analytical chemistry

ARTICLE

pubs.acs.org/ac

Nano-LC–MS/MS of Glycopeptides Produced by Nonspecific Proteolysis Enables Rapid and Extensive Site-Specific Glycosylation Determination

John W. Froehlich,[†] Mariana Barboza,[†] Caroline Chu,[†] Larry A. Lerno, Jr.,[†] Brian H. Clowers,[‡] Angela M. Zivkovic,[§,∥] J. Bruce German,[§,∥] and Carlito B. Lebrilla*[,†,⊥]

[†]Department of Chemistry, University of California, Davis, California 95616, United States

[‡]Pacific Northwest National Laboratory, Richland, Washington 99352, United States

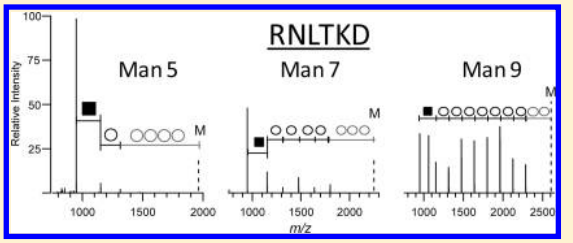
[§]Department of Food Science, University of California, Davis, California 95616, United States

[∥]Foods for Health Institute, University of California, Davis, California 95616, United States

[⊥]Department of Biochemistry and Molecular Medicine, University of California, Davis, California 95616, United States

Supporting Information

ABSTRACT: Given the biological importance of glycosylation on proteins, the identification of protein glycosylation sites is integral to understanding broader biological structure and function. Unfortunately, the determination of the microheterogeneity at the site of glycosylation still remains a significant challenge. Nanoflow liquid chromatography with tandem mass spectrometry provides both separation of glycopeptides and the ability to determine glycan composition and site-specific glycosylation. However, because of the size of glycopeptides, they are not often amenable to tandem MS. In this work, proteins are digested with multiple proteases to produce glycopeptides that are of suitable size for tandem MS analysis. The conditions for collision-induced dissociation are optimized to obtain diagnostic ions that maximize glycan and peptide information. The method is applied to glycoproteins with contrasting glycans and multiple sites of glycosylation and identifies multiple glycan compositions at each individual glycosylation site. This method provides an important improvement in the routine determination of glycan microheterogeneity by mass spectrometry.



Protein glycosylation is a critical post-translational modification that influences many key biochemical and cellular processes including cell–cell signaling, protein stability, and protein function.[1−3] Despite the importance of protein glycosylation to biological processes, relatively few methodologies exist to support large, high-throughput analysis of site-specific glycosylation. Notably, large scale proteomic assays exist to determine the sites of glycosylation in a complex biological mixture after removal of the glycan moiety.[4] While valuable, these assays are incapable of identifying or providing any information regarding the glycans present at each individual glycosylation site (called the glycosylation microheterogeneity) as the glycan moiety is removed before analysis. Site-specific glycosylation analysis requires the analysis of intact glycopeptides or glycoproteins, rather than formerly glycosylated peptides. Glycoproteins often have multiple sites of glycosylation, which greatly complicates analysis of the intact protein. The typical method of reducing this complexity is to enzymatically digest the protein and then analyze the individual sites separately. This method has recently been employed successfully in the analysis of tryptic glycopeptides by Marshall and co-workers,[5] in combination with high resolution, high mass accuracy Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS). One challenge to this approach is that glycosylation sites are not always located close to the cleavage sites of the standard proteolytic enzymes, potentially resulting in glycopeptides that are too large for effective tandem MS. Those preparations that use specific proteases also generally require chemical, chromatographic, or affinity based enrichment of glycopeptides prior to analysis in order to avoid suppression of the glycopeptide ions by the cogenerated, nonglycosylated peptides.[6−10]

Non- or broadly specific proteases have also been utilized for the determination of glycosylation microheterogeneity.[11,12] These proteases digest the amino acid backbone of a glycoprotein to small (<4 amino acid) peptides and amino acids, except in regions where glycosylation is present to inhibit digestion. As peptides are not generated, the enrichment step prior to analysis is unnecessary, reducing the total time for sample preparation and analysis. Proteinase K[12] and Pronase are examples of such proteases that have been used successfully.

Received: February 19, 2011
Accepted: June 10, 2011
Published: June 10, 2011

An alternative strategy has been developed in recent years. This approach utilizes immobilized Pronase, which is a collection of proteases which exhibit highly nonspecific digestion of proteins. This sample preparation has been applied in concert with high mass accuracy, high resolution MALDI and ESI FT-ICR measurements to determine site-specific glycosylation of several standard glycoproteins[13] as well as mixtures of glycoproteins.[14] However, high mass accuracy alone is sometimes insufficient to provide unambiguous glycopeptide identification. Several possible glycopeptide combinations often are within experimental mass tolerances due to glycan heterogeneity and a lack of protease specificity. The approach developed in this study combines these strategies with an automated, microfluidic nano-LC separation which chromatographically resolves glycopeptides, reducing sample complexity and minimizing ion suppression.

Tandem MS of glycopeptides has been successfully employed to identify compositions from mixtures of glycoproteins and single glycoproteins.[15] Previously, infrared multiphoton dissociation, electron capture dissociation, and collision-induced dissociation (CID) have been utilized.[16−26] CID is the most widespread MS/MS technique available, yet glycans and glycopeptides are relatively labile under typical energies, which lessens the quality of the MS/MS spectra attained. For instrumentation with both CID and ECD methods available, complementary structural information may be obtained by utilizing both techniques.[19]

In the current approach, observed masses are assigned to putative glycopeptides based on matching high-mass accuracy precursor measurements to a combinatorial library of possible glycopeptides. The time-of-flight mass analyzer is routinely capable of <10 ppm mass accuracies, which greatly decreases the number of possible glycopeptide combinations matching a given precursor ion. High-accuracy tandem MS (MS/MS) is then employed to confidently identify glycopeptide compositions. This is often a necessary step, because the diversity of glycan compositions potentially present on a glycopeptide require high-quality MS/MS spectra for confident assignment.

In this study, we tune the CID energies for small glycopeptides by analyzing each of three glycoprotein standards under several distinct CID energies. The MS/MS spectra of each of these runs were compared, and the conditions which yielded the most extensive fragment coverage of the glycan residues were noted. An optimal fragmentation energy setting was then developed based on these results, with the goal of maximizing the glycan fragmentation information by LC−MS/MS. Despite the distinct glycan and peptide compositions among the glycoprotein standards, the optimal settings facilitate the acquisition of high-quality MS/MS in each instance. As expected, glycopeptide size, charge, and glycan type all influence the observed fragmentation patterns and each of these were considered in the final optimal conditions. With the use of this approach, over one hundred glycopeptides are identified from a single data-dependent nano-LC−MS/MS analysis, without prior enrichment or fractionation, representing more than a 2-fold improvement as compared to prior results from our laboratory. In this manner, the method developed in this work furthers the robust and high-throughput investigation of the microheterogeneity of protein glycosylation by greatly increasing the number of glycopeptide assignments. This method lays the groundwork for more systematic global glycoproteomic studies involving complex protein mixtures.

EXPERIMENTAL SECTION

Materials. All water used was generated by milli-Q filtration and was measured as 18 MΩ water. Lactoferrin from human milk, ribonuclease B, high-purity grade IV chicken egg albumin, formic acid (HPLC grade), and Pronase E were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile (Honeywell Burdick and Jackson) was purchased from VWR (West Chester, PA). Cyanogen bromide activated 4B sepharose beads (90 μm average diameter) were purchased from GE Healthcare (Piscataway, NJ).

Sample Preparation. Pronase-immobilized beads were prepared in similar fashion as detailed previously.[13] Briefly, 900 mg of beads were weighed and washed with 10 mM HCl with gentle agitation for 30 min. The solution was discarded, and phosphate buffer (pH 7.4) was used to wash the beads in triplicate. Washed beads were then coupled for 4 h with 24 mg of Pronase E. Following coupling, the supernatant was discarded and residual binding sites were blocked with ethanolamine. Pronase beads (PB) were then extensively washed with 25 mM ammonium bicarbonate and stored at 4 °C until digestion.

A total of 60 nmol of human lactoferrin (hLF), chicken egg albumin (CEA), and bovine ribonuclease B (RnB) were added to separate centrifuge tubes for digestion. To each, 300 μL of PB were added, and samples were incubated at 37 °C with gentle mixing. Aliquots were taken at 0.5, 1, 2, 4, 8, and 12 h time points and stored at −20 °C. When the digestion was complete, the separate time points for each protein were pooled for analysis. After concentration by vacuum centrifugation, the equivalent of 0.3 nmol was utilized for each nano-LC−MS/MS analysis. The heterogeneity generated by pooling different digestion times was resolved during the analytical separation.

Mass Spectrometry. An Agilent 6520 QTOF MS system equipped with an autosampler, chip-cube interface, and nano and capillary pump was utilized for all experiments. Samples were loaded onto porous graphitized carbon chips by the capillary pump, operating at 3 μL/min, and briefly flushed to remove residual salts. The gradient was 0%−10% B, 0.0−5.0 min; 10%−55% B, 5.0−45 min; 45%−90% B, 45.0−48.0 min, and held for 5.0 min at 100% B with an equilibration time of 7.0 min at 0% B. Channels A and B were 3% acetonitrile; 0.1% formic acid and 90% acetonitrile; 0.1% formic acid, respectively. The capillary (spray) voltage was held between 1650 and 1800 V. Source gas was heated to 325 °C, with a flow of 5.0 L/min. Data were acquired in centroid mode, with scan times of 400 ms per MS scan and 800 ms per MS/MS scan. Nitrogen was the collision gas, with a cell pressure of 18 psi. A data-dependent method was used, with the top five most intense precursors selected for subsequent fragmentation. Dynamic exclusion was utilized after acquiring an MS/MS spectrum, with a period of 15 s. Peaks below 525 m/z were excluded from fragmentation.

To optimize collision energies (CE), static fragmentation voltages (defined with respect to the entrance of the collision cell) of 0, 5, 10, 15, 20, 25, and 30 V were utilized in separate LC−MS/MS analyses. Following examination of MS/MS spectra at each of these constant settings, it was determined that the optimal glycan fragment ion coverage CEs (in volts) for the Pronase glycopeptides observed in this study vary with the m/z and the charge state of the precursor as follows:

	charge state		
m/z	2	3	4
100	18	10	9
2000	28	18	14

For comparison, the optimal peptide fragmentation values for this instrument are 3.6 V/100m/z −4.8 V.

