Glyco-Analytical Multispecific Proteolysis (Glyco-AMP): A Simple Method for Detailed and Quantitative Glycoproteomic Characterization

Serenus Hua,‡§ Chloe Y. Hu,§ Bum Jin Kim,¶ Sarah M. Totten,§ Myung Jin Oh,∥ Nayoung Yun,¶ Charles C. Nwosu,§ Jong Shin Yoo,¶,‖ Carlito B. Lebrilla,§,* and Hyun Joo An*¶,∥,†

INTRODUCTION

Proteins are commonly decorated with long carbohydrate chains, known as glycans, during normal biosynthesis. Over 60% of the known proteome is believed to be glycosylated.1 As the importance of glycosylation toward protein structure and function becomes ever more apparent, glycomic methods for the importance of glycosylation toward protein structure and function have undergone extensive innovation and development.2−4 However, glycoproteomic methods that characterize both glycosylation site and glycan structure are still in their infancy. Until now, the most commonly used methods have been adapted by their linear sequence, glycans are branched, with different linkage possibilities between each monosaccharide residue. When peptide and glycan moieties both exist on a single molecule, as with glycopeptides, peptide sequence isomers and glycosylation site isomers further increase the already significant structural complexity.8 The abundance of isomeric glycopeptides encountered during glycoproteomic analysis is problematic for mass spectrometry, which cannot distinguish between isomers without the assistance of additional analytical technologies. Consequently, glycoproteomic methods incorporating isomer-specific chromatographic separation and/or tandem MS are crucial for effective glycoprotein characterization.9

When analyzing peptides with post-translational modifications such as glycosylation, peptide length must be considered. Large glycopeptides are not only harder to detect by mass spectrometry (due to decreased ionization efficiency as well as instrumental mass limitations) but can also incorporate multiple glycosylation sites, severely complicating or obfuscating site-specific analysis.10−12 Proteases with high substrate specificity, such as trypsin, are especially prone to creating these large glycopeptides, since their specificity severely limits

Received: May 10, 2013
Published: September 9, 2013
potential cleavage sites. This is exacerbated by the well-known phenomenon of glycoprotein trypsin resistance, wherein large glycosyl modifications sterically hinder protease access to a glycoprotein cleavage site, resulting in a missed cleavage.

While the amino acid sequences of some glycoproteins may be fortuitously rich in tryptic cleavage sites, many are not so easily dissected. The glycoproteomics community is therefore in great need of alternative proteases and proteolytic digestion strategies that can be broadly applied to enhance and/or replace conventional trypsin-based methodology. One such strategy is multispecific (or nonspecific) proteolysis, which utilizes proteases or protease cocktails with multiple substrate specificities to hydrolyze peptide bonds at a number of different sites on a glycoprotein. The resulting digest contains glycopeptides of a roughly consistent size, irrespective of the glycoprotein’s amino acid sequence or steric hindrance by the glycan moiety.

Early forays into glycoprotein characterization via multispecific proteolysis involved MS-only analysis of single glycoprotein digests. Sensitivity and specificity were relatively low, due to interference from ion-suppressing unglycosylated peptides. Later iterations of the strategy used various chemical methods to immobilize the multispecific proteases on solid supports, removing a significant amount of the peptide interference originating from protease autolysis. However, the most dramatic improvements in glycoproteomics have come just recently, with the incorporation of nano-LC separation into established mass spectrometric methods.

Use of nano-LC–MS not only boosts sensitivity by vastly reducing interference from ion-suppressing peptides, but also enables differentiation of isobaric or even isomeric molecules. Additionally, condensation of the in-spectrum dynamic range enables increased use of MS/MS to supplement accurate mass MS, adding an extra layer of certainty to glycopeptide identifications.

In this study, isomer-sensitive porous graphitized carbon nano-LC–MS and nano-LC–MS/MS are used to interrogate glycopeptide mixtures digested in solution by various multispecific proteases. Because the glycoprotein digestions are carried out in solution, rather than by bead-immobilized proteases, sample preparation is vastly simplified while experimental variation is minimized. Additionally, the in-solution protease digestions inherently proceed faster than bead-immobilized protease digestions due to more favorable kinetics. Although in-solution protease digest do contain higher levels of signal-suppressing peptides than immobilized protease digestes, nano-LC successfully separates the peptide contaminants from the glycopeptide analytes—in essence, solving the problem with analytical technology rather than chemical derivatization.

Whereas previous studies have concentrated on the analytical performance of individual multispecific proteases, commonly pronase, we demonstrate the general applicability of glycoanalytical multispecific proteolysis (Glyco-AMP) by performing detailed and quantitative characterization of site-specific protein glycosylation using a variety of multispecific proteases. The accuracy and sensitivity of the nano-LC–MS profiling methods are evaluated. Basic digestion kinetics and activity of multispecific proteases subtilisin, pronase, and proteinase K are tracked and compared. Site-specific glycoform characterization is demonstrated to be effective on multiply glycosylated glycoproteins as well as glycoprotein mixtures. Isomer separation by porous graphitized carbon nano-LC, combined with MS/MS structural elucidation, reveals fine details about the structure and originating site of the glycopeptide.

### MATERIALS AND METHODS

#### Materials and Reagents

All multispecific proteases (porcine elastase, papain, porcine pepsin, pronase, proteinase K, subtilisin from Bacillus licheniformis, and thermolysin) were obtained from Sigma-Aldrich (St Louis, MO), as were glycoproteins bovine ribonuclease B and human plasma vitronectin. Prostate-specific antigen was obtained from Lee BioSolutions (St Louis, MO). Influnixam was obtained from Janssen Biotech (Horsham, PA). Darbepoetin alfa was obtained from Amgen (Thousand Oaks, CA). Graphitized carbon cartridges were obtained from Grace Davison (Deerfield, IL). Solvents were of LC–MS grade. All other materials and reagents were of analytical grade or higher.

#### Digestion of Glycoproteins by Multispecific Proteolysis

In general, 50 µg of glycoprotein and 50 µg of protease were dissolved in 100 µL of 100 mM phosphate buffer (pH 7.5) and incubated at 37 °C for 20 min. Exceptions occurred during: (a) initial protease screening, for which the optimal digestion buffer and temperature for each protease was specified by the manufacturer’s product notes; (b) kinetics studies, for which multiple digestion time points were taken; and (c) sensitivity studies, for which the amount of glycoprotein was varied.

#### Glycopeptide Enrichment with Graphitized Carbon SPE

Digested glycopeptides were enriched by graphitized carbon solid-phase extraction according to previously optimized procedures. Briefly, an automated liquid handler (Gilson, Middleton, WI) conditioned graphitized carbon cartridges with water; loaded aqueous glycopeptide solutions; washed with water; then eluted glycopeptides with 40% acetonitrile and 0.05% trifluoroacetic acid (v/v) in water. Samples were dried in vacuo.

#### Chromatographic Separation and MS Analysis

Samples were reconstituted in water and analyzed using an HPLC-Chip Quadrupole Time-of-Flight (Chip/Q-TOF) MS system (Agilent Technologies, Santa Clara, CA) comprising an autosampler (maintained at 6 °C), capillary pump, nano pump, HPLC-Chip/MS interface, and a 6520 Q-TOF MS detector. The chip used consisted of a 9 × 0.075 mm i.d. enrichment column and a 43 × 0.075 mm i.d. analytical column, both packed with 5 µm porous graphitized carbon as the stationary phase, with an integrated nano-ESI spray tip. Chromatographic separation was performed according to previously optimized procedures.

Briefly, following sample loading, a rapid glycopeptide elution gradient was delivered at 0.4 µL min⁻¹ using solutions of (A) 3.0% acetonitrile and 0.1% formic acid (v/v) in water, and (B) 90.0% acetonitrile and 0.5% formic acid (v/v) in water, at the following proportions and time points: 5% to 32.8% B, 0 to 13.3 min; and 32.8% to 35.9% B, 13.3 to 16.5 min. The columns were then flushed with 100% B, and the analytical column was re-equilibrated with 5% B, while the enrichment column was re-equilibrated with 0% B. The drying gas temperature was set at 325 °C with a flow rate of 4 L min⁻¹ (2 L of filtered nitrogen gas and 2 L of filtered dry compressed air).

MS spectra were acquired in positive ionization mode over a mass range of m/z 500–2000 with an acquisition time of 1.5 s per spectrum. MS/MS spectra were acquired in positive ionization mode over a mass range of m/z 100–3000 with an...
acquisition time of 1.5 s per spectrum. Following an MS scan, precursor compounds were automatically selected for MS/MS analysis by the acquisition software based on ion abundance and charge state \((z = 2, 3, \text{ or } 4)\) and isolated in the quadrupole with a mass bandpass fwhm (full width at half-maximum) of 1.3 \(m/z\). In general, collision energies for CID fragmentation were calculated for each precursor compound based on the following formula:

\[
v_{\text{collision}} = 3.6 \sqrt{\frac{m/z}{100 \text{ Da}}} - 4.8 \text{ V}
\]

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.\(^8\)

**LC–MS Data Processing and Glycopeptide Identification**

After data acquisition, raw LC–MS data were processed using the Molecular Feature Extractor algorithm included in the MassHunter Qualitative Analysis software (version B.04.00 SP2, Agilent Technologies). MS peaks were filtered with a signal-to-noise ratio of 5.0 and parsed into individual ion species. Using expected isotopic distribution, charge state information, and retention time, all ion species associated with a single compound (e.g., the doubly protonated ion, the triply protonated ion, and all associated isotopologues) were summed together, and the neutral mass of the compound was calculated. Using this information, a list of all compound peaks in the sample was generated, with abundances represented by chromatographic peak areas.

Glycopeptides were identified using an in-house software tool based on the GlycoX algorithm.\(^27\) Taking into account the mass of a potential glycopeptide, the amino acid sequence(s) of the protein or proteins it could be derived from, the type of glycosylation (N or O), and a given mass tolerance (here, 5 ppm), the algorithm generated a list of potential amino acid and glycan compositions associated with a given compound. Results were further filtered based on known glycosylation patterns and/or other biological rules.\(^28\)

**RESULTS AND DISCUSSION**

**Screening for Multispecific Proteases with High Glycopeptide-Generating Activity**

A number of common, commercially available proteases were initially screened for their ability to digest glycoproteins into (detectable) glycopeptides. Elastase, papain, pepsin, pronase, proteinase K, subtilisin, and thermolysin were each evaluated by combining 50 \(\mu g\) of protease with 50 \(\mu g\) of well-characterized glycoprotein ribonuclease B (at a total digestion volume of 100 \(\mu L\)). Digestion buffers and conditions were chosen based on the manufacturers’ suggested protocols\(^29\)–\(^35\)—for example, the thermolysin digestion took place at 75 °C in 50 mM Tris and 0.5 mM calcium chloride, while the pepsin digestion took place at 37 °C in 10 mM hydrochloric acid. Digestions were carried out for 20 min each, after which the digestes were cleaned by graphitized carbon solid-phase extraction and analyzed by nano-LC–MS/MS.

The glycopeptide content of each digest was evaluated by screening MS/MS spectra for glycopeptide-associated fragment ions with \(m/z\) 163.06 (Hex + H), \(m/z\) 204.09 (HexNAc+H), and \(m/z\) 366.14 (HexNAc2Hex+H). Spectra with significant levels (over 50% base peak abundance) of these fragment ions were assumed to be of glycopeptides. The number and ion intensity of glycopeptide spectra were taken as indicative of the number of glycopeptides.

On the basis of these preliminary evaluations, the subtilisin, pronase, and proteinase K digests were found to contain the highest abundances of glycopeptides. Subtilisin and proteinase K are serine proteases (of bacterial and fungal origin, respectively) that are evolutionarily unrelated yet show a high degree of structural homology, particularly at the catalytic triad.\(^36\) Pronase is an enzyme cocktail of bacterial origin containing mostly serine proteases but also some additional endo- and exoproteases. None of these proteases are themselves glycosylated. All are stable over a wide pH and temperature range, with optimal activity in slightly alkaline conditions at around 40 °C.

**Sensitive and Quantitative Glycoform Profiling through Multispecific Proteolysis**

To evaluate the quantitative accuracy of Glyco-AMP, subtilisin, pronase, and proteinase K were each applied to analyze bovine ribonuclease B (RNase B). RNase B is a well-characterized glycoprotein with a known glycosylation profile, thus providing a convenient means of validation. The relative abundances of each RNase B glycoform were compared after parallel 20 min digestions by subtilisin, pronase, and proteinase K. Overall profiles in Figure 1a are obtained by summing up the abundances of observed glycopeptides associated with each glycoform. To minimize differences in glycopeptide ionization efficiencies, the peptide moieties used for quantitation are kept constant across glycoforms.\(^8\) High mannose type glycan Man\(_n\)GlcNAc\(_z\) (55%) is the most abundant glycan attached to RNase B, followed by Man\(_n\)GlcNAc\(_z\) (27%), Man\(_n\)GlcNAc\(_z\) (9%), Man\(_n\)GlcNAc\(_z\) (7%), and Man\(_n\)GlcNAc\(_z\) (3%); these results correlate highly with previous studies quantifying the glycoforms of RNase B.\(^37\)–\(^39\) All protease digestes result in similar profiles, with the relative abundances of individual glycosylations varying by less than 2% between proteases. For subtilisin, major peptide moieties (exceeding 10% abundance) consist of NLTK (66%) and NLT (13%), while the remaining 21% consists of minor peptide moieties SRNL, KSRNL, NL, and SRNLNT (in decreasing order of abundance). For pronase, major peptide moieties consist of SRN (52%) and RN (21%), while the remaining 27% consists of minor peptide moieties NLT, NLTKDR, N, and NL. For proteinase K, major peptide moieties consist of NLTK (40%), SRNLNT (27%), NLT (14%), and SRNL (13%), while the remaining 6% consists of minor peptide moieties SRNLTKD and NL. Here, as earlier with infliximab, the subtilisin appears to exhibit the highest
cleavage specificity (resulting in a simpler peptide mixture), while the pronase and proteinase K both exhibit much lower cleavage specificity (resulting in a more complex peptide mixture).

The sensitivity of Glyco-AMP was also tested by digesting different amounts of RNase B with subtilisin. The relative abundances of each RNase B glycoform were compared after parallel 20 min digestions from initial glycoprotein amounts of 10 ng, 100 ng, 1 μg, and 50 μg. All initial glycoprotein amounts result in similar profiles, summarized in Figure 1b, with the relative abundances of individual glycoforms varying by less than 5% between digests. As in Figure 1a, these results correlate highly with previous studies quantifying the glycoforms of RNase B.37–39

Since RNase B has a molecular weight of approximately 15 kDa, 10 ng of RNase B corresponds to 670 femtomoles. Additionally, not all of the sample was injected into the LC–MS system—because the autosampler needle could not physically reach the bottom of the sample vial, only 8 μL (out of 25 μL) of glycopeptide solution were drawn from the vial for analysis. Despite all of this, RNase B glycoform Man9GlcNAc2 was still detected at 5% relative abundance, corresponding to the detection (from within a mixture) of just 10 femtomoles of the glycoform.

Characterization of Subtilisin, pronase, And Proteinase K Enzyme Activity

To better understand the glyco-analytical performance of subtilisin, pronase, and proteinase K, in-depth kinetic studies were performed using infliximab (Remicade) as a test substrate. Infliximab is a chimeric monoclonal antibody (mAb) made up of a mouse-derived antigen-binding (Fab) region and a human IgG1-derived constant (Fc) region. Glycosylation is present exclusively on the IgG1-derived constant region, and closely imitates (but exhibits some significant differences from) the glycosylation pattern of IgG. Multispecific proteolysis of infliximab is kinetically similar to multispecific proteolysis of pure IgG1 (without interference from IgG isotypes 2–4) or, more relevantly, other mAbs with an IgG1-derived constant region. Thus, the conclusions derived from these experiments may be applied directly to future glycoproteomic analyses of IgG1-derived mAbs such as adalimumab (Humira), bevacizumab (Avastin), or rituximab (Rituxan), all of which possess perfectly conserved IgG1-like amino acid sequences around the N-glycosylation site. These experiments also provide a rough indication of how each protease will perform when applied to glycoproteomic analyses of more complex glycoproteins with multiple glycosylation sites.
To evaluate and characterize each protease, the abundances of the generated glycopeptides were tracked across six time points: 20 min, 60 min, 2 h, 4 h, 8 h, and 18 h. These six time points were tested in each of the three selected proteases: pronase, proteinase K, and subtilisin. All 18 digests were conducted, processed, and analyzed in parallel. Glycopeptides were identified by nano-LC−MS/MS and quantified by nano-LC−MS.

Figure 2 shows the overlaid chromatograms of glycopeptides separated by nano-LC (and identified by accurate mass MS and MS/MS). Figure 2a shows the glycopeptides generated by 20 min of digestion by subtilisin, whereas Figure 2b shows the glycopeptides generated by 18 h of digestion by subtilisin. These results demonstrate one of the unique advantages of multispecific proteolysis; that is, the ability to modulate digestion time (as well as enzyme concentration) in order to customize the size or length of the peptide moiety on each glycopeptide. The 20 min digestion (Figure 2a) results predominantly in longer glycopeptides with peptide moiety sequences of QYNST (the major product) and QYNSTY (a minor product), identifying the originating protein as IgG1-like rather than IgG2-like. Shorter glycopeptides with peptide moiety sequences of NST and NSTY are also present, supporting this identification. In contrast, the 18 h digestion (Figure 2b) results almost entirely in short glycopeptides possessing the NST peptide moiety, with an extremely minor contribution from glycopeptides with a peptide moiety sequence of NSTY.

Figure 3 provides a preliminary glimpse of the kinetics of a subtilisin digestion, showing the abundance of glycopeptide QYNST-Hex6HexNAc4Fuc (i.e., QYNST-G1F) decreasing over time while the abundance of glycopeptide NST-Hex6HexNAc4Fuc (i.e., NST-G1F) increases over time. Figure 3b plots the absolute abundances of these two glycopeptides at six different points during the 18-h digestion, while Figure 3, parts a and c, shows their chromatograms at these points.

The chromatograms of glycopeptide NST-G1F (Figure 3c) highlight the ability of porous graphitized carbon to separate glycopeptide isomers (provided that they are sufficiently small). Here, the only difference between the two glycopeptide isomers is their G1F glycan moiety. Previous glycan structural studies using the same isomer-sensitive porous graphitized carbon stationary phase indicate that on the earlier-eluting G1F isomer, the antennal galactose is attached to the alpha-1,6-linked branch of the glycan, whereas on the later-eluting G1F isomer, the antennal galactose is attached to the alpha-1,3-linked branch. On the basis of these results, the predominant G1F isomer present on infliximab is the one in which the antennal galactose is attached to the alpha-1,6-linked branch.

In contrast to glycopeptide NST-G1F, glycopeptide QYNST-G1F (Figure 3a) appears to have too long of a peptide moiety for effective isomer separation by porous graphitized carbon nano-LC. However, as mentioned earlier, the presence of a longer peptide moiety (such as QYNST) is still desirable for certain applications because it increases the specificity of the glycosylation site assignment—in this case, by differentiating between IgG isotypes. Detection of glycopeptides with QYNST peptide moieties differentiates glycoforms of infliximab (which is an IgG1 homologue) from glycoforms of IgG2 homologues such as panitumumab (Vectibix) or denosumab (Prolia/Xgeva). In instances of multiply glycosylated proteins or protein mixtures where multiple sites of glycosylation exist, precise control of peptide moiety length enables researchers to select digestion times that would create glycopeptide markers with high specificity for a particular site of glycosylation. These markers could then be used to quantify selected glycoforms of a specific glycoprotein, with implications for biosimilar batch analysis or biomarker research.

Figure 4 tracks the progression over time of infliximab digestion by subtilisin (Figure 4a), pronase (Figure 4b), and proteinase K (Figure 4c). For each digestion time point, the percent contributions of each peptide moiety to the overall glycopeptide population are plotted. Minor peptide moieties (right axis) are denoted by lines connected each time point; major peptide moieties (left axis) do not have lines connecting the time points.
more closely imitating the reaction plots of QYNST and QYNSTY. Presumably, the initial rise is fed largely by digestion of QYNSTY peptide moieties (or even larger precursors) to form NSTY. Simultaneously, NSTY peptide moieties are being digested to form NST. Approximately two hours later, the rate of NSTY formation slows (due to lack of substrate) and is overtaken by the rate of NSTY digestion, resulting in the observed decreases in abundance of NSTY peptide moieties. By the 18 h digestion time point, 96% of all infliximab glycopeptides have NST peptide moieties, while the remaining 4% have NSTY peptide moieties. During the observed course of the reaction, total detected glycopeptide abundances (in ion counts) increase only slightly, by just 4%, indicating that, after the 20 min time point, subtilisin is mainly acting on already-digested short glycopeptides, and very little undigested infliximab remains.

In pronase (Figure 4b), the major peptide moieties are initially NST and YNST, with minor contributions from TKPREEQYNST and EEQYNST. As the digestion progresses, abundances of TKPREEQYNST and EEQYNST drop drastically; abundances of YNST decrease at a slower rate; and abundances of NST increase correspondingly. The slower rate of decrease for YNST peptide moieties may be due in part to digestion of TKPREEQYNST and EEQYNST to form new YNST; however, abundances of NST appear to decrease quite slowly even after abundances of TKPREEQYNST and EEQYNST peptide moieties have dropped below detectable levels. A more likely explanation is that, due to enzyme autolysis, the pronase is losing activity as the digestion progresses. This would explain the lack of significant decreases in YNST abundances (as well as the lack of corresponding increases in NST abundances) after the four hour digestion time point. Even after 18 h of digestion by pronase, only 84% of all infliximab glycopeptides have YNST peptide moieties, while the remaining 16% have NST peptide moieties. During the observed course of the reaction, total detected glycopeptide abundances (in ion counts) increase only slightly, by just 8%, indicating that, after the 20 min time point, pronase is mainly acting on already-digested short glycopeptides, and very little undigested infliximab remains.

In proteinase K (Figure 4c), the major peptide moieties are initially QYNST, TKPREEQYNST, and NST, with a minor contribution from YNST. As the digestion progresses, abundances of QYNST and TKPREEQYNST generally decrease, while abundances of NST and YNST generally increase. Due to the curious behavior of the digestion during the initial two hours, additional time points have been added to the analysis. In all, time points are shown for 20 min, 40 min, 60 min, 90 min, 2 h, 3 h, 4 h, 6 h, 8 h, 12 h, and 18 h of digestion.

In contrast to the subtilisin and pronase digestions, total detected glycopeptide abundances (in ion counts) increase dramatically, by 150%, during the initial two hours of the proteinase K digestion, indicating that proteinase K is still digesting either whole infliximab or large pieces thereof up until this time point. After two hours, there are no further increases in total detected glycopeptide abundance, indicating that from this time point on, proteinase K is mainly acting on already-digested short glycopeptides, and that very little undigested infliximab remains.

The initial rise in abundance of QYNST and TKPREEQYNST peptide moieties is most likely related to the continuing digestion of whole infliximab during this period. After two hours, all of the whole infliximab has been digested into (glyco)peptides of various sizes, and abundances of QYNST and TKPREEQYNST begin to drop. Correspondingly, the rate of generation of NST and YNST peptide moieties increases. After 18 h of digestion by proteinase K, the rate of change has slowed, but not stopped. The dominant peptide moiety at this time is NST, comprising 89% of all infliximab glycopeptides; however, detectable abundances of YNST (8%) and QYNST (3%) peptide moieties are still present.

When selecting the appropriate multispecific protease for an application, sensitivity may also be a consideration. Total detected glycopeptide abundances (in ion counts) were used to evaluate the yield, and thus sensitivity, afforded by each protease digest. For subtilisin, the total detected glycopeptide abundance was $4.2 \times 10^7$ counts at 20 min, increasing to $4.4 \times 10^7$ counts by the end of the digestion. For pronase, the total detected glycopeptide abundance was $1.0 \times 10^7$ counts at 20 min, increasing rapidly to $2.5 \times 10^7$ counts after two hours and remaining steady until the end of the digestion.

When determining the optimal digestion conditions for a particular application, multiple factors come into play. Specificity of the peptide moiety for a single site of glycosylation; digestion yield (i.e., sensitivity); and digestion purity (i.e., the degree of signal splitting) should all be considered. For example, at the 18 h digestion time points in Figure 4, all three proteases offer their own unique advantages. Proteinase K (Figure 4c) offers the highest degree of glycosite specificity, since the digest contains glycopeptides with three overlapping peptide moieties, two of which cross-confirm that the originating IgG-like protein is homologous to the IgG1 isotype, rather than the IgG2 isotype. Pronase (Figure 4b) offers the highest glycopeptide yield, increasing the sensitivity of the analysis and decreasing the amount of analyte glycoprotein necessary. Subtilisin (Figure 4a) offers the highest glycopeptide purity, with almost all of the glycopeptides condensed to a single peptide moiety, geometrically reducing the complexity of the digest. Depending on the specific application, researchers have an unlimited menu of options with which to customize their Glyco-AMP digestion conditions for optimal results.

Site-Specific Profiling with Protein Mixtures and Multiple Glycosylation Sites

The main purpose of using multispecific proteolysis to digest glycoproteins is to obtain site-specific information about their glycosylation, particularly when there is more than one glycosylation site involved. This can occur when more than one glycoprotein is present in a mixture (e.g., as a result of coprecipitation or coelution); or, when a glycoprotein has multiple glycosylation sites. In-solution multispecific proteolysis is demonstrated to be effective in both of these scenarios.

Figure 5 shows the chromatograms of glycopeptides digested by subtilisin from a mixture of RNase B and prostate-specific antigen (PSA). PSA is a commonly used clinical biomarker for prostate cancer; however, in recent years, its specificity has been called into question. Closer examination of its glycosylation may provide a means of increasing the specificity of PSA as a biomarker. Unfortunately, since PSA is present at relatively low levels in serum, isolation by either immunoprecipitation or chromatography may well result in contamination.
by coprecipitating or coeluting proteins, necessitating site-specific analysis to confirm the origin of glycosylation. Here, site-specific peptide moieties unambiguously identify each glycopeptide as originating from either RNase B or PSA. PSA glycoforms detected during this analysis (Figure 6) support and expand upon previous studies of PSA glycosylation. For example, multiple regioisomeric glycoforms of PSA are baseline-resolved by porous graphitized carbon nano-LC, providing structural specificity to the analysis. These regioisomers most likely arise from differential linkages of the terminal sialic acids (which can be either $\alpha$-2,3- or $\alpha$-2,6-linked to the preceding galactoses) and may be con

Figure 5. Overlaid chromatograms of glycopeptides digested by subtilisin from a mixture of RNase B and prostate-specific antigen (PSA). Glycopeptide identities are confirmed by MS/MS.

Part of the di

Figure 6. Glycoforms of prostate-specific antigen (PSA). Quantities are reported relative to the most abundant glycoform. A putative structure is suggested for each glycoform composition observed. Chromatographically separated regioisomers are separately quantified. Retention times are reported, in minutes, for glycopeptides attached to an NK peptide moiety.

by coprecipitating or coeluting proteins, necessitating site-specific analysis to confirm the origin of glycosylation. Here, site-specific peptide moieties unambiguously identify each glycopeptide as originating from either RNase B or PSA. PSA glycoforms detected during this analysis (Figure 6) support and expand upon previous studies of PSA glycosylation. For example, multiple regioisomeric glycoforms of PSA are baseline-resolved by porous graphitized carbon nano-LC, providing structural specificity to the analysis. These regioisomers most likely arise from differential linkages of the terminal sialic acids (which can be either $\alpha$-2,3- or $\alpha$-2,6-linked to the preceding galactoses) and may be con

Figure 7. Quantitative, isomer-specific, site-specific glycosylation profile of vitronectin. Quantities are reported relative to the most abundant glycoform. A putative structure is suggested for each glycoform composition observed. Chromatographically separated regioisomers are separately quantified. Each color represents a different glycosylation site.

P2' position. Steric hindrance of trypsin at these sites would cause missed cleavages, leading to excessively large glycopeptides that would be difficult to detect or characterize.

In Figure 7, multispecific proteolysis is employed to sidestep the issue by creating glycopeptides that are small enough for easy detection as well as isomeric separation by porous graphitized carbon nano-LC, yet still retain their specificity for a single site of glycosylation. For each site of glycosylation, glycan occupation was confirmed by at least two overlapping peptide moieties. For $^{169}$Asn, glycopeptides were detected with peptide moieties NN (91%), NNAT (6%), and NAT (3%). For $^{169}$Asn, glycopeptides were detected with peptide moieties NGS (79%) and NIS (21%). Multiple regioisomeric glycan moieties, arising from differences in either monosaccharide linkage or position, were chromatographically separated and differentiated; however, de novo identification was not possible based on CID MS/MS data alone. As this is the first site-specific analysis of vitronectin glycosylation with any kind of isomer separation, further experimentation with linkage-specific glycosidases (or other structural elucidation methods) will be necessary in order to determine the exact structural characteristics of the isomeric glycoforms.

Isomer-Specific LC–MS/MS Characterization of Site-Specific Glycoforms

One of the major advantages of multispecific proteolysis is its ability to generate small glycopeptides that can be isomERICally separated by nano-LC and then structurally characterized by MS/MS. Figure 8 further explores this concept by displaying isomer-specific MS/MS spectra of two chromatographically separated O-glycopeptide isomers originating from a pronase digest of darbepeotin alfa (aka novel erythropoiesis stimulating protein, or NESP). Darbepeotin alfa is a commonly used biotherapeutic analogue of erythropoietin, the glycoprotein that stimulates red blood cell production.

The glycopeptide isomers in Figure 8 are each composed of a tri-O-acetylated Hex$_x$HexNAc$_y$NeuAc$_z$ O-glycan moiety attached to peptide moiety SPPDAASAAP. Extensive CID fragmentation of both the glycan and peptide moieties not only confirms the composition of the glycopeptides, but also reveals stark differences in the MS/MS fingerprint of each

Figure 8 further explores this concept by displaying isomer-specific MS/MS spectra of two chromatographically separated O-glycopeptide isomers originating from a pronase digest of darbepeotin alfa (aka novel erythropoiesis stimulating protein, or NESP). Darbepeotin alfa is a commonly used biotherapeutic analogue of erythropoietin, the glycoprotein that stimulates red blood cell production.

The glycopeptide isomers in Figure 8 are each composed of a tri-O-acetylated Hex$_x$HexNAc$_y$NeuAc$_z$ O-glycan moiety attached to peptide moiety SPPDAASAAP. Extensive CID fragmentation of both the glycan and peptide moieties not only confirms the composition of the glycopeptides, but also reveals stark differences in the MS/MS fingerprint of each
isomer. For example, the fragments at \( m/z \) 204.09 (HexNAc +H)+, \( m/z \) 274.09 (NeuAc-H2O+H)+, \( m/z \) 292.10 (NeuAc +H)+, \( m/z \) 366.14 (Hex,HexNAc+H)+, \( m/z \) 657.23 (Hex, HexNAc,NeuAc+H)+, and \( m/z \) 1461.62 (SPPDAASAAP+HexNAc,NeuAc,+2OAc+H)+) are present in high abundance in MS/MS of the isomer eluting at 14.5 min (Figure 8a), yet conspicuously absent in MS/MS of the isomer eluting at 15.6 min (Figure 8b).

Closer examination suggests that discrepancies between the isomers in Figure 8, parts a and b, stem from differing degrees of O-acetylation on their sialic acids. In Figure 8a, the fragments at \( m/z \) 1461.62 (SPPDAASAAP+HexNAc,NeuAc,+2OAc+H)+ and \( m/z \) 1346.54 (SPPDAASAAP+HexNAc,NeuAc,+2OAc +H)+ suggest the HexNAc-attached NeuAc as the site of two out of three total O-acetyl modifications on this isomer. These fragments are completely absent from Figure 8b; however, Figure 8b does show a high abundance of the fragment at \( m/z \) 538.18 (Hex,NeuAc,+2OAc+H)+, which conversely suggests the Hex-attached NeuAc as the site of two out of three total O-acetyl modifications on this isomer. On the basis of these data, we propose that the isomer shown in Figure 8a has one O-acetyl modification on the Hex-attached NeuAc and two O-acetyl modifications on the HexNAc-attached NeuAc, while the isomer shown in Figure 8b has two O-acetyl modifications on the Hex-attached NeuAc and one O-acetyl modification on the HexNAc-attached NeuAc.

Previous studies of darbepoetin alfa glycosylation have found a maximum of two O-acetyl modifications per sialic acid for both N- and O-glycosylation, in concurrence with the proposed structures for the two isomers shown here. Of course, additional structural differences are also possible between these two isomers—for instance, either of the two sialic acids on each glycopeptide might be connected to the rest of the molecule by a different linkage. Complementary structural elucidation methods such as exoglycosidase digestion and/or nonergodic MS/MS may provide further information.

O-acetylation increases the half-life and potency of recombinant erythropoietin by interfering with in vivo enzymatic desialylation of the glycans present on the drug. Additionally, O-acetylated sialic acids act as binding substrates for several viruses, including common cold viruses such as influenza C as well as certain coronaviruses. Thus, the level of fine structural detail provided by nano-LC—MS/MS of multispecific protease digests could be of great import to both producers and regulators of biopharmaceutical drugs and vaccines.

**CONCLUSIONS**

Although subtilisin, pronase, and proteinase K have previously been described as “nonspecific” proteases, detailed and extensive characterization reveals that these proteases neither lack specificity, nor digest proteins at random. Instead, they may best be described as “multispecific” proteases, possessing multiple substrate specificities, such that they hydrolyze peptide bonds at a large, but finite number of sites on a (glyco)protein. As demonstrated, the interplay of these different substrate specificities can be analytically exploited to optimize glyco-protein digestion and glycopeptide generation, providing a valuable new class of tools for the burgeoning field of glycoproteomics.

Analogously to the thousands of restriction enzymes available to genomicsists, a large protease toolbox provides greater flexibility and experimental customizability to the field of proteomics, and particularly to the characterization of post-translational modifications such as glycosylation. Glyco-AMP packages this flexibility into a rapid, easy-to-use methodology that can be broadly applied to a variety of complex glycoproteins as well as glycoprotein mixtures. Using the simple, in-solution digestion protocol, researchers avoid laborious protease immobilization reactions while increasing quantitative reproducibility. In addition, researchers can control peptide moiety length by adjusting the multispecific digestion kinetics, allowing them to choose the degrees of glycosylation site specificity and glycan structure specificity necessary for their experiment.

Multispecific proteases, as well as the glyco-AMP strategy in general, find maximum utility in detailed, structure-specific studies of the glycoproteome. In contrast to trypsin, which is most effective at sketching an outline of whole glycoproteomes, glyco-AMP and related strategies are crucial for filling in the fine details about specific targeted glycoproteins or mixtures. As a result, glyco-AMP is an ideal method for detailed characterization of potential glycoprotein biomarkers as well as glycosylated biopharmaceutical products, providing unparalleled site- and structure-specific information.

**AUTHOR INFORMATION**

**Corresponding Author**

*Tel: 82 42-821-8547; fax: 82 42-821-8551; e-mail: hjan@cnu.ac.kr (H.J.A.). Tel: 1 530-752-0504; fax: 1 530-752-8995; e-mail: cblebrilla@ucdavis.edu (C.B.L.).

**Author Contributions**

H.J.A. and C.B.L. contributed equally to this manuscript.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We are grateful for the support provided by the 2012 University-Institute Cooperation Program via the National Research Foundation of Korea as well as the Converging Research Center Program (2012K001505 for H.J.A.) via the Ministry of Education, Science and Technology.
ABBREVIATIONS:
Glyco-AMP, glyco-lysyl proteolytic release; Hex, hexose; HexNAc, N-acetylhexosamine; Man, mannoside; GlcNAc, N-acetylgalactosamine; Fuc, fucose; G1F, HexHexNAcFuc; NeuAc, N-acetylneuraminic acid; OAc, O-acetylation

REFERENCES:


