Technical Note

Identification of N-Linked Glycosylation Sites Using Glycoprotein Digestion with Pronase Prior to MALDI Tandem Time-of-Flight Mass Spectrometry

Ying Qing Yu, Jennifer Fournier, Martin Gilar, and John C. Gebler

Anal. Chem., **2007**, 79 (4), 1731-1738 • DOI: 10.1021/ac0616052

Anal. Chem. **2007,** *79,* 1731−1738

Identification of N-Linked Glycosylation Sites Using Glycoprotein Digestion with Pronase Prior to MALDI Tandem Time-of-Flight Mass Spectrometry

Ying Qing Yu,* Jennifer Fournier, Martin Gilar, and John C. Gebler

Life Sciences R&D, Waters Corporation, Milford, Massachusetts 01757

Glycopeptides are typically prepared by cleaving the proteins with specific proteolytic enzymes, such as trypsin. The resulting glycopeptides tend to have weak mass spectrometry ion signals (ESI or MALDI) due to their relatively large molecular weight. The identification of glycosylation sites with tandem mass spectrometry is further complicated by fragmentation of both the peptide backbone and the glycan moiety. We explored a method using a nonspecific enzyme, pronase, to generate small glycopeptides (between two and six amino acids). These glycopeptides were enriched and desalted using a microscale hydrophilic interaction chromatography extraction device prior to MALDI QTof MS analysis. MALDI matrix, 2, 5-dihydroxybenzoic acid, doped with ammonium triscitrate, was utilized for analysis. Sodiated ions were observed as minor ions, while protonated ions were enhanced dramatically with this matrix. Collision-induced dissociation was performed on both the protonated and sodiated ions. MS/MS fragmentation spectra reveal that proton has greater affinity for the peptide moiety, while the sodium cation tends to associate with the sugar moiety. Characteristic fragment patterns allowed for identifications of glycosylation sites for both the protonated and the sodiated precursor ions. Model proteins, horseradish peroxidase and α1-acid glycoproteins, were analyzed to illustrate the identification of N-linked glycosylation sites and data interpretation algorithm.

Several mass spectrometry-based strategies have been developed recently to facilitate glycoprotein characterization using both liquid chromatography−mass spectrometry (LC−MS) and matrix-assisted laser desorpotion/ionization (MALDI) MS.[1−6] The sample preparation step typically involves the use of a specific protease to enzymatically digest glycoproteins into glycopeptides. In order to improve the selectivity and sensitivity of MS analysis, the glycopeptides are often selectively enriched with affinity chromatography sorbent, such as immobilized lectin or antibody chromatography.[7,8] Tandem mass spectrometry analysis (MS/MS) on the glycopeptides is typically employed to reveal the glycosylation sites.[9,10]

Trypsin, a highly specific protease, is commonly used for characterization of proteins. The disadvantage of tryptic digestion of glycoproteins is that the glycopeptides of interest are often large in size (mass contribution of glycans and amino acid sequence). Therefore, it is difficult to acquire MS/MS fragmentation spectra of sufficient quality for unambiguous identification of glycosylation sites.[11,12] Data interpretation is further complicated by the simultaneous fragmentation of peptide bonds and more labile glycosylic bonds.

Recently, an alternative method has been proposed for analysis of glycosylation sites, applying a nonspecific enzyme pronase.[13,14] Pronase hydrolyzes all peptide bonds, but its cleavage efficiency is hindered by the proximity of bulky glycan groups. Therefore, the resulting pronase-generated glycopeptides are two to six amino acids in length, which can be easily analyzed by MALDI or ESI MS.[13,14]

Glycosylation site assignment on N-linked glycoproteins has been achieved using Fourier transform ion cyclotron eesonance MS with high mass accuracy.[13] The molecular ion mass of the glycopeptides and the glycans released by pronase digestion were determined separately. The backbone peptide mass was calculated by subtracting the mass of N-linked glycopeptides and corresponding glycans. Since the pronase-generated peptides have short chains, only the limited number of amino acid combinations matches the known exact mass. Therefore, the interpretation of MS data is rather simple.[13]

* To whom correspondence should be addressed. Tel.: (508)482-2650. Fax (508)482-3625. E-mail: ying_qing_yu@waters.com.

(1) Ethier, M.; Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2003**, *19*, 721−727.
(2) Yang, Y.; Bergmann, C.; Benen, J.; Orlando, R. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1257−1262.
(3) Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, J.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2020−2030.
(4) Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P. *J. Proteome Res.* **2004**, *3*, 556−566.
(5) Albach, C.; Damoc, E.; Denzinger, T.; Schachner, M.; Przybylski, M.; Schmitz, B. *Anal. Bioanal. Chem.* **2004**, *378*, 1129−1135.
(6) Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Sorensen, E. S.; Zubarev, R. A. *Anal. Chem.* **2003**, *75*, 2355−2361.
(7) Kumamoto, Y.; Higashi, N.; Denda-Nagai, K.; Tsuiji, M.; Sato, K.; Crocker, P. R.; Irimura, T. *J. Proteome Res.* **2004**, *3*, 841−850.
(8) Ghosh, D.; Krokhin, O.; Antonovici, M.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. *J. Proteome Res.* **2004**, *4*, 841−850.
(9) Harazono, A.; Kawasaki, N.; Kawanishi, T.; Hayakawa, T. *Glycobiology* **2005**, *5*, 447−462.
(10) Wada, Y.; Tajiri, M.; Yoshida, S. *Anal. Chem.* **2004**, *76*, 6560−6565.
(11) Wuhrer, M.; Koeleman, C. A.; Hokke, C. H.; Deelder, A. M. *Anal. Chem.* **2005**, *77*, 886−894.
(12) Larsen, M. R.; Hojrup, P.; Roepsorff, P. *Mol. Cell. Proteomics* **2005**, *4*, 107−119.
(13) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **2003**, *75*, 5628−5637.
(14) Juhasz, P.; Martin, S. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169/170*, 217−230.