Data Analysis. Compound identification, deconvolution, and data analysis were performed using Masshunter Qual software and Microsoft Excel. The presence of glycopeptide marker ions was utilized to screen for glycopeptide candidates, similar to a previous approach.[25] An in-house developed software utilized a combinatorial approach to generate a list of all potential glycopeptide compositions matching putative glycopeptide precursors. This approach generates a list of all peptides below a user-defined size (15 amino acids in this case), which contain a consensus site of N-glycosylation. This list was then combined with possible glycan compositions to yield a list of all possible glycopeptides generated during PB digestion. Glycopeptide assignments were made by hand annotation of MS/MS spectra and comparison to the software-generated list.

RESULTS

After MS analysis, the individual MS/MS spectra are then screened for the presence of characteristic glycopeptide fragments, as determined previously. These fragments include protonated hexose, *N*-acetyl hexosamine, fucose, and sialic acid monosaccharides as well as di- and trisaccharide ions. Table 1 in the Supporting Information lists a number of diagnostic peaks that are useful for identifying saccharide residues as well as the peptide moiety. Each precursor mass that yields one of these marker ions is compared to a combinatorial library of all possible glycans and potential sites of glycosylation for that sample. Those precursor masses that are within a user-defined error tolerance (10 ppm) are compared with the MS/MS spectra to assign glycopeptide compositions. The glycopeptides assigned utilizing optimal CEs possess complex, hybrid, and high mannose N-glycans and range from ~1500 to 4100 Da, exhibiting a very high mass range. It should be noted that very small (<1800 Da) glycopeptides require similar fragmentation energies as peptides. The largest glycopeptide observed in this study, however, required only 40% of the equivalent optimal peptide collision energy. At typical peptide CEs, glycopeptide MS/MS spectra are uninformative and yield no useful structural information, only oxonium fragments in the low m/z range.

The presence of one of the marker ions indicated is not sufficient to assign a glycopeptide composition. An example of the need for high-quality MS/MS spectra to make these assignments is summarized in Figure 1. The extreme heterogeneity of glycosylation often results in multiple putative glycopeptide compositions matching to a single precursor mass, as is the case with the doubly charged precursor ion at m/z 1006.408 shown in Figure 1. This species is derived from chicken egg albumin and has two possibilities based on exact mass, namely, EEKYNL-$Hex_5HexNac_2$ and YNLT-$Hex_3HecNac_5$. The glycoprotein CEA is known to possess both complex and high mannose type glycans. Accurate mass (<2 ppm) is insufficient to assign the correct glycopeptide composition in this case, as these species differ by only 0.2 mDa. Fragmentation at 10 V yielded a single hexose loss, which is insufficient for full assignment, as this could be from a terminal mannose residue of a high mannose glycan or a galactose residue derived from a complex-type glycan. Glycan fragmentation coverage improves at the 15 V setting, but the most informative spectrum is obtained at 30 V, showing full fragmentation of the glycan moiety and thereby verifying the assignment of EEKYNL-Man_5.

As expected, the charge state of the precursor ion greatly affects the observed fragmentation pattern. An example of this is shown in Figure 2, which depicts the deconvoluted MS/MS spectra from the triply protonated (left) and quadruply protonated (right) charge states of the ribonuclease B (RnB) glycopeptide SRNLTKR-Man_5. The optimal CE for the quadruply charged species occurs at 10 V, while the triply charged species yields only a single hexose loss at that setting. Instead, the optimal CE for the triply protonated species is 20 V, demonstrating a fragmentation pattern from the precursor to the peptide with a single HexNac residue remaining on the peptide. At this setting, the quadruply charged species still yields informative fragments, but the signal of these peaks is diminished. This trend toward lower CE for greater charge states was observed for all glycopeptides regardless of the glycan moiety, as expected. As a result of this trend the CEs were optimized separately for 2+, 3+, and 4+ charge states.

Glycan size is also an important factor in the fragmentation behavior. An example of this is shown in Figure 3, which shows the deconvoluted MS/MS spectra of different glycoforms of the RNLTKD peptide from ribonuclease B (RnB). The peptide is shown modified by three different high mannose N-glycans, Man_5, Man_7, and Man_9. At a CE of 10 V, each glycoform

