Anal Bioanal Chem (2012) 403:1291–1302
DOI 10.1007/s00216-011-5109-x

ORIGINAL PAPER

Site-specific protein glycosylation analysis with glycan isomer differentiation

Serenus Hua · Charles C. Nwosu · John S. Strum · Richard R. Seipert · Hyun Joo An · Angela M. Zivkovic · J. Bruce German · Carlito B. Lebrilla

Received: 14 March 2011 / Revised: 5 May 2011 / Accepted: 13 May 2011 / Published online: 8 June 2011

Abstract Glycosylation is one of the most common yet diverse post-translational modifications. Information on glycan heterogeneity and glycosite occupancy is increasingly recognized as crucial to understanding glycoprotein structure and function. Yet, no approach currently exists with which to holistically consider both the proteomic and glycomic aspects of a system. Here, we developed a novel method of comprehensive glycosite profiling using nanoflow liquid chromatography/mass spectrometry (nano-LC/MS) that shows glycan isomer-specific differentiation on specific sites. Glycoproteins were digested by controlled non-specific proteolysis in order to produce informative glycopeptides. High-resolution, isomer-sensitive chromatographic separation of the glycopeptides was achieved using microfluidic chip-based capillaries packed with graphitized carbon. Integrated LC/MS/MS not only confirmed glycopeptide composition but also differentiated glycan and peptide isomers and yielded structural information on both the glycan and peptide moieties. Our analysis identified at least 13 distinct glycans (including isomers) corresponding to five compositions at the single *N*-glycosylation site on bovine ribonuclease B, 59 distinct glycans at five *N*-glycosylation sites on bovine lactoferrin, 13 distinct glycans at one *N*-glycosylation site on four subclasses of human immunoglobulin G, and 20 distinct glycans at five *O*-glycosylation sites on bovine κ-casein. Porous graphitized carbon provided effective separation of glycopeptide isomers. The integration of nano-LC with MS and MS/MS of non-specifically cleaved glycopeptides allows quantitative, isomer-sensitive, and site-specific glycoprotein analysis.

Keywords Site-specific glycosylation · Glycopeptide · Non-specific protease · Isomer · Quantitation · LC/MS

Published in the special issue *High-Resolution Mass Spectrometry* with guest editors Hans H. Maurer and David C. Muddiman.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-5109-x) contains supplementary material, which is available to authorized users.

S. Hua · C. C. Nwosu · J. S. Strum · R. R. Seipert · C. B. Lebrilla
Department of Chemistry, University of California,
Davis, CA 95616, USA

H. J. An
Graduate School of Analytical Science and Technology,
Chungnam National University,
Daejeon 305-764, South Korea

A. M. Zivkovic · J. B. German
Department of Food Science and Technology,
University of California,
Davis, CA 95616, USA

A. M. Zivkovic · J. B. German · C. B. Lebrilla
Foods for Health Institute, University of California,
Davis, CA 95616, USA

C. B. Lebrilla (✉)
Department of Biochemistry and Molecular Medicine,
University of California,
Davis, CA 95616, USA
e-mail: cblebrilla@ucdavis.edu

Abbreviations

bLF	Bovine lactoferrin
CID	Collision-induced dissociation
CNBr	Cyanogen bromide
ECC	Extracted compound chromatogram
Fuc	Fucose
Gal	Galactose
Glc	Glucose
Hex	Hexose
HexNAc	*N*-acetylhexosamine

IgG	Immunoglobulin G
Man	Mannose
Man*X* (where *X*=5 to 9)	High mannose glycan of composition $GlcNAc_2Man_X$
NeuAc	*N*-acetylneuraminic acid
PGC	Porous graphitized carbon
Q-TOF	Quadrupole time-of-flight
RNAse B	Ribonuclease B
TIC	Total ion chromatogram

Introduction

Glycoproteins are involved in a variety of important biological processes ranging from structural functions to cellular signaling and molecular transport [1, 2]. Information on glycoprotein structure is crucial to our understanding of their biological roles. However, glycan heterogeneity significantly complicates glycoprotein analysis. The current paradigm in glycoproteomics is characterized by proteocentric methods in which glycoproteins are simply deglycosylated, and the glycan moiety largely ignored [3–7]. Yet, protein glycosylation is one of the most common as well as most complex post-translational modifications, with more than half of the proteome exhibiting at least some degree of glycosylation, and scores of glycan structures possible at each glycosylation site [8].

The variety and complexity of glycosylation mean that any comprehensive analysis of protein glycosylation must also take into account glycan heterogeneity and glycosite occupancy. Glycans are produced by a set of over 200 competing glycosyltransferases that modify the nascent glycan by adding a specific saccharide via a specific linkage [9]. For example, a standard hexose has five different sites at which it may be connected to other saccharide units. When anomeric character is considered, the possibilities increase to ten unique potential linkages. Furthermore, each glycosyltransferase has a multitude of glycoprotein substrates, and glycosylation at any given site is by no means homogenous or even obligatory. Thus, a glycoprotein containing just four sites of glycosylation, with five possible glycans at each site, can result in more than a thousand different glycoforms.

Conventional methods for analyzing site-specific glycosylation involve digestion of the glycoprotein with a single enzyme, commonly trypsin [10–13]. However, extended spans between protease active sites often result in excessively large glycopeptides. In addition, steric effects from glycans are known to significantly hinder nearby protease activity, especially that of trypsin, resulting in missed cleavages [14–16]. Consequently, the glycopeptides produced by single-enzyme digest often have long peptide moieties that contain multiple sites of glycosylation and thus severely complicate or obfuscate site-specific analysis.

In contrast, protease cocktails enable the preparation of glycopeptides with scalable peptide lengths. Our group has previously developed methods using Pronase E, a mixture of proteolytic enzymes isolated from *Streptomyces griseus*, to prepare glycopeptides suitable for site-specific glycosylation analysis [17–20]. Pronase nonspecifically hydrolyzes all peptide bonds in a protein, with the exception of those in the immediate vicinity of a glycan. By varying digestion time, the length of the peptide moieties may be scaled up or down as desired. This not only allows single glycosylation sites to be isolated on a glycopeptide but also reduces signal suppression from non-glycosylated peptides by breaking them down into their component amino acids. We have further refined the technique by immobilizing Pronase on a solid substrate such as beaded agarose, minimizing contamination due to enzyme autolysis and facilitating the preparation of glycopeptide-enriched digests [18–20]. Combined with high-resolution, accurate-mass mass spectrometry (MS), Pronase digestion allows us to consider site-specific glycoproteomic analysis as a relatively simple computational exercise rather than a convoluted chemical/biochemical puzzle as it has traditionally been portrayed.

Mass spectrometric analysis of fractionated glycopeptide mixtures has been performed by our group in the past [17–20]. However, glycoprotein heterogeneity necessitates a method of analysis that can separate and differentiate a number of isomeric glycoforms. Chromatographic separation of tryptic glycopeptides by C18 has been reported previously; however, such methods are optimized for the long peptide/glycopeptide chains commonly produced by tryptic digestion and have shown extremely limited ability with respect to separation of isomeric glycoforms [10–13]. Meanwhile, porous graphitized carbon (PGC), amine/amide-based, and anion-exchange stationary phases have been used to successfully separate isomeric glycans [21, 22], but their application to glycopeptide separation has thus far been limited [11, 23, 24].

