research articles Journal of proteome research

Exploiting Differential Dissociation Chemistries of O-Linked Glycopeptide Ions for the Localization of Mucin-Type Protein Glycosylation

Richard R. Seipert, Eric D. Dodds, and Carlito B. Lebrilla*

Department of Chemistry and School of Medicine, University of California Davis, One Shields Avenue, Davis, California 95616

Received September 4, 2008

From a glycoproteomic perspective, the unambiguous localization of O-linked oligosaccharide attachment sites is fraught with analytical obstacles. Because no consensus protein sequence exists for O-glycosylation, there is potential for glycan attachment at numerous serine and threonine residues of a given protein. The well-established tendency for O-glycan attachment to occur within serine and threonine rich domains adds further complication to site-specific assignment of mucin-type glycosylation. In addition to the complexities contributed by the polypeptide chain, the O-linked carbohydrate modifications themselves are exceedingly diverse in both compositional and structural terms. This work is aimed at contributing an improved fundamental understanding of the chemistry that dictates dissociation of O-glycopeptide ions during tandem mass spectrometry (MS/MS). Infrared multiphoton dissociation (IRMPD) has been applied to an assortment of O-linked glycopeptide ions encompassing various compositions and charge states. Protonated O-glycopeptides were found to undergo a combination of glycosidic bond cleavage (complete coverage) and peptide bond cleavage (partial coverage). In contrast to previous observations of N-linked glycopeptide dissociation, the sodiated O-glycopeptides did not yield significantly different information as compared to the corresponding protonated ions. IRMPD of deprotonated O-glycosylated peptides provided informative side chain losses from nonglycosylated serine and threonine residues, which indirectly implicated sites of glycan attachment. In this manner, the combination of positive mode and negative mode MS/MS was found to provide conclusive assignment of O-glycosites.

Keywords: ESI • FT-ICR MS • glycopeptides • glycoproteomics • IRMPD • MS/MS • O-linked glycosylation • protein glycosylation • site-specific glycosylation analysis

Introduction

To further understand the various biological roles of proteins, it is essential to characterize their post-translational modifications (PTMs) and determine the influence of these modifications on function. Glycosylation is one of the most common PTMs and affects more than 50% of proteins.[1] Unlike other PTMs, specific molecular details on the interplay between glycosylation and protein function are only beginning to be revealed on a protein specific basis.[2−5] The relatively poor understanding of structure−function relationships involving protein glycosylation is largely attributable to a dearth of analytical capability for investigation of glycosylation in both a high-throughput and site-specific manner.

In recent years, site-specific glycosylation analysis using mass spectrometry (MS) has seen important methodological advancement; however, most of this work has been directed at N-linked glycoproteins.[6−10] The site-specific investigation of O-linked glycosylation has seen far less progress due to additional complications not shared by N-linked glycosylation. For example, N-linked glycans are attached only to asparagine residues which occur in the context of a specific consensus sequence, whereas O-linked glycosylation can potentially occur at any serine or threonine residue. In addition, all N-linked glycans share a common core structure, whereas eight core structures exist for O-linked glycans. Several of these O-linked cores may be present on a single protein.[11] Another confounding characteristic of O-glycosylation is the tendency of this modification to occur within domains rich in serine, threonine, alanine, and proline.[12] The presence of multiple flanking serine and threonine residues can introduce significant complications in assigning O-glycan modification to a specific site.

Standard methods for the investigation of site-specific glycosylation involve digestion of the glycoprotein with a specific enzyme such as trypsin followed by MS analysis of the peptide and glycopeptide mixture. This method has proven somewhat useful for determining sites of N-linked glycosylation; however, this approach is far less suitable for the analysis of O-linked glycosylation. Again, O-linked oligosaccharide modification usually occurs in serine, threonine, alanine, and proline rich regions which are often void of the lysine and arginine residues

* To whom correspondence should be addressed. Telephone: 1-530-752-6364. Fax: 1-530-754-5609. E-mail: cblebrilla@ucdavis.edu.

10.1021/pr8007072 CCC: $40.75

Published on Web 12/09/2008

necessary for tryptic cleavage. Moreover, glycosylation confers protease resistance to nearby peptide bonds, thus resulting in O-glycopeptides that can contain greater than 35 amino acid residues.[13−16] To identify sites of glycosylation on such glycopeptides, tandem mass spectrometry (MS/MS) with very high peptide sequence coverage is essential.

Recently, the use of a protease mixture called Pronase has been illustrated for site-specific glycosylation analysis of both N- and O-linked glycoproteins.[17] This method takes advantage of the nonspecific nature of Pronase to hydrolyze essentially every peptide bond of a protein, except for those around the site of glycosylation. This process eliminates nonglycosylated peptides which otherwise could prevent the detection of the lower abundance and less efficiently ionized glycopeptides. The length of the peptide tag, or "footprint," around the glycosylation site has also been shown to be tunable (2−10 amino acid residues) via alteration of digestion time. This allows preparation of shorter glycosylated peptides harboring fewer potential glycosylation sites as compared to large tryptic glycopeptides. This approach has proven very useful for preparation of informative glycopeptides; however, there are often numerous potential glycopeptide compositions that correspond to a given mass spectral signal. This is especially true for larger glycopeptides such as those formed via specific proteolysis.[18,19] Because of the vast number of theoretical glycopeptides, conflicting glycopeptide assignments may occur even for very accurately measured glycopeptide masses.

Even with the advantages of Pronase digestion, there is still a great need for MS/MS analysis in order to confirm glycopeptide composition and more importantly characterize sites of glycosylation.[20] Glycoproteomic site characterization should both identify the protein and localize the attached glycans to their respective sites on that protein. It has been illustrated that low energy collision-induced dissociation (CID) or infrared multiphoton dissociation (IRMPD) of multiply charged glycopeptide ions mainly results in glycosidic bond fragmentation.[20,21] By contrast, electron capture dissociation (ECD) or electron transfer dissociation (ETD) primarily induce fragmentation at peptide bonds.[21,22] Fragmentation of N-linked glycopeptide ions via IRMPD has recently been illustrated to have high dependence upon charge carrier, charge state, and overall glycopeptide composition with both glycosidic and peptide backbone cleavages being accessible depending on the chemistry of a given precursor ion.[10]

Early studies of O-linked glycopeptide fragmentation were conducted by Medzihradszky et al. utilizing high energy CID of BOC-Tyr derivatized glycopeptides.[23] Other synthetic or modified O-linked glycopeptides have also been investigated via post source decay and more recently with CID and ECD.[24−27] ECD has proven to be a useful method for the localization of glycosylation sites, but in the majority of these analyses, sialic acid residues were removed to enhance ECD efficiency. This practice precludes the site-specific assignment of glycans containing sialic acid residues. Marshall and coworkers also recently showed that both ECD and activated ion ECD (AI-ECD) of a 16 amino acid O-linked glycopeptide resulted in the predominant loss of the sugar residues with no peptide backbone cleavages observed; however, a glycopeptide with 38 amino acid residues yielded up to 50% peptide sequence coverage.[14]

On the basis of these studies and others, it has become apparent that the dissociation chemistries of glycosylated peptide ions are sufficiently complex to prevent a simplistic generalization of their fragmentation behavior. Nonetheless, it should be possible to exploit the rich and varied fragmentation chemistry of glycopeptide ions to obtain more informative MS/MS data. A detailed understanding of these dissociation pathways will be essential for the further advancement of MS-based glycoproteomics. This is especially true of O-glycoproteomics, which encompasses a greater diversity of carbohydrate structures and compositions, and involves glycosylation at relatively unpredictable sites. This work strives to address the need for an improved fundamental understanding of O-linked glycopeptide fragmentation, especially with regard to those glycopeptides obtained via Pronase digestion.