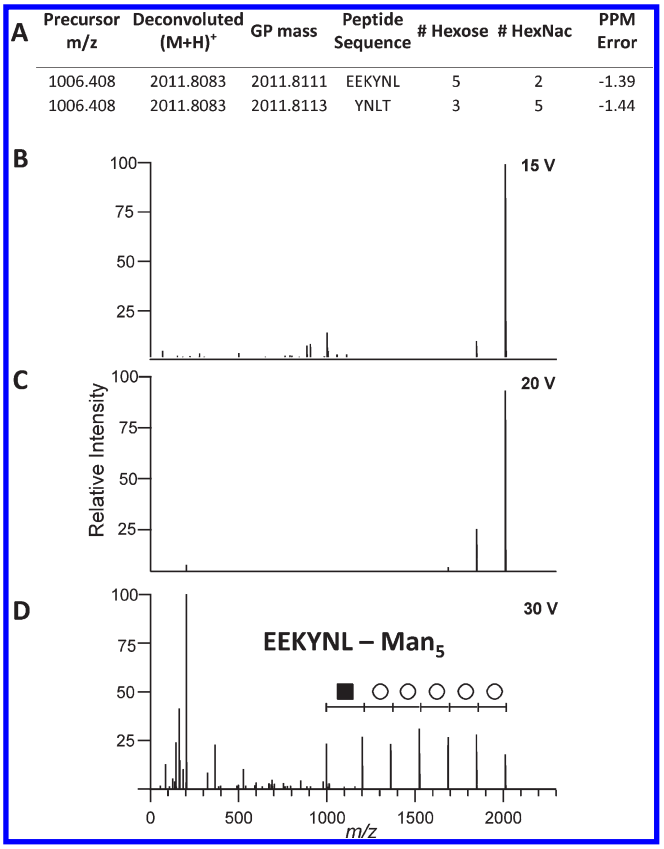


Figure 1. An example of a glycopeptide assignment that requires significant glycan coverage for a confident assignment. A combinatorial approach identifies two possible glycopeptide compositions for a doubly charged chicken egg albumin glycopeptide ion at 1006.408 m/z (A). Upon inspection of the deconvoluted CID spectra obtained at 15 V collision offset, a single hexose loss is observed, which does not aid in structural assignment (B). At 20 V, two consecutive hexose fragments are observed, suggesting the unknown species is the EEKYNL-Man_5 glycopeptide, although the $-Hex_2$ fragment is low in abundance (C). The 30 V analysis yields more complete fragmentation coverage, leading to a confident assignment (D).

dx.doi.org/10.1021/ac2003888 |*Anal. Chem.* 2011, 83, 5541–5547

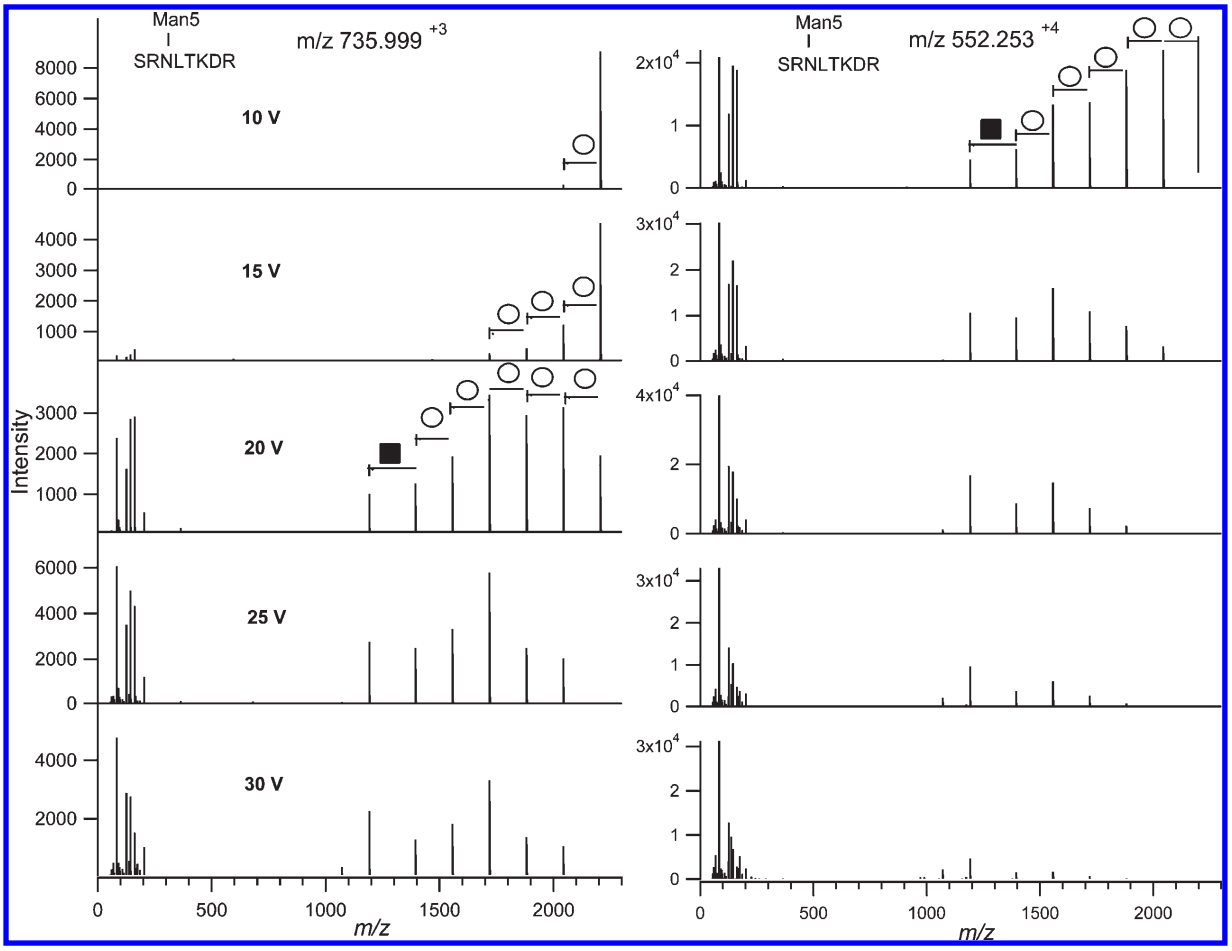


Figure 2. The charge state of the precursor greatly impacts the fragmentation behavior of Pronase glycopeptides. As an example of this, deconvoluted fragmentation spectra of triply (at left) and quadruply (at right) protonated Man5-SRNLTKDR glycopeptides from ribonuclease B are shown. The oxonium fragment ions below m/z 300, extent of glycan fragmentation, and precursor ion intensities are all profoundly changed between these charge states.

