Supplementary Information

for

Towards Automated *N*-glycopeptide Identification in Glycoproteomics

Ling Y. Lee[1], Edward S. X. Moh[1], Benjamin L. Parker[2], Marshall Bern[3], Nicolle H. Packer[1], and Morten Thaysen-Andersen*[1]

[1]*Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia.*

[2]*Charles Perkins Centre, School of Molecular Bioscience, The University of Sydney, Sydney, Australia*

[3]*Protein Metrics Inc., San Carlos, California 94070, United States*

Running title: Towards automated *N*-glycopeptide identification in glycoproteomics

Keywords: *N*-glycosylation, LC-MS/MS, glycopeptide, glycoproteomics, basigin, glycoprofiling, glycomics, automated glycopeptide identification, Byonic.

***Corresponding author**:

Dr. Morten Thaysen-Andersen

Department of Chemistry and Biomolecular Sciences

Macquarie University

Sydney, NSW-2109 - Australia

Phone / Fax: (+61) 2 9850 7487 / (+61) 2 9850 6192

E-mail: morten.andersen@mq.edu.au

Table of contents

Abbreviations

ACN, acetonitrile; AGC, automatic gain control; CID, collision induced dissociation; EIC, extracted ion chromatogram; ESI, electrospray ionisation; Fuc, fucose; GlcNAc, *N*-acetylglucosamine; HCD, higher-energy collision dissociation; Hex, hexose; HexNAc, *N*-acetylhexosamine; HEK, human embryonic kidney; LC-MS/MS, liquid chromatography tandem mass spectrometry; Man, mannose; NeuAc, *N*-acetylneuraminic acid; NeuGc, *N*-glycolylneuraminic acid; PGC, porous graphitised carbon; QE, Q-Exactive.

FIGURE LEGENDS

Supplementary Fig. S-1. Canonical amino acid sequence of human basigin (UniProtKB: BAS_HUMAN, P35613). The three potential *N*-glycosylation sites (sequons) are indicated in red i.e. Asn160: NDS, Asn268: NGS and Asn302: NGT.

Supplementary Fig. S-2. The *N*-glycan type distribution of basigin as evaluated quantitatively by the *N*-glycome profiling. The majority of the identified *N*-glycans belonged to the complex/hybrid type i.e. 96.9 ± 4.1% often containing core fucosylation and occasionally NeuAc-type sialylation. The remaining *N*-glycans were of the paucimannosidic i.e. 1.3 ± 0.1% or high mannosidic type i.e. 1.9 ± 0.1%. See also Supplementary Table S-1.

Supplementary Fig. S-3. The analysis of basigin peptide mixtures on (A) ion trap CID-MS/MS and (B) high resolution QE Orbitrap HCD-MS/MS platforms showed several basigin glycopeptide clusters. The top panels shows the base peak chromatogram (*m/z* 400-2,000) and bottom panels show the EIC of *m/z* 366.1 ± 0.2 (for ion trap) and *m/z* 204.08 ± 0.01 (for QE Orbitrap), which were used to extract spectra containing the HexHexNAc and HexNAc saccharide oxonium ions, respectively. The glycopeptide clusters covering the individual sites of basigin are indicated i.e. Asn160 (blue), Asn268 (green) and Asn302 (red). In particular Asn268 glycopeptides showed extensive peptide heterogeneity, whereas Asn160 glycopeptides eluted as a single cluster without much peptide heterogeneity. Asn302 glycopeptides showed little, if any, reversed phase LC retention and were thus excluded from further analysis.

Supplementary Fig. S-4. Example of manual annotation of ion trap CID-MS/MS (A) and QE Orbitrap HCD-MS/MS of a basigin Asn160 glycopeptide (^{153}ILLT**C**SL**N**DSATEVTGHR^{170}) of *m/z* 1076.8^{3+} carrying the *N*-glycan composition, $HexNAc_3Hex_3Fuc_1$. Confident identification of the peptide carrier and the *N*-glycan monosaccharide composition was facilitated by i) ensuring an accurate match of the observed and theoretical glycopeptide mass, ii) performing a thorough annotation of the relevant fragments in the fragment spectra i.e. saccharide oxonium ions, the y- and b- ion series (red) and the B-/C- and Y-/Z-ion series (blue) and, finally, iii) ensuring that the

glycopeptide glycoforms of interest (black EIC trace) eluted within a cluster of other related glycopeptide glycoforms (red EIC traces).

Supplementary Fig. S-5. Schematic presentation of human basigin and the main *N*-glycosylation features of the three *N*-glycosylation sites (Asn160, Asn268 and Asn302). Varying degrees of *N*-glycan micro- and macro-heterogeneity were observed across the sites; The most abundant glycoforms based on the established site-specific reference glycoprofiles, predominantly the GlcNAc-terminating *N*-glycans, are shown. * The site-occupancy was not determined for the Asn302 site.

Supplementary Fig. S-6. The unique retention time of the heavily truncated (chitobiose-type) *N*-glycoforms of basigin Asn160 glycopeptides (^{153}ILLT**C**SL**N**DSATEVTGHR^{170}) as evaluated by EICs of several related glycopeptides sharing the same peptide backbone (i.e. peptide + $HexNAc_1$, *m/z* 730.8^{3+} and peptide + $HexNAc_1Fuc_1$, *m/z* 779.5^{3+}) showed that these unusual species are not artificially arising from *in-source* fragmentation from larger more abundant glycopeptide forms (peptide + $HexNAc_5Hex_3Fuc_1$, *m/z* 909.7^{4+}). These truncated species also eluted differently compared to the corresponding non-glycosylated peptide (*m/z* 663.1^{3+}). *Indicates contaminating signals.

Supplementary Fig. S-7. The distribution of Met267 oxidation and carbamidomethylation on non-glycosylated (left) and glycosylated (right) Asn268 glycopeptides (^{265}ALMNGSESR^{273}) derived from basigin as evaluated based on manual annotation of QE Orbitrap HCD-MS/MS data. The distribution of oxidation and carbamidomethylation was clearly dependent of the Asn268 occupancy status.

Supplementary Fig. S-8. Examples of incorrect (**A**) and correct (**B**) Byonic-based annotation of two similar QE Orbitrap HCD-MS/MS spectra of the same precursor i.e. Asn160 glycopeptide (^{153}ILLTCSLNDSATEVTGHR^{170}) at *m/z* 1012.441^{4+} due to inconsistent precursor isotope selection. The monosaccharide composition ($HexNAc_4Hex_5Fuc_3$) was incorrectly identified based on incorrect isotopic precursor selection (labelled 'x'). The same peptide with the more likely composition $HexNAc_4Hex_5Fuc_1NeuAc_1$ was correctly assigned following correct monoisotopic peak picking (y). Both spectra clearly showed presence of NeuAc-type sialylation i.e. oxonium ions of *m/z* 292.128 and

the fact that the Byonic data output indicated a '±1 error' for all the glycopeptides annotated with $HexNAc_4Hex_5Fuc_3$ supported a correct assignment of a sialoglycopeptide in the bottom spectrum.

Supplementary Fig. S-9. The accuracy and coverage of Byonic-driven glycoprofiling were determined for Asn160 glycosylation using searches with (brown trace) and without (dotted blue trace) the inclusion of a non-standard variable peptide modification i.e. Met carbamidomethylation. Allowing the non-standard peptide heterogeneity yielded, as expected, higher coverages, but at the cost of lower accuracies relative to the standard search criteria for the lower confidence thresholds i.e. Byonic scores 50-100, see insert for zoom.

Supplementary Fig. S-10. Examples of QE Orbitrap HCD-MS/MS spectra of basigin-derived Asn268 glycopeptides (^{265}AL**MN**GSESR^{273}) displaying significant neutral losses of (A-B) 105.1300 Da arising from decomposition of Met carbamidomethylation (CH_3-S-CH_2CO-NH_2) and (C-D) 64.1079 Da arising from decomposition of Met oxidation (CH_3-S-OH). Both Byonic (A, C) and manual (B, D) annotations of the same HCD-MS/MS spectra (paired) are provided. A) Using Byonic, the *m/z* 1139.967^{2+} precursor was incorrectly identified to carry Met267 oxidation and the Asn268 monosaccharide composition $HexNAc_4Hex_3$. B) The manual annotation of the same spectrum showed that Met267 was instead carbamidomethylated with the *N*-glycan composition $HexNAc_3Hex_4$. This incorrect identification by Byonic was caused by the lack of annotation of the abundant neutral loss of 105.1300 Da arising from almost all fragments of the Met267 carbamidomethylated peptide. C) Although Byonic also did not annotate the neutral losses of 64.1079 Da of the Met267 oxidised peptide, it identified the correct glycopeptide, but with lower Byonic score. D) The manual annotation showed that Byonic was able to perform the identification since part of the original fragments ions (non-neutral loss ions e.g. intact Y_1) were still present in the HCD-MS/MS.

Supplementary Text S-1 (Experimental Section).

Materials

HEK293-derived recombinant human basigin (UniProtKB: P35613) fused with a C-terminal His tag isolated to >95% purity as determined by SDS-PAGE was purchased from ACRObiosystems (product CD7-H5222) (Newark, NE). Ultrapure water (Millipore) was used in all aqueous solutions. Unless otherwise stated, all chemicals were from Sigma. Dithiothreitol, iodoacetamide, Coomassie Blue G250 and AG 50W X8 cation exchange resin (Biorad), NuPAGE 4-12% Bis-Tris precast polyacrylamide gels and molecular weight markers (Invitrogen), sequencing-grade porcine trypsin (Promega) and *Flavobacterium meningospecticum* peptide:*N*-glycosidase F (Roche) were purchased.

Preparation of basigin *N*-glycans

N-glycans were released from approximately 5 µg basigin as previously described (1, 2). Briefly, basigin was reduced (5 mM dithiothreitol) for 10 min at 70°C and alkylated (10 mM iodoacetamide, both final concentrations) for 30 min at room temperature in the dark, then immobilised on a methanol-activated PVDF membrane (Millipore). After overnight drying, the membrane-bound protein was incubated with 2.5 U peptide:*N*-glycosidase F for 16 h at 37°C to ensure complete release of *N*-glycans. Released *N*-glycans were incubated with 100 mM ammonium acetate (pH 5) for 1 h at room temperature and subsequently dried by vacuum centrifugation. Reduction of *N*-glycans was performed with 20 µl 1 M sodium borohydride in 50 mM potassium hydroxide for 3 h at 50°C. Reduced samples were quenched with 2 µl glacial acetic acid and desalted on columns of AG 50W X8 cation exchange resin packed on top of ZipTip C18 columns (Millipore) as described. The desalted *N*-glycans were dried by vacuum centrifugation and residual borate was removed by repeated cycles of methanol addition and evaporation in a vacuum centrifuge. Further desalting was performed on small home-made columns packed with porous graphitised carbon (PGC) resin (Grace) on top of ZipTip C18 columns as described. The purified *N*-glycans were eluted with 40% (v/v) acetonitrile (ACN) containing 0.1% (v/v) trifluoroacetic acid, dried and stored at -80°C if not analysed immediately.

Preparation of basigin peptides and a complex peptide mixture

Basigin (~10 µg) was reduced and alkylated as described above and loaded on a 4-12% Bis-Tris polyacrylamide gel for electrophoresis at 200 V for 20 min. The gel was fixed in 40% (v/v) ethanol and 10% (v/v) acetic acid for 1 h, stained for 3 h with Coomassie Blue G250 and then destained in water. The basigin containing band was excised and sliced into approximately 1 mm^3 cubes. The gel pieces were destained with 50% (v/v) ACN in 50 mM ammonium bicarbonate until clear, dehydrated in 100% (v/v) ACN and dried. Protein digestion was performed overnight at 37°C using 1:30 (enzyme/protein substrate, w/w) trypsin. The resulting peptide mixture was extracted from the gel plugs, divided into aliquots, dried and stored at -80°C until LC-MS/MS analysis.

Crude complex peptide mixtures were prepared by digesting the entire complement of secreted proteins from serum-free culture media of a MDA157 human breast epithelial cancer cell line after cysteine reduction and alkylation (as above) using trypsin at 1:30 (w/w) ratio and incubation overnight at 37°C. The resulting complex peptide mixture was aliquoted, dried and stored at -20°C, if not used immediately. One pmol of basigin peptides equating to approximately 40 ng was spiked into ~2 µg of the complex peptide mixture resuspended in 0.1% formic acid (v/v) in a total volume of 10 µl prior to LC-MS/MS analysis. This ratio was chosen since it theoretically forms a 1:1 molar abundance of basigin and 50 other medium-high abundance glycoproteins in the complex mixture. We argue that for the purpose of this analytical exercise, this is an appropriate ratio since basigin is typically identified in cell lysates and secreted fractions amongst the most abundant proteins.

LC-MS/MS of basigin *N*-glycans

Basigin *N*-glycans alditols were taken up in water, transferred to high recovery vials (Agilent) and injected onto a PGC LC column (5 µm particle size, Hypercarb KAPPA, 100 mm length x 200 µm inner diameter, 250 Å pore size, Thermo Scientific) on a 1260 Infinity LC system (Agilent) connected directly to an ESI-MS/MS 3D ion trap mass spectrometer (LC/MSD Trap XCT Plus Series 1100, Agilent Technologies). Separation was performed for 85 min over a linear gradient of 0-45% (v/v) ACN/10 mM ammonium bicarbonate at a constant flow rate of 2 µl/min. Data were acquired in negative ion polarity mode with three scan events at a scan speed of 8,100 *m/z*/s: a MS full scan (*m/z*

400-2,200) and data-dependent MS/MS scans after isolation and resonance activation CID fragmentation of the top two most intense precursor ions with an absolute intensity threshold >30,000 and a relative intensity threshold >5% relative to the base peak. The *m/z* axis calibration of the mass spectrometer was performed using a well-defined tune mix (Agilent) prior to acquisition. *N*-glycans released in parallel from bovine fetuin (Sigma) served as a control for the sample preparation and the LC-MS/MS performance. Differences between observed and theoretical precursor and fragment masses were generally better than 0.2 Da. LC-MS/MS technical duplicates of the basigin *N*-glycans were acquired.

LC-MS/MS of basigin peptides with and without a complex peptide mixture

Aliquots containing basigin peptide mixtures were taken up in 0.1% (v/v) formic acid, transferred to high recovery vials (Agilent) and injected onto LC systems connected to a HCT 3D ion trap (Bruker Daltonics) or a Q-Exactive (QE) Orbitrap (Thermo Fisher) mass spectrometer.

For the ion trap analysis, basigin peptide mixtures were separated on a C18 reversed-phase column (ProteCol HQ303, 3 µm particle size, 100 mm length, 300 µm inner diameter, 300 Å pore size, SGE, Australia) on an Ultimate 3000 LC system (Dionex). The column was heated to 45°C during the separation. Separation was performed using a binary gradient solvent system made up of solvent A (aqueous 0.1% (v/v) formic acid) and solvent B (0.1% (v/v) formic acid in ACN). The flow rate was constant at 5 µl/min during the 95 min gradient: Isocratic 100% solvent A over 8 min, 0-30% solvent B over 60 min; 30-60% solvent B over 7 min; 60-80% solvent B over 1 min; isocratic 80% solvent B over 10 min and isocratic 100% solvent A over 9 min. Basigin peptide mixtures were injected in triplicates (8 pmol/injection) and spectra were recorded in the positive ion mode with four scan events at a scan speed of 8,100 *m/z*/s: an MS full scan (*m/z* 300-2,200) followed by three data-dependent MS/MS scans after isolation and resonance activation CID fragmentation of the three most intense precursor ions with an absolute intensity threshold >30,000 and a relative intensity threshold >5% relative to the base peak. The *m/z* axis calibration of the mass spectrometer was performed using a well-defined tune mix (Agilent) prior to acquisition. Tryptic digests of bovine serum albumin (Sigma)

were used to ensure high LC-MS/MS performance prior to data acquisition. Differences between observed and theoretical precursor and product ion masses were generally better than 0.2 Da.

For the QE Orbitrap analysis, basigin peptides with and without a background of a complex peptide mixture were loaded onto a custom-made C18 reversed-phase LC column packed with Halo C18 resin (2.7 µm particle size, 100 mm length, 75 µm inner diameter, 160 Å pore size) (Advance Materials Technology, Wilmington, DE). The LC was equipped with an LC trapping column (35 mm length, 100 µm inner diameter) made from the same resin. Separation of peptides was performed using an identical gradient as for the ion trap analysis (described above) with a constant flow rate of 300 nl/min. Samples were injected in triplicates (1 pmol/injection) and spectra were recorded with a resolution of 35,000 in the positive polarity mode over the range of *m/z* 350-2,000 and an automatic gain control (AGC) target value of 1×10^6. The ten most intense precursor ions in each full scan were isolated for HCD-MS/MS fragmentation and fragment mass analysis using the following settings: normalised collision energy, 35%; AGC target, 2×10^5; isolation window *m/z* 3.0, and dynamic exclusion enabled, fragment resolution was set to 17,500. Yeast enolase was used between samples to monitor the LC-MS/MS performance. Differences between observed and theoretical precursor and product ion masses for the QE Orbitrap analyses were generally better than 10 ppm.

Data analysis of basigin *N*-glycome

LC-MS/MS raw data files of the basigin *N*-glycome were viewed and manually analysed using Data Analysis v4.0 (Bruker Daltonics). Observed monoisotopic masses were used after deconvolution to search for possible glycan monosaccharide compositions using GlycoMod (http://web.expasy.org/glycomod/). The monosaccharide compositions were subsequently verified manually by *de novo* sequencing of the corresponding CID-MS/MS spectra and by the absolute and relative PGC-LC retention time (3). The identified *N*-glycan monosaccharide compositions were all assumed to have a trimannosylchitobiose core i.e. two *N*-acetylglucosamine (GlcNAc) and three mannose (Man) residues unless they were truncated to paucimannosidic or chitobiose core type structures. Core fucosylated *N*-glycans were distinguished from antenna fucosylated structures by the presence of the diagnostic ions (*m/z* 350/368) corresponding to the Y_1-/Z_1-fragments comprising an

α1,6-linked fucose (Fuc) attached to the reducing end GlcNAc residue. Other isomeric structures such as those with tri/tetra-antennary or bisecting GlcNAc branching features were left undetermined. The distribution of the basigin *N*-glycan monosaccharide compositions was calculated using their relative abundances as measured by the relative peak areas of their corresponding extracted ion chromatograms (EICs) of all the observed *N*-glycan charge states. The relative *N*-glycan distribution was calculated as an average of a duplicate profiling experiment.

REFERENCES

1. Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protocols.* **2012**, *7*, 1299-1310.

2. Lee, L. Y.; Thaysen-Andersen, M.; Baker, M. S.; Packer, N. H.; Hancock, W. S.; Fanayan, S. Comprehensive N-Glycome Profiling of Cultured Human Epithelial Breast Cells Identifies Unique Secretome N-Glycosylation Signatures Enabling Tumorigenic Subtype Classification. *Journal of Proteome Research.* **2014**, *13*, 4783-4795.

3. Pabst, M.; Bondili, J. S.; Stadlmann, J.; Mach, L.; Altmann, F. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem.* **2007**, *79*, 5051-5057.

Supplementary Table S-1. Global and site-specific distribution of *N*-glycoforms of human basigin mapped using manual annotation of ion trap CID-MS/MS or QE Orbitrap HCD-MS/MS data. The relative abundances (in %) of the individual glycoforms are presented as mean ± SD (n = 3).

***N*-glycan monosaccharide composition**	***N*-glycome profiling** Ion trap CID-MS/MS, rel. abundance, %	**Intact *N*-glycopeptide profiling** (relative abundance in %)					
		Asn160		**Asn268**		**Asn302**	
		Ion trap CID-MS/MS	QE HCD-MS/MS	Ion trap CID-MS/MS	QE HCD-MS/MS	Ion trap CID-MS/MS	QE HCD-MS/MS
$HexNAc_1$		1.4 ± 0.4	1.4 ± 0.2		3.9 ± 0.2		
$HexNAc_1Fuc_1$		1.6 ± 0.4	1.2 ± 0.1				
$HexNAc_2Hex_3$	0.9 ± 0.1					2.0 ± 0.4	
$HexNAc_2Hex_3Fuc_1$	0.3 ± 0.0		0.5 ± 0.1		0.2 ± 0.0		
$HexNAc_2Hex_4$	0.4 ± 0.0		0.2 ± 0.1		0.7 ± 0.1		
$HexNAc_2Hex_5$	1.5 ± 0.1	1.5 ± 0.4	5.8 ± 2.0	7.7 ± 1.6	18.1 ± 0.5	1.8 ± 0.7	7.6 ± 0.9
$HexNAc_3Hex_3$	5.8 ± 0.2	4.3 ± 0.5	6.1 ± 0.2	2.8 ± 0.7	4.9 ± 0.5	28.8 ± 3.0	6.0 ± 0.6
$HexNAc_3Hex_3Fuc_1$	1.5 ± 0.1	9.4 ± 1.2	12.8 ± 0.3	2.0 ± 0.9	7.5 ± 0.4		
$HexNAc_3Hex_4$	0.2 ± 0.0	1.1 ± 0.4	0.3 ± 0.2		1.0 ± 0.1		
$HexNAc_3Hex_4Fuc_1$			0.3 ± 0.2		0.9 ± 0.1		
$HexNAc_3Hex_5$	0.2 ± 0.2						
$HexNAc_3Hex_5Fuc_1$			0.6 ± 0.0		0.3 ± 0.1		
$HexNAc_3Hex_5NeuAc_1$	0.5 ± 0.1						
$HexNAc_3Hex_6NeuAc_1$					0.03 ± 0.0		
$HexNAc_3Hex_6$	0.8 ± 0.2		0.3 ± 0.1				
$HexNAc_3Hex_6Fuc_1$	0.5 ± 0.1		0.6 ± 0.1				
$HexNAc_4Hex_3$	10.3 ± 0.6	2.4 ± 0.4	1.5 ± 0.3	1.5 ± 0.6	4.0 ± 0.4	42.1 ± 2.5	26.7 ± 5.3
$HexNAc_4Hex_3Fuc_1$	10.4 ± 0.3	11.4 ± 1.1	14.0 ± 0.5	9.3 ± 1.1	9.0 ± 0.1		
$HexNAc_4Hex_4$	5.0 ± 0.2	0.7 ± 0.5	2.1 ± 0.6		3.0 ± 0.3		
$HexNAc_4Hex_4Fuc_1$	6.4 ± 0.2	6.4 ± 0.1	6.2 ± 0.6		3.1 ± 0.1		
$HexNAc_4Hex_4NeuAc_1$	1.7 ± 0.1		0.8				
$HexNAc_4Hex_4Fuc_1NeuAc_1$	1.4 ± 0.1	2.3 ± 0.5	0.9 ± 0.1		0.1 ± 0.1		
$HexNAc_4Hex_5$	0.9 ± 0.0		0.5 ± 0.2				
$HexNAc_4Hex_5Fuc_1$	1.3 ± 0.0	0.7 ± 0.3	1.9 ± 0.5	6.6 ± 0.3	0.3 ± 0.0		
$HexNAc_4Hex_5NeuAc_1$	0.7 ± 0.2						
$HexNAc_4Hex_5Fuc_1NeuAc_1$	0.6 ± 0.3	1.6 ± 0.2	1.9 ± 0.3				
$HexNAc_5Hex_3$	15.0 ± 0.3	7.6 ± 2.8	6.8 ± 1.0	17.5 ± 1.9	10.4 ± 0.4	18.0 ± 2.0	20.0 ± 2.5
$HexNAc_5Hex_3Fuc_1$	13.6 ± 0.4	14.4 ± 1.7	8.7 ± 2.1	14.5 ± 1.4	7.6 ± 0.4		
$HexNAc_5Hex_4$	4.0 ± 0.1	1.1 ± 0.3	0.5 ± 0.2		2.7 ± 0.1		
$HexNAc_5Hex_4Fuc_1$	3.5 ± 0.0	6.8 ± 0.7	4.9 ± 0.6	0.7 ± 0.3	2.5 ± 0.1		
$HexNAc_5Hex_4NeuAc_1$	0.2 ± 0.0						
$HexNAc_5Hex_4Fuc_1NeuAc_1$	0.1 ± 0.0	2.3 ± 0.8	0.7 ± 0.1				
$HexNAc_5Hex_5Fuc_1$	0.9 ± 0.0		1.9 ± 0.3		0.2 ± 0.0		
$HexNAc_5Hex_5Fuc_1NeuAc_1$		1.4 ± 0.5	0.2 ± 0.0				
$HexNAc_5Hex_6Fuc_1$	0.2 ± 0.0		0.4 ± 0.2				
$HexNAc_6Hex_3$	6.8 ± 0.2	6.8 ± 0.4	3.3 ± 0.5	16.8 ± 1.8	9.5 ± 0.3	7.3 ± 0.6	39.7 ± 3.9
$HexNAc_6Hex_3Fuc_1$	3.2 ± 0.1	8.8 ± 0.4	8.8 ± 0.4	17.0 ± 1.4	7.8 ± 0.3		
$HexNAc_6Hex_4$	0.2 ± 0.0	1.5 ± 0.2	1.3 ± 0.1	1.9 ± 0.6	1.9 ± 0.0		
$HexNAc_6Hex_4Fuc_1$	0.9 ± 0.0	5.3 ± 0.7	1.5 ± 0.1	1.7 ± 0.6	0.9 ± 0.2		
$HexNAc_6Hex_4Fuc_1NeuAc_1$	0.03 ± 0.0						
$HexNAc_6Hex_5Fuc_1$			0.9 ± 0.3				
$HexNAc_7Hex_3$	0.2 ± 0.0						
$HexNAc_7Hex_3Fuc_1$			1.0 ± 0.2				
Sum	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Supplementary Table S-2. Observed peptide heterogeneity of tryptic glycopeptides derived from human basigin and their retention times on the ion trap and QE Orbitrap LC-MS/MS platforms. Modified amino acid residues are in bold underscore and the specific modifications are indicated. See Supplementary Fig. S-3 for the base peak chromatograms showing the glycopeptide clusters. *Carbamid., carbamidomethylation; Deamid., deamidation; Ox., oxidation. **ND, Not detected.

Basigin peptide	**Glyco-syla-tion site**	**Observed peptide heterogeneity***	**Retention time** (min)	
			Ion trap	**QE**
ILLT**C**SL**N**DSATEVTGHR	Asn160	Carbamid. (Cys), *N*-glycan (Asn)	43-45	45-46
ILLT**C**SL**N**DSATEVTGHR	Asn160	Carbamid. (Cys), Deamid. (Asn)	44-45	46-47
ILLT**C**SLNDSATEVTGHR	Asn160	Carbamid. (Cys)	44-45	47-48
ALM**N**GSESR	Asn268	*N*-glycan (Asn)	19-20	26-28
AL**M****N**GSESR	Asn268	Ox. (Met), *N*-glycan (Asn)	16-17	22-24
AL**M****N**GSESR	Asn268	Carbamid. (Met), *N*-glycan (Asn)	15-16	23-24
AL**M****N**GSESR	Asn268	Ox. (Met), Deamid. (Asn)	16-17	23-24
AL**M****N**GSESR	Asn268	Carbamid. (Met), Deamid. (Asn)	14-15	24-26
ALM**N**GSESR	Asn268	Deamid. (Asn)	19-20	28-29
AL**M**NGSESR	Asn268	Ox. (