Experimental Section

Materials and Reagents. Bovine fetuin A (BF; Swiss-Prot accession number P12763) was purchased from CalBiochem (San Diego, CA). Bovine kappa-casein (κ-CN; Swiss-Prot accession number P02668), Pronase E, ammonium acetate, ethanolamine, hydrochloric acid, formic acid, ammonium acetate, and cyanogen bromide activated sepharose 4B beads (average particle diameter, 90 μm) were obtained from Sigma-Aldrich (St. Louis, MO). All other reagents were of analytical grade or higher.

Glycoprotein Digestion. Pronase E was immobilized on sepharose beads following a modified procedure originally described by Clowers et al.[17] In short, 150 mg of cyanogen bromide activated sepharose beads was hydrated with HCl (1 mM), rinsed with phosphate buffer (100 mM, pH 7.4), combined with Pronase (1 mg), and allowed to couple for 2 h. The remaining active sites on the sepharose beads were then blocked with ethanolamine (1 M, pH 9.0), and rinsed with ammonium acetate buffer (100 mM, pH 7.4). The Pronase beads were then stored in ammonium acetate buffer with sodium azide (0.05%) at 4 °C until needed. Stock glycoprotein solutions were prepared at a concentration of 50 μM in ammonium acetate buffer. For nonspecific proteolysis, the glycoprotein stock solution (100 μL) and fresh ammonium acetate buffer (200 μL) were added to the Pronase-coupled beads after washing with fresh ammonium acetate to remove the sodium azide. The digestion was allowed to proceed at 37 °C with gentle mixing. The supernatant was sampled at multiple time points ranging from 90 min to 24 h in order to obtain glycopeptides with varying peptide lengths. To avoid the transfer of any Pronase-immobilized beads, the digestion solution was centrifuged before sampling the supernatant. Glycopeptide solutions were then analyzed immediately or stored at −20 °C prior to analysis.

No sample cleanup was performed on the digestion supernatants; instead, they were diluted into an electrospray-compatible solvent system for direct MS analysis. All digests were diluted at least 10-fold to a final composition of 50% aqueous acetonitrile buffered with either formic acid (0.1%, pH 3.5) or ammonium acetate (1.0 mM, pH 7.4), depending on the desired precursor ion and/or polarity of mass spectral analysis. Multiple ionic forms of each glycopeptide were observed with each buffer system; however, the lower pH solutions favored the formation of protonated species while the higher pH solutions favored the formation of sodiated species. After removal of the supernatant, the Pronase-coupled beads were washed and stored for future digestions.

Instrumentation and Analysis. Mass spectral analyses were performed using a 9.4 T Fourier transform ion cyclotron resonance mass spectrometer (IonSpec QFT, Lake Forest, CA)

equipped with a Picoview nano-ESI stage (New Objective, Woburn, MA). Samples were introduced via a 1.0 μL full loop injection using a NanoLC-1D microfluidic pump and injector port (Eksigent Technologies, Dublin, CA), or via direct infusion using a syringe pump. In both cases, flow rates of 100 nL/min were applied. The infused solutions were charged via a liquid junction immediately before the nanoelectrospray tip and held at ±1500−2000 V with respect to the sample cone interface. Ions entered the system through a 390 μm aperture and were externally accumulated in a RF-only storage hexapole for up to 5 s prior to injection into the ICR cell.

Individual glycopeptide ions were selected within the ICR cell using stored-waveform inverse Fourier transform (SWIFT) isolation prior to infrared multiphoton dissociation (IRMPD). For IRMPD experiments, the infrared radiation was supplied by a 10.6 μm, 20 W CO_2 laser (Parallax Laser, Inc., Waltham, MA). The infrared laser beam was expanded to a diameter of 0.5 cm with an inline beam expander (Synrad Laser, Mukilteo, WA) to ensure the isolated ion cloud received maximum exposure during each infrared laser pulse. The fragmentation was optimized by varying the IRMPD laser pulse between 500 and 1500 ms. Irradiation time was increased until the majority of the precursor ion was dissociated. Up to 50 mass spectral scans were acquired and averaged to enhance signal-to-noise ratio. Spectra were externally calibrated using the IonSpec Omega data station. The *m*/*z* error threshold used for all glycopeptide and fragment ion assignments was 10 ppm.

Results and Discussion

Multiple O-linked glycopeptides were produced via the digestion of κ-CN and BF with immobilized Pronase. κ-CN is a well-characterized protein that is known to contain five glycosylation sites. Glycosylation at these sites has recently been shown to occur in a mostly ordered fashion, with only T^{152} occupied in the singly glycosylated glycoform, T^{163} and T^{154} becoming occupied in the doubly and triply glycosylated forms, respectively, after which sites T^{157} and T^{142} are occupied without specificity.[28] BF is a glycoprotein containing both N-linked and O-linked glycans at N^{99}, N^{156}, N^{176}, S^{271}, T^{280}, and S^{282}, respectively.[16,23,29,30] The N-linked glycans are large, complex type, multiply sialylated di- or triantennary glycans, whereas the O-linked glycans are a mixture of core 1 and core 2 types. Because of the nonspecific nature of Pronase digestion, glycopeptides containing between four and eight amino acid residues were formed by modulating the digestion time. The nomenclature for glycopeptide fragmentation used here follows the conventions of Roepstorff and Fohlman for peptide cleavages, and the conventions of Domon and Costello for glycan cleavages.[31,32] For differentiation between peptide and glycan fragments, all peptide fragments are marked in lower case letters and all glycan fragments are marked in upper case letters. The monosaccharides are represented in the figures according to the symbolic nomenclature of the Consortium for Functional Glycomics (http://glycomics.scripps.edu/CFGnomenclature.pdf). According to this scheme, the following monosaccharides are indicated by the following respective symbols: galactose (Gal), yellow circle; mannose (Man), green circle; *N*-acetylgalactosamine (GalNAc), yellow square; *N*-acetylglucosamine (GlcNAc), blue square; and *N*-acetyIneuraminic acid (NeuAc), purple diamond. In the tables, the more general monosaccharide abbreviations Hex (hexose) and HexNAc (*N*-acetylhexosamine) are used for simplicity.

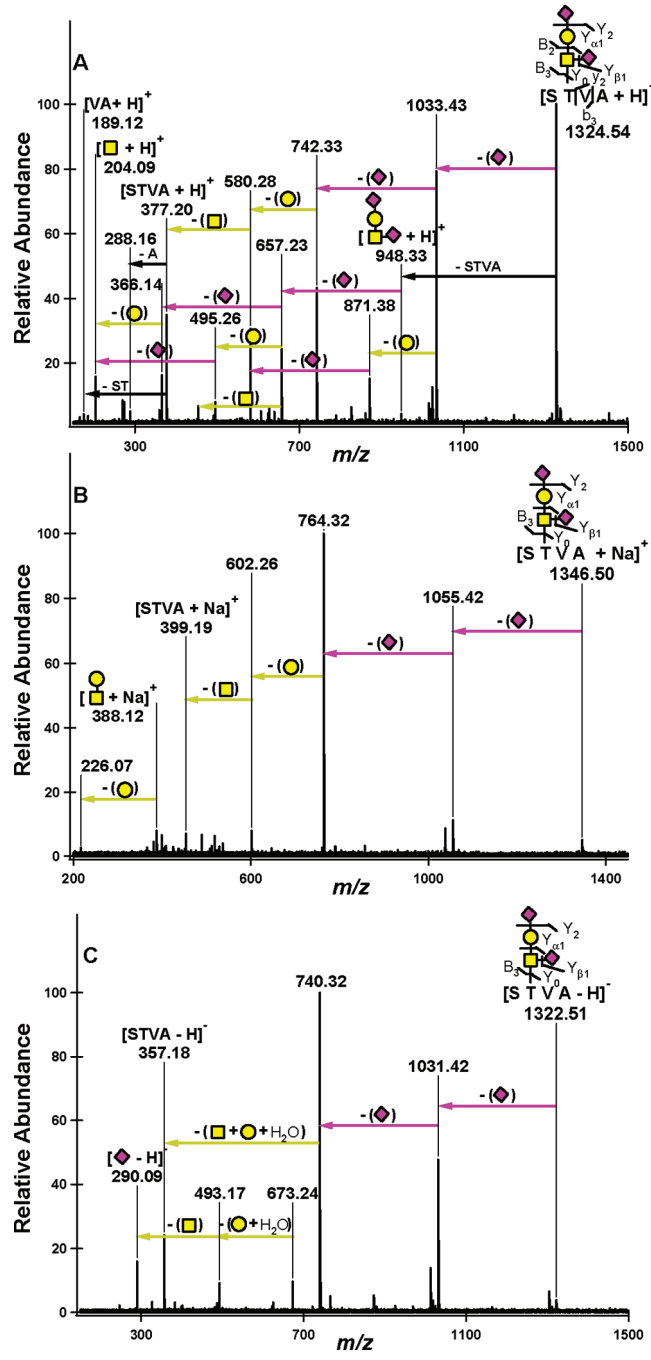

Figure 1. IRMPD nano-ESI FT-ICR mass spectra of the O-glycopeptide ^{162}ST*VA^{165} + Hex_1 $HexNAc_1$ $NeuAc_2$ derived from κ-CN with multiple charge carriers: (A) singly charged, protonated glycopeptide *m*/*z* 1324.54 (10 scans); (B) singly charged, sodium coordinated glycopeptide *m*/*z* 1346.50 (15 scans); (C) singly charged, deprotonated glycopeptide *m*/*z* 1322.51 (10 scans).

Protonated O-Linked Glycopeptides. Upon investigation of multiple O-linked glycopeptide ions with IRMPD, it was observed that the charge carrier greatly affected their fragmentation behavior. This result was not unexpected, as similar dependence on both charge state and charge carrier was previously demonstrated for N-linked glycopeptides ions.[10] Figure 1 shows the IRMPD spectra of the glycopeptide STVA + Hex_1 $HexNAc_1$ $NeuAc_2$ in three different ionic forms. The singly protonated form of this glycopeptide was cleaved along both glycosidic and peptide bonds to reveal the connectivity

of the molecule (Figure 1A). The most abundant fragments resulted from cleavage of the labile sialic acid residues at m/z 1033.43 and 742.33. It is unclear which sialic residue is lost first; however, due to the presence of m/z 495.26 and 871.38, it is clear that there is a sialic residue linked to the HexNAc with the other NeuAc residing on the hexose. After these losses, the remaining glycosidic bonds fragmented in a sequential manner to yield the protonated peptide. The peptide then further dissociated, allowing confirmation of the amino acid composition and sequence. The location of the glycosylation site in this case remained ambiguous due the lack of peptide fragmentation prior to glycan loss.

The IRMPD mass spectra for the protonated glycopeptides PTST + Hex_1 $HexNAc_1$ $NeuAc_2$ and GEPTSTPT + Hex_1 $HexNAc_1$ $NeuAc_2$ are shown in Figure 2A and 3A, respectively. In a parallel manner to the previous example, these glycopeptides first lost the labile sialic acid residues; however, unlike the previous example, the entire glycan was not lost from the peptide prior to the cleavage of peptide bonds. The fragment ion at m/z 489.22, which was very abundant in both spectra, corresponds to PTS + HexNAc. This fragment narrows down the site of glycosylation to either the remaining serine or threonine residue. The site was conclusively assigned based on additional information gained from the negative mode spectrum (as will be discussed below). Complete sequence coverage of the short peptide PTST was obtained as shown in Figure 2A. The glycopeptide GEPTSTPT + Hex_1 $HexNAc_1$ $NeuAc_2$ had several abundant peptide fragment ions, including y_6 and b_6 which both occur N-terminal to proline residues with one also C-terminal to a glutamic acid residue (Figure 3A). These were expected to be major product ions due to the well-known "proline effect" and "aspartic acid effect."[33] These fragmentation effects lead to the assignment of the before mentioned ion m/z 489.22. On the basis of accurate mass, this ion could have been assigned to PTS + HexNAc or STP + HexNAc; however, due to the proline effect, the observation of STP + HexNAc is highly unlikely since cleavage N-terminal to proline is much more labile than both of the cleavages that would lead to this product ion.

A core 2 type glycan occurring on the protonated glycopeptide GPTPSA + Hex_2 $HexNAc_2$ $NeuAc_2$ derived from BF fragmented in a similar manner to the core 1 glycan containing glycopeptides previously discussed (Figure 4A). The core structure was confirmed by m/z 935.42, which corresponded to the intact peptide connected to two HexNAc residues. This ion was complemented by two fragment ions at m/z 713.26 and 551.21, both of which contain two HexNAc residues after release from the peptide backbone indicating that this was a single glycan moiety with HexNAc to HexNAc connectivity. The peptide moiety of this glycopeptide dissociated to yield b_3 and y_3 fragment ions, both of which were N-terminal to proline residues. This glycopeptide had the lowest peptide sequence coverage of all the glycopeptides investigated.