demonstrates acceptable glycan fragment coverage, although the most intense fragment peak in each case is different. While the Man_5 glycoform fragments to yield intense signal for the RNLTKD-HexNac and RNLTKD-$HexNac_2$ species, these species are much lower in intensity for the Man_7 glycoform. This trend continues for the Man_9 glycoform. This trend is consistent with predicted behavior in which smaller species would fragment more extensively at a constant CE. The 30 V experiment yields an identical trend toward more extensive fragmentation with smaller glycopeptides.

In addition to size and charge state, the monosaccharides present on glycoproteins influence the observed fragmentation pattern. Hexose and *N*-acetylhexosamine residues form the backbone of the glycan yield predictable fragmentation patterns in this study. Fucose and sialic acid residues, however, are terminal species and are generally much more labile than hexose or *N*-acetylhexosamine residues. This effect is demonstrated for the doubly protonated human lactoferrin (hLF) glycopeptides FNQ-$Hex_5HexNac_4Fuc_2$ and FNQ-$Hex_5HexNac_4NeuAc_1$ in Figure 4. At a CE of 15 V, the precursor is the most abundant species for the sialylated species, and losses of the sialic acid residue and the *N*-acetyl lactosamine are observed at approximately equal abundances.

An example of a challenging assignment of a hLF glycopeptide made using optimized CEs is shown in Figure 5. A single precursor, corresponding to a neutral glycopeptide mass of 2566.976 yields 11 putative glycopeptide compositions within a 10 ppm mass error tolerance (Figure 5A). The presence of marker ions for both sialylation and fucosylation are present, eliminating seven of the putative compositions. However, the remaining four entries are structurally similar, necessitating a high-quality MS/MS spectrum to make a confident assignment. As can be seen in Figure 5B, near-complete coverage of the oligosaccharide moiety is achieved with the optimized CE. The marker ion of peptide + HexNac is useful for making a final assignment and has also been utilized with an MS^3 approach recently to aid in structural assignment.[17] However, a full coverage of the glycan moiety is preferable for a confident assignment. Furthermore, the fragmentation pattern observed suggests that one of the two fucose residues is on a core *N*-acetyl glucosamine (as indicated by the fragment ion at 913.35 m/z), while the other is located on one of the antennae.

As an example of the depth of qualitative analysis afforded by this technique, a list of the glycopeptides identified in a single, optimized CE, LC−MS/MS analysis of hLF is shown in Table 2 in the Supporting Information. There are 103 species listed, representing 15 distinct glycan compositions distributed over

dx.doi.org/10.1021/ac2003888 |*Anal. Chem.* 2011, 83, 5541–5547

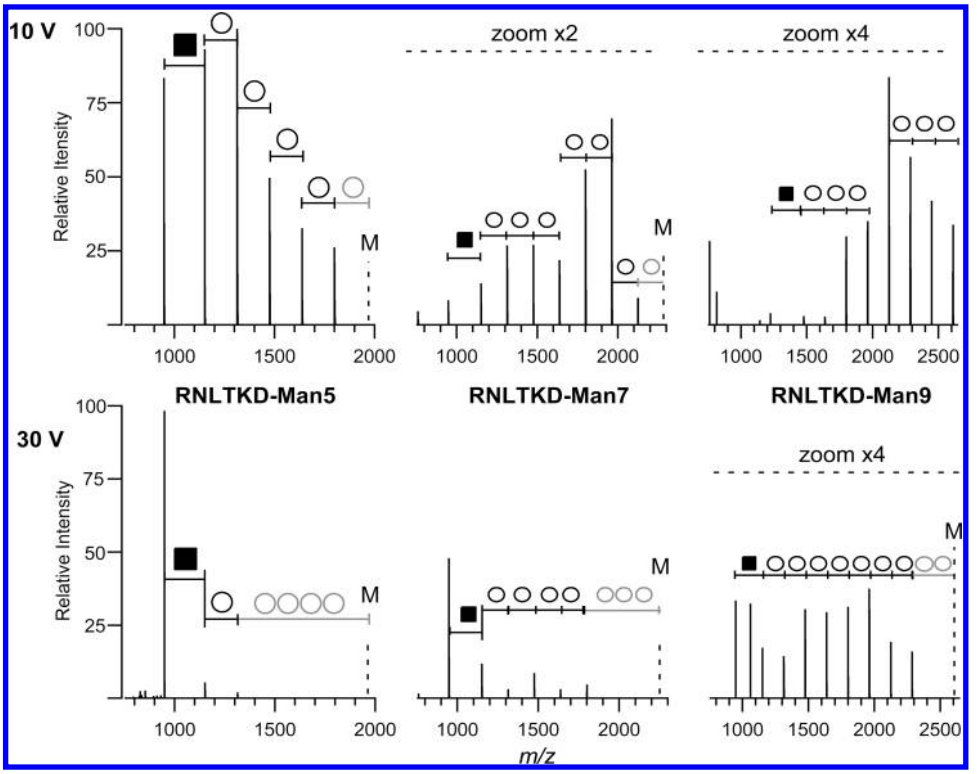


Figure 3. Fragmentation of RNLTKD glycopeptides modified by Man5, Man7, or Man 9 glycans (at left, center, and right) with either 10 (at top) or 30 V (at bottom) of collision potential applied. Increasing the collision voltage for a given glycan size increases the intensities of the lower mass ions. Increasing the glycan size for a given collision potential decreases the overall extent of fragmentation.