10.1021/ac0616052 CCC: $37.00
Published on Web 01/23/2007

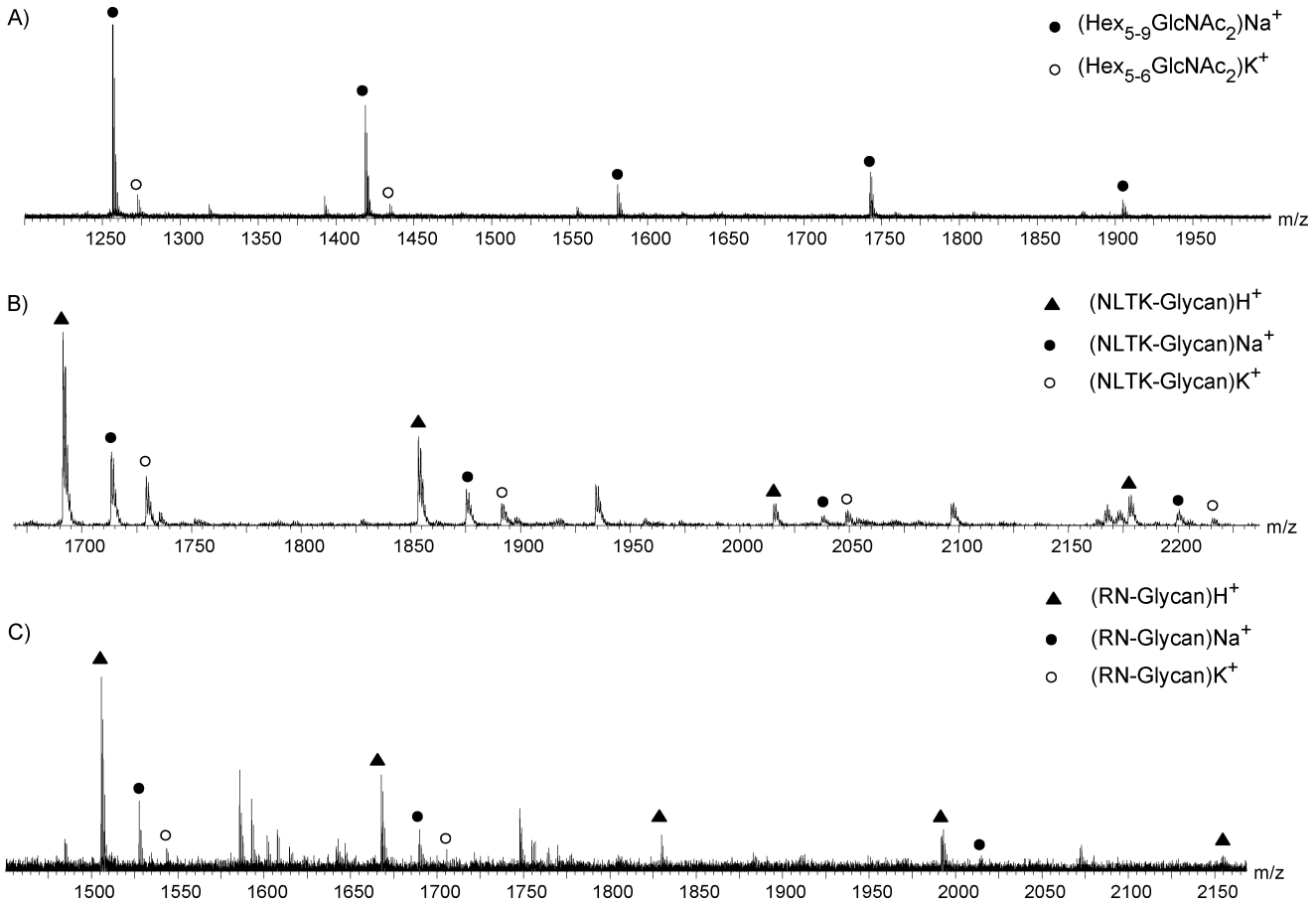


Figure 1. (A) RNase B glycans removed using PNGase F, (B) glycopeptides generated from trypsin digestion of RNase B, and (C) glycopeptides generated from pronase digestion of RNase B analyzed by MALDI QTof MS using DHB matrix.

