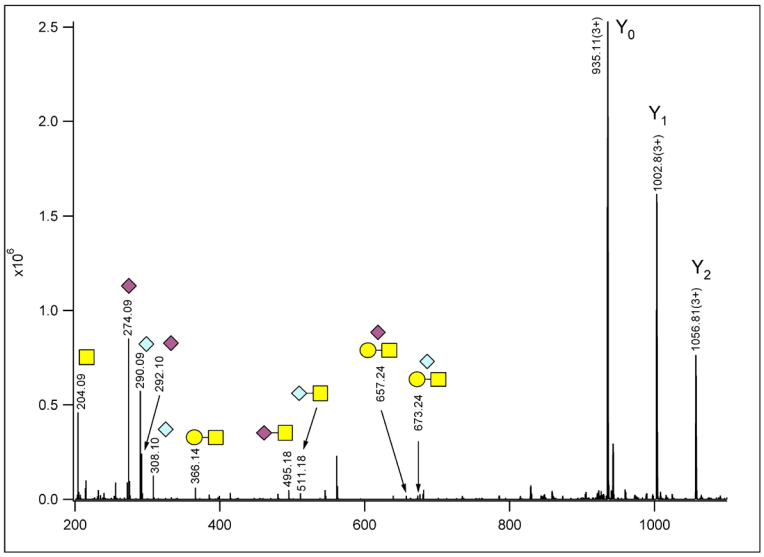

Figure 5.

HCD spectrum of a glycopeptide acquired as part of a phosphopeptide analysis study [23]. The precursor ion was m/z 942.4100(4+). The presence of both NeuAc and NeuGc was ascertained from their diagnostic fragment ions (labeled). The glycan fragmentation allows determination that this is an O-linked glycopeptide featuring two isomeric glycan structures: NeuAcGal(NeuGc)GalNAc and NeuGcGal(NeuAc)GalNAc. The sialic acid positions within the trisaccharide fragments could not be determined, so are shown 'disconnected'. The identity of the peptide could not be determined from these data.

Table 1

Glycoforms of Nucleobindin-1 O-linked glycopeptide ^{31}AAPPQEDSQAgTETPDTGLYYHR^{52} detected in a tryptic digest of a mouse liver sample.

Glycan	mass added	RT*	MH^+_{cal}	$(3+)_{cal}$	$(3+)_{mes}$	Δ[ppm]
NeuAcGalGalNAc	656.1976	38.210	3103.3386	1035.1177	1035.1195	2
$NeuAc_2$GalGalNAc	698.2082	40.030	3145.3491	1049.1212	1049.1224	1
NeuAcGal(NeuAc)GalNAc	947.2930	41.102	3394.4340	1132.1495	1132.1510	1
NeuAcGal($NeuAc_2$)GalNAc	989.3036	44.638	3436.4446	1146.1530	1146.1542	1
NeuAcGal(NeuGc)GalNAc	963.2879	41.070	3410.4289	1137.4811	1137.4800	−1
$NeuAc_2$Gal(NeuGc)GalNAc	1005.2985	43.975	3452.4394	1151.4847	1151.4836	1
NeuAcGal(NeuGc,Ac)GalNAc	1005.2985	43.330	3452.4394	1151.4847	1151.4838	1
$NeuAc_2$Gal(NeuGc,Ac)GalNAc	1047.3091	45.248	3494.4500	1165.4882	1165.4907	2

RT indicates the timepoint when the (3+) ions listed were selected for MS/MS analysis $NeuAc_2$ stands for Neu5,xAc_2, since the site of O-acetylation cannot be determined from these data

Table 2

Diagnostic sugar oxonium ions observed in the low mass region of glycopeptide beam-type CID (HCD) spectra.

Sugar Composition	Diagnostic Oxonium Ions (m/z)	Sugar Composition	Diagnostic Oxonium Ions (m/z)
HexNAc	204.087[a]	HexNAcHex	366.139
NeuAc	292.103, 274.092	HexNAcNeuAc	495.182
NeuGc	308.098, 290.087	HexNAcNeu5Gc	511.177
$NeuAc_2$[b]	334.113, 316.103	HexNAcNeuAc,Ac	537.193
NeuGc,Ac[b]	350.108, 332.098	HexNAcNeu5Gc,Ac	553.188

[a] while this oxonium ion is identical for GalNAc and GlcNAc, they can be distinguished from lower mass fragments in beam-type CID spectra [30]

[b] $NeuAc_2$ and NeuGc, Ac indicate O-acetylation without assigning its position, since this cannot be determined from these data

Table 3

Complex glycoforms detected for haptoglobin peptide ^{177}VVLHPNHSVVDIGLIK^{192}, from a mouse liver sample.

Glycan structure	mass added	RT	MH^+_{cal}	$(4+)_{cal}$	$(4+)_{mes}$	Δ[ppm]
biantennary NeuGc2	2236.7622	51.37	3976.7789	994.9502	994.9534	3
biantennary NeuGcNeuGc,Ac	2278.7728	51.92	4018.7895	1005.4528	1005.4536	1
biantennary NeuGc2Fuc	2382.8201	51.19	4122.837	1031.4646	1031.4651	0
biantennary NeuGc3	2543.8525	58.60	4283.869	1071.7227	1071.7299	7
triantennary NeuAcNeuGc2	2892.9898	57.74	4633.007	1159.0071	1159.0103	3
triantennary NeuGc3	2908.9847	57.46	4649.001	1163.0058	1163.0089	3
triantennary NeuGc2NeuGc,Ac	2950.9953	58.09	4691.0120	1173.5084	1173.5094	1

RT indicates the timepoint when the (4+) ions listed were selected for MS/MS analysis