Table 1 summarizes the fragmentation spectra of 12 distinct glycopeptide ions derived form BF and κ-CN in the protonated form, including the four specific cases detailed above. The glycosidic bond cleavages were not listed in the table because all of these glycopeptide ions dissociated to yield a complete series of either B or Y ions, which was sufficient to verify connectivity (but not linkage). The glycopeptides exhibited differing amounts of peptide fragmentation, with some yielding 100% sequence coverage and others only 20% (53% on average). Since these are relatively small peptide moieties, it is more likely

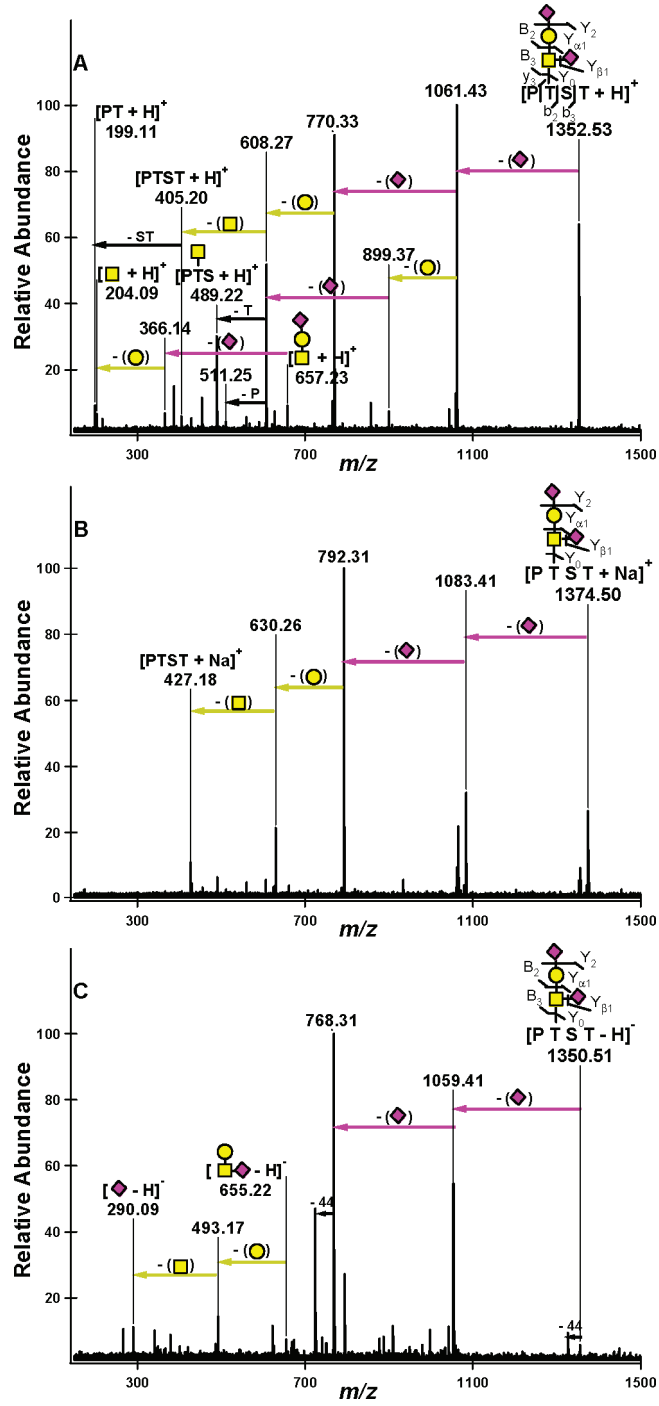

Figure 2. IRMPD nano-ESI FT-ICR mass spectra of the O-glycopeptide ^{151}PT*ST^{154} + Hex_1 $HexNAc_1$ $NeuAc_2$ derived from κ-CN with multiple charge carriers: (A) singly charged, protonated glycopeptide m/z 1352.53 (25 scans); (B) singly charged, sodium coordinated glycopeptide m/z 1374.50 (25 scans); (C) singly charged, deprotonated glycopeptide m/z 1350.51 (30 scans).

that the site can be determined despite mediocre sequence coverage. This is an important contrast to glycopeptides with large peptide moieties containing many potential glycosites, which would require very high sequence coverage for site localization. Another notable feature was tendency of the peptide moiety to act as the more dominant charge carrier, as the majority of the fragments contained the peptide portion of the molecule while comparatively few fragments only

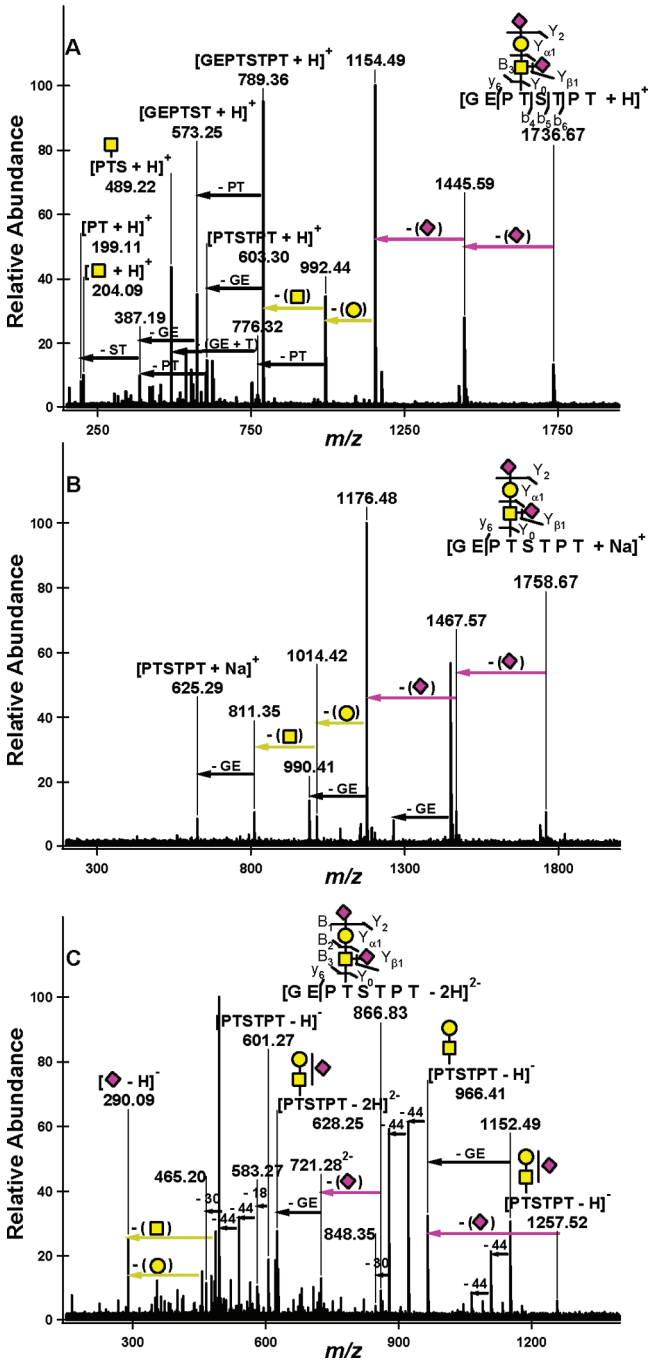

Figure 3. IRMPD nano-ESI FT-ICR mass spectra of the O-glycopeptide ^{149}GEPT*STPT^{156} + Hex_1 $HexNAc_1$ $NeuAc_2$ derived from κ-CN with multiple charge carriers: (A) singly charged, protonated glycopeptide *m/z* 1736.67 (25 scans); (B) singly charged, sodium coordinated glycopeptide *m/z* 1758.67 (50 scans); (C) doubly charged, doubly deprotonated glycopeptide *m/z* 866.83 (10 scans).

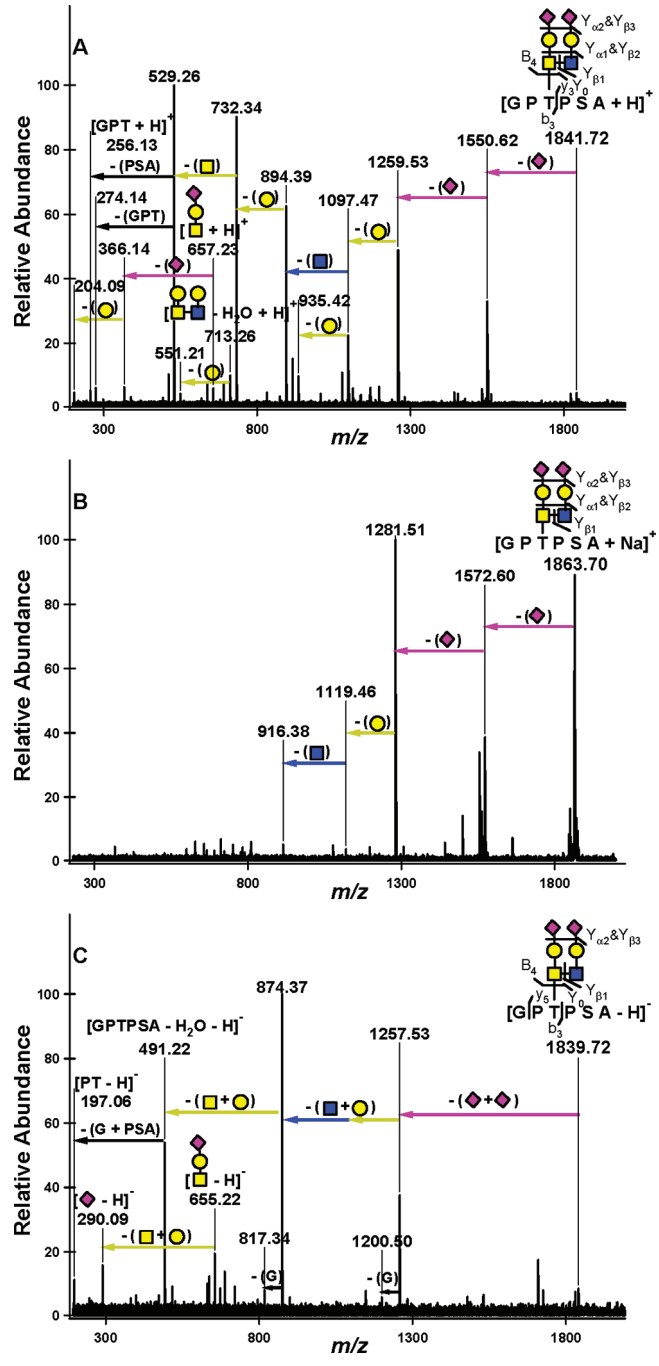

Figure 4. IRMPD nano-ESI FT-ICR mass spectra of the O-glycopeptide ^{278}GPT*PSA^{283} + Hex_2 $HexNAc_2$ $NeuAc_2$ derived from BF with multiple charge carriers: (A) singly charged, protonated glycopeptide *m/z* 1841.72 (20 scans); (B) singly charged, sodium coordinated glycopeptide *m/z* 1863.70 (50 scans); (C) singly charged, deprotonated glycopeptide *m/z* 1839.72 (20 scans).

contained the glycan. A multiply glycosylated glycopeptide ion GEPTSTPT + $(Hex_1$ $HexNAc_1$ $NeuAc)_2$ was found to fragment in a manner similar to the singly glycosylated variant. No additional peptide fragment ions were observed for this glycopeptide; however, another series of glycan fragments was observed due to the additional glycan moiety. All fragment ions observed below *m/z* 1175 were identical to those shown in Figure 3A.

Sodiated O-Linked Glycopeptides. The dissociation of the sodium coordinated STVA + Hex_1 $HexNAc_1$ $NeuAc_2$ yielded *m/z* 764.32 as the most abundant fragment (Figure 1B). After these neutral losses of the two sialic acid residues, low-abundance fragments were observed for both the sodiated peptide, *m/z* 399.19, and the sodiated glycan void of sialic acid residues, *m/z* 388.12 (Hex_1 $HexNAc_1$). This was the only sodiated O-glycopeptide that was found to dissociate in a manner that yielded a glycan only fragment. The sodium coordinated glycopeptide

Table 1. Protonated O-Linked Glycopeptide Ions Subjected to IRMPD[a]

m/*z*	peptide	glycan	peptide cleavages	seq. cov.
1114.45	GPT*PS	Hex_1 $HexNAc_1$ $NeuAc_1$	b_3	25%
1253.49	ST*V	Hex_1 $HexNAc_1$ $NeuAc_2$	b_2	50%
1312.55	APS*AVPD	Hex_1 $HexNAc_1$ $NeuAc_1$	y_6, b_3, b_5	50%
1324.54	ST*VA	Hex_1 $HexNAc_1$ $NeuAc_2$	y_2, b_3	66%
1352.53	PT*ST	Hex_1 $HexNAc_1$ $NeuAc_2$	y_3, b_2, b_3	100%
1425.57	ST*VAT	Hex_1 $HexNAc_1$ $NeuAc_2$	y_4, b_3, b_4	75%
1445.59	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_1$	y_6, b_4, b_5, b_6	57%
1453.56	EST*VA	Hex_1 $HexNAc_1$ $NeuAc_2$	y_4	25%
1603.65	APS*AVPD	Hex_1 $HexNAc_1$ $NeuAc_2$	y_6, b_3, b_5	50%
1736.67	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_2$	y_6, b_4, b_5, b_6	57%
1841.72	GPT*PSA	Hex_2 $HexNAc_2$ $NeuAc_2$	b_3, y_3	20%
2101.82	GEPT*ST*PT	$(Hex_1$ $HexNAc_1$ $NeuAc)_2$	y_6, b_4, b_5, b_6	57%

[a] Asterisks appear after glycosylated residues.

Table 2. Sodium Coordinated O-Linked Glycopeptide Ions Subjected to IRMPD[a]

m/*z*	peptide	glycan	peptide cleavages	seq. cov.
1346.50	ST*VA	Hex_1 $HexNAc_1$ $NeuAc_2$		0%
1374.50	PT*ST	Hex_1 $HexNAc_1$ $NeuAc_2$		0%
1467.57	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_1$	y_6	14%
1625.63	APS*AVPD	Hex_1 $HexNAc_1$ $NeuAc_2$		0%
1758.67	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_2$	y_6	14%
1863.70	GPT*PSA	Hex_2 $HexNAc_2$ $NeuAc_2$		0%

[a] Asterisks appear after glycosylated residues.

PTST + Hex_1 $HexNAc_1$ $NeuAc_2$ demonstrated similar fragmentation, with an abundant loss of both NeuAc residues, *m*/*z* 792.31 (Figure 2B). The intact sodium coordinated peptide was also observed at *m*/*z* 427.18.

The fragmentation of sodium coordinated GEPTSTPT + Hex_1 $HexNAc_1$ $NeuAc_2$ was found to closely mimic the behavior of the other sodiated glycopeptides analyzed, with the loss of the two sialic residues yielding the most abundant fragment (Figure 3B). The y_6 peptide cleavage was also observed in multiple product ions (i.e., in combination with various glycan losses) due to the labile cleavage of the peptide between the glutamic acid and proline residues. The intact peptide was observed at *m*/*z* 811.35, as was the y_6 ion at *m*/*z* 625.29. The IRMPD mass spectrum for the sodium coordinated glycopeptide ion GPTPSA + Hex_2 $HexNAc_2$ $NeuAc_2$ showed dissociation exclusively along glycosidic bonds, with the desialylated glycopeptide, *m*/*z* 1281.51, being the most abundant fragment ion (Figure 4B). Since this glycopeptide contained the larger core 2 type glycan, it was predicted that the glycan moiety or some part thereof would be a favored product ion due to the increased ability of a large glycan to coordinate with sodium. In practice, this glycopeptide behaved in a parallel manner to the other O-linked glycopeptides discussed, as well as the additional examples listed in Table 2. In general, the sodium coordinated O-glycopeptide ions fragmented almost exclusively along glycosidic bonds, with the desialylated glycopeptide being the most abundant fragment. Sodiated peptide moieties were present to a lesser extent; hence, no additional information was gained by dissociation of sodiated precursor ions as compared to the protonated precursor ions.

Although the fragmentation behavior of the sodium coordinated O-linked glycopeptides was distinct from that of the protonated forms, the difference in dissociation character was not as dramatic as that recently observed by Lebrilla and coworkers for N-linked glycopeptide ions.[10] In that work, the N-linked chitobiose core cleavage was the dominant fragment in both ionic forms; however, in the protonated form, the peptide + GlcNAc was observed while the glycan − GlcNAc was observed in the sodium coordinated form. Unlike the N-linked glycopeptides, the sodium coordinated O-linked glycopeptides investigated in this study did not exhibit a large number of glycan only fragments. This difference is proposed to be a result of two effects. First, O-glycopeptides lack the labile chitobiose core, which easily divides the molecule in the N-linked case. Second, the smaller O-glycans appear less capable to coordinate solely with the smaller glycan moiety. Thus, part of the peptide moiety is required for stable multidentate coordination of sodium. As a result, facile separation of the glycan and peptide moieties is impeded.

Deprotonated O-Linked Glycopeptides. Upon investigation of the deprotonated glycopeptide ions, it was observed that this ionic form dissociated more readily, requiring on average 75% of the irradiation time when compared to both the protonated and sodiated forms. When subjected to IRMPD, the deprotonated glycopeptide STVA + Hex_1 $HexNAc_1$ $NeuAc_2$ was found to dissociate exclusively along glycosidic bonds, with the desialylated glycopeptide as the most abundant fragment at *m*/*z* 740.32 (Figure 1C). Deprotonated NeuAc, *m*/*z* 290.09, was also observed in addition to other glycan only fragments at *m*/*z* 493.17 and *m*/*z* 673.24. The fragments containing NeuAc were anticipated to be more abundant due to their acidic nature and previously observed trends for sialic acid containing oligosaccharides.[34,35] These results indicate that the peptide moiety, with the acidic C-terminus, can effectively compete with sialic acid as the charge carrier in negative mode.

The deprotonated form of PTST + Hex_1 $HexNAc_1$ $NeuAc_2$ replicated the behavior of the previous example, yielding the desailylated glycopeptide, *m*/*z* 768.31, and the glycan moiety fragment ions at *m*/*z* 290.09, *m*/*z* 493.17, and *m*/*z* 655.22 (Figure 2C). Moreover, the presence of two peptide side chain cleavages for threonine and serine add substantial site-specific information to that obtained in the positive ion mode. From all of the peptide containing fragment ions, the loss of 44.03 Da was also observed. In one case, this loss was combined with the loss of 30.01 Da for a total loss of 74.04 Da. These losses have previously been described to dominate deprotonated peptide spectra, and correspond to side chain losses with 44.03 and 30.01 Da corresponding to the loss of C_2H_4O and CH_2O from the side chains of threonine and serine, respectively.[36] Each native, nonglycosylated threonine residue was found to yield an abundant side chain loss (Table 3). Because of the high fidelity of threonine side chain fragmentation, it was possible

Table 3. Deprotonated O-Linked Glycopeptide Ions Subjected to IRMPD[a]

m/z	peptide	glycan	charge carrier	peptide cleavages	seq. cov.	other
721.28	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_1$	$[M-2H]^{2-}$	y_6	14%	−44, −44, −30
866.83	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_2$	$[M-2H]^{2-}$	y_6	14%	−44, −44, −30
1322.51	ST*VA	Hex_1 $HexNAc_1$ $NeuAc_2$	$[M-H]^-$		0%	
1350.51	PT*ST	Hex_1 $HexNAc_1$ $NeuAc_2$	$[M-H]^-$		0%	−44, −30
1443.58	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_1$	$[M-H]^-$	y_6	14%	−44, −44, −30
1734.67	GEPT*STPT	Hex_1 $HexNAc_1$ $NeuAc_2$	$[M-H]^-$	y_6	14%	−44, −44, −30
1839.72	GPT*PSA	Hex_2 $HexNAc_2$ $NeuAc_2$	$[M-H]^-$	b_3, y_5	40%	
2099.80	GEPT*ST*PT	$(Hex_1$ $HexNAc_1$ $NeuAc_1)_2$	$[M-H]^-$	y_6	14%	−44, −30

[a] Asterisks appear after glycosylated residues. Informative side chain losses are listed under "other," with −44 indicating loss of C_2H_4O from threonine and −30 corresponding to the loss of CH_2O from serine.

to both identify the number of nonglycosylated threonine residues and hence localize the site of glycosylation to either serine or threonine. For example, PTST + Hex_1 $HexNAc_1$ $NeuAc_2$ yielded only one threonine side chain loss, thus providing evidence for glycosylation at one of the threonine residues but not the serine residue. The information gained from the presence or absence of the side chain cleavage of serine could also be used to confirm the assignment. The knowledge that the glycan occupied a threonine residue combined with the information obtained from the positive mode spectra allowed the site of glycosylation to be conclusively assigned to the threonine C-terminal to proline.

The lack of threonine side chain loss for the glycopeptide STVA + Hex_1 $HexNAc_1$ $NeuAc_2$ was also used to assign the glycan to the threonine residue (Figure 1C). The utility of side chain losses was further illustrated upon investigation of the doubly deprotonated glycopeptide GEPTSTPT + Hex_1 $HexNAc_1$ $NeuAc_2$ (Figure 3C). Two threonine side chain losses were clearly present and were observed in multiple series together with a single threonine side chain loss. Once again, this information localized the site of glycosylation to a threonine residue, and when combined with information from the protonated spectra (Figure 3A), the site was unambiguously assigned. The majority of the fragment ions including the side chain losses were observed in the charge reduced form due to the loss of the labile sialic acid residues. The y_6 ion corresponding to the loss of GE was observed as it was in both the protonated and sodium coordinated spectra. The core 2 type glycopeptide GPTPSA + Hex_2 $HexNAc_2$ $NeuAc_2$ was also investigated as a deprotonated ion (Figure 4C). This glycopeptide also fragmented mainly along glycosidic bonds with the majority of the fragment ions containing the intact peptide moiety. Some peptide fragments ions were observed including the loss of glycine and deprotonated PT at *m/z* 197.06 both of which were due to cleavages N-terminal to proline.[37]

N-Linked Glycopeptides Containing NeuAc. For comparison, an N-linked glycopeptide containing multiple sialic acid residues was investigated. The doubly charged protonated and sodium coordinated N-linked glycopeptide SNGS + Hex_6 $HexNAc_5$ $NeuAc_4$ derived from BF was subjected to IRMPD (Figure 5A). From the doubly charged precursor, the loss of four sialic acid residues was observed sequentially at *m/z* 1624.07, *m/z* 1478.52, *m/z* 1332.97, and *m/z* 1187.42. After the loss of these labile residues, the rest of the glycan was cleaved in a charge reduced state. The larger glycan fragment ions were sodium coordinated, while the peptide containing fragments were mainly protonated. The most abundant fragment ion at *m/z* 915.30 was the result of chitobiose core cleavage combined with the loss of glycan antennae. The complement to this cleavage, the peptide moiety plus HexNAc, was also observed

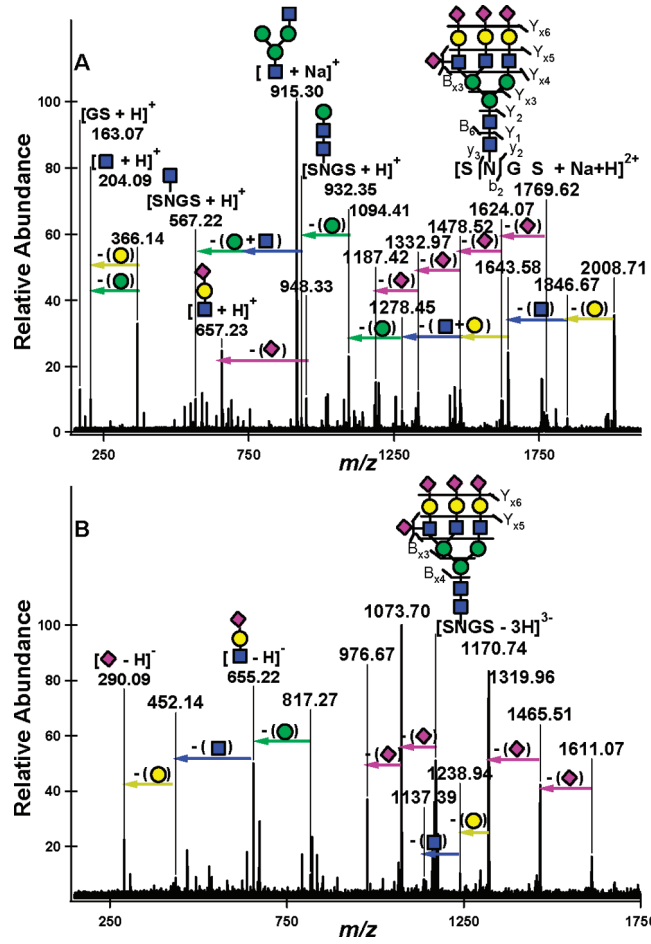

Figure 5. IRMPD nano-ESI FT-ICR mass spectra of the N-glycopeptide ^{175}SNGS^{178} + Hex_6 $HexNAc_5$ $NeuAc_4$ derived from BF with multiple charge carriers: (A) doubly charged, protonated and sodium coordinated glycopeptide *m/z* 1769.62 (20 scans); (B) triply charged, triply deprotonated glycopeptide *m/z* 1170.74 (20 scans).

except with a proton as the charge carrier. This further demonstrates the lability of the chitobiose core and the preferred charge carriers for both the glycan and peptide moieties.

The triply deprotonated IRMPD mass spectra for this same glycopeptide is presented in Figure 5B. The fragment ions corresponding to the loss of one and two sialic acid residues were observed in the triply charged state at *m/z* 1073.70 and 976.67 and in the doubly charged state at *m/z* 1611.07 and 1455.51. A few fragment ions resulting from other glycan losses were observed in the doubly charged state at *m/z* 1319.96, *m/z*

1238.94, and m/z 1137.39. The neutral losses of 30.01 and 60.02 Da were observed, which arise from two sequential serine side chain losses. These side chain losses complement the assignment of the N-glycopeptide; however, they do not assist in the assignment of the glycosylation site since N-glycans only occur at specific glycosylation motifs which do not involve the modification of serine or threonine side chains. The only other fragment ion series observed corresponded to the deprotonated intact antennae down to the deprotonated NeuAc residue. These results further indicate that N-linked and O-linked glycopeptides fragment with distinct behaviors, and that this distinction is not solely due to the presence of labile and anionic sialic acid residues on the O-glycopeptides.

Conclusions

The characterization of O-linked glycosylation sites has lagged behind the characterization of N-linked glycosites due to the increased complexity inherent to O-linked glycosylation. The lack of a consensus sequence and the large number of serine and threonine residues present in O-glycosylated regions of proteins contribute a complicated problem. The utilization of specific enzymes such as trypsin does not alleviate these complications due the large tryptic peptides produced in these regions (greater than 40 amino acid residues is not uncommon). Such large glycopeptides with so many potential glycosylation sites necessitates almost complete peptide sequence coverage to identify glycosylation sites.

The use of nonspecific proteolysis combined with low energy tandem mass spectrometry has been illustrated to simplify this complex problem. Nonspecific proteolysis affords the tunability of peptide footprints around each glycosylation site, thus, allowing for the creation of smaller and more manageable glycopeptides with reduced numbers of potential glycosylation sites. IRMPD of the deprotonated form of O-glycopeptides has been shown to produce informative side chain cleavages from unoccupied serine and threonine residues, allowing determination of whether a serine or threonine residue is glycosylated. Uniting this information with the sequence coverage produced via IRMPD of the protonated glycopeptide leads to unambiguous site identification. These dissociation chemistries inherent to O-glycopeptide ions have also been observed via CID and should be transferable for use on an LC time scale.

Acknowledgment. The authors would like to acknowledge the following funding sources: California Dairy Research Foundation (06 LEC-01-NH), Dairy Management Incorporated, University of California Discovery Grant (05GEB01NHB), and the National Institutes of Health (GM 49077).

References

(1) Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002**, *12* (4), 43R–56R.
(2) Walsh, G.; Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **2006**, *24* (10), 1241–1252.
(3) Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **2007**, *25*, 21–50.
(4) Varki, A. Biological roles of oligosaccharides—all of the theories are correct. *Glycobiology* **1993**, *3* (2), 97–130.
(5) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. Glycosylation and the immune system. *Science* **2001**, *291* (5512), 2370–2376.
(6) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *J. Proteome Res.* **2006**, *5* (7), 1792–1802.
(7) Wang, Y. H.; Wu, S. L.; Hancock, W. S. Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry. *Glycobiology* **2006**, *16* (6), 514–523.
(8) Nakano, M.; Nakagawa, T.; Ito, T.; Kitada, T.; Hijioka, T.; Kasahara, A.; Tajiri, M.; Wada, Y.; Taniguchi, N.; Miyoshi, E. Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *Int. J. Cancer* **2008**, *122* (10), 2301–2309.
(9) Wada, Y.; Azadi, P.; Costello, C. E.; Dell, A.; Dwek, R. A.; Geyer, H.; Geyer, R.; Kakehi, K.; Karlsson, N. G.; Kato, K.; Kawasaki, N.; Khoo, K. H.; Kim, S.; Kondo, A.; Lattova, E.; Mechref, Y.; Miyoshi, E.; Nakamura, K.; Narimatsu, H.; Novotny, M. V.; Packer, N. H.; Perreault, H.; Peter-Katalinic, J.; Pohlentz, G.; Reinhold, V. N.; Rudd, P. M.; Suzuki, A.; Taniguchi, N. Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **2007**, *17* (4), 411–422.
(10) Seipert, R. R.; Dodds, E. D.; Clowers, B. H.; Beecroft, S. M.; German, J. B.; Lebrilla, C. B. Factors that influence fragmentation behavior of N-linked glycopeptide ions. *Anal. Chem.* **2008**, *80* (10), 3684–3692.
(11) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* **2006**, *6* (14), 3993–4015.
(12) Wilson, I. B. H.; Gavel, Y.; Vonheijne, G. Amino-acid distributions around o-linked glycosylation sites. *Biochem. J.* **1991**, *275*, 529–534.
(13) Settineri, C. A.; Medzihradszky, K. F.; Masiarz, F. R.; Burlingame, A. L.; Chu, C.; Georgenascimento, 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.* **1990**, *19* (11), 665–676.
(14) Renfrow, M. B.; Mackay, C. L.; Chalmers, M. J.; Julian, B. A.; Mestecky, J.; Kilian, M.; Poulsen, K.; Emmett, M. R.; Marshall, A. G.; Novak, J. Analysis of O-glycan heterogeneity in IgA1 myeloma proteins by Fourier transform ion cyclotron resonance mass spectrometry: implications for IgA nephropathy. *Anal. Bioanal. Chem.* **2007**, *389* (5), 1397–1407.
(15) Pouria, S.; Corran, P. H.; Smith, A. C.; Smith, H. W.; Hendry, B. M.; Challacombe, S. J.; Tarelli, E. Glycoform composition profiling of O-glycopeptides derived from human serum IgA1 by matrix-assisted laser desorption ionization-time of flight-mass spectrometry. *Anal. Biochem.* **2004**, *330* (2), 257–263.
(16) Peterman, S. M.; Mulholland, J. J. A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (2), 168–179.
(17) Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B. Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and Fourier transform ion cyclotron resonance mass spectrometry. *J. Proteome Res.* **2007**, *6* (10), 4032–4040.
(18) Ren, J. M.; Rejtar, T.; Li, L.; Karger, B. L. N-Glycan structure annotation of glycopeptides using a linearized glycan structure database (GlyDB). *J. Proteome Res.* **2007**, *6* (8), 3162–73.
(19) Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. Automated N-glycopeptide identification using a combination of single- and tandem-MS. *J. Proteome Res.* **2007**, *6* (10), 3995–4005.
(20) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr., B* **2007**, *849* (1−2), 115–128.
(21) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal. Chem.* **2001**, *73* (18), 4530–4536.
(22) Wu, S. L.; Huehmer, A. F. R.; Hao, Z. Q.; Karger, B. L. On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J. Proteome Res.* **2007**, *6* (11), 4230–4244.

(23) Medzihradszky, K. F.; Gillececastro, B. L.; Settineri, C. A.; Townsend, R. R.; Masiarz, F. R.; Burlingame, A. L. Structure determination of O-linked glycopeptides by tandem mass-spectrometry. *Biomed. Environ. Mass Spectrom.* **1990**, *19* (12), 777–781.
(24) Goletz, S.; Leuck, M.; Franke, P.; Karsten, U. Structure analysis of acetylated and non-acetylated O-linked MUC1-glycopeptides by post-source decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11* (13), 1387–1398.
(25) Deguchi, K.; Ito, H.; Baba, T.; Hirabayashi, A.; Nakagawa, H.; Fumoto, M.; Hinou, H.; Nishimura, S. I. Structural analysis of O-glycopeptides employing negative- and positive-ion multi-stage mass spectra obtained by collision-induced and electron-capture dissociations in linear ion trap time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21* (5), 691–698.
(26) Mormann, M.; Paulsen, H.; Peter-Katalinic, J. Electron capture dissociation of O-glycosylated peptides: radical site-induced fragmentation of glycosidic bonds. *Eur. J. Mass Spectrom.* **2005**, *11* (5), 497–511.
(27) Zaia, J. Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **2004**, *23* (3), 161–227.
(28) Holland, J. W.; Deeth, H. C.; Alewood, P. F. Analysis of O-glycosylation site occupancy in bovine kappa-casein glycoforms separated by two-dimensional gel electrophoresis. *Proteomics* **2005**, *5* (4), 990–1002.
(29) Edge, A. S. B.; Spiro, R. G. Presence of an O-glycosidically linked hexasaccharide in fetuin. *J. Biol. Chem.* **1987**, *262* (33), 16135–16141.
(30) Green, E. D.; Adelt, G.; Baenziger, J. U.; Wilson, S.; Vanhalbeek, H. The asparagine-linked oligosaccharides on bovine fetuin—structural-analysis of N-glycanase-released oligosaccharides by 500-Megahertz H-1-NMR spectroscopy. *J. Biol. Chem.* **1988**, *263* (34), 18253–18268.
(31) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.
(32) Domon, B.; Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in Fab-MS MS spectra of glycoconjugates. *Glycoconjugate J.* **1988**, *5* (4), 397–409.
(33) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24* (4), 508–548.
(34) Sagi, D.; Peter-Katalinic, J.; Conradt, H. S.; Nimtz, M. Sequencing of tri- and tetraantennary N-glycans containing sialic acid by negative mode ESI QTOF tandem MS. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (9), 1138–1148.
(35) Harvey, D. J.; Martin, R. L.; Jackson, K. A.; Sutton, C. W. Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 2997–3007.
(36) Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)(-) parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. *Mass Spectrom. Rev.* **2002**, *21* (2), 87–107.
(37) Harrison, A. G.; Young, A. B. Fragmentation reactions of deprotonated peptides containing proline. The proline effect. *J. Mass Spectrom.* **2005**, *40* (9), 1173–1186.

PR8007072