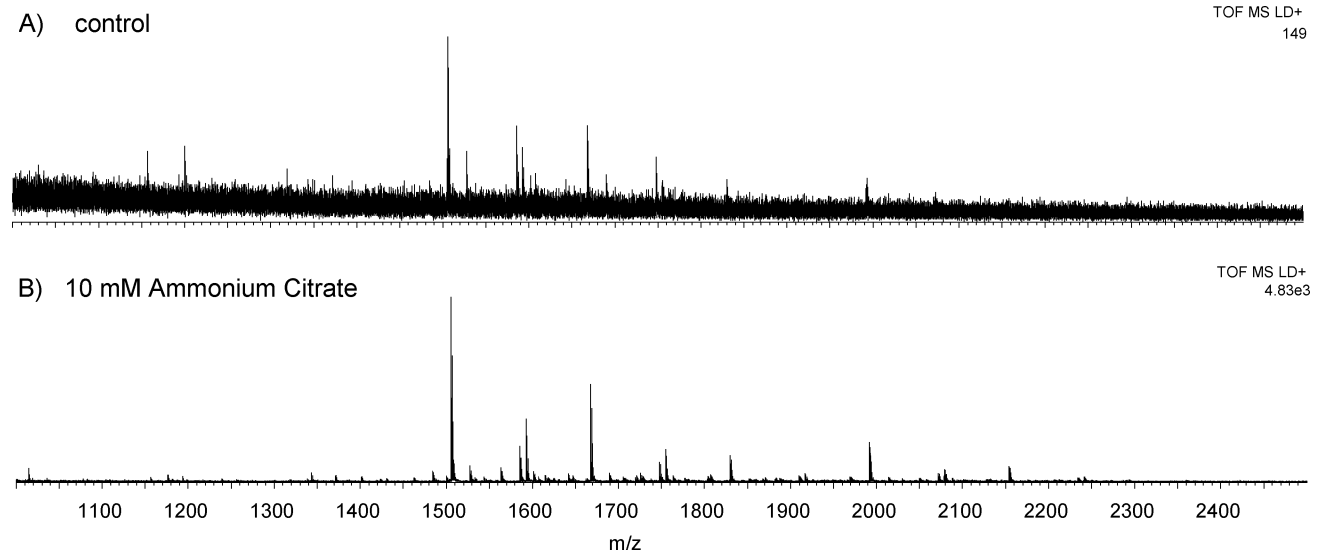


Figure 2. MALDI QTof MS analysis of glycopeptides generated from HRP. (A) The DHB matrix was solubilized in 100% ethanol without any additives. (B) DHB matrix contains 10 mM ammonium citrate.

Wuhrer at al. also utilized pronase to produce glycopeptides. The sample cleanup step was performed by normal-phase nanoLC separation, and the assignment of glycosylation sites was accomplished by multistage MS fragmentations.[11]

In this work, we describe an improved, simple method for assignment of N-linked glycosylation sites of glycoproteins using pronase digestion. The sample cleanup is based on microscale hydrophilic interaction chromatography solid-phase extraction (HILIC SPE), which desalts and isolates the glycopeptides.[15] Enriched glycopeptides are subsequently analyzed by MALDI QTof tandem MS analysis. An algorithm for N-linked glycosylation sites assignment is proposed.

(15) Yu, Y. Q.; Gilar, M.; Kaska, J.; Gebler, J. C. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1−6.

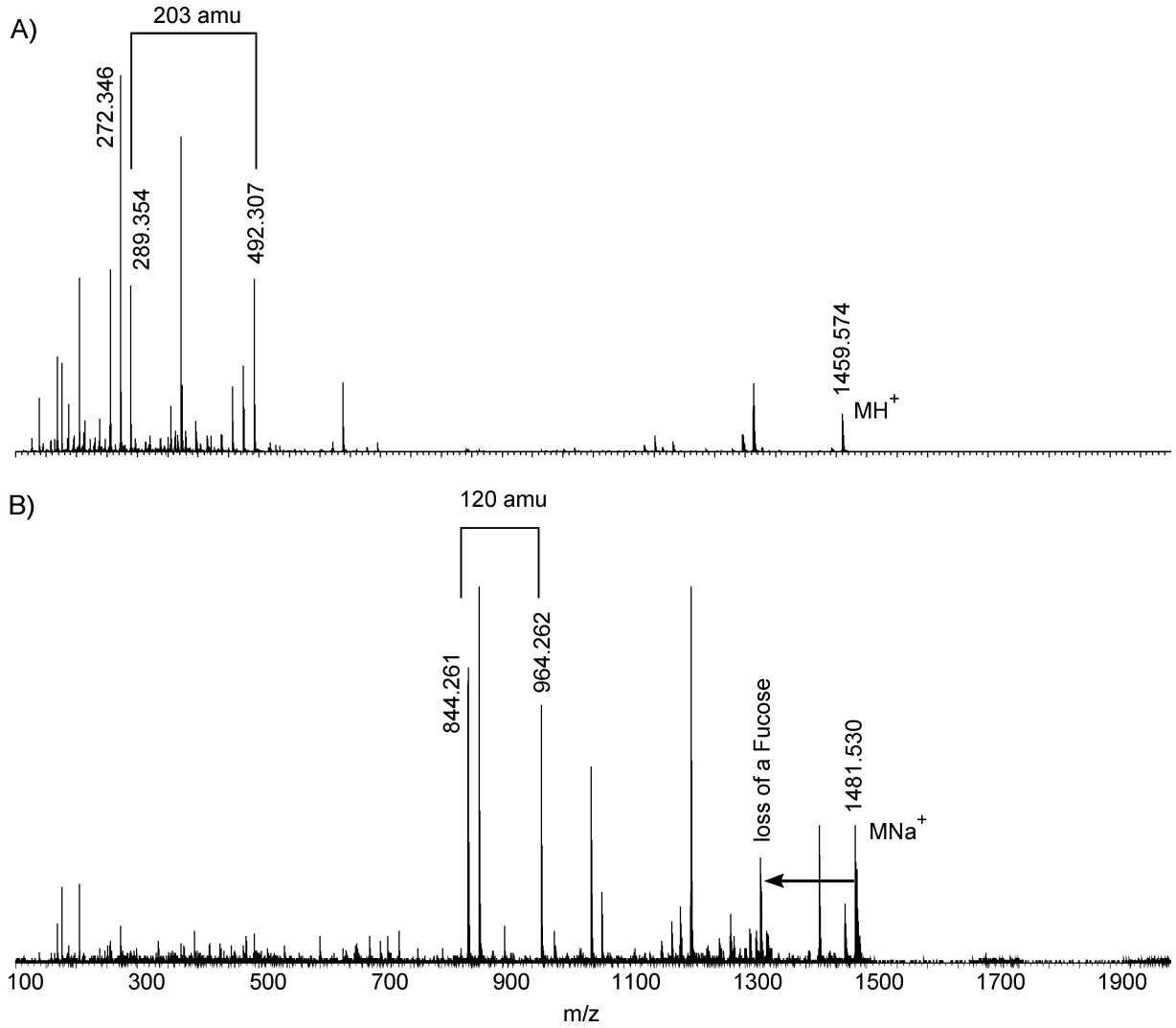


Figure 3. MS/MS fragmentation of a HRP glycopeptide. (A) protonated precursor ion; (B) sodiated precursor ion. The characteristic pairs of fragments are labeled.

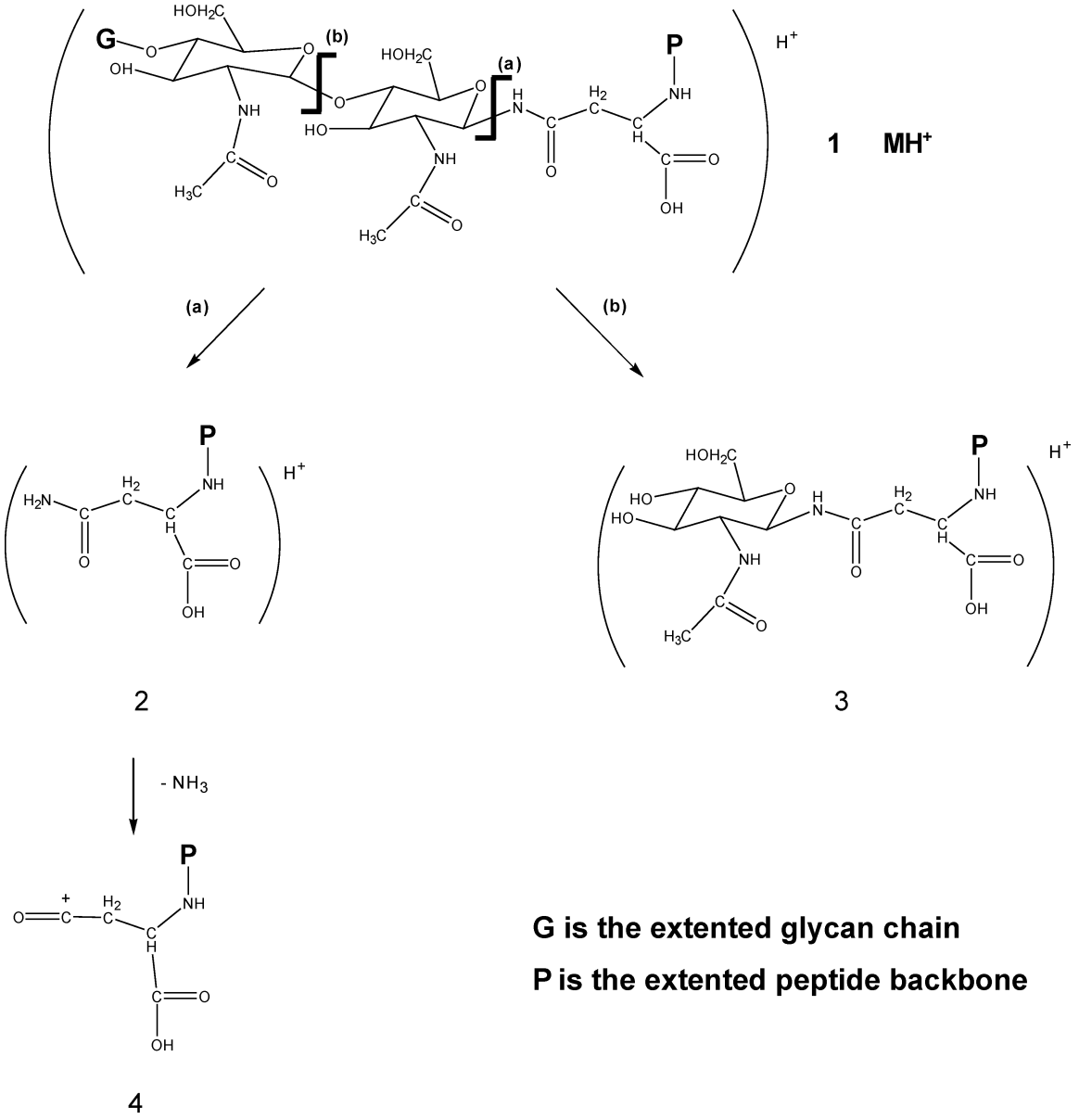
EXPERIMENTAL SECTION

Materials and Reagents. Glycoproteins, ribonuclease B (RNase B), α_1-acid glycoprotein (AGP), horseradish peroxidase (HRP), pronase, and α_2(3,6,8,9)-neuraminidase were purchased from Sigma (St. Louis, MO) and used without further purification. Sequencing grade trypsin was supplied by Promega. The detergent used to solubilize proteins, RapiGest SF, was obtained from Waters Corp. (Milford, MA). Microscale HILIC SPE and MALDI matrix 2,5-dihydroxybenzoic acid (DHB) were also supplied by Waters Corp. under the trade name of MassPREP. All other chemicals were purchased from Sigma, unless specified otherwise.

Pronase Digestion. Glycoproteins (0.5−2 mg) were solubilized in 0.1% (w/v) RapiGest solution in 50 mM ammonium bicarbonate. Proteins were reduced with 10 mM dithiothretol and subsequently alkylated using 15 mM iodoacetamide. The pronase to protein mass was 1:50 (w/w). Enzymatic digestion was performed at 37 °C overnight. In the case of AGP, the sialic acid residues were removed by addition of 1 unit of α_2(3,6,8,9)-neuraminidase after the pronase digestion was accomplished (the additional incubation time was 12 h). Glycans from RNase B were released using peptide N-glycosidase (PNGase F). PNGase F was added to the intact protein in a 1:50 mass ratio, and the protein−enzyme was incubated at 37 °C overnight. Tryptic digestion was also performed using the same conditions as the pronase digestion. Portions of samples were used for HILIC SPE extractions; the rest was stored at −20 °C.

HILIC SPE. Prior to the MALDI MS analysis, the glycopeptides were desalted using HILIC SPE. If not removed, salts and detergent severely interfere with MALDI analysis and no useful data are obtained. A detailed HILIC SPE protocol was published elsewhere.[15] Briefly, a 96-well microelution plate packed with 5 mg of sorbent per well was conditioned with 200 μL of water followed by the same volume of 90% acetonitrile (MeCN). The glycopeptide samples (10 μg) were reconstituted in 200 μL of 90% MeCN and loaded onto the SPE plate. The wells were washed with 200 μL of 90% MeCN, and eluted with 25−50 μL of 25% MeCN or 10 mM ammonium citrate in 25% MeCN. The SPE plate was operated using a vacuum manifold. Short pronase-generated glycopeptides were recovered with a high yield. Some loss was observed when applying HILIC SPE for desalting of tryptic (longer) glycopeptides.

Scheme 1. Proposed Fragmentation Pattern of Protonated MH^+ Glycopeptide Ions



MALDI-QTof MS Analysis. A Micromass MALDI-QTof Ultima (Waters) was equipped with a fixed nitrogen laser source. The laser was operated at 10 Hz. Each spectrum was collected over 2-s scans, and spectra were accumulated over 2 min. Argon gas was used as the collision gas. The typical collision energy used was in the range of 70−120 V depending on the size of the analyte ions. The instrument was controlled by MassLynx 4.0 software. FindPept tool (SwissProt web site) was used to identify the possible peptide sequences by submitting the peptide masses and searching against the primary amino acid sequence.

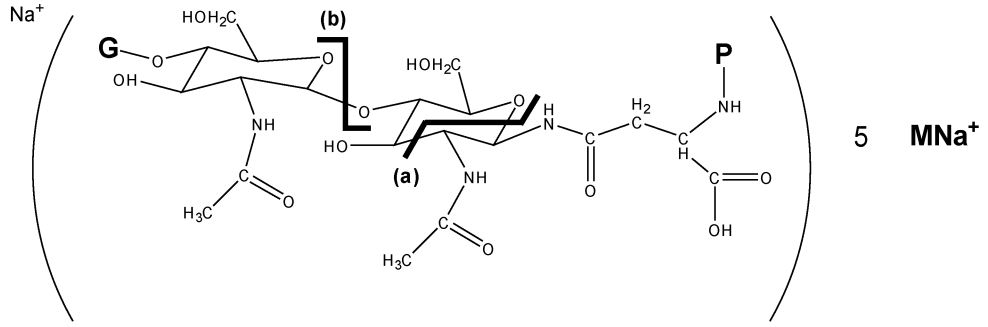
DHB matrix was prepared in ethanol with a final concentration of 20 mg/mL. The DHB solution was tested with several additives, ammonium citrate (10 mM), NaCl (5 mM), or formic acid (0.1%). One microliter of sample (~0.5 μg of protein digest) was mixed with 1 μL of matrix directly on the target and dried at ambient temperature. After the sample was crystallized with matrix, pure ethanol (0.8 μL) was added to the same well to recrystallize the sample.

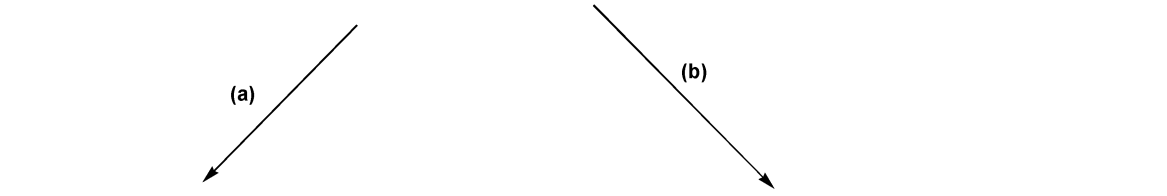
RESULTS AND DISCUSSION

Impact of MALDI Matrix on Glycopeptide Analysis. Pronase is a nonspecific endoprotease that can cleave at every peptide bond. However, due to the steric hindrances, the amino acids adjacent to the glycosylation site(s) are ineffectively cleaved by the enzyme. As a result, a series of short glycopeptides are generated for a given glycosylated site. The uncleaved peptide motifs are usually two to six amino acids long. Because of the hybrid nature of the resulting glycopeptides (part peptide and part glycan), their ionization property differs from glycans or tryptic glycopeptides. Our first goal was to optimize the MALDI signals for the pronase-released glycopeptides, so that sufficiently abundant precursor ions are generated for subsequent MS/MS fragmentation.

The glycans are composed of saccharides that do not have basic functional groups. Therefore, the protonated glycan ions are rarely observed. Instead, the ionization of the glycans is achieved

Scheme 2. Proposed Fragmentation Pattern of Sodiated MNa^+ Glycopeptide Ions





$(G + GlcNAc + C_4H_8O_4)Na^+$

6

$(G + GlcNAc)Na^+$

7

G is the extented glycan chain

P is the extented peptide backbone

via sodium or potassium ion adductation.[16] On the contrary, tryptic glycopeptides containing lysine or arginine tend to form protonated ions.[17] If the glycopeptides generated by pronase digestion contain basic amino acid residues like lysine or arginine, the protonated ions are expected to dominate the MALDI MS spectrum. However, if no basic residues are present in the peptide sequence, both protonated and sodiated ions are likely to be formed.[18,19]

In order to verify the ionization mechanism, we carried out an experiment using RNase B as a test sample. RNase B glycosylation has been well documented; there are five high-mannose oligosaccharides linked to a single Asn_{34} site. We investigated the ionization of three samples: (i) the glycans cleaved from RNase B using PNGase F, (ii) RNase B glycopeptides generated by tryptic cleavage, and (iii) RNase B glycopeptides generated by pronase digestion. All three samples were desalted using the HILIC SPE method prior to MALDI QTOF MS analysis. Importantly, SPE sample treatment removes other peptides (that would otherwise complicate MALDI analysis), and only glycans and glycopeptides were recovered. The samples were analyzed using 20 mg/mL DHB matrix dissolved in pure ethanol (no ammonium citrate was added).

As expected, the MALDI MS spectra of the glycans show exclusively sodiated adduct ions (Figure 1A). Traces of potassium adducts were also detected. MALDI MS analysis of tryptic glycopeptides shows both protonated and alkali metal adduct ions for each of the glycopeptide variants (Figure 1B). Similarly, the pronase-generated glycopeptides show protonated and sodiated ions (Figure 1C). The protonated ions are dominant in the spectra B and C due to the fact that both trypsin and pronase digested glycopeptides contain basic amino acids K or R. Noticeably, the pronase-digested glycopeptides have the lowest ion intensity using DHB matrix (Figure 1C). Ammonium citrate has been reported to improve the peptide ion signals when added to a matrix solution.[14,20] We observed at least a 10-fold increase of ion intensity when 10 mM ammonium citrate was incorporated into the DHB matrix (Figure 2). Other additives such as 0.1% formic acid were also tested; however, only moderate signal improvement was observed (data not shown.)

We observed that ammonium citrate-doped DHB matrix improved certain glycopeptide signals more than others. The ionization of glycopeptides containing basic amino acids (K or R) was significantly enhanced (MH^+ ions), while the signal of glycopeptides lacking the basic amino acids improved only moderately. Similar signal enhancement was observed for alkali metal adducted ions.

Tandem MALDI-QTof MS/MS Analysis of Glycopeptides. MALDI-QTof was used to perform collision-induced dissociation (CID) of pronase-generated glycopeptides. Intriguingly, distinct fragment ion pairs were repeatedly observed for glycopeptides subjected to CID. The detailed inspection revealed that MS/MS spectra of protonated precursors always contain a distinct pair of

(16) Harvey, D. *Mass Spectrom. Rev.* **1999**, *18*, 349−451.
(17) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399−1406.
(18) Asara, J. M.; Allison, J. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 35−44.
(19) Zhu, X. P.; Papayannopoulos, I. A. *J. Biomol. Tech.* **2003**, *14*, 298−307.
(20) Juhasz, P.; Martin, S. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169/170*, 217−230.

fragments with a mass difference of 203 amu. Similar pairs of fragments (120 amu difference) were observed in the fragment ions generated from sodiated precursors. This is illustrated in Figure 3 using a selected HRP glycopeptide. These characteristic 203 or 120 amu pairs are typically present among the most intense ions that were recognized easily.

As discussed earlier, for protonated precursor ions, the proton is likely attached to the peptide (basic) moiety rather than the sugar moiety. Therefore, the fragmentation is likely caused by proton-induced dissociation, as suggested in Scheme 1.[16] Pathway a in Scheme 1 shows the cleavage between the Asn and the first GlcNAc in the core glycan structure, producing fragment 2; pathway b shows the glycosidic cleavage between the first and the second GlcNAc, producing fragment 3. The mass difference between fragments 2 and 3 is 203 amu, corresponding to a single GlcNAc group. Presumably the fragmentation pathways are thermodynamically preferred and result in the rise of the distinct ion pair (2 and 3) that is labeled in the spectrum in Figure 3A. Further loss of NH_3 from structure 2 is also observed in the spectrum (fragment 4, Scheme 1). The loss of NH_3 becomes a dominant product ion when a basic amino acid residue (R or K) is present in the peptide backbone. The same sequential bond cleavage was observed for tryptic glycopeptides that were protonated.[3,21]

The molecular weight of the peptide can be easily deduced from the MS/MS spectrum by locating the above-described pair of fragments. The smaller fragment ion 2 (Scheme 1) directly represents the mass of the protonated peptide. The accurate mass of peptides can be used to identify the glycosylation site, provided that the glycoprotein primary sequence is known. We submitted the peptide mass to the FindPept program and searched against the glycoprotein sequence. Since the sequence is relatively short, the number of candidate amino acid sequences that includes Asn (N-linked) is often limited to a single hit. For example, a horseradish peroxidase (HRP) peptide (MH^+) at 1459.574 amu (Figure 3A) yielded the peptide fragment mass of 288.35. When searched against the HRP sequence, the only match was NR motif with the Asn_{188} glycosylation site.

MS/MS fragmentation of sodiated ions also generates a distinct pair of fragments (Δ120 amu). The formation of this ion pair is rationalized in Scheme 2. The glycosidic bond cleavage between the first and the second GlcNAc and the cross ring fragmentation at the first GlcNAc generates a pair of ions that differs by 120 amu (ion 6 and 7). The exact structures of the fragment ions were not listed, since they are merely inferred. The peptide molecular weight is derived by subtracting the mass of fragment ion 7 and one GlcNAc (203 Da) from the precursor glycopeptide ion 5 in Scheme 2. If a fucose is also attached with the innermost GlcNAc (Figure 3B), its mass (146 Da) should be subtracted from the precursor ions as well.

The example of a glycopeptide site assignment is shown in Figure 3B. The sodiated precursor with m/z of 1481.53 fragmented into a characteristic pair differing by 120 Da (844.26 and 964.26 Da). A peptide mass was calculated according to rules outlined above (1481.53 − 844.26 − 203.08 (GlcNAc) − 146.06 (fucose) = 288.13); the mass of 288.13 points to the same glycosylation site of $N_{188}R$ as the corresponding MH^+ ion (Figure 3A).

Table 1. List of Pronase-Digested AGP Peptide Residues Derived from the MS/MS Fragment Ion Pair Masses

glycopeptide precursor ion $(MH)^+$ (m/z)	peptide mass (m/z)		peptide-sequence	sequence number
	experiment	actual		
1883.90	260.18	260.15	$\mathbf{N}_{56}$K	56−57
				72−73
2066.90	357.30	357.20	P$\mathbf{N}_{72}$K	71−73
2287.04	304.30	304.20	T$\mathbf{N}_{33}$A	32−34
			$\mathbf{N}_{33}$AT	33−35
2929.38	575.33	575.27	RE$\mathbf{N}_{103}$GT	101−105
2959.47	605.42	605.27	$\mathbf{N}_{72}$KTED	72−76
	605.42	605.32	FTP$\mathbf{N}_{72}$K	69−73

Table 2. List of Pronase-Digested HRP Peptide Residues Derived from the MS/MS Fragment Ion Pair Masses

glycopeptide precursor ion $(MH)^+$ (m/z)	peptide mass (m/z)		peptide sequence	sequence number
	experiment	actual		
1459.54	288.30	288.15	$\mathbf{N}_{188}$R	188−189
1513.55[a]	320.06	320.13	$\mathbf{N}_{298}$ST	298−300
1525.50[a]	332.11	332.12	CP$\mathbf{N}_{43}$	41−43
	332.11	332.17	$\mathbf{N}_{244}$LS	244−246
1527.50[a]	334.11	334.15	$\mathbf{N}_{87}$TT	87−89
			$\mathbf{N}_{228}$TT	228−230
1584.50[a]	391.11	391.17	A$\mathbf{N}_{298}$ST	297−300
1660.59	401.30	401.19	P$\mathbf{N}_{285}$AT	284−287
	401.30	401.24	L$\mathbf{N}_{188}$R	187−189
1629.67	458.30	458.26	GL$\mathbf{N}_{188}$R	186−189
1651.56	488.30	488.22	SP$\mathbf{N}_{285}$AT	283−287
	488.30	488.27	L$\mathbf{N}_{188}$RS	187−190
1810.60	617.25	617.26	P$\mathbf{N}_{285}$ATDT	284−289
1920.76	575.29	575.25	SSP$\mathbf{N}_{285}$AT	282−287
	575.29	575.30	L$\mathbf{N}_{188}$RSS	187−191

[a] The precursor ion is sodiated instead of protonated.

The ion pairs resulting from MS/MS fragmentation of the protonated or the sodiated precursors provide complementary information and mutually validate the mass assignment of glycopeptide amino acid sequence. This is important, since the successful identification of primary amino acid sequence is based solely on this knowledge. The corresponding peptide sequence motif candidates are found via a search against a known parent glycoprotein sequence. Therefore, the method is applicable for a confirmation of known or suspected glycosylation sites of known proteins but not for unknown proteins of unknown sequence.

The reliability of the proposed method was tested using two proteins, HRP (SwissProt P00433) and AGP (SwissProt P02763). Both proteins have multiple glycosylation sites, which may potentially complicate the glycosylation site assignments. The glycopeptides generated from a pronase digestion of each protein were purified by HILIC SPE as described in the Experimental Section. Since sialic acid-containing glycopeptides do not ionize well in positive ion mode, the dominant fragment ions are those with a loss of sialic acid.[16,22] Therefore, the sialic acid in AGP was removed using neuraminase prior to pronase digestion.

The glycopeptide precursor ions with significant ions counts (>20 counts/scan) were subjected to MS/MS fragmentation. The

(21) Bykova, N. V.; Rampitsch, C.; Krokhin, O.; Standing, K. G.; Ens, W. *Anal. Chem.* **2006**, *78*, 1093−1103.

(22) Morelle, W.; Slomianny, M. C.; Diemer, H.; Schaeffer, C.; van Dorsselaer, A.; Michalski, J. C. *Rapid Commun. Mass Spectrom.* **2004**, *22*, 2637−2649.

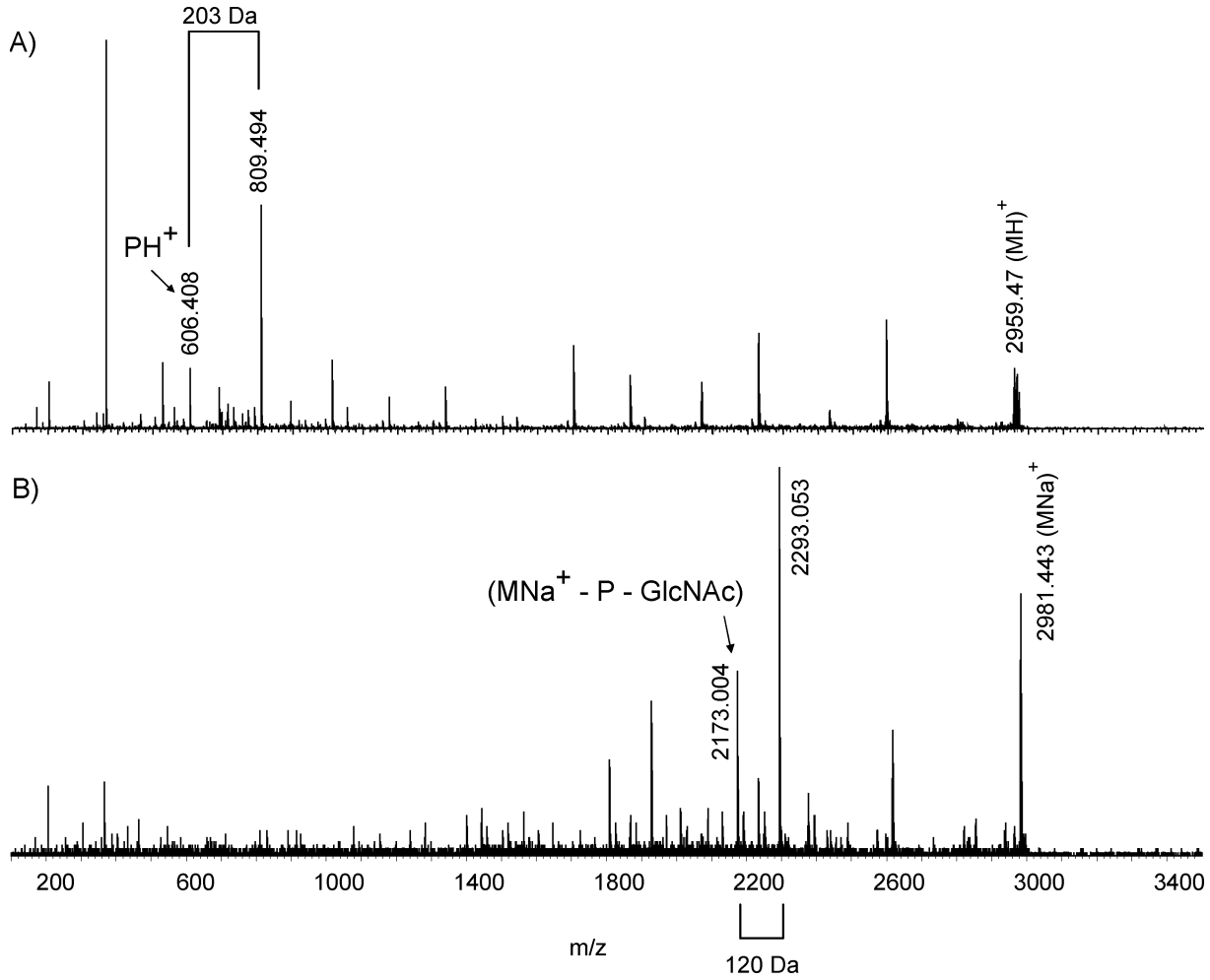


Figure 4. MS/MS spectra of an AGP glycopeptide. (A) protonated precursor ion; (B) sodiated precursor ion. The characteristic pairs of fragments are labeled.

characteristic fragment ion pairs from the protonated or the sodiated ions were located manually and used to calculate the corresponding peptide molecular weight. The accurate mass was submitted to FindPept to identify the candidate amino acid sequences of the glycosylation sites. Occasionally, the search for specific accurate mass returns motifs that do not contain Asn or do not correspond to the known N-linked glycosylation site motif -Asn-X-S/T. The Asn-X-S/T was used as an additional filter; all peptides lacking the motif were treated as false identifications. The results shown in Table 1 and 2 are compiled from the MS/MS analysis of the most intense glycopeptide precursor ions. A total of seven glycosylation sites were identified from HRP and four for AGP.

Tables 1 and 2 list the peptides with N-glycosylation from AGP and HRP. The majority of glycosylation sites were redundantly identified via the multiple glycoforms attached to the same peptide motifs; Tables 1 and 2 show only the unique peptides. Certain glycosylation sites were identified by several peptide variants all related to the same peptide backbone motif. Since the pronase is a nonspecific enzyme, it generates several different glycopeptides centered on the particular site (e.g., the glycosylation site 188 identified by five different peptide fragments in Table 2).

In rare cases, the two N-linked peptides have nearly identical masses. For example, Figure 4A shows a MS/MS fragmentation of a protonated ion with a mass to charge ratio of 2959.47; the Figure 4B spectrum depicts the MS/MS fragmentation of the sodiated version of the same glycopeptide. FindPept finds two possible AGP peptides, NKTED and FTPNK (Table 1). Because the mass difference between the candidate peptides is only 0.05 Da, the mass accuracy of the MALDI Q Tof instrument with external mass calibration was not sufficient to unambiguously assign the correct sequence. However, in this particular case, the two sequences overlap at the same glycosylation site, Asn_{72}.

Because the glycopeptide ion intensity is related to the molar concentration of the respective glycans, the generation of multiple different cleavage products dilutes the signal and potentially undermines the limit of detection. On the other hand, the identification of redundant peptides is very useful for the internal validation of the glycosylation sites.

According to SwissProt, there are five potential glycosylation sites within AGP and eight possible sites for HRP. In our experiment, we detected only four and seven sites for AGP and HRP, respectively. AGP glycopeptides containing Asn_{93} were observed at very low ion intensity in the MALDI QTof spectrum. Interference caused by more abundant neighboring ions during CID prevented a confident assignment of this glycosylation site. The site for HRP that was not detected was Asn_{216}. The unidentified glycosylation sites for AGP and HRP had both -YN- motifs. The significance of this observation is under investigation.

In addition to the proposed method for glycosylation sites identification, the glycan molecular weight can be deduced from the data. However, the glycan structure information cannot be readily obtained from the CID spectra of glycopeptides. The combined fragmentation in peptide and sugar moieties complicates the glycan structure assignment.

CONCLUSIONS

A method for glycosylation site assignment based on pronase digestion of glycoprotein was developed. The pronase-generated glycopeptides are hydrophilic species since they are mostly glycans linked with small piece of peptides; therefore, HILIC SPE were able to retain and clean up these glycopeptides prior to tandem MALDI QTof fragmentation. The characteristic MS/MS fragmentation ion pairs were observed for both protonated and sodiated precursor ions and were utilized for the corresponding peptide mass calculation. The knowledge of peptide accurate mass and the candidate protein primary sequence was sufficient for unambiguous assignment of N-linked glycosylation sites in a single experiment. The known motif of the N-linked glycosylation site (Asn-X-S/T) serves as an efficient filter to eliminate false positives. In some cases, multiple possibilities (assignments) arise due to nearly identical peptide masses, especially for the larger peptides. The analytical approach presented is useful for assigning glycosylation sites for proteins with known primary sequence, such as recombinant proteins, but not for unknown proteins or proteins from complex mixtures.

Received for review August 28, 2006. Accepted December 15, 2006.

AC0616052