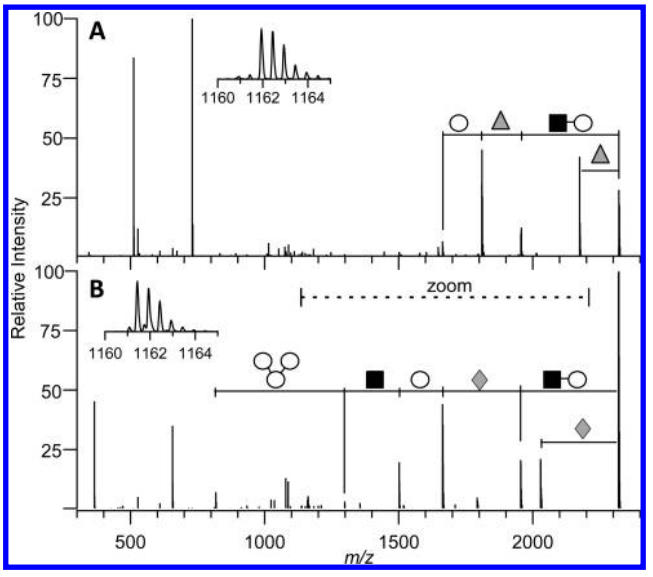


Figure 4. Deconvoluted fragmentation spectra of the FNQ peptide, modified by a glycan moiety of $Hex_5HexNac_4Fuc_2$ (A) and $Hex_5HexNac_4Sia_1$ (B) are shown, each fragmented at 15 V. The labile fucose and sialic acid residues are removed at low collision energies. As two fucose residues differ from a sialic acid residue by only 1 Da, glycopeptides possessing both fucose and sialic acid residues will require chromatographic resolution for proper elucidation by tandem MS. Elution times are 12.5 and 16.4 min for the difucosylated and sialylated species, respectively. The indicated region is zoomed by a factor of 2 for the fucosylated and 5 for the sialylated glycopeptides.

two glycosylation sites. On the average, each glycan is assigned to a site with 4.6 distinct peptides, each verified by tandem MS, supporting each assignment by multiple identifications. Peptide fragmentation is rarely observed. The assignments are also represented graphically in the Supporting Information.

DISCUSSION

In the field of proteomics, extensive work has established protein templates, elucidated peptide fragmentation mechanisms,[27] characterized highly specific proteases that enable reproducible generation of peptides,[28] and established automated peptide identification methods.[29] Comparable fundamental challenges to large-scale glycoproteomics have still not been met. A major limiting factor in the development of the field of glycoproteomics is the lack of a genome-based template for glycan structure, meaning that creating a genome-wide glycopeptide database is not possible. In addition, limited data has been published regarding the fragmentation behavior of nontryptic glycopeptides.[30,31] Furthermore, a robust and broadly specific glycopeptide preparation has not yet emerged among published reports.

The absence of a database of possible glycopeptides necessitates hand-annotation of tandem spectra for glycopeptide identifications. The analytical complexity introduced by glycosylation demands extremely high-quality tandem spectra for correct glycopeptide identification in complex samples. The difficulty in attaining the needed high-quality MS/MS spectra has been another major analytical challenge to the determination of large-scale site-specific glycoproteomics. However, recent MS advances have led to instruments that are routinely capable of delivering

dx.doi.org/10.1021/ac2003888 |*Anal. Chem.* 2011, 83, 5541–5547

A

Entry #	Peptide Mass	Site	Peptide	OS mass	Calc Gly mass	error	Hex	Fuc	HexNac	NeuAc	ppm error
1	896.40284	156	NWTGPPEP	1670.5732	1670.5914	-0.0182	6	2	2		7.1
2	652.28165	497	FNQTGS	1914.6944	1914.6974	-0.0031	5	2	4		1.2
3	418.18122	497	NQTG	2148.7948	2148.7713	0.0235	5	3	3	1	9.2
4	**361.15976**	**497**	**NQT**	**2205.8162**	**2205.7928**	**0.0234**	**5**	**2**	**4**	**1**	9.1
5	976.36685	642	GRNGSDCPD	1590.6091	1590.5918	0.0173	3	2	4		6.7
6	919.34539	642	RNGSDCPD	1647.6306	1647.6133	0.0173	3	1	5		6.7
7	636.29796	642	FGRNGS	1930.678	1930.6923	-0.0143	6	1	4		5.6
8	547.23503	642	RNGSD	2019.741	2019.7287	0.0123	6	3	3		4.8
9	402.19752	642	GRNG	2164.7785	2164.7662	0.0123	6	2	3	1	4.8
10	345.17606	642	GRN	2221.7999	2221.7877	0.0122	6	1	4	1	4.8
11	345.17606	642	RNG	2221.7999	2221.7877	0.0122	6	1	4	1	4.8

B

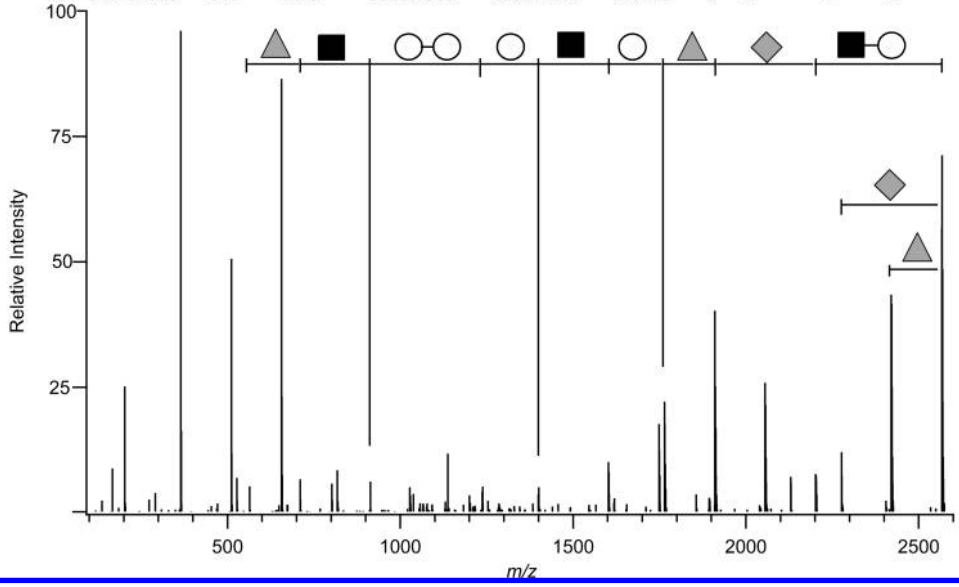


Figure 5. An example of a complex glycopeptide assignment. The putative glycopeptide from human lactoferrin matches 11 biologically plausible glycopeptide compositions by accurate mass with a 10 ppm mass error tolerance (A). Despite the large number of putative compositions, the optimized MS/MS spectrum enables a confident assignment (B), NQT-$Hex_5Fuc_2HexNac_4NeuAc_1$. Reporter ions below 700 m/z indicate the presence of fucose (nominal mass 512) and sialic acid (nominal masses 657, 292, and 274).

high-resolution, high-mass accuracy MS/MS data. In addition, the known specificity of glycosyltransferases responsible for the creation of the glycan moieties may also be used to restrict the positions of specific monosaccharide residues. Sialic acid residues appear in terminal positions of N-glycans linked to galactose residues typically with an α-2,3 or α-2,6 linkage. Fucose residues may appear attached to the asparagine-linked *N*-acetyl glucosamine or on one of the branches of complex-type N-glycans attached to galactose or *N*-acetyl glucosamine residues.

Some general observations apply while annotating MS/MS spectra of glycopeptides generated by nonspecific digestion. Typically, a series of monosaccharide losses from the nonreducing end of the glycopeptide are observed. After an extensive series of these losses, a loss of either 486.15 m/z, corresponding to Hex_3, or 689.22 m/z, corresponding to $Hex_3HexNac$, is observed. The lability of the HexNac−HexNac bond on the reducing end of the glycan favors fragmentation at this site. The conserved N-glycan core linked to an asparagine contributes approximately 1024 Da. Any MS/MS peak below this threshold with a corresponding peak 486.15 or 689.22 Da higher is a putative peptide + HexNac peak. When the peptide + HexNac ion is observed, it may also be followed by sequential losses of 120 and 83 Da, corresponding to cross ring cleavage of the last HexNac residue, yielding the unmodified protonated peptide. Peptide fragmentation is rarely observed under these conditions as the glycosidic bonds are much more labile than peptide bonds; however, there are indications that yet higher energies will yield eventual peptide fragmentation. Notably, the peptide + HexNac marker ion is not always required to make an accurate assignment due to other compositional information attained via optimized CE analyses.

Chromatographic separations are especially useful in reducing ion suppression as the glycan and protein complexity increases. Chromatographic separation is also critical when glycopeptides bearing multiple fucose and sialic acid residues are present in a given sample. A sialic acid residue has an accurate mass of 291.0954 Da, while two fucose residues add to 292.1158 Da, a difference of only 1.0204 Da. In the absence of a separation or fractionation before analysis, these species would share the same isotopic distribution, greatly confounding their isolation for subsequent MS/MS. Fortunately, LC separation using PGC can achieve chromatographic resolution of N-glycans,[32] and we observe similar resolution in the chromatographic separation of small, N-linked glycopeptides. In general, sialic acid containing glycopeptides have later eluting retention times than the corresponding difucosylated glycopeptides, and the isotopic issue mentioned previously is negated by the chromatographic separation. As no tri- or quadruply sialylated glycopeptides were identified in this study, it remains to be seen whether further

chromatographic and/or CE optimization may be required for the analysis of highly sialylated glycopeptides.

The optimized CEs utilized in this study increased the number of glycopeptides identified in a single analysis by a factor of 2. In addition, spectra containing complete glycan coverage are readily obtained at these collision energies, although they may not be optimal for all analytes. For lactoferrin, over 85% of all putative glycopeptides observed present sufficient fragmentation information under the optimized conditions to make an unambiguous assignment. This metric is improved in the analyses of RnB and CEA. The combination of high-mass accuracy, nano-LC separation, and robust fragmentation characteristics afforded by this technique enables glycopeptide characterization from a wide range of glycan sizes, charge states, and peptide footprints. Furthermore, the most abundant glycoforms of the glycoproteins observed in these analyses coincide with previously published reports in each case.

CONCLUSIONS

A robust methodology has been developed utilizing nonspecific protease digestion, accurate mass measurements, chromatographic separations, and optimized collision energies to determine the microheterogeneity of protein glycosylation. This methodology has been applied to three standard glycoproteins possessing a variety of high mannose, hybrid, and complex type N-glycosylation. Under optimized conditions, approximately 80% of all putative glycopeptides were assigned to a unique composition, and over 100 glycopeptides were identified in a single nano-LC−MS/MS analysis of a model glycoprotein. Furthermore, as this analytical approach yields peptide footprints of differing sizes, multiple independent glycopeptide assignments support the assignment of glycan microheterogeneity at each site of glycosylation.

ASSOCIATED CONTENT

Supporting Information. A method overview, a list of glycopeptide marker ions utilized in this study, a list of lactoferrin glycopeptides observed, a graphic of lactoferrin microheterogeneity observed, and a chart detailing initial factors influencing LC retention of glycopeptides. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cblebrilla@ucdavis.edu.

REFERENCES

(1) Varki, A. *Glycobiology* **1993**, *3* (2), 97–130.
(2) O'Connor, S. E.; Imperiali, B. *Chem. Biol.* **1996**, *3* (10), 803–812.
(3) Wormald, M. R.; Dwek, R. A. *Structure* **1999**, *7* (7), R155–R160.
(4) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G., 2nd; Monroe, M. E.; Moore, R. J.; Smith, R. D. *J. Proteome Res.* **2005**, *4* (6), 2070–2080.
(5) Wang, X.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **2010**, *82* (15), 6542–8.
(6) Ito, S.; Hayama, K.; Hirabayashi, J. *Methods Mol. Biol.* **2009**, *534*, 195–203.
(7) Wohlgemuth, J.; Karas, M.; Eichhorn, T.; Hendriks, R.; Andrecht, S. *Anal. Biochem.* **2009**, *395* (2), 178–188.
(8) Wada, Y.; Tajiri, M.; Yoshida, S. *Anal. Chem.* **2004**, *76* (22), 6560–6565.
(9) Alvarez-Manilla, G.; Atwood, J., 3rd; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. *J. Proteome Res.* **2006**, *5* (3), 701–708.
(10) Snovida, S. I.; Bodnar, E. D.; Viner, R.; Saba, J.; Perreault, H. *Carbohydr. Res.* **2010**, *345* (6), 792–801.
(11) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **2003**, *75* (20), 5628–5637.
(12) Zauner, G.; Koeleman, C. A.; Deelder, A. M.; Wuhrer, M. *J. Sep. Sci.* **2010**, *33* (6−7), 903–910.
(13) Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B. *J. Proteome Res.* **2007**, *6* (10), 4032–4040.
(14) Dodds, E. D.; Seipert, R. R.; Clowers, B. H.; German, J. B.; Lebrilla, C. B. *J. Proteome Res.* **2008**, *8* (2), 502–512.
(15) Desaire, H.; Hua, D. *Int. J. Mass Spectrom.* **2009**, *287* (1−3), 21–26.
(16) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. *Anal. Chem.* **2001**, *73* (18), 4530–4536.
(17) Wu, Y.; Mechref, Y.; Klouckova, I.; Mayampurath, A.; Novotny, M. V.; Tang, H. *Rapid Commun. Mass Spectrom.* **2010**, *24* (7), 965–972.
(18) Segu, Z. M.; Mechref, Y. *Rapid Commun. Mass Spectrom.* **2010**, *24* (9), 1217–1225.
(19) Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V. *Rapid Commun. Mass Spectrom.* **2009**, *23* (1), 161–170.
(20) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2007**, *849* (1−2), 115–128.
(21) Henning, S.; Peter-Katalinic, J.; Pohlentz, G. *J. Mass Spectrom.* **2007**, *42* (11), 1415–1421.
(22) Harazono, A.; Kawasaki, N.; Kawanishi, T.; Hayakawa, T. *Glycobiology* **2005**, *15* (5), 447–462.
(23) Ethier, M.; Krokhin, O.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Perreault, H. *Rapid Commun. Mass Spectrom.* **2005**, *19* (5), 721–727.
(24) Wuhrer, M.; Hokke, C. H.; Deelder, A. M. *Rapid Commun. Mass Spectrom.* **2004**, *18* (15), 1741–1748.
(25) Jebanathirajah, J.; Steen, H.; Roepstorff, P. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (7), 777–784.
(26) Itoh, S.; Kawasaki, N.; Ohta, M.; Hayakawa, T. *J. Chromatogr., A* **2002**, *978* (1−2), 141–152.
(27) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(28) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3* (6), 608–614.
(29) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20* (18), 3551–3567.
(30) Seipert, R. R.; Dodds, E. D.; Lebrilla, C. B. *J. Proteome Res.* **2009**, *8* (2), 493–501.
(31) Seipert, R. R.; Dodds, E. D.; Clowers, B. H.; Beecroft, S. M.; German, J. B.; Lebrilla, C. B. *Anal. Chem.* **2008**, *80* (10), 3684–3692.
(32) Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B. *Proteomics* **2009**, *9* (7), 1939–1951.