In order to develop a method of rapid and thorough glycoproteomic analysis with glycan isomeric and site-specific information, we have combined nonspecific proteolysis, chip-based nanoflow liquid chromatography (nano-LC) separation, in silico screening by accurate mass, and tandem MS analysis to differentiate and quantify the isomeric glycoforms of glycoproteins. PGC allows isomer-specific separation of *N*- and *O*-glycopeptides modulated by both peptide and glycan structure. High mass accuracy data from single-stage MS is used to assign glycopeptide compositions. Tandem MS provides supplementary data that confirms compositional assignments and determines glycopeptide structure over a dynamic range of five orders of magnitude. Bovine ribonuclease B (RNAse B), bovine lactoferrin (bLF), human immunoglobulin G (IgG), and bovine κ-casein were analyzed as representative glycoproteins exhibiting *N*-linked

high-mannose, *N*-linked complex, and *O*-linked glycosylation. Hundreds of glycopeptides were separated by nano-LC and identified by accurate mass. Tandem MS confirmed glycopeptide assignments and elucidated the structure of chromatographically separated isomers. The combination of LC/MS and LC/MS/MS glycopeptide data allowed us to identify 13 site-specific glycans at one *N*-glycosylation site on bovine RNAse B, 59 site-specific glycans at five *N*-glycosylation sites on bLF (including one previously unreported site), 13 site-specific glycans at one *N*-glycosylation site on four subclasses of human IgG, and 20 site-specific glycans at five *O*-glycosylation sites on bovine κ-casein. Using this data, glycan heterogeneity at each site of glycosylation was described and quantified. Site-specific, isomer-specific glycoprotein profiling will dramatically advance our understanding of glycosylation and glycosylated systems, and may one day be applied to the discovery of new disease biomarkers.

Methods

Materials and reagents

Pronase E, bovine RNAse B, bLF, bovine κ-casein, human IgG, and cyanogen bromide (CNBr)-activated Sepharose 4B beads were obtained from Sigma-Aldrich (St. Louis, MO, USA). Graphitized carbon cartridges were obtained from Grace Davison (Deerfield, IL, USA). Solvents were of LC-MS grade. All other reagents were of analytical grade or higher.

Immobilization of pronase enzymes

Pronase E enzyme cocktail was covalently linked to CNBr-activated Sepharose 4B (S4B) beads of 40–165-μm diameter using well-established coupling chemistry optimized by our group [18–20]. S4B beads were reconstituted and washed in 1-mM hydrochloric acid, then conditioned with 100 mM phosphate buffer (pH 7.4). For enzyme coupling, Pronase E was combined with the S4B beads in 100 mM phosphate buffer at a ratio of 6.7 mg Pronase per gram of dry S4B. The coupling reaction proceeded under gentle agitation for 16 h at room temperature. Excess enzyme was removed by washing with 100 mM phosphate buffer, after which 1 M aqueous ethanolamine (pH 9.0) was added to block all remaining active CNBr sites on the beads. To remove excess ethanolamine, the beads were washed with 100 mM ammonium acetate (pH 7.2).

Glycoprotein digestion by immobilized Pronase

Glycoproteins were dissolved in 100 mM ammonium acetate and mixed with bead-immobilized Pronase at a glycoprotein–Pronase ratio of 1:1 (*w*/*w*). Digestions were allowed to proceed under gentle agitation for 12–24 h at 37 °C. Bead-immobilized Pronase was separated from the resulting glycopeptide solution by centrifugation.

Glycopeptide enrichment with graphitized carbon SPE

Digested glycopeptides were purified by graphitized carbon solid-phase extraction. Graphitized carbon cartridges were washed with 0.10% (*v*/*v*) trifluoroacetic acid in 80% acetonitrile/water (*v*/*v*) followed by conditioning with water. The glycopeptide solution was loaded onto each cartridge and washed with water at a flow rate of approximately 1 mL/min to remove salts and buffer. Glycopeptides were eluted with 0.05% (*v*/*v*) trifluoroacetic acid in 40% acetonitrile/water (*v*/*v*) and dried in vacuo prior to MS analysis.

Chip-based nano-LC/MS analysis

Glycopeptides were reconstituted in water at concentrations corresponding to between 40 and 200 ng of original glycoprotein (depending on the extent and rate of glycosylation) per 2-μL injection. Nano-LC/MS and nano-LC/MS/MS analyses were performed on an Agilent HPLC-Chip Quadrupole Time-of-Flight (Q-TOF) MS system equipped with a microwell-plate autosampler (maintained at 6 °C), capillary sample loading pump, nano pump, HPLC-Chip/MS interface, and Agilent 6520 Q-TOF MS detector. The chip used consisted of a 9×0.075-mm i.d. enrichment column and a 150×0.075-mm i.d. analytical column, both packed with 5-μm PGC as the stationary phase. For sample loading, the capillary pump delivered 0.1% formic acid in 3.0% acetonitrile/water (*v*/*v*) isocratically at 4.0 μL/min. Following sample injection, a nano pump gradient was delivered at 0.4 μL/min using (A) 0.1% formic acid in 3.0% acetonitrile/water (*v*/*v*) and (B) 0.1% formic acid in 90.0% acetonitrile/water (*v*/*v*). Samples were eluted with 0% B, 0.00–2.50 min; 0 to 16% B, 2.50–10.00 min; 16 to 44% B, 10.00–30.00 min; 44 to 100% B, 30.00–35.00 min; and 100% B, 35.00–45.00 min. This was followed by equilibration at 0% B, 45.01–65.00 min. The drying gas temperature was set at 325 °C with a flow rate of 5.0 L/min (2.5 L of filtered nitrogen gas and 2.5 L of filtered dry compressed air).

Single-stage MS spectra were acquired in the positive ionization mode over a mass range of *m*/*z* 400–3,000 with an acquisition time of 1 s per spectrum. Mass correction was enabled using reference masses of *m*/*z* 322.048, 622.029, 922.010, 1,221.991, and 1,521.971 (ESI-TOF Calibrant Mix G1969-85000, Agilent Technologies, Santa Clara, CA, USA).

MS/MS spectra were acquired in the positive ionization mode over a mass range of *m*/*z* 100–3,000 with an acquisition time of 250 ms per spectrum at a rate of four MS/MS spectra per single MS spectrum. In general, collision energies were

calculated for each compound based on the following formula:

$$V_{\text{collision}} = 3.6\text{V}\left(\frac{m/z}{100\text{Da}}\right) - 4.8\text{V}$$

Here, $V_{\text{collision}}$ is the potential difference across the collision cell. The slope and offset values of the energy-m/z ramp could be changed as needed to produce more or less fragmentation. Following detection during MS mode (using the conditions described previously for single-stage MS acquisition), compounds matching a specified mass-to-charge ratio, charge state, and retention time range were isolated in the quadrupole with a mass bandpass FWHM (full width at half maximum) of 1.3 m/z.

Results and discussion

In order to demonstrate the efficacy and wide applicability of our methods, several representative *N*- and *O*-glycoproteins were selected for site-specific analysis. These were bovine RNAse B, bLF, and human IgG, which exhibit *N*-glycosylation, and bovine κ-casein, which exhibits *O*-glycosylation. RNase B and IgG are well-characterized glycoproteins whose glycosylation profiles are known and thus provide a means of validation. κ-casein and bLF, on the other hand, have multiple glycosylation sites that until now have been only partially characterized.

Each glycoprotein was individually digested by immobilized Pronase, chromatographically separated by PGC nano-LC, and analyzed by single-stage MS as well as tandem MS/MS. Theoretical sites of glycosylation were examined, and glycan heterogeneity at each confirmed glycosite was profiled and quantified.

Site-specific determination of protein glycosylation by accurate mass

Single-stage MS data was used for preliminary identification of the glycopeptides resulting from each glycoprotein digest. Raw LC/MS data was filtered with a signal-to-noise ratio of 5.0 and analyzed using the Molecular Feature Extractor algorithm included in the MassHunter Qualitative Analysis software (version B.03.01, Agilent Technologies). Taking into account expected charge carriers, potential neutral mass losses, and a predicted isotopic distribution, the total ion chromatogram (TIC) was divided into a number of extracted compound chromatograms (ECCs). Each ECC represents the summed chromatograms of all ion species associated with a single compound (e.g., the singly protonated, doubly protonated, or singly dehydrated ions). Thus, each individual ECC peak could be taken to represent the total ion count associated with a single distinct compound.

