Journal of proteome research

Article

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 Chuks Nwosu, Jong Shin Yoo, Carlito B. Lebrilla, and Hyun Joo An

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/pr400442y • Publication Date (Web): 09 Sep 2013

Glyco-analytical multispecific proteolysis (Glyco-AMP): A simple method for detailed and quantitative glycoproteomic characterization

Serenus Hua,[1,2] Chloe Y. Hu,[3] Bum Jin Kim,[1] Sarah M. Totten,[3] Myung Jin Oh,[1] Nayoung Yun,[1] Charles C. Nwosu,[3] Jong Shin Yoo,[1,4] Carlito B. Lebrilla,[3*†] and Hyun Joo An[1,2*†]

1. *Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon, Korea*
2. *Cancer Research Institute, Chungnam National University, Daejeon, Korea*
3. *Department of Chemistry, University of California, Davis, CA*
4. *Division of Mass Spectrometry Research, Korea Basic Science Institute, Ochang, Korea*

[*] To whom correspondence should be addressed:

Hyun Joo An, email: hjan@cnu.ac.kr, Tel: 82 42-821-8547, Fax: 82 42-821-8551

Carlito B. Lebrilla, email: cblebrilla@ucdavis.edu, Tel: 1 530-752-0504, Fax: 1 530-752-8995

[†] Hyun Joo An and Carlito B. Lebrilla contributed equally to this paper as co-corresponding authors.

KEYWORDS: multispecific proteases, non-specific proteases, site-specific glycosylation, glycoproteomics, biopharmaceutical glycoproteins, glycan isomers

ABBREVIATIONS: Glyco-AMP, glyco-analytical multispecific proteolysis; Hex, hexose; HexNAc, N-acetylhexosamine; Man, mannose; GlcNAc, N-acetylglucosamine; Fuc, fucose; G1F, $Hex_4HexNAc_4Fuc$; NeuAc, N-acetylneuraminic acid; OAc, O-acetylation

ABSTRACT

Despite recent advances, site-specific profiling of protein glycosylation remains a significant analytical challenge for conventional proteomic methodology. To alleviate the issue, we propose glyco-analytical multispecific proteolysis (Glyco-AMP) as a strategy for glycoproteomic characterization. Glyco-AMP consists of rapid, in-solution digestion of an analyte glycoprotein (or glycoprotein mixture) by a multispecific protease (or protease cocktail). Resulting glycopeptides are chromatographically separated by isomer-specific porous graphitized carbon nano-LC, quantified by high-resolution MS, and structurally elucidated by MS/MS. To demonstrate the consistency and customizability of Glyco-AMP methodology, the glyco-analytical performances of multispecific proteases subtilisin, pronase, and proteinase K were characterized in terms of quantitative accuracy, sensitivity, and digestion kinetics. Glyco-AMP was shown be effective on glycoprotein mixtures as well as glycoproteins with multiple glycosylation sites, providing detailed, quantitative, site- and structure-specific information about protein glycosylation.

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 towards protein structure and function becomes ever more apparent, glycomic methods for profiling the detached glycan component of a glycoprotein 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 from traditional proteomics[5-7] and often do not sufficiently address the complexities of characterizing a post-translational modification as intricate as glycosylation.

Whereas peptides can be identified and characterized simply 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 potential cleavage sites.[13] 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.[14, 15]

While the amino acid sequences of some glycoproteins may be fortuitously rich in tryptic cleavage sites, many are not so easily dissected.[16, 17] 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 non-specific) 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.[18, 19]

Early forays into glycoprotein characterization via multispecific proteolysis involved MS-only analysis of single glycoprotein digests.[20] 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.[21, 22] However, the most dramatic improvements in glycoproteomics have come just recently, with the incorporation of nano-LC separation into established mass spectrometric methods.[8, 9, 23] 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.[24, 25] Though in-solution protease digests do contain higher levels of signal-suppressing peptides than immobilized protease digests, 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 glyco-analytical 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.

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). Infliximab 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 ºC for 20 minutes. 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 timepoints 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.[26] 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 ^{o}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 [2]. 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 min to 13.3 min; and 32.8% to 35.9% B, 13.3 min 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 seconds 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 seconds 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, 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:

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

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.[8]

LC/MS data processing and glycopeptide identification

After data acquisition, raw LC/MS data was 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 µg of protease with 50 µg of well-characterized glycoprotein ribonuclease B (at a total digestion volume of 100 µL). Digestion buffers and conditions were chosen based on the manufacturers' suggested protocols[29-35]— for example, the thermolysin digestion took place at 75 ºC in 50 mM Tris and 0.5 mM calcium chloride, while the pepsin digestion took place at 37 ºC in 10 mM hydrochloric acid. Digestions were carried out for 20 minutes each, after which the digests 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 $(HexNAc_1Hex_1+H)^+$. 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.

Based on 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 ºC.

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 minute 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 $Man_5GlcNAc_2$ (55%) is the most abundant glycan attached to RNase B, followed by $Man_6GlcNAc_2$ (27%), $Man_8GlcNAc_2$ (9%), $Man_7GlcNAc_2$ (7%), and $Man_9GlcNAc_2$ (3%); these results correlate highly with previous studies quantifying the glycoforms of RNase B.[37-39] All protease digests result in similar profiles, with the relative abundances of individual glycoforms 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 SRNLT (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, NLTK, NLTKDR, N, and NL. For proteinase K,

major peptide moieties consist of NLTK (40%), SRNLT (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 minute 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 $Man_9GlcNAc_2$ 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 minutes of digestion by subtilisin, whereas **Figure 2b** shows the glycopeptides generated by 18 hours 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.[21, 22] The 20 minute 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 hour 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-$Hex_4HexNAc_4Fuc$ (i.e. QYNST-G1F) decreasing over time while the abundance of glycopeptide NST-$Hex_4HexNAc_4Fuc$ (i.e. NST-G1F) increases over time. **Figure 3b** plots the absolute abundances of these two glycopeptides at six different points during the 18-hour digestion, while **Figures 3a** and **3c** show 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.[40] Based on 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 homolog) from glycoforms of IgG2 homologs 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 timepoint, the percent contributions of each peptide moiety to the overall glycopeptide population are plotted. These charts may be used to estimate the rate of glycopeptide production at a given time point and can help optimize a digestion for generation of a target glycopeptide.

In subtilisin (**Figure 4a**), as discussed earlier, the major peptide moieties are initially QYNST and NST, with minor contributions from QYNSTY and NSTY. As the digestion progresses, abundances of QYNST and QYNSTY drop drastically, eventually becoming undetectable, while abundances of NST rise correspondingly. However, the contribution from NSTY is more interesting. For the first two hours of the digestion, abundances of NSTY rise,

very nearly imitating the reaction plot of NST (proportionately, at much lower levels). However, after the first two hours, abundances of NSTY drop, 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 hour digestion timepoint, 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 minute timepoint, 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 YNST appears 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 timepoint. Even after 18 hours of digestion by pronase, only 84% of all

infliximab glycopeptides have NST peptide moieties, while the remaining 16% have YNST 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 minute timepoint, 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 timepoints have been added to the analysis. In all, timepoints 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 timepoint. After two hours, there are no further increases in total detected glycopeptide abundance, indicating that from this timepoint 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 hours 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 4.2×10^7 counts at 20 minutes, increasing to 4.4×10^7 counts by the end of the digestion. For pronase, the total detected glycopeptide abundance was 7.7×10^7 counts at 20 minutes, increasing to 8.4×10^7 counts by the end of the digestion. For proteinase K, the total detected glycopeptide abundance was 1.0×10^7 counts at 20 minutes, increasing rapidly to 2.5×10^7 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 hour digestion timepoints 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 co-precipitation or co-elution); 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;[41, 42] however, in recent years, its specificity has been called into question.[43, 44] 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 co-precipitating or co-eluting 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.[45, 46] 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 confirmed in future experiments by linkage-specific glycosidase digestion.[40]

Pronase digestion of vitronectin, a glycoprotein with three sites of glycosylation, reveals the detailed glycosylation profile shown in **Figure 7**. Recent research has indicated that vitronectin, an abundant serum glycoprotein, may be involved in the early stages of cancer.[47, 48] However, few studies have examined vitronectin glycosylation,[49] and none have been able to isolate glycoforms at specific sites of glycosylation until now. Part of the difficulty may have to do with the proteomics community's traditional reliance on trypsin— three of vitronectin's glycosylation sites are located in the near vicinity of tryptic cleavage sites, with two in a P1' position, immediately adjacent to a tryptic cleavage site, and another one located in a 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 ^{86}Asn, glycopeptides were detected with peptide moieties NN (91%), NNAT (6%), and NAT (3%). For ^{169}Asn, glycopeptides were detected with peptide moieties NGS (89%) and KNGS (3%). For ^{242}Asn, glycopeptides were detected with peptide moieties NISDG (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 darbepoetin alfa (aka novel erythropoiesis stimulating protein, or NESP). Darbepoetin 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 comprised of a tri-O-acetylated $Hex_1HexNAc_1NeuAc_2$ 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 $(NeuAc\text{-}H_2O+H)^+$, *m*/*z* 292.10 $(NeuAc+H)^+$, *m*/*z* 366.14 $(Hex_1HexNAc_1+H)^+$, *m*/*z* 657.23 $(Hex_1HexNAc_1NeuAc_1+H)^+$, and *m*/*z* 1461.62 $(SPPDAASAAP+HexNAc_1NeuAc_1+2OAc+H)^+$) are present in high abundance in MS/MS of the isomer eluting at 14.5 minutes (**Figure 8a**), yet conspicuously absent in MS/MS of the isomer eluting at 15.6 minutes (**Figure 8b**).

Closer examination suggests that discrepancies between the isomers in **Figures 8a** and **8b** stem from differing degrees of O-acetylation on their sialic acids. In **Figure 8a**, the fragments at

m/*z* 1461.62 $(SPPDAASAAP+HexNAc_1NeuAc_1+2OAc+H)^+$ and *m*/*z* 1346.54 $(SPPDAASAA+HexNAc_1NeuAc_1+2OAc+H)^+$ 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 $(Hex_1NeuAc_1+2OAc+H)^+$, which conversely suggests the *Hex*-attached NeuAc as the site of two out of three total O-acetyl modifications on this isomer. Based on 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,[50, 51] 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 non-ergodic MS/MS may provide further information.[52, 53]

O-acetylation increases the half-life and potency of recombinant erythropoietins by interfering with *in vivo* enzymatic desialylation of the glycans present on the drug.[54] 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.[55, 56] 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.

CONCLUSION

Though subtilisin, pronase, and proteinase K have previously been described as "non-specific" 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 glycoprotein 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 genomicists, 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.

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. An) via the Ministry of Education, Science and Technology.

REFERENCES

1. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999,** 1473, (1), 4-8.
2. Hua, S.; Williams, C. C.; Dimapasoc, L. M.; Ro, G. S.; Ozcan, S.; Miyamoto, S.; Lebrilla, C. B.; An, H. J.; Leiserowitz, G. S., Isomer-Specific Chromatographic Profiling Yields Highly Sensitive And Specific Potential N-Glycan Biomarkers For Epithelial Ovarian Cancer. *J Chromatogr A* **2013**, (0).
3. Hua, S.; An, H. J.; Ozcan, S.; Ro, G. S.; Soares, S.; DeVere-White, R.; Lebrilla, C. B., Comprehensive native glycan profiling with isomer separation and quantitation for the discovery of cancer biomarkers. *Analyst* **2011,** 136, (18), 3663-3671.
4. Hua, S.; Lebrilla, C.; An, H. J., Application of nano-LC-based glycomics towards biomarker discovery. *Bioanalysis* **2011,** 3, (22), 2573-2585.
5. Ahn, Y.; Shin, P.; Ji, E.; Kim, H.; Yoo, J., A lectin-coupled, multiple reaction monitoring based quantitative analysis of human plasma glycoproteins by mass spectrometry. *Analytical and Bioanalytical Chemistry* **2012,** 402, (6), 2101-2112.
6. Miyoshi, E.; Nakano, M., Fucosylated haptoglobin is a novel marker for pancreatic cancer: Detailed analyses of oligosaccharide structures. *PROTEOMICS* **2008,** 8, (16), 3257-3262.
7. Kurogochi, M.; Amano, M.; Fumoto, M.; Takimoto, A.; Kondo, H.; Nishimura, S.-I., Reverse Glycoblotting Allows Rapid-Enrichment Glycoproteomics of Biopharmaceuticals and Disease-Related Biomarkers. *Angewandte Chemie International Edition* **2007,** 46, (46), 8808-8813.

8. Hua, S.; Nwosu, C.; Strum, J.; Seipert, R.; An, H.; Zivkovic, A.; German, J. B.; Lebrilla, C., Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal Bioanal Chem* **2012,** 403, (5), 1291-1302.

9. Hua, S.; An, H. J., Glycoscience aids in biomarker discovery. *BMB Rep* **2012,** 45, (6), 323-30.

10. Dallas, D. C.; Martin, W. F.; Hua, S.; German, J. B., Automated glycopeptide analysis—review of current state and future directions. *Briefings in Bioinformatics* **2012**.

11. Kim, J. Y.; Kim, S.-K.; Kang, D.; Moon, M. H., Dual Lectin-Based Size Sorting Strategy to Enrich Targeted N-Glycopeptides by Asymmetrical Flow Field-Flow Fractionation: Profiling Lung Cancer Biomarkers. *Analytical Chemistry* **2012,** 84, (12), 5343-5350.

12. Schlosser, A.; Vanselow, J. T.; Kramer, A., Mapping of Phosphorylation Sites by a Multi-Protease Approach with Specific Phosphopeptide Enrichment and NanoLC−MS/MS Analysis. *Analytical Chemistry* **2005,** 77, (16), 5243-5250.

13. Pompach, P.; Chandler, K. B.; Lan, R.; Edwards, N.; Goldman, R., Semi-Automated Identification of N-Glycopeptides by Hydrophilic Interaction Chromatography, nano-Reverse-Phase LC–MS/MS, and Glycan Database Search. *Journal of Proteome Research* **2012,** 11, (3), 1728-1740.

14. Wu, Z. L.; Ethen, C.; Hickey, G. E.; Jiang, W., Active 1918 pandemic flu viral neuraminidase has distinct N-glycan profile and is resistant to trypsin digestion. *Biochemical and Biophysical Research Communications* **2009,** 379, (3), 749-753.

15. Fujihara, J.; Yasuda, T.; Kunito, T.; Fujii, Y.; Takatsuka, H.; Moritani, T.; Takeshita, H., Two N-Linked Glycosylation Sites (Asn18 and Asn106) Are Both Required for Full Enzymatic Activity, Thermal Stability, and Resistance to Proteolysis in Mammalian

Deoxyribonuclease I. *Bioscience, Biotechnology, and Biochemistry* **2008,** 72, (12), 3197-3205.

16. Giménez, E.; Ramos-Hernan, R.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V., Analysis of recombinant human erythropoietin glycopeptides by capillary electrophoresis electrospray–time of flight-mass spectrometry. *Analytica Chimica Acta* **2012,** 709, (0), 81-90.
17. Wang, D.; Hincapie, M.; Rejtar, T.; Karger, B. L., Ultrasensitive Characterization of Site-Specific Glycosylation of Affinity-Purified Haptoglobin from Lung Cancer Patient Plasma Using 10 μm i.d. Porous Layer Open Tubular Liquid Chromatography−Linear Ion Trap Collision-Induced Dissociation/Electron Transfer Dissociation Mass Spectrometry. *Analytical Chemistry* **2011,** 83, (6), 2029-2037.
18. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *Journal of Proteome Research* **2010,** 9, (3), 1323-1329.
19. Seipert, R. R.; Dodds, E. D.; Lebrilla, C. B., Exploiting Differential Dissociation Chemistries of O-Linked Glycopeptide Ions for the Localization of Mucin-Type Protein Glycosylation. *Journal of Proteome Research* **2008,** 8, (2), 493-501.
20. An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B., Determination of N-Glycosylation Sites and Site Heterogeneity in Glycoproteins. *Analytical Chemistry* **2003,** 75, (20), 5628-5637.
21. Clowers, B. H.; Dodds, E. D.; Seipert, R. R.; Lebrilla, C. B., Site Determination of Protein Glycosylation Based on Digestion with Immobilized Nonspecific Proteases and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of Proteome Research* **2007,** 6, (10), 4032-4040.

22. Dodds, E. D.; Seipert, R. R.; Clowers, B. H.; German, J. B.; Lebrilla, C. B., Analytical Performance of Immobilized Pronase for Glycopeptide Footprinting and Implications for Surpassing Reductionist Glycoproteomics. *Journal of Proteome Research* **2008,** 8, (2), 502-512.

23. Froehlich, J. W.; Barboza, M.; Chu, C.; Lerno, L. A.; Clowers, B. H.; Zivkovic, A. M.; German, J. B.; Lebrilla, C. B., Nano-LC–MS/MS of Glycopeptides Produced by Nonspecific Proteolysis Enables Rapid and Extensive Site-Specific Glycosylation Determination. *Analytical Chemistry* **2011,** 83, (14), 5541-5547.

24. He, P.; Greenway, G.; Haswell, S., Development of enzyme immobilized monolith micro-reactors integrated with microfluidic electrochemical cell for the evaluation of enzyme kinetics. *Microfluid Nanofluid* **2010,** 8, (5), 565-573.

25. Calleri, E.; Temporini, C.; Gasparrini, F.; Simone, P.; Villani, C.; Ciogli, A.; Massolini, G., Immobilized trypsin on epoxy organic monoliths with modulated hydrophilicity: Novel bioreactors useful for protein analysis by liquid chromatography coupled to tandem mass spectrometry. *Journal of Chromatography A* **2011,** 1218, (49), 8937-8945.

26. Kronewitter, S. R.; de Leoz, M. L. A.; Peacock, K. S.; McBride, K. R.; An, H. J.; Miyamoto, S.; Leiserowitz, G. S.; Lebrilla, C. B., Human Serum Processing and Analysis Methods for Rapid and Reproducible N-Glycan Mass Profiling. *Journal of Proteome Research* **2010,** 9, (10), 4952-4959.

27. An, H. J.; Tillinghast, J. S.; Woodruff, D. L.; Rocke, D. M.; Lebrilla, C. B., A New Computer Program (GlycoX) To Determine Simultaneously the Glycosylation Sites and Oligosaccharide Heterogeneity of Glycoproteins. *Journal of Proteome Research* **2006,** 5, (10), 2800-2808.

28. Kronewitter, S. R.; An, H. J.; de Leoz, M. L.; Lebrilla, C. B.; Miyamoto, S.; Leiserowitz, G. S., The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. *PROTEOMICS* **2009,** 9, (11), 2986-2994.

29. Bieth, J.; Spiess, B.; Wermuth, C. G., The synthesis and analytical use of a highly sensitive and convenient substrate of elastase. *Biochemical Medicine* **1974,** 11, (4), 350-357.

30. Glazer, A. N.; Smith, E. L., 14 Papain and Other Plant Sulfhydryl Proteolytic Enzymes. In *The Enzymes*, Paul, D. B., Ed. Academic Press: 1971; Vol. Volume 3, pp 501-546.

31. Ryle, A., Pepsins, gastricsins and their zymogens. *Methods of enzymatic analysis* **1984**, 228-233.

32. Juhasz, P.; Martin, S. A., The utility of nonspecific proteases in the characterization of glycoproteins by high-resolution time-of-flight mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **1997,** 169–170, (0), 217-230.

33. Sweeney, P. J.; Walker, J. M., Proteinase K (EC 3.4. 21.14). *Enzymes of Molecular Biology* **1993,** 16, 305-317.

34. Guntelberg, A.; Ottesen, M., Purification of the proteolytic enzyme from Bacillus subtilis. *Comptes rendus des travaux du Laboratoire Carlsberg. Série chimique* **1954,** 29, (3-4), 36.

35. Hanzawa, S.; Kidokoro, S.-I.; Flickinger, M. C., Thermolysin. In *Encyclopedia of Industrial Biotechnology*, John Wiley & Sons, Inc.: 2009.

36. Fischer, D.; Wolfson, H.; Lin, S. L.; Nussinov, R., Three-dimensional, sequence order-independent structural comparison of a serine protease against the crystallographic database reveals active site similarities: Potential implications to evolution and to protein folding. *Protein Science* **1994,** 3, (5), 769-778.

37. Thobhani, S.; Yuen, C.-T.; Bailey, M. J. A.; Jones, C., Identification and quantification of N-linked oligosaccharides released from glycoproteins: An inter-laboratory study. *Glycobiology* **2009,** 19, (3), 201-211.

38. Thaysen-Andersen, M.; Mysling, S.; Højrup, P., Site-Specific Glycoprofiling of N-Linked Glycopeptides Using MALDI-TOF MS: Strong Correlation between Signal Strength and Glycoform Quantities. *Analytical Chemistry* **2009,** 81, (10), 3933-3943.

39. Fu, D.; Chen, L.; O'Neill, R. A., A detailed structural characterization of ribonuclease B oligosaccharides by 1H NMR spectroscopy and mass spectrometry. *Carbohydrate Research* **1994,** 261, (2), 173-186.

40. Aldredge, D.; An, H. J.; Tang, N.; Waddell, K.; Lebrilla, C. B., Annotation of a Serum N-Glycan Library for Rapid Identification of Structures. *Journal of Proteome Research* **2012,** 11, (3), 1958-1968.

41. Catalona, W. J.; Richie, J. P.; Ahmann, F. R.; Hudson, M. A.; Scardino, P. T.; Flanigan, R. C.; deKernion, J. B.; Ratliff, T. L.; Kavoussi, L. R.; Dalkin, B. L., Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *The Journal of Urology* **1994,** 151, (5), 1283-1290.

42. Partin, A. W.; Catalona, W. J.; Southwick, P. C.; Subong, E. N. P.; Gasior, G. H.; Chan, D. W., Analysis of percent free prostate-specific antigen (PSA) for prostate cancer detection: Influence of total psa, prostate volume, and age. *Urology* **1996,** 48, (6, Supplement 1), 55-61.

43. Schröder, F. H.; Hugosson, J.; Roobol, M. J.; Tammela, T. L. J.; Ciatto, S.; Nelen, V.; Kwiatkowski, M.; Lujan, M.; Lilja, H.; Zappa, M.; Denis, L. J.; Recker, F.; Berenguer, A.; Määttänen, L.; Bangma, C. H.; Aus, G.; Villers, A.; Rebillard, X.; van der Kwast, T.;

Blijenberg, B. G.; Moss, S. M.; de Koning, H. J.; Auvinen, A., Screening and Prostate-Cancer Mortality in a Randomized European Study. *New England Journal of Medicine* **2009,** 360, (13), 1320-1328.

44. Heidenreich, A.; Bellmunt, J.; Bolla, M.; Joniau, S.; Mason, M.; Matveev, V.; Mottet, N.; Schmid, H.-P.; van der Kwast, T.; Wiegel, T.; Zattoni, F., EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Treatment of Clinically Localised Disease. *European Urology* **2011,** 59, (1), 61-71.
45. Peracaula, R.; Tabarés, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R., Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **2003,** 13, (6), 457-470.
46. Tabarés, G.; Radcliffe, C. M.; Barrabés, S.; Ramírez, M.; Aleixandre, R. N.; Hoesel, W.; Dwek, R. A.; Rudd, P. M.; Peracaula, R.; de Llorens, R., Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA. *Glycobiology* **2006,** 16, (2), 132-145.
47. Kenny, H. A.; Kaur, S.; Coussens, L. M.; Lengyel, E., The initial steps of ovarian cancer cell metastasis are mediated by MMP-2 cleavage of vitronectin and fibronectin. *The Journal of Clinical Investigation* **2008,** 118, (4), 1367-1379.
48. Hurt, E. M.; Chan, K.; Duhagon Serrat, M. A.; Thomas, S. B.; Veenstra, T. D.; Farrar, W. L., Identification of Vitronectin as an Extrinsic Inducer of Cancer Stem Cell Differentiation and Tumor Formation. *STEM CELLS* **2010,** 28, (3), 390-398.
49. Ogawa, H.; Yoneda, A.; Seno, N.; Hayashi, M.; Ishizuka, I.; Hase, S.; Matsumoto, I., Structures of the N-Linked Oligosaccharides on Human Plasma Vitronectin. *European Journal of Biochemistry* **1995,** 230, (3), 994-1000.

50. Stübiger, G.; Marchetti, M.; Nagano, M.; Grimm, R.; Gmeiner, G.; Reichel, C.; Allmaier, G., Characterization of N- and O-glycopeptides of recombinant human erythropoietins as potential biomarkers for doping analysis by means of microscale sample purification combined with MALDI-TOF and quadrupole IT/RTOF mass spectrometry. *Journal of Separation Science* **2005,** 28, (14), 1764-1778.
51. Oh, M. J.; Hua, S.; Kim, B. J.; Jeong, H. N.; Jeong, S. H.; Grimm, R.; Yoo, J. S.; An, H. J., Analytical platform for glycomic characterization of recombinant erythropoietin biotherapeutics and biosimilars by MS. *Bioanalysis* **2013,** 5, (5), 545-559.
52. Du, Y.; Wang, F.; May, K.; Xu, W.; Liu, H., LC–MS analysis of glycopeptides of recombinant monoclonal antibodies by a rapid digestion procedure. *Journal of Chromatography B* **2012,** 907, (0), 87-93.
53. Dodds, E. D., Gas-phase dissociation of glycosylated peptide ions. *Mass Spectrometry Reviews* **2012,** 31, (6), 666-682.
54. Llop, E.; Gutiérrez-Gallego, R.; Segura, J.; Mallorquí, J.; Pascual, J. A., Structural analysis of the glycosylation of gene-activated erythropoietin (epoetin delta, Dynepo). *Analytical Biochemistry* **2008,** 383, (2), 243-254.
55. Rogers, G. N.; Herrler, G.; Paulson, J. C.; Klenk, H. D., Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. *Journal of Biological Chemistry* **1986,** 261, (13), 5947-5951.
56. Schultze, B.; Herrler, G., Bovine coronavirus uses N-acetyl-9-O-acetylneuraminic acid as a receptor determinant to initiate the infection of cultured cells. *Journal of General Virology* **1992,** 73, (4), 901-906.

FIGURES

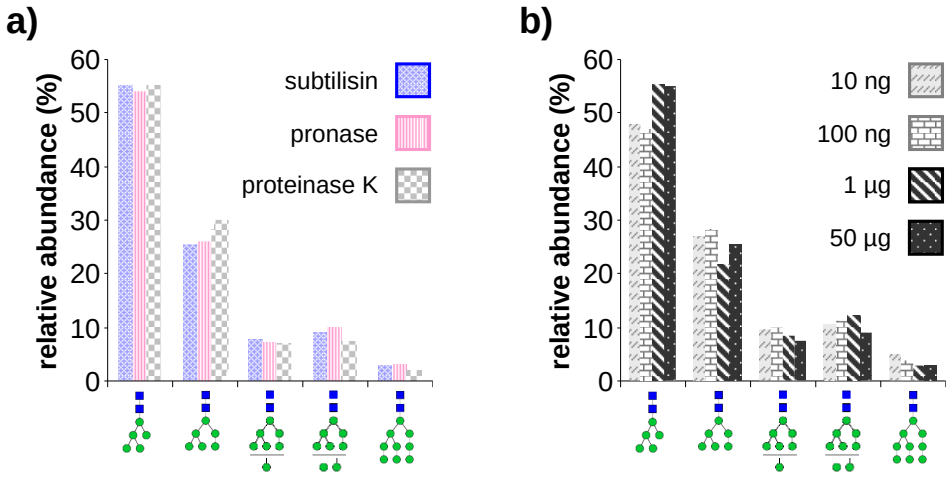

FIGURE 1: Glycoform profiles of standard glycoprotein ribonuclease B obtained following a) digestion with multispecific proteases subtilisin, pronase, and proteinase K; and b) subtilisin digestion using 50 µg, 1 µg, 100 ng, and 10 ng of initial glycoprotein.

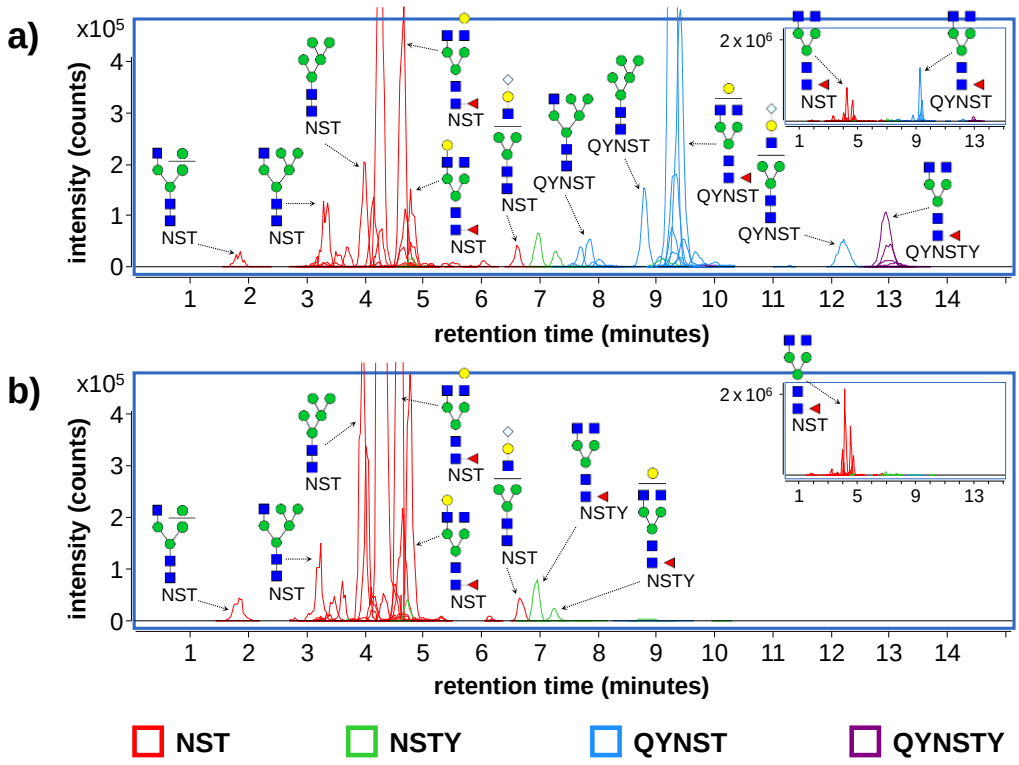

FIGURE 2: Overlaid chromatograms of glycopeptides generated by digestion of infliximab with subtilisin for a) 20 minutes, and b) 18 hours. Glycopeptide identities are confirmed by MS/MS.

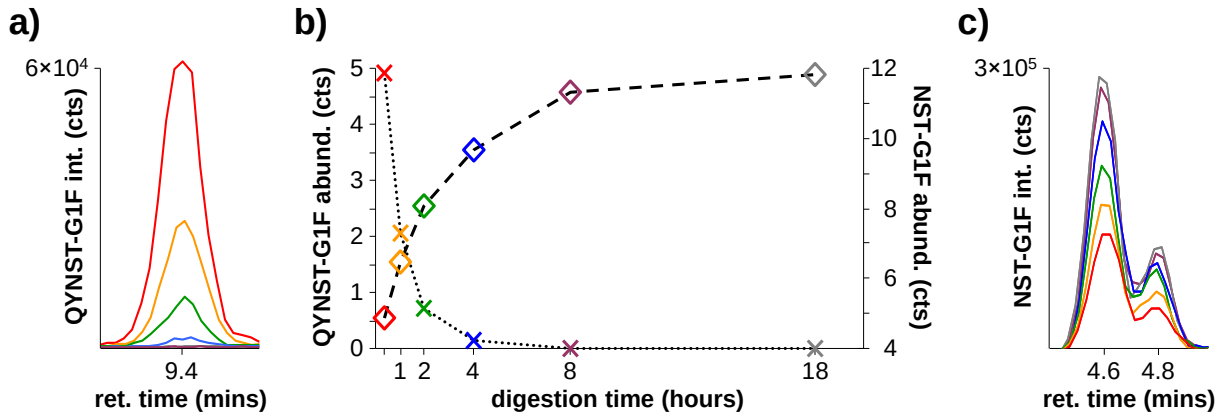

FIGURE 3: Digestion of infliximab glycoform $Hex_4HexNAc_4Fuc$ (G1F) by subtilisin over the course of 18 hours. a) Chromatograms of QYNST-G1F decreasing in size over time; b) Plot of QYNST-G1F and NST-G1F abundances over time; and c) Chromatograms of NST-G1F increasing in size over time. Colors denote progression of time, with red marking the earliest timepoint (20 minutes) and gray marking the latest timepoint (18 hours).

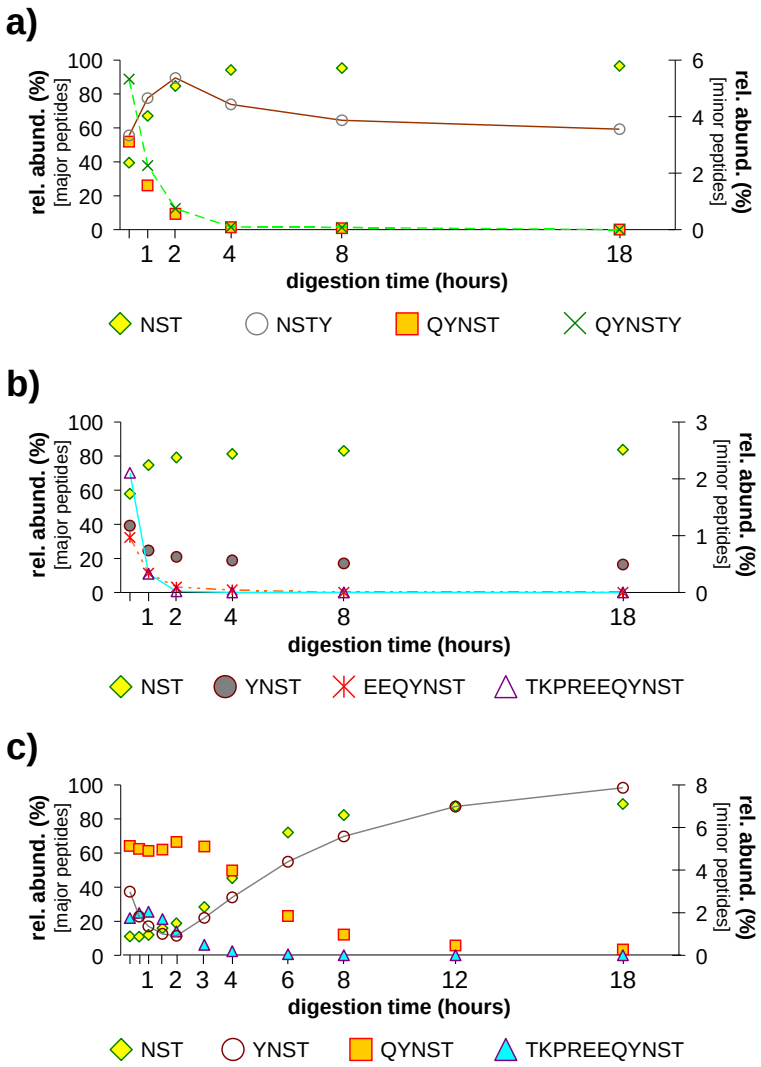

FIGURE 4: Progression of infliximab digestion by a) subtilisin; b) pronase; and c) proteinase K at different intervals over the course of 18 hours. For each digestion timepoint, the percent contributions of each peptide moiety to the overall glycopeptide population are plotted. Minor peptide moieties (right axis) are denoted by lines connected each timepoint; major peptide moieties (left axis) do not have lines connecting the timepoints.

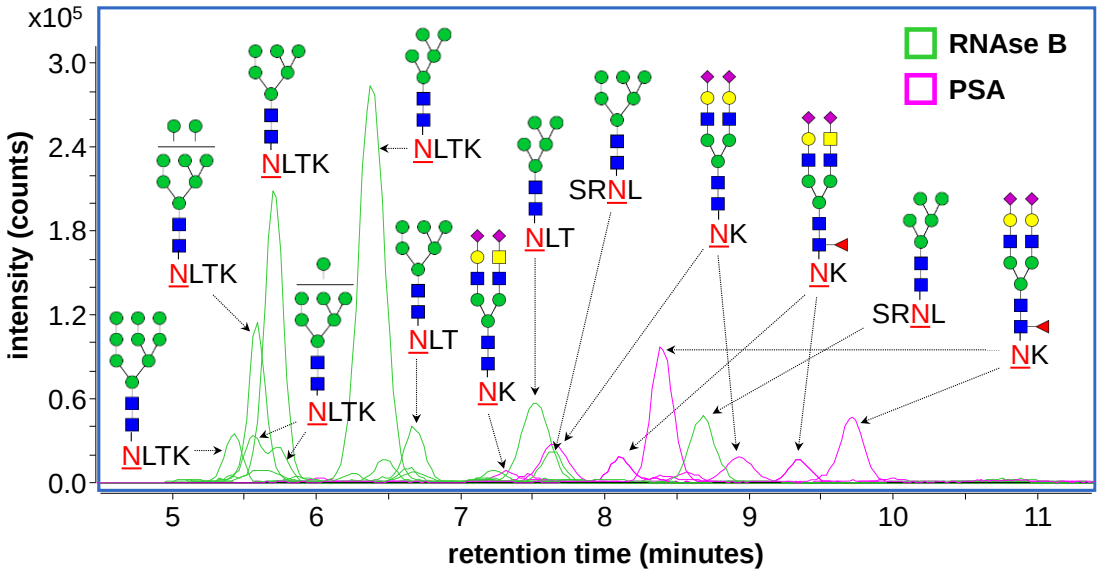

FIGURE 5: Overlaid chromatograms of glycopeptides digested by subtilisin from a mixture of RNase B and prostate-specific antigen (PSA). Glycopeptide identities are confirmed by MS/MS.

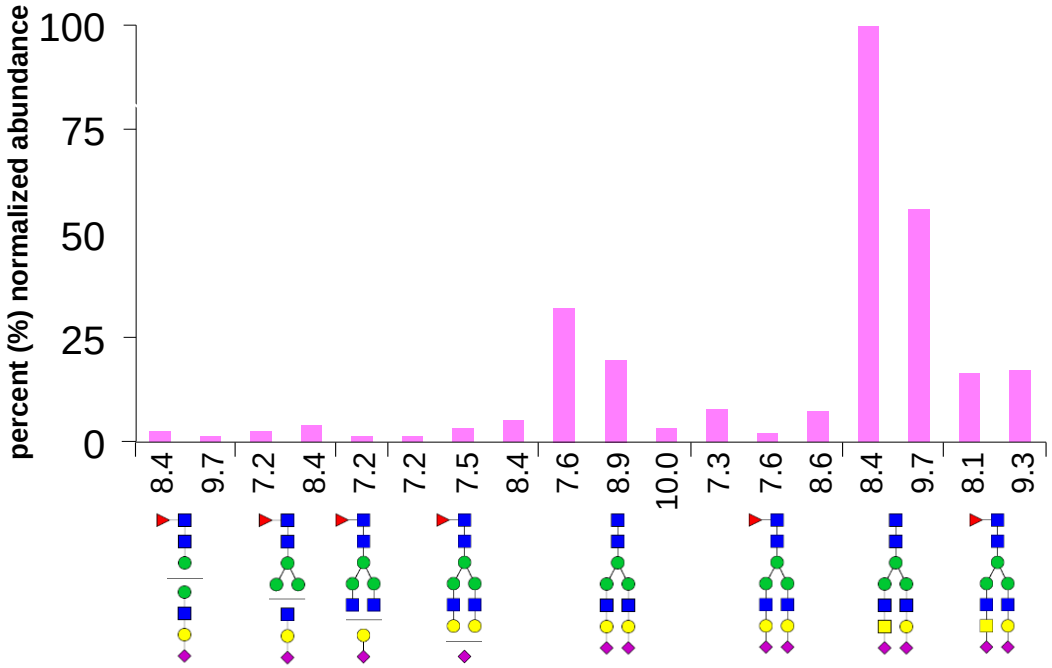

FIGURE 6: Glycoforms of prostate-specific antigen (PSA). Quantities are reported relative to the most abundant glycoform. A putative structure is suggested for each glycoform composition observed. Chromatographically-separated regioisomers are separately quantified. Retention times are reported, in minutes,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

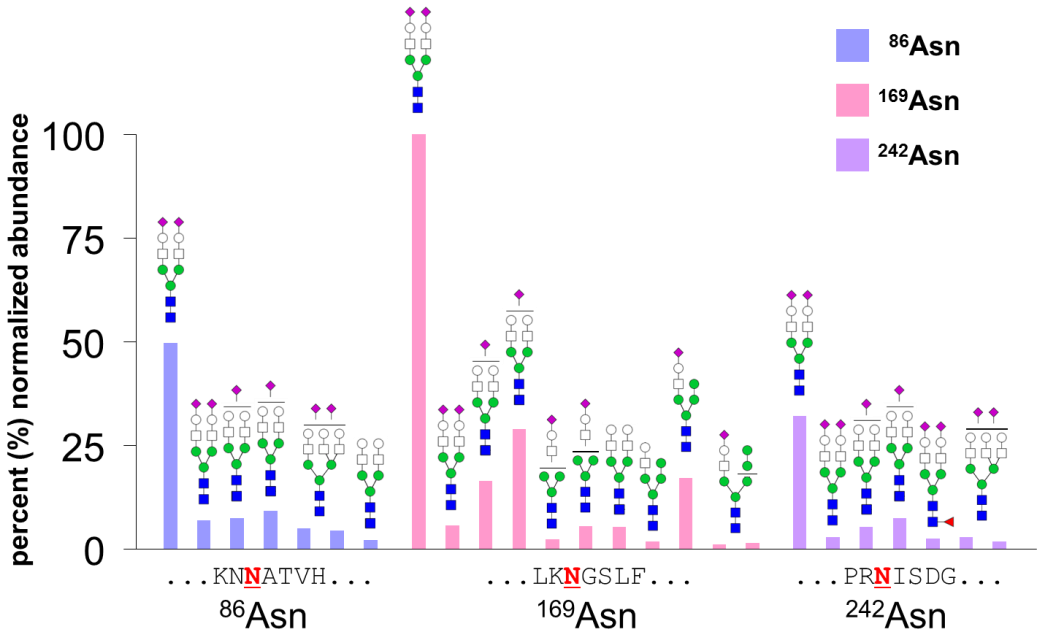

FIGURE 7: Quantitative, isomer-specific, site-specific glycosylation profile of vitronectin. Quantities are reported relative to the most abundant glycoform. A putative structure is suggested for each glycoform composition observed. Chromatographically-separated regioisomers are separately quantified.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

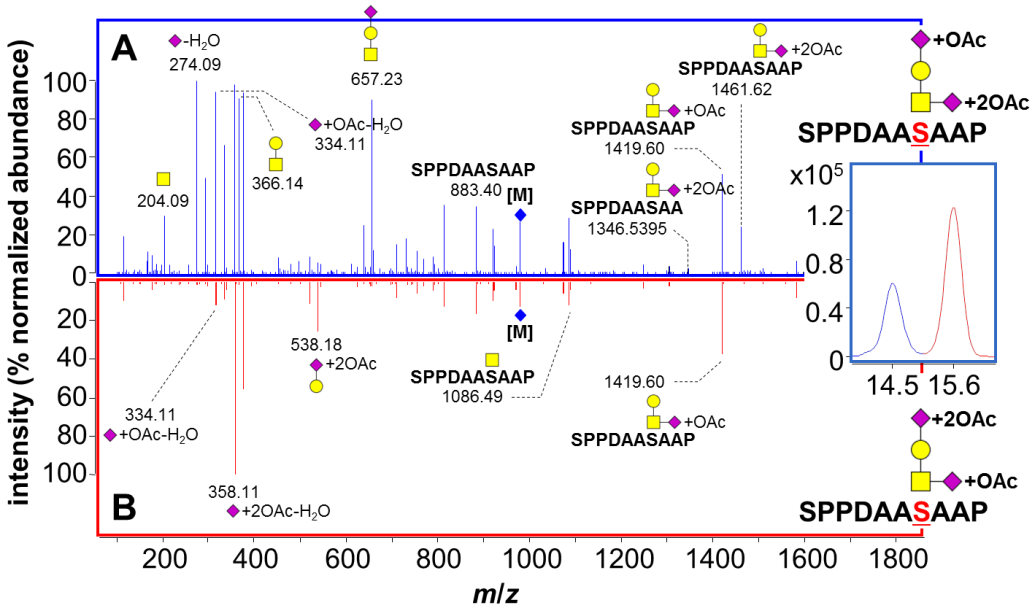

FIGURE 8: Positive ion CID MS/MS spectra of isomeric glycopeptides digested from the O-glycosylation site of recombinant erythropoietin and separated by porous graphitized carbon nano-LC. On top, a) the isomer of SPPDAASAAP+$Hex_1HexNAc_1NeuAc_2$+3OAc eluting at 14.5 minutes; and on the bottom, b) the isomer eluting at 15.6 minutes.

Graphical Abstract: