



# HHS Public Access

Author manuscript

*Anal Chem.* Author manuscript; available in PMC 2016 January 21.

Published in final edited form as:

*Anal Chem.* 2015 December 15; 87(24): 11948–11951. doi:10.1021/acs.analchem.5b03886.

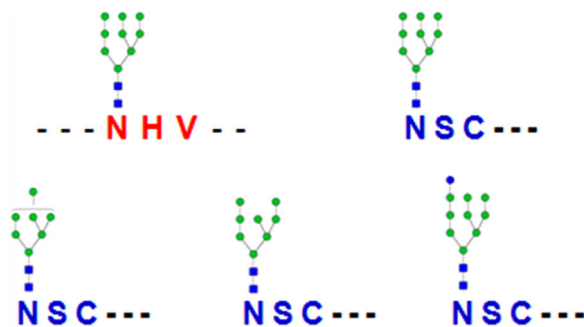
## Identification and Validation of Atypical N-Glycosylation Sites

Shisheng Sun\* and Hui Zhang

Department of Pathology, Johns Hopkins University, 400 North Broadway, Smith Building, Room 4001, Baltimore, Maryland 21287, United States

### Abstract

It is well-known that N-linked glycans usually attach to asparagine residues in the N-X-S/T motifs of proteins. However, accumulating evidence indicates that N-glycosylation could also possibly occur at other atypical motifs. In this study, we tried to identify atypical N-glycosylation sites using our recently developed solid-phase extraction of the N-linked Glycans And Glycosite-containing peptides (NGAG) method. Peptides with deamidation sites at asparagine residues but lacking a typical asparagine-X-serine/threonine sequons (N-X-S/T, X is any amino acid except proline) motif were identified from deglycosylated peptide data as potentially atypical glycosite-containing peptides. These atypical glycosites were verified by the presence of glycans on their intact glycopeptides and further confirmed by specific inhibition of cells with an N-linked glycosylation inhibitor, tunicamycin. From this study, two atypical N-linked glycosylation sites with N-X-C and N-X-V motifs were identified and validated from an ovarian cancer cell line (OVCAR-3).



N-Glycosylation at Atypical Glycosites

Glycosylation is one of the most important protein modifications. N-Linked glycosylation is a common feature shared by a large fraction of transmembrane proteins, cell surface proteins, and proteins secreted in body fluids.<sup>1</sup> N-Linked glycosylation usually occurs at the consensus of asparagine-X-serine/threonine sequons (N-X-S/T, X is any amino acid except proline).<sup>1,2</sup> In recent years, several atypical sequons, such as N-X-C,<sup>3</sup> N-X-V,<sup>4,5</sup> and N-G,<sup>5</sup> have also been reported as N-glycosylation motifs. Except for the N-X-C motif, which has been confirmed in several known glycoproteins,<sup>3</sup> all other atypical motifs were only

\*Corresponding Author, Phone: (410) 955-3561. Fax: (443) 287-6388. sunsglyco@gmail.com.

The authors declare no competing financial interest.

identified on the basis of the deamidation of asparagine (N) residues in the peptides after PNGase F treatment (with/without 18O labeling) using mass spectrometry-based glycoproteomics.<sup>4,5</sup> However, the atypical sites identified on the basis of deamidation (N) are potentially false positives as deamidation (N) could occur naturally or be induced during sample preparation.<sup>6,7</sup>

Recently, we developed a new N-linked Glycans And Glycosite-containing peptides (NGAG) method for comprehensive analysis of N-linked glycoproteins by simultaneous analyses of N-linked glycans, glycosite-containing peptides, and intact glycopeptides with glycans attached. In this method, N-linked glycans and glycosite-containing peptides were sequentially isolated by solid-phase based extraction and identified by mass spectrometry. Identified glycans and glycosite-containing peptides were then used as the sample-specific database for intact glycopeptide identification. Using the NGAG method, we identified 85 N-linked glycans, 2044 glycosite-containing peptides (with typical N-X-T/S motifs), and 1562 intact glycopeptides from an ovarian cancer cell line (OVCAR-3). From the same study, we also identified peptides that contain deamidation (N) sites at asparagine, but lack the typical N-X-S/T sequon. These deamidated peptides could result from the deglycosylation of N-linked glycopeptides with atypical sites, but they could also be caused by chemical deamidation.<sup>6,7</sup>

In order to determine whether these peptides with deamidation (N) but lacking a typical N-X-S/T sequon are derived from *N*-glycopeptides or from chemical deamidation, we first tried to identify their intact glycopeptides from HILIC-enriched samples. Accordingly, we first constructed a new *N*-glycopeptide candidate database by combining each of these atypical sequon-containing peptides with all glycans identified from OVCAR-3 cells. The intact glycopeptide MS/MS spectra were extracted from the glycopeptide data based on the presence of the glycopeptide specific oxonium ions in the spectra.<sup>8</sup> The oxonium ion-containing MS/MS spectra were then matched to the candidate database on the basis of the accurate masses of their precursors and peptide/peptide + HexNAc fragment ions using GPQuest.<sup>9</sup> By using the same parameters and filters as used in our previous report, we identified five new intact glycopeptides that belong to two unique atypical glycosites. Among these glycopeptides, LVA<sup>146</sup>N#HVASDISLTGGSVVQR from Protein sel-1 homologue 1 (SEL1L) was modified by the glycan Man<sub>9</sub> (Figure 1), and <sup>156</sup>N#SCNNFIYGGCR from Kunitz-type protease inhibitor 2 (SPINT2) was modified by four different oligo-mannose glycans (HexNAc<sub>2</sub>Hex<sub>7</sub>-HexNAc<sub>2</sub>Hex<sub>10</sub>, Figure 2). The peptide LVA<sup>146</sup>N#HVASDISLTGGSVVQR contains an N-X-V motif, and <sup>156</sup>N#SCNNFIYGGCR has an N-X-C motif.

Among the spectra that were matched to these five atypical sequon-containing glycopeptides, all contain at least five oxonium ions and both peptide and peptide + HexNAc ions. The presence of oxonium ions with high quality of MS/MS spectra containing at least 13 b- and y-ions indicated the spectra were from these glycopeptides, and the detected peptide and peptide + HexNAc ions further improved the confidence of intact glycopeptides in higher-energy collisional dissociation (HCD) fragmentation mode.<sup>10</sup> In addition, it was observed that different glycopeptides share similar fragmentation patterns if they contain the same peptide backbone but differ only by their glycan compositions (Figure

2). Similar  $y$ -ion fragment profiles of the peptide portion (prior to the glycosite) were also observed between glycosylated and deglycosylated forms of the same glycosite-containing peptides as we reported previously<sup>11</sup> (Figures 1 and 2). These features enhanced the reliability of the identification of these atypical glycosite-containing glycopeptides.

To evaluate whether the atypical site assignment could come from the false assignment of oxonium ion-containing spectra that belong to typical N-X-S/T motif-containing glycopeptides but the glycosites were not identified in the study, we searched the oxonium ion-containing spectra against a theoretical human N-glycopeptide database. The database was comprised of all N-X-S/T sequon-containing peptides predicted from a RefSeq human protein database<sup>12</sup> (up to 2 missed cleavage sites) and all known human glycans from Byonic software.<sup>13</sup> By using the same assignment parameters as described above, we did not identify any typical N-X-S/T motif-containing peptide that matched better to these spectra than the identified atypical sequon-containing peptides. These results confirmed that these five atypical glycosite-containing glycopeptides are the best matches to the spectra.

To further investigate whether the two atypical sequon-containing glycosites are N-linked, we measured glycosylation alterations at these glycosites in tunicamycin-treated cells. Tunicamycin (TM) is a specific inhibitor for the initial attachment of N-linked glycans; therefore, glycosylation occupancy at the N-linked glycosylation sites is expected to be reduced in TM-treated cells if they were *N*-glycosylated. By comparing relative alterations of the atypical glycosite-containing peptides and their corresponding proteins in TM-treated vs untreated OVCAR-3 cells using SILAC quantification, we found that the glycosylation occupancies at both atypical glycosites were decreased (Figure 3A), which confirmed these two glycosites are *N*-glycosylated and that the deamidation (N) of the peptides was not an artifact of sample preparation. In addition, on the basis of the relative ratios of the deglycosylated and nonglycosylated forms of NSCNNFIYGGCR as well as the ratios of the protein SPINT2 between the TM-treated and untreated cells, the glycosylation occupancies of the glycosite in both  $\pm$  TM-treated cells were estimated using our previously reported method<sup>14</sup> (Figure 3B).

SEL1L is an endoplasmic reticulum membrane protein. It is involved in the endoplasmic reticulum quality control system by playing roles in ER-associated degradation (ERAD) of misfolded proteins.<sup>15</sup> In our study, we identified three N-X-S/T motif-containing glycosites<sup>8</sup> (Asn272, Asn431, and Asn608) in addition to one atypical glycosite (Asn146). SPINT2 is a protease inhibitor, which can inhibit the activities of hepatocyte growth factor and kallikrein (UniprotKB, [www.uniprot.org](http://www.uniprot.org)). Our results showed that our NGAG method identified two typical N-glycosites<sup>8</sup> (Asn57 and Asn94) and one atypical glycosite (Asn156) from SPINT2.

In conclusion, two atypical *N*-glycosites with N-X-C and N-X-V motifs were identified and further confirmed using the NGAG method. Compared to previously reported atypical *N*-glycosites that were identified on the basis of the deamidation of asparagine residues after PNGase F treatment,<sup>4,5</sup> this study further validated the identified atypical motif glycosites by directly identifying their intact glycopeptides. Since the deamidation (N) can occur naturally and during sample preparation, direct identification of the intact glycopeptides

with glycan attached is particularly important for confirmation of these atypical N-glycosylation sites detected by mass spectrometry-based glycoproteomics.

## METHOD SECTION

### MS Data

The mass spectrometry data of glycosite-containing peptides (deglycopeptides) and intact glycopeptides from OVCAR-3 cells generated by the NGAG method can be downloaded from the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>)<sup>16</sup> using the data set identifier PXD001571. The sample preparation and mass spectrometric analysis are described in our previous publication.<sup>8</sup>

### Identification of Atypical Sequon-Containing Peptide Candidates

The glycosite-containing peptide data were searched against RefSeq human protein databases<sup>12</sup> (downloaded from NCBI Web site July 29th, 2013) using Proteome Discoverer V1.4 (Thermo Fisher Scientific). The search parameters were set as follows: up to two missed cleavage were allowed for semi-C-terminal trypsin digestion (nonspecific cleavage at N-termini of peptides); the mass tolerance for precursor and fragment ions were set as 10 ppm and 0.06 Da, respectively; carbamidomethylation (C, +57.0215 Da) was set as a static modification; deamination (N, +0.98 Da), guanidination (K, +42.0218 Da), aniline label (D/E/K/R, +75.0473 Da), guanidination + aniline label (K, +117.0691 Da), and oxidation (M, +15.9949 Da) were set as dynamic modifications. The results were filtered with a 1% FDR. The identified peptides with a deamidated (N1) site but lacking a typical N-X-S/T glycosylation sequon were considered as atypical sequon-containing peptide candidates.

### Identification of Atypical Sequon-Containing Glycopeptides

The precursor mass matching approach in the GPQuest software was used to identify intact glycopeptides.<sup>8,9</sup> Briefly, the oxonium ion-containing MS/MS spectra were extracted from the glycopeptide data and were matched to the atypical sequon-containing N-glycopeptide candidate database (comprised of the identified N-glycan and atypical sequon-containing peptides) to assign the atypical sequon-containing glycopeptides. All other parameters and filters were the same as previously described.<sup>8</sup>

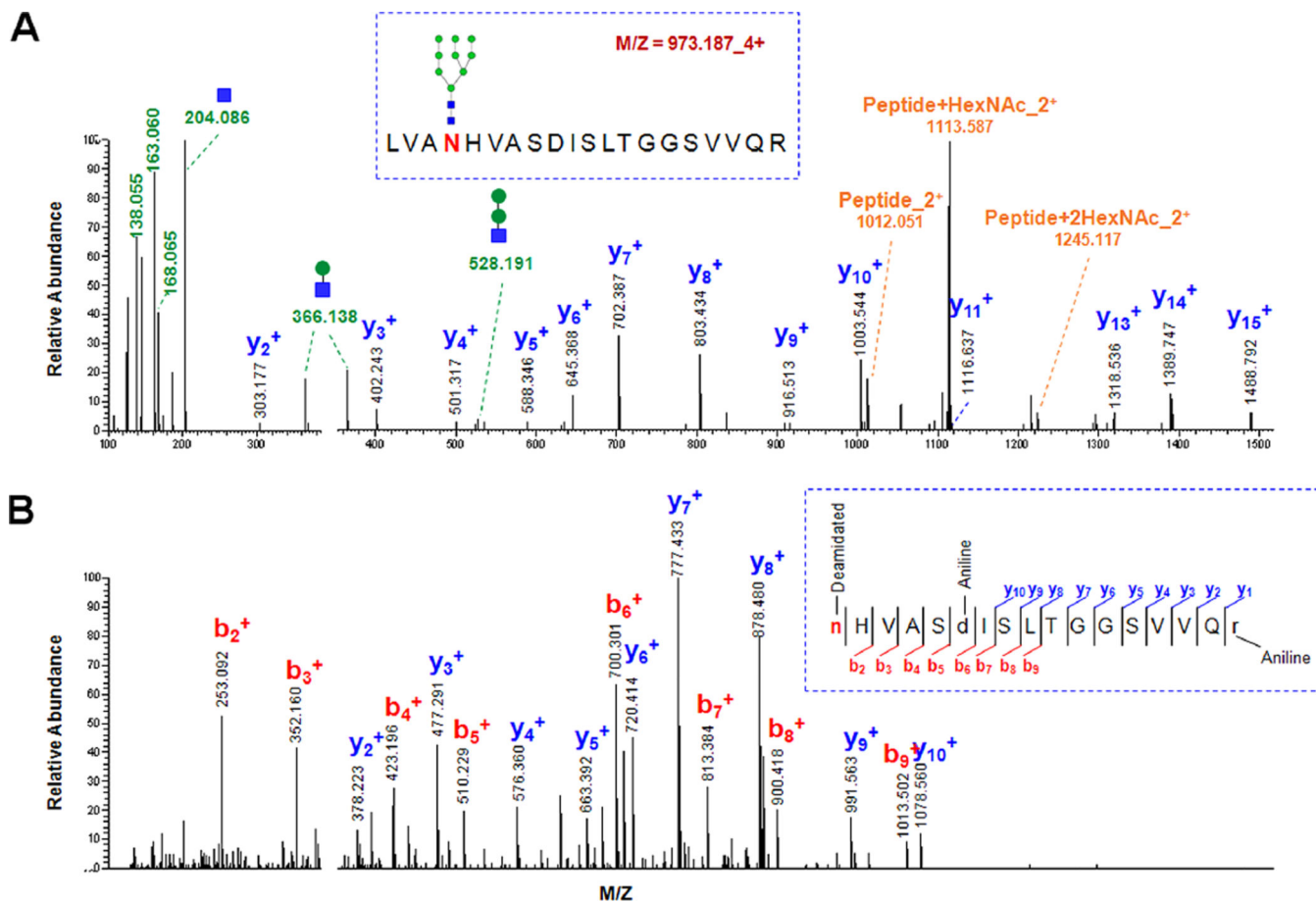
## Acknowledgments

This work was supported by the National Institutes of Health, National Cancer Institute, Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA160036), the Early Detection Research Network (EDRN, U01CA152813 and U24CA115102) and R01CA112314 and by the National Institutes of Health, National Heart Lung and Blood Institute Programs of Excellence in Glycosciences (PEG, P01HL107153) and the Johns Hopkins Proteomics Center (N01-HV-00240). We also acknowledge Punit Shah and Shadi Toghi Eshghi for their help in technical assistance and data analysis.

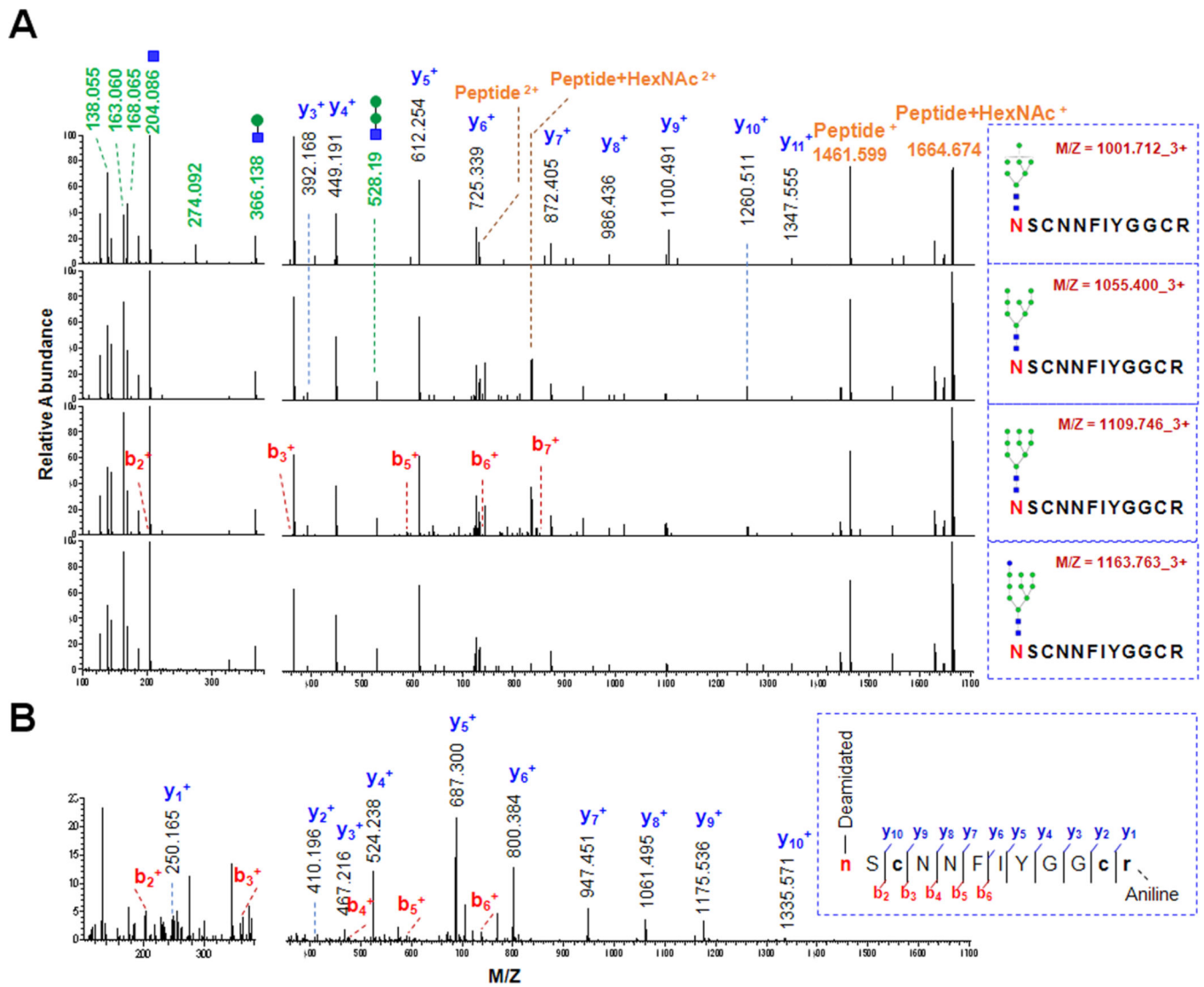
## REFERENCES

1. Ohtsubo K, Marth JD. *Cell*. 2006; 126:855–867. [PubMed: 16959566]
2. Ajit, V.; Richard, C.; Jeffrey, E.; Hudson, F.; Gerald, H.; Jamey, M. *Essentials of Glycobiology*. 2nd. New York: Cold Spring Harbor Laboratory Press; 2009.
3. Bause E, Legler G. *Biochem. J*. 1981; 195:639–644. [PubMed: 7316978]

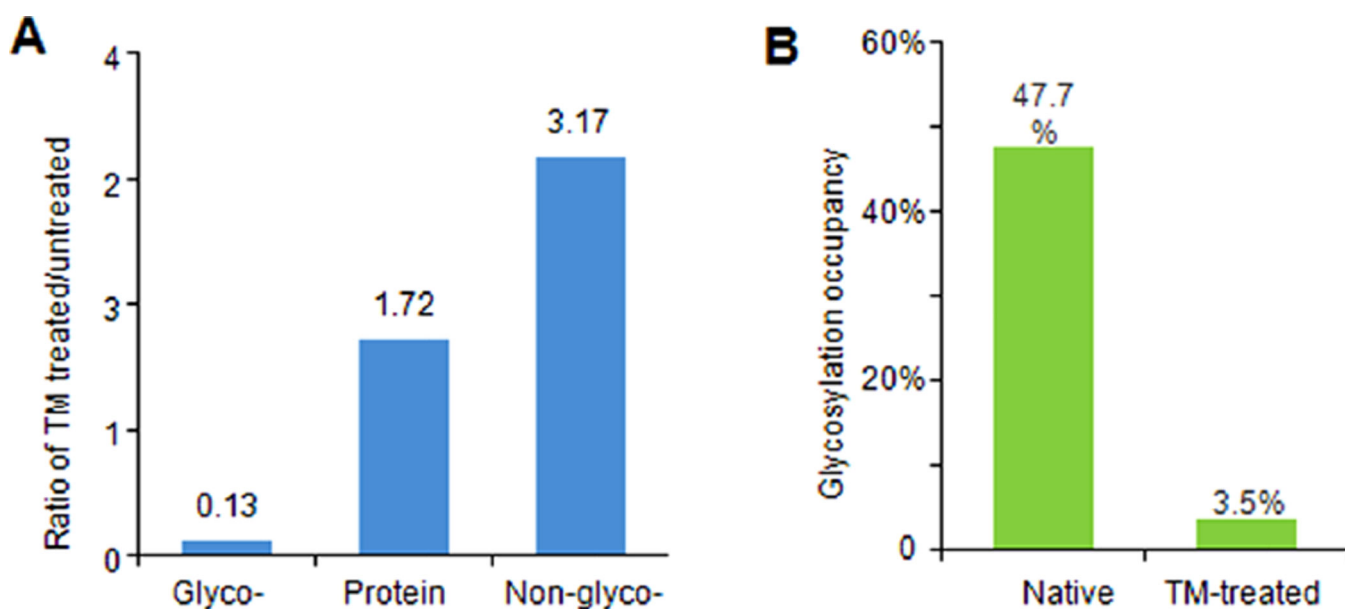
4. Baycin-Hizal D, Tian Y, Akan I, Jacobson E, Clark D, Wu A, Jampol R, Palter K, Betenbaugh M, Zhang H. *Anal. Chem.* 2011; 83:5296–5303. [PubMed: 21591763]
5. Zielinska DF, Gnad F, Wi niewski JR, Mann M. *Cell.* 2010; 141:897–907. [PubMed: 20510933]
6. Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR. *J. Proteome Res.* 2012; 11:1949–1957. [PubMed: 22256963]
7. Hao P, Ren Y, Alpert AJ, Sze SK. *Mol. Cell. Proteomics.* 2011; 10:O111.009381. [PubMed: 21784994]
8. Sun S, Shah P, Eshghi ST, Yang W, Trikannad N, Yang S, Chen L, Aiyetan P, Hoti N, Zhang Z, Chan DW, Zhang H. *Nat. Biotechnol.* 2015
9. Toghi Eshghi S, Shah P, Yang W, Li X, Zhang H. *Anal. Chem.* 2015; 87:5181–5188. [PubMed: 25945896]
10. Hart-Smith G, Raftery M. *J. Am. Soc. Mass Spectrom.* 2012; 23:124–140. [PubMed: 22083589]
11. Yang W, Shah P, Toghi Eshghi S, Yang S, Sun S, Ao M, Rubin A, Jackson JB, Zhang H. *Anal. Chem.* 2014; 86:6959–6967. [PubMed: 24941220]
12. Pruitt KD, Tatusova T, Maglott DR. *Nucleic Acids Res.* 2007; 35:D61–D65. [PubMed: 17130148]
13. Bern M, Kil YJ, Becker C. *Current Protocols in Bioinformatics.* 2012; 40:13.20.1–13.20.14.
14. Sun S, Zhang H. *Anal. Chem.* 2015; 87:6479–6482. [PubMed: 26066578]
15. Hosokawa N, Wada I. *FEBS J.* 2015
16. Vizcaíno JA, Côté RG, Csordas A, Dianas JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, et al. *Nucleic Acids Res.* 2013; 41:D1063–D1069. [PubMed: 23203882]



**Figure 1.** Identification and validation of an N-X-V motif-containing glycosite using the NGAG method. (A) A spectrum of the intact glycopeptide LVAN<sup>#</sup>HVASDISLTGGSVVQR + Man<sub>9</sub> from SEL1L. The oxonium ions (green) were used to extract the intact glycopeptide spectrum, and the masses of the precursor and peptide/peptide + HexNAc ions (orange) as well as the b/y-ions of the peptide portion (b-ions: blue; y-ions: red) were used for intact glycopeptide identification. (B) A spectrum of the deglycosylated form of the peptide LVAN<sup>#</sup>HVASDISLTGGSVVQR. The only asparagine residue in the peptide was identified as the glycosylation site based on the deamidation.



**Figure 2.** Identification and validation of an N-X-C motif-containing glycosite using the NGAG method. (A) The spectra of the intact glycopeptides N<sup>#</sup>SCNNFIYGGCR + (HexNAc<sub>2</sub>Hex<sub>7</sub>-HexNAc<sub>2</sub>Hex<sub>10</sub>). (B) A spectrum of the deglycosylated form of the peptide N<sup>#</sup>SCNNFIYGGCR. The first asparagine residue was identified as the glycosylation site based on the deamidation.



**Figure 3.** Glycosylation occupancy alteration at the glycosite Asn156 of SPINT2 ( $^{156}\text{N}\#\text{SCNNFIYGGCR}$ ) with tunicamycin treatment. (A) Tunicamycin-treated/untreated ratios of the glycosylated and nonglycosylated forms of the peptide NSCNNFIYGGCR as well as the protein SPINT2. (B) The glycosylation occupancies at the glycosite Asn156 in both native and tunicamycin treated OVCAR-3 cells. The occupancies were calculated on the basis of the three relative ratios shown in (A) using our previously reported method.<sup>14</sup>