The deconvoluted mass, retention time, abundance (in summed ion counts), and observed charge states associated with each extracted compound were exported for further analysis. Glycopeptides were identified using an in-house software tool known as Glycopeptide Finder (version 1.0.9). Based on the mass of a potential glycopeptide, the amino acid sequence(s) of the protein or proteins it may be derived from [25], the type of glycosylation (*N* or *O*), and a given mass tolerance (here, 20 ppm), the Glycopeptide Finder program presented a list of potential amino acid and glycan compositions associated with a given compound. Results are further filtered based on known glycosylation patterns and/or other biological rules [26].

The ECCs of abundant Pronase-digested glycopeptides from RNase B are shown in Fig. 1. RNase B is a well-characterized glycoprotein with a single *N*-glycosylation site at ^{60}Asn and a heterogeneous population of high-mannose glycoforms containing between five and nine mannose residues. Glycopeptide compositional assignment by accurate mass showed the exclusive presence of high-mannose type glycan moieties linked at ^{60}Asn to peptides of lengths ranging between three and six amino acid residues. Though the search algorithm used a mass error tolerance of 20 ppm, actual mass errors stayed well within tolerance. The root mean square deviation (RMSD) of the experimental mass was 4.31 ppm, confirming the high mass accuracy of the MS analysis. Peak assignments were subsequently confirmed by MS/MS.

Nonspecific digestion of glycoproteins is a key element of effective site-specific glycosylation analysis. Whereas specific proteases such as trypsin return fixed glycopeptide products which may or may not elucidate glycan and glycosite heterogeneity, Pronase activity may be modulated by digestion time and enzyme concentration in order to produce informative glycopeptides. In addition, due to the nonspecific nature of Pronase activity, glycopeptides are typically produced in sets of related families differing from each other by only a few amino acids. The peptide moieties of the RNAse B glycopeptides in Fig. 1, for example, may easily be grouped, with ^{60}NLT, ^{60}NLTK, ^{60}NLTKD, and ^{60}NLTKDR forming a clearly related family with identical N-terminal cleavages but differing C-terminal cleavages. Such glycopeptide families serve to confirm and reinforce the accuracy of our compositional assignments.

Glycopeptide separation by porous graphitized carbon (PGC)

Glycopeptide digests are complex mixtures with large structural diversity and dynamic range. Incorporation of chromatographic separation into established mass spectral methods of glycoproteomic analysis [17–20] allows us to address these issues by minimizing peptide ion suppression as well as distinguishing between isobaric and isomeric compounds of the same mass and/or composition. In

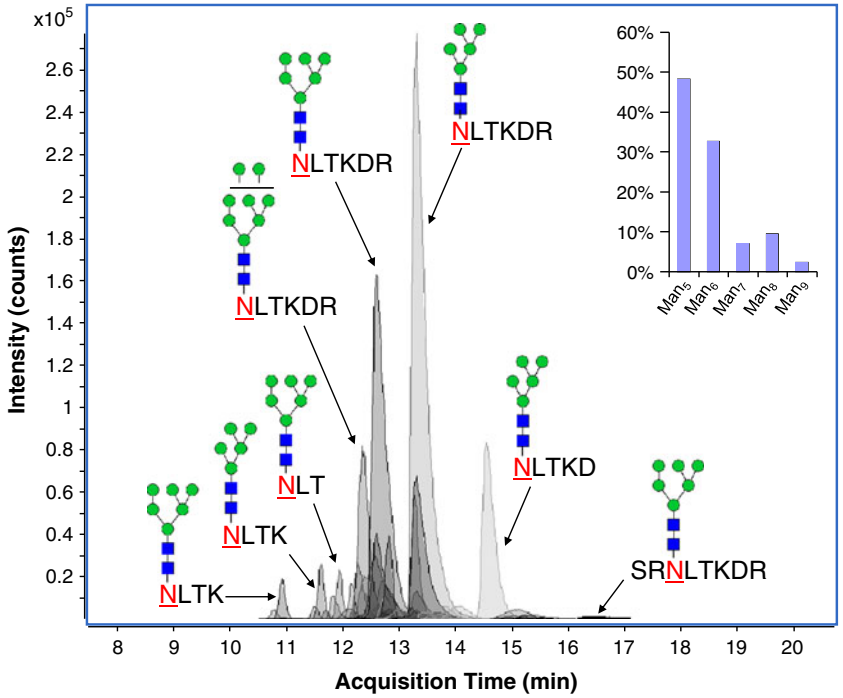

Fig. 1 Overlaid chromatograms and associated structural assignments of glycopeptides from bovine ribonuclease B. *Inset*, relative abundances of each glycoform of bovine ribonuclease B

particular, microfluidic chip-based nano-LC allows picogram sensitivity, while the TOF-MS detector provides greater than five orders of magnitude of instrumental dynamic range [21, 27]. To take full advantage of these attributes, our separation method used PGC as the stationary phase.

PGC chromatographically separates glycans and glycoconjugates based on size, polarity, and three-dimensional structure [24]. PGC offers several important advantages over stationary phases such as C18—for instance, effective separation of hydrophilic compounds, as well as the ability to resolve stereo- and regio-isomers. We have shown previously the application of microfluidic chip-based PGC nano-LC towards the separation of enzyme-released glycans [21, 27]. However, the utility and performance of PGC nano-LC in glycopeptide isomer separation has heretofore gone unreported.

Pronase-digested glycopeptides are uniquely suited for separation by PGC due to their stronger carbohydrate character. Compared with tryptic glycopeptides, which typically contain long, hydrophobic peptide moieties, Pronase-digested glycopeptides contain relatively short peptide moieties conjugated to a hydrophilic glycan structure. Thus, interactions with PGC are influenced more or less equally by the peptide and glycan portions of the glycopeptide.

Glycopeptide separation by PGC revealed a number of consistent trends. In general, more polar molecules were more strongly retained by PGC. While this effect has been previously observed in the separation of native glycans [27], this is the first time it has been observed with glycopeptides. In particular, the presence of acidic amino acid residues as well as glycan sialylation significantly increased glycopeptide retention times. In digests of IgG (Fig. 2), for example, glycopeptides NST+$Hex_4HexNAc_4Fuc$ (12.0 min) and NST+$Hex_5HexNAc_4Fuc$ (12.3 min) are also present with the addition of a terminal NeuAc (13.9 and 14.1 min, respectively). With all else held equal, the sole addition of a single acidic NeuAc residue increased the retention time of the compound by nearly 2 min. Similar observations may be made between glycopeptides NST+$Hex_4HexNAc_4Fuc$ (12.0 min), EEQFNST+$Hex_4HexNAc_4Fuc$ (18.7 min), and EEQYNST+$Hex_4HexNAc_4Fuc$ (19.1 min), all of which have identical glycans but different peptide moieties. Notably, the two later-eluting glycopeptides contain two glutamic acid (E) residues, contributing to the overall polarity of the molecule and thus the stronger retention exhibited by the PGC. The slight difference in retention time between these two nearly identical compounds may be attributed to the substitution of phenylalanine (F) in the earlier-eluting compound (originating from IgG subclasses 2 and 4) with tyrosine (Y) in the later-eluting compound (originating from IgG subclasses 1 and 3). While these two aromatic amino acid residues are structurally very similar, tyrosine is slightly more polar than phenylalanine due to the presence of a single additional hydroxyl group.

In addition to polarity, three-dimensional structure also makes a major contribution to the overall retention of a molecule by PGC [24, 28]; in fact, this is the basis of PGC's ability to separate isomeric glycopeptides. Structural differences in either the peptide or glycan moiety of a glycopeptide result in modified interactions with PGC. This is acutely evident in digests of *O*-glycoproteins such as κ-casein (Fig. 3). For example,

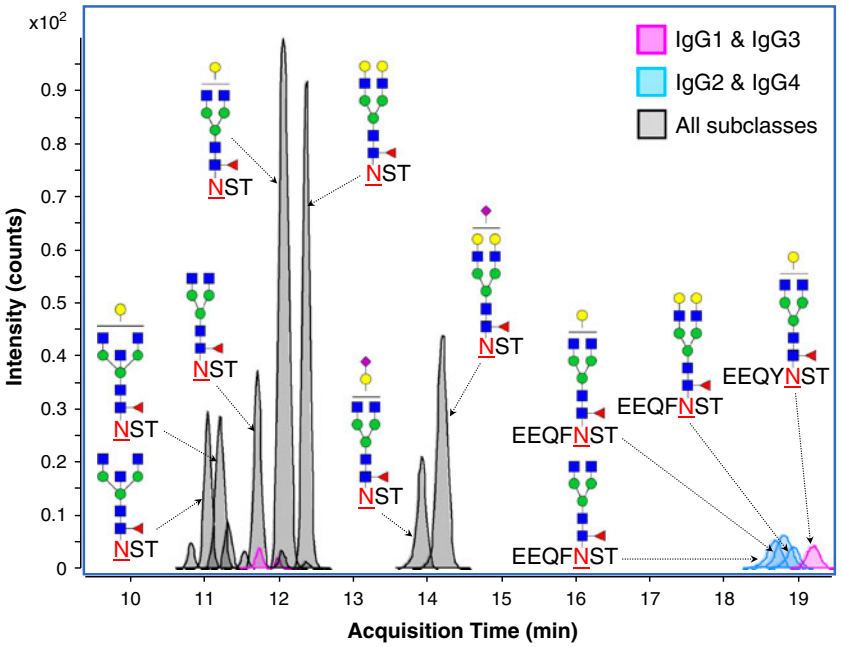

Fig. 2 Overlaid chromatograms and associated structural assignments of glycopeptides from human immunoglobulin G. *Color* denotes the IgG subclasses from which the glycopeptide originated

the isomeric glycopeptides of composition STVAT+$GalNAc_1Gal_1NeuAc$, eluting at 12.7 and 13.6 min, differ only by the glycan connectivity in the glycan moiety. In the earlier-eluting isomer, the NeuAc is conjugated to the core Gal, whereas in the later-eluting isomer, the NeuAc is conjugated to the core GalNAc. Similar examples may be seen with the isomeric glycopeptides of composition STVA+$GalNAc_1Gal_1NeuAc$ (12.3 and 13.3 min), as well as those of composition GEPTSTPT+$GalNAc_1Gal_1NeuAc$ (15.4 and 16.5 min; 19.0 and 19.9 min). In addition to glycan connectivity, differences in glycan-peptide connectivity are fully resolved by PGC. Isomeric glycopeptides of composition GEPTSTPT+$GalNAc_1Gal_1NeuAc$, eluting at 15.4 and 19.0 min, have identical glycan structures but differ in their point of attachment to the peptide moiety. The earlier-eluting isomer is glycosylated at ^{152}Thr, whereas the later-eluting isomer is glycosylated at ^{154}Thr. In each of these cases, differences in the structure of the glycopeptide affected retention by PGC independently of molecular composition.

A combination of these effects may be observed in digests of RNAse B (Fig. 1). Glycopeptide interaction with PGC was modulated by both the polarity and the steric nature of the molecule. Comparison of glycopeptides with identical glycosylation but different peptide moieties revealed that each peptide moiety imparts a different

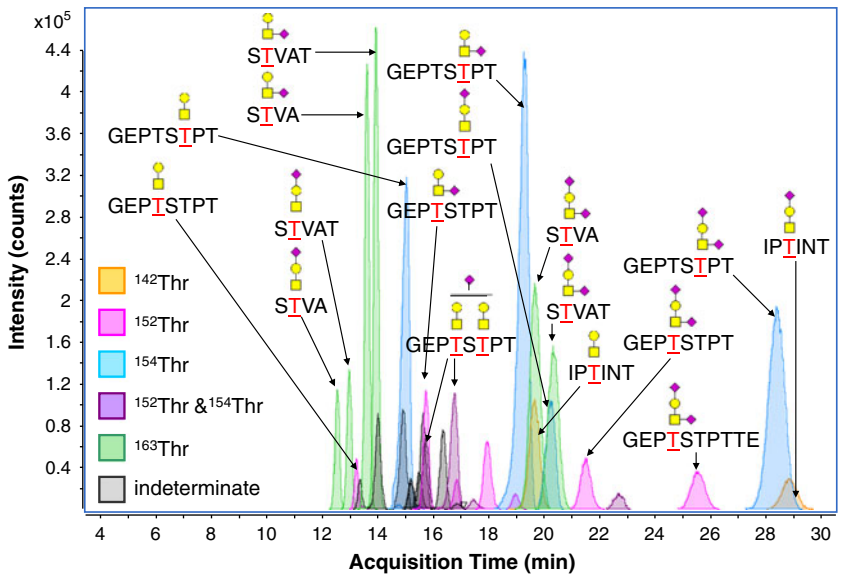

Fig. 3 Overlaid chromatograms and associated structural assignments of glycopeptides from bovine κ-casein. *Color* denotes the site(s) of glycosylation from which the glycopeptide originated

retention time shift and/or scale factor. For example, glycopeptides NLTKDR+$GlcNAc_2Man_5$ (13.402 min) and NLTKD+$GlcNAc_2Man_5$ (14.644 min) differ only by the presence of a basic arginine (R) in the earlier-eluting compound. The basicity of the arginine may offset the acidity of the adjacent aspartic acid (D), lowering the retention exhibited by the PGC. Similarly, the additional presence of a basic lysine in glycopeptide NLTK+ $GlcNAc_2Man_6$ (10.997 min) may cause it to elute earlier than its non-lysinated counterpart NLT+$GlcNAc_2Man_6$ at (11.914 min).

The separative ability of PGC was crucial for differentiating both *N*- and *O*-linked glycopeptide isomers. Numerous high-mannose glycopeptide isomers were resolved from RNAse B and bLF (Table 1). Notably, multiple isomers of Man6 and Man9 were observed. Traditional perceptions of Man6 and Man9 are of a single structure rather than a heterogeneous mix of isomers. However, Electronic Supplementary Material Figures S1 and S2 show Man9 isomers at ^{564}Asn and ^{252}Asn on bLF, while Electronic Supplementary Material S3 shows Man6 isomers at ^{252}Asn. These results support and expand on observations by Liang et al. [29] and Prien et al. [30] on isomers of RNAse B glycoforms. Accurate mass MS/MS confirms the composition of these compounds (Fig. 4) but does not provide sufficient information for de novo structural assignment.

PGC exhibited similar isomer separation abilities with glycopeptides from *O*-linked κ-casein. As seen in Fig. 3, glycopeptides from κ-casein were fully resolved based not only on glycan structure (Gal- or GalNAc-conjugated NeuAc) but also site of glycosylation (e.g., ^{152}Thr vs. ^{154}Thr). Isomeric pairs and quadruplets were routinely separated and elucidated. Tandem MS was used to identify and assign structures to these isomers (Figs. 5 and 6).

Our nano-LC/MS method separated and detected 91 glycopeptides across four orders of magnitude in the RNAse B digest (Fig. 1 and Electronic Supplementary Material Figure S1), 403 glycopeptides across five orders of magnitude in the bLF digest (Fig. 7 and Electronic Supplementary Material Figure S2), 48 glycopeptides across four orders of magnitude in the IgG digest (Fig. 2 and Electronic Supplementary Material Figure S3), and 947 glycopeptides across five orders of magnitude in the κ-casein digest (Fig. 3 and Electronic Supplementary Material Figure S4). The heterogeneity of the glycoprotein digests varied based on the number of glycosites exhibited as well as glycan heterogeneity and dynamic range at each glycosite. These results, in conjunction with tandem data, allowed us to identify 13 site-specific glycans at one *N*-glycosylation site on RNAse B, 59 site-specific glycans at five *N*-glycosylation sites on bLF, 13 site-specific glycans at one *N*-glycosylation site on four subclasses of human IgG, and 20 site-specific glycans at five *O*-glycosylation sites on bovine κ-casein.

Table 1 Number of isomeric glycoforms at each site of glycosylation for each high-mannose glycan composition on bovine ribonuclease B and bovine lactoferrin

	Glycan	Isomers
Bovine ribonuclease B (RNAse B)		
Asn 60	Man5	2
	6	2
	7	3
	8	4
	9	2
Bovine lactoferrin (bLF)		
Asn 252	Man5	3
	6	5
	7	3
	8	3
	9	2
Asn 300	Man5	2
	6	2
	7	–
	8	2
	9	1
Asn 387	Man5	2
	6	3
	7	2
	8	3
	9	1
Asn 495	Man5	2
	6	3
	7	3
	8	3
	9	1
Asn 564	Man5	2
	6	4
	7	3
	8	3
	9	2

Results are a conservative estimate based on observed glycopeptide isomers

Confirmation of site-specific analysis by LC/MS/MS

To supplement information on glycoprotein digests gained from single-stage accurate-mass MS, observation of chromatographic tendencies, and knowledge of Pronase activity, tandem MS experiments were performed in order to independently confirm glycopeptide identities. This allowed rapid differentiation between isobaric glycopeptide assignments with similar accurate masses but different

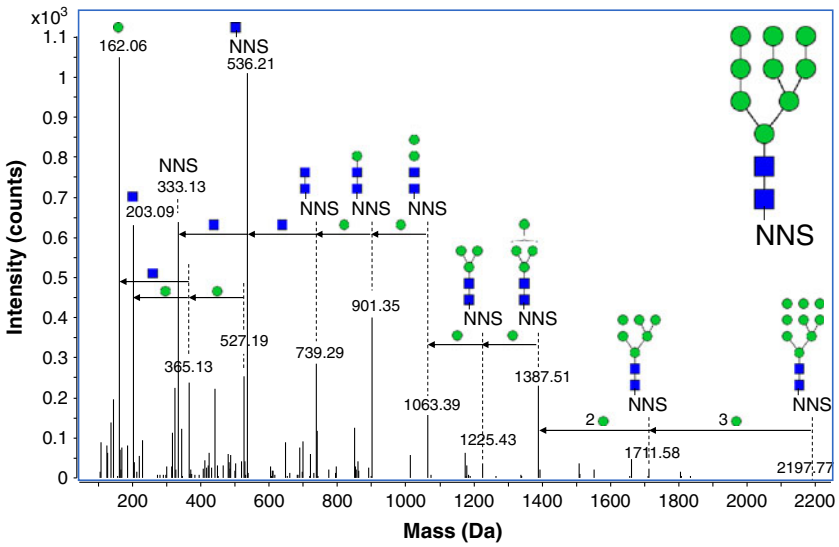

Fig. 4 Deconvoluted tandem MS spectrum of a bovine lactoferrin glycopeptide. The presence of lone peptide as well as HexNAc-conjugated peptide fragment peaks identify ^{252}NNS+Man9 as the correct glycopeptide composition

compositions. In addition, use of LC/MS/MS was invaluable when dealing with sets of stereo- and regio-isomeric glycopeptides.

Glycopeptides were selected for LC/MS/MS analysis based on single-stage LC/MS data analyzed by MassHunter Qualitative Analysis and Glycopeptide Finder. Collision-induced dissociation (CID) was performed at different collision cell voltages so as to produce informative fragments from both the glycan and peptide moieties. Lower energy CID (collision cell voltages of 5–15 V for glycopeptides of *m/z* 500–1,500 Da) produced a variegated selection of glycan fragments that helped to differentiate structural isomers of the glycan moiety. In contrast, higher energy CID (collision cell voltages of 15–50 V for glycopeptides of *m/z* 500–1,500 Da) resulted in peptide fragmentation, which helped to differentiate glycosite isomers as well as isobaric peptide moieties. Resulting tandem spectra were filtered using 5.0 signal-to-noise ratio.

Compositional assignment by tandem MS was based largely upon identification of product ions corresponding to the lone peptide moiety as well as the HexNAc-conjugated peptide. Due to the relative lability of glycosidic bonds in comparison with peptide bonds, these two fragments were consistently found in high abundance in tandem spectra of *N*-linked high mannose, *N*-linked complex, and *O*-linked glycopeptides across all tested glycoprotein digests. Based on the *m/z* of these fragments, the accurate mass of the peptide moiety was determined, providing an easy method of differentiating isobaric glycopeptides.

Tandem MS was also crucial when determining sites of glycosylation in *O*-linked glycoproteins. *O*-glycosylation tends to occur only in protein domains that are rich in proline, serine, threonine, and alanine; thus, *O*-glycoproteins often have multiple sites of *O*-glycosylation within close proximity to each other [31]. Whereas Pronase digestion

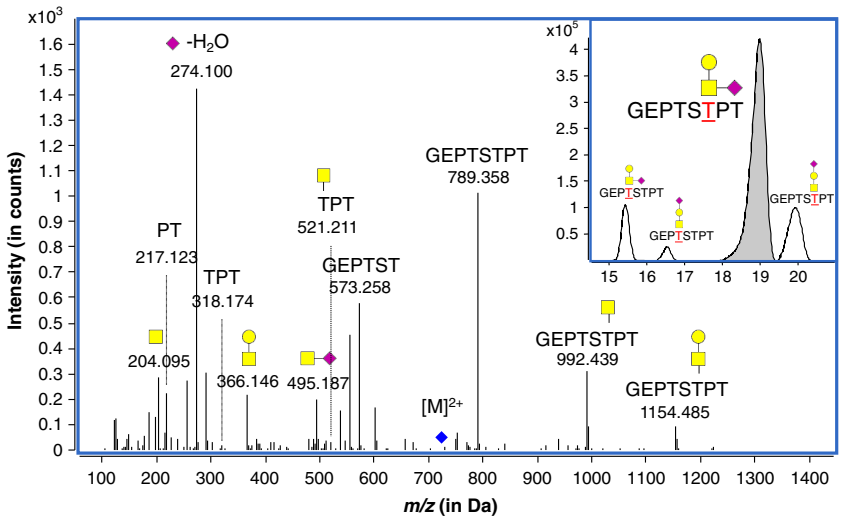

Fig. 5 Tandem MS spectrum of a bovine κ-casein glycopeptide with three isomers. Extensive peptide and glycan fragmentation enable complete site- and structure-specific assignment. *Inset*, chromatograms and structures of the glycopeptide (*shaded*) and its three isomers

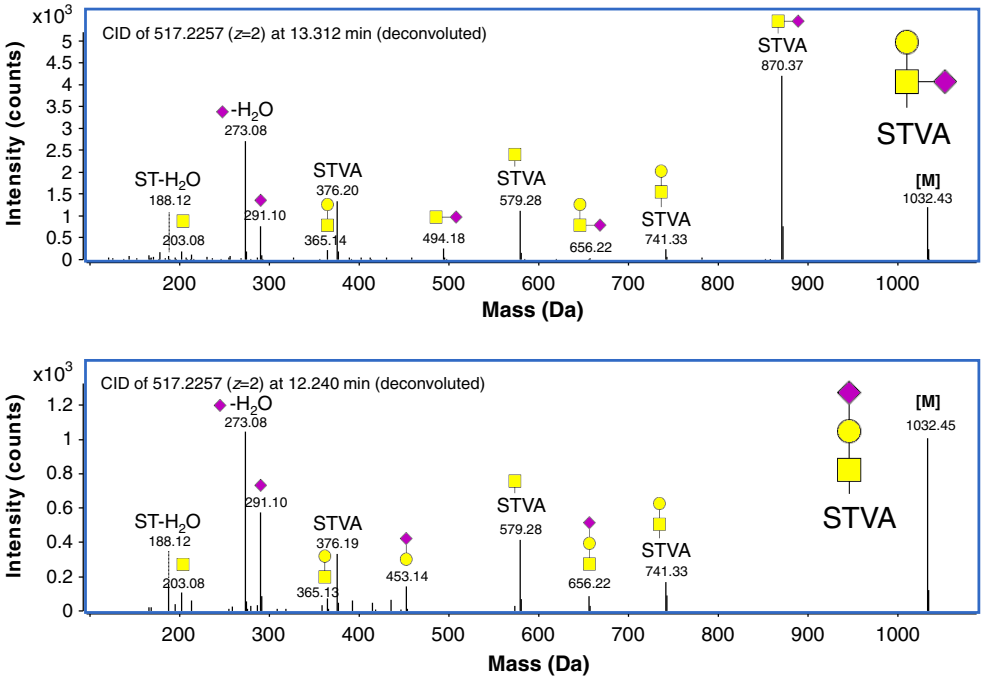

Fig. 6 Deconvoluted tandem MS spectra of two glycopeptide isomers found in bovine κ-casein. Diagnostic Y-ions at 494.18 Da (*top*) and 453.14 Da (*bottom*) as well as B-ion 870.37 Da (top) confirm the structural assignment

could be relied upon to produce *N*-glycopeptides containing only a single glycosite, Pronase-digested *O*-glycopeptides often contained multiple glycosites on a single glycopeptide. Pronase digests of *O*-glycoprotein κ-casein, for example, produced glycopeptides that were four to ten amino acids long containing between one and three glycosites. This was still vastly preferable to trypsin, which would have produced a single 53-amino acid-long glycopeptide containing all six potential sites of *O*-glycosylation; however, the possibility of more than one glycosite on a glycopeptide necessitated MS/MS in order to pinpoint the actual site(s) of glycosylation on the glycopeptide in question.

Figure 5 shows extensive peptide fragmentation in the MS/MS spectrum of the κ-casein glycopeptide GEP^{152}TS^{154}TPT+GalNAc$_1$Gal$_1$NeuAc at retention time 19.0 min. Peptide b and y fragment ions corresponding to ^{154}TPT, PT, PT-H$_2$O, GEP^{152}TS^{154}T, and GEP^{152}TS^{154}T-H$_2$O were observed in the MS/MS data. In addition, glycosite-informative glycopeptide fragments corresponding to ^{152}TS^{154}T+GalNAc (*m/z*511.2), S^{154}TP+GalNAc-H$_2$O (*m/z*489.2), and ^{154}TPT+GalNAc (*m/z*521.2) were observed. The only point of overlap in the fragmented peptide moieties, ^{154}T, was identified as the site of glycosylation for the isomer of GEP^{152}TS^{154}TPT+GalNAc$_1$Gal$_1$NeuAc

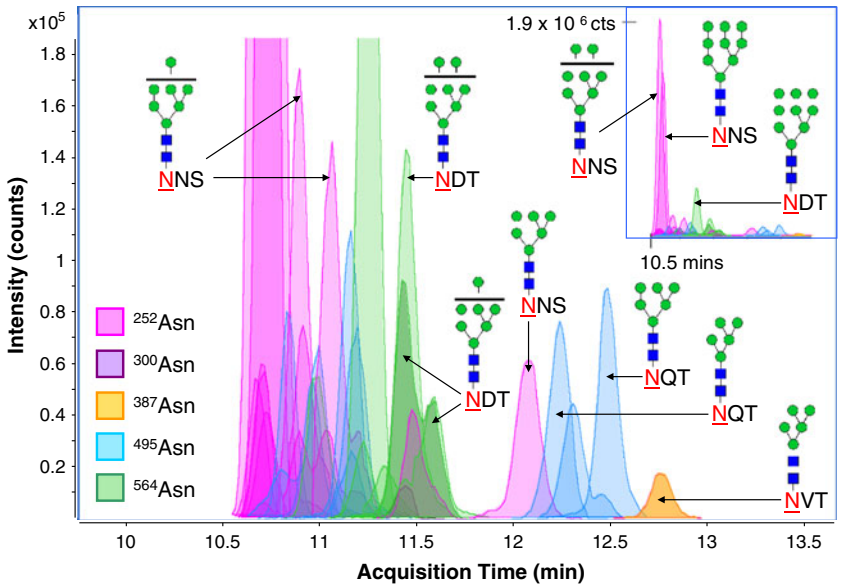

Fig. 7 Overlaid chromatograms and associated structural assignments of glycopeptides from bovine lactoferrin. *Inset*, a zoomed-out view of the same chromatograms, showing the dynamic range of the glycopeptide mixture. *Color* denotes the site of glycosylation from which the glycopeptide originated

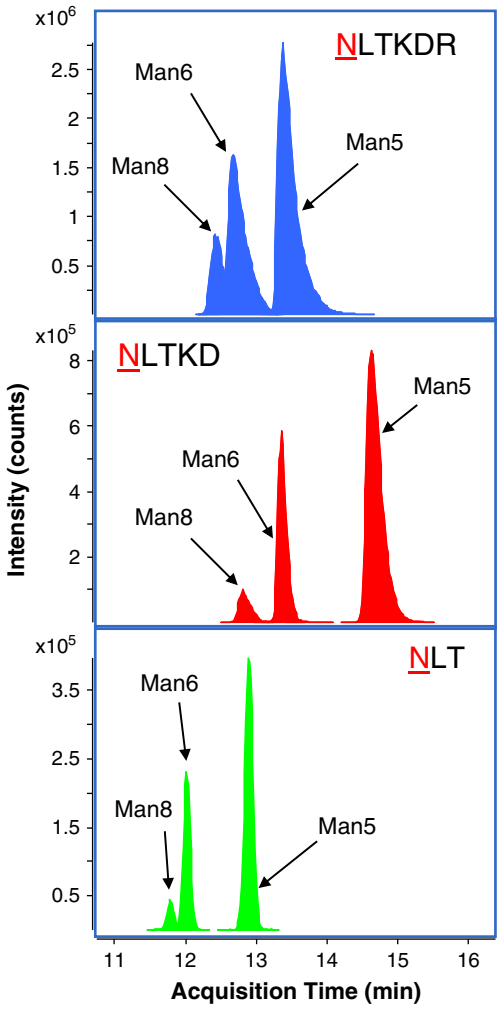

Fig. 8 Chromatograms of the glycopeptides from bovine ribonuclease B associated with each of the three most abundant peptide moieties: ^{60}NLTKDR (*top*), ^{60}NLTKD (*middle*), and ^{60}NLT (*bottom*). Glycoform elution order remains unchanged despite differences in peptide moiety

eluting at 19.0 min. Note that, while biological information about potential glycosites was available in this case, the information provided by LC/MS/MS would have been sufficient even for de novo glycosite determination.

Most intriguingly, isomeric glycopeptides of identical mass and composition but slightly different connectivity in the glycan moiety, once separated by nano-LC, could be differentiated by MS/MS. Figure 6 shows the MS/MS spectra of two *O*-glycopeptide isomers found in κ-casein. Both glycopeptides are monosialylated core 1 type *O*-linked with composition $S^{163}TVA+GalNAc_1Gal_1NeuAc$, but the isomers differ in the attachment site of the terminal NeuAc (either Gal- or GalNAc-conjugated). Diagnostic glycan B-ions corresponding to Gal_1NeuAc (494.2 Da) and $GalNAc_1NeuAc$ (453.1 Da) were used to differentiate the isomers. The high abundance in the isomer at 13.3 min of Y-ion 870.4 Da, corresponding to loss of a terminal Gal, combined with its absence in the isomer at 12.2 min, confirmed the structural assignment.

Quantitative site-specific glycan profiling

The glycopeptide information gained from our LC/MS and LC/MS/MS methods, when pooled together, allow us to easily determine which glycan structures are present at each site of glycosylation. In addition, ion counting adds a quantitative aspect to our glycoproteomic analysis. Nonspecific Pronase digestions create glycopeptides of similar (and short) peptide moiety lengths, minimizing potential differences in ionization efficiency. Since the glycan moieties being compared are of similar size, structure, and acidity, their ionization efficiencies are also similar [32], enabling comparison of glycopeptide ion abundances.

Validation of the quantitative profiling method was performed on well-characterized protein RNAse B. Glycopeptides with a common peptide moiety were grouped together into glycopeptide series consisting of several chromatographic peaks which differed in mass by some multiple of 162 Da, i.e., a mannose residue.

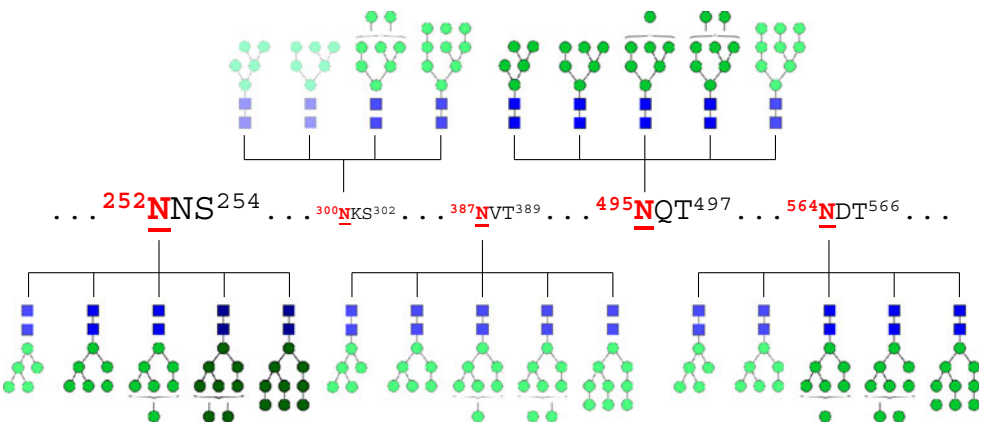

Fig. 9 Glycosite occupancy and heterogeneity in bovine lactoferrin. Font size of each glycosite (*in red*) is directly correlated with glycan abundance at that site. Glycans drawn with *darker shading* are present in higher abundance, while those with *lighter shading* are less abundant

Cross-comparison of different glycopeptide series revealed similar relative abundances for a given glycan moiety (Fig. 8). Man5 was consistently the most abundant glycan moiety in each series, accounting for an average of 51.3% of total ion abundance within each series, followed by Man6, Man8, Man7, and finally Man9. Relative glycoform ratios within a series were also fairly consistent. The glycopeptide series associated with peptide moiety ^{60}NLTKD showed a Man5:Man6:Man8 ratio of 1:2.4:10.7, while the series associated with ^{60}NLT showed a similar ratio of 1:2.0:10.8. Minor series-to-series variations in relative abundances may be attributed to the different steric effects of each glycan structure during protease digestion.

Overall relative quantitation of the glycoforms of RNAse B was performed by summing up the glycopeptide abundances associated with each high-mannose glycan, irrespective of peptide. For the five high-mannose glycoforms Man5-9 of RNAse B, it was determined that Man5 was most abundant, comprising 48% of glycosylated RNAse B, followed by Man6 at 33%, Man8 at 9%, Man7 at 7%, and finally, Man9 at 2%. These results, summarized in Fig. 1 (inset), mirror the relative abundances exhibited by each individual peptide series and are supported by previous works [17, 33].

The quantitative profiling method was applied to several *N*- and *O*-glycoproteins. *N*-glycoproteins profiled included bLF, which exhibits mainly high-mannose N-glycosylation at five potential *N*-linked sites. *O*-glycoproteins profiled included κ-casein, which exhibits core 1 *O*-glycosylation at up to six sites of glycosylation [25].

In *N*-glycosylated protein bLF, high-mannose glycosylation was found to be most abundant at ^{252}Asn, with 65% of all high-mannose glycans originating from this glycosylation site, followed by ^{495}Asn (16%), ^{564}Asn (15%), ^{387}Asn (3%), and ^{300}Asn (1%). Site-to-site differences in the relative abundances of each high-mannose glycan were also quantified. The most abundant glycan overall was Man_8 (38%), due to the high frequency of Man_8 glycosylation at ^{252}Asn. Other sites had different patterns of high-mannose glycosylation, with Man_5 favored at ^{495}Asn and Man_9 favored at ^{564}Asn. Results are summarized in Fig. 9. System-wide levels of high-mannose glycosylation in bLF spanned five orders of magnitude, with spans of up to three orders of magnitude at any one site of glycosylation.

In *O*-glycosylated protein κ-casein, quantitative profiling showed that glycosylation was most abundant at ^{154}Thr, with 41% of all glycosylation occurring at this site. This was followed by ^{163}Thr (29%), ^{152}Thr (14%), ^{142}Thr (7%), and ^{157}Thr (0.1%). An additional 9% of total glycosylation occurred at ^{152}Thr, ^{154}Thr, or ^{157}Thr; however, the specific site of glycosylation could not be narrowed down further due to the extreme proximity of these glycosites. The most abundant glycoform overall was the monosialylated core 1 type glycan (49%), followed by the disialylated core 1 type glycan (26%) and the neutral, unsialylated core 1 type glycan (18%). Due to ambiguities associated with having multiple glycans conjugated to a single peptide, the final 6% were assigned simply as sialylated core 1 type glycans, without distinction between mono- or disialylation.

Conclusions

The results presented demonstrate for the first time site-specific, isomer-specific profiling of *N*- and *O*-linked glycosylation. Chromatographic separation of complex glycopeptide mixtures and subsequent identification of structure by both accurate mass and tandem MS allow us to quantitatively describe glycan heterogeneity in a site-specific manner. Glycoforms of *N*- and *O*-linked glycoproteins were comprehensively profiled. The quantitative nature of nano-LC/MS analysis allowed us to obtain a complete picture of glycan abundances and varieties at each glycosite. In addition, the isomer separation properties of PGC allowed the separate quantification of each isomer.

The present study underscores the value of nonspecific proteolysis methods in glycoproteomic analysis. The shorter peptide chains obtained by Pronase digestion allowed greater retention of glycan character and thus effective separation by PGC. Modulation of digestion time enabled scalable peptide tag size and facilitated identification of glycopeptide composition based on accurate mass. Work is ongoing to develop computer algorithms and automated data processing methods which will streamline the data analysis and allow rapid interpretation of vast quantities of mass spectral data.

Site-specific glycoproteomic analysis has a number of important analytical and biological applications. Changes in glycosylation have been linked to various pathologic states including cancer [34, 35], atherosclerosis [36], and rheumatoid arthritis [37], as well as a number of infectious diseases [1, 2, 38]. The extra information afforded by site-specific analysis could prove invaluable to the discovery of new disease biomarkers. Ultimately, the ability to track changes in glycoprotein heterogeneity during a biological process will help provide a new paradigm for understanding the role of the glycoproteome.

Acknowledgments We would like to thank Ning Tang and Keith Waddell (Agilent Technologies Inc.) for instrumentation and technical support. Financial support was provided by the University of California Discovery Grant Program, the California Dairy Research Foundation, the NIEHS Superfund Research Program (P42 ES02710), and the CHARGE Study (P01 ES11269).

References

1. Ohtsubo K, Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. Cell 126:855–867
2. Dennis JW, Granovsky M, Warren CE (1999) Protein glycosylation in development and disease. BioEssays 21:412–421
3. Heo S-H, Lee S-J, Ryoo H-M, Park J-Y, Cho J-Y (2007) Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. Proteomics 7:4292–4302
4. Pan S, Wang Y, Quinn JF, Peskind ER, Waichunas D, Wimberger JT, Jin J, Li JG, Zhu D, Pan C, Zhang J (2006) Identification of glycoproteins in human cerebrospinal fluid with a complementary proteomic approach. J Proteome Res 5:2769–2779
5. Liu T, Qian W-J, Gritsenko MA, Camp DG, Monroe ME, Moore RJ, Smith RD (2005) Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J Proteome Res 4:2070–2080
6. Sun B, Ranish JA, Utleg AG, White JT, Yan X, Lin B, Hood L (2007) Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. Mol Cell Proteomics 6:141–149
7. Qiu R, Regnier FE (2005) Use of multidimensional lectin affinity chromatography in differential glycoproteomics. Anal Chem 77:2802–2809
8. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat Biotech 24:1241–1252
9. Baum LG (2002) Developing a taste for sweets. Immunity 16:5–8
10. Harazono A, Kawasaki N, Itoh S, Hashii N, Matsuishi-Nakajima Y, Kawanishi T, Yamaguchi T (2008) Simultaneous glycosylation analysis of human serum glycoproteins by high-performance liquid chromatography/tandem mass spectrometry. J Chromatogr B 869:20–30
11. Alley WR Jr, Mechref Y, Novotny MV (2009) Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. Rapid Commun Mass Spectrom 23:495–505
12. Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F (2008) Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. Proteomics 8:2858–2871
13. Huddleston MJ, Bean MF, Carr SA (1993) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. Anal Chem 65:877–884
14. Bezouska K, Sklenár J, Novák P, Halada P, Havlícek V, Kraus M, Tichá M, Jonáková V (1999) Determination of the complete covalent structure of the major glycoform of DQH sperm surface protein, a novel trypsin-resistant boar seminal plasma O-glycoprotein related to pB1 protein. Protein Sci 8:1551–1556
15. Schnitzer JE, Carley WW, Palade GE (1988) Albumin interacts specifically with a 60-kDa microvascular endothelial glycoprotein. Proc Natl Acad Sci USA 85:6773–6777
16. Cauchi MR, Henchal EA, Wright PJ (1991) The sensitivity of cell-associated dengue virus proteins to trypsin and the detection of trypsin-resistant fragments of the nonstructural glycoprotein NS1. Virology 180:659–667
17. An HJ, Peavy TR, Hedrick JL, Lebrilla CB (2003) Determination of N-glycosylation sites and site heterogeneity in glycoproteins. Anal Chem 75:5628–5637
18. Clowers BH, Dodds ED, Seipert RR, Lebrilla CB (2007) Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and Fourier transform ion cyclotron resonance mass spectrometry. J Proteome Res 6:4032–4040
19. Dodds ED, Seipert RR, Clowers BH, German JB, Lebrilla CB (2008) Analytical performance of immobilized pronase for glycopeptide footprinting and implications for surpassing reductionist glycoproteomics. J Proteome Res 8:502–512
20. Seipert RR, Dodds ED, Lebrilla CB (2008) Exploiting differential dissociation chemistries of O-linked glycopeptide ions for the localization of mucin-type protein glycosylation. J Proteome Res 8:493–501
21. Ninonuevo M, An H, Yin H, Killeen K, Grimm R, Ward R, German B, Lebrilla C (2005) Nanoliquid chromatography-mass spectrometry of oligosaccharides employing graphitized carbon chromatography on microchip with a high-accuracy mass analyzer. Electrophoresis 26:3641–3649
22. Packer NH, Lawson MA, Jardine DR, Redmond JW (1998) A general approach to desalting oligosaccharides released from glycoproteins. Glycoconj J 15:737–747
23. Hardy MR, Townsend RR (1988) Separation of positional isomers of oligosaccharides and glycopeptides by high-performance anion-exchange chromatography with pulsed amperometric detection. Proc Natl Acad Sci USA 85:3289–3293
24. Davies M, Smith KD, Harbin A-M, Hounsell EF (1992) High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. J Chromatogr A 609:125–131
25. Consortium TU (2010) The universal protein resource (UniProt) in 2010. Nucl Acids Res 38:D142–D148
26. Kronewitter SR, An HJ, de Leoz ML, Lebrilla CB, Miyamoto S, Leiserowitz GS (2009) The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. Proteomics 9:2986–2994
27. Chu CS, Niñonuevo MR, Clowers BH, Perkins PD, An HJ, Yin H, Killeen K, Miyamoto S, Grimm R, Lebrilla CB (2009) Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. Proteomics 9:1939–1951
28. Koizumi K, Okada Y, Fukuda M (1991) High-performance liquid chromatography of mono- and oligo-saccharides on a graphitized carbon column. Carb Res 215:67–80
29. Liang C-J, Yamashita K, Kobata A (1980) Structural study of the carbohydrate moiety of bovine pancreatic ribonuclease B. J Biochem 88:51–58
30. Prien J, Ashline D, Lapadula A, Zhang H, Reinhold V (2009) The high mannose glycans from bovine ribonuclease B isomer characterization by ion trap MS. J Amer Soc Mass Spectrom 20:539–556
31. Wilson IB, Gavel Y, Von Heijne G (1991) Amino-acid distributions around o-linked glycosylation sites. Biochem J 275:529–534
32. Cancilla MT, Wong AW, Voss LR, Lebrilla CB (1999) Fragmentation reactions in the mass spectrometry analysis of neutral oligosaccharides. Anal Chem 71:3206–3218
33. Juhasz P, Martin SA (1997) The utility of nonspecific proteases in the characterization of glycoproteins by high-resolution time-of-flight mass spectrometry. Int J Mass Spectrom Ion Proc 169–170:217–230
34. Fuster MM, Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. Nat Rev Cancer 5:526–542
35. Dube DH, Bertozzi CR (2005) Glycans in cancer and inflammation—potential for therapeutics and diagnostics. Nat Rev Drug Discov 4:477–488
36. Ge J, Jia Q, Liang C, Luo Y, Huang D, Sun A, Wang K, Zou Y, Chen H (2005) Advanced glycosylation end products might promote atherosclerosis through inducing the immune maturation of dendritic cells. Arterioscler Thromb Vasc Biol 25:2157–2163
37. Hänsler M, Kötz K, Häntzschel H (1995) Detection of immunoglobulin G glycosylation changes in patients with rheumatoid arthritis by means of isoelectric focusing and lectin-affinoblotting. Electrophoresis 16:811–812
38. Cooke CL, An HJ, Kim J, Canfield DR, Torres J, Lebrilla CB, Solnick JV (2009) Modification of gastric mucin oligosaccharide expression in rhesus macaques after infection with *Helicobacter pylori*. Gastroenterol 137:1061–1071, 1071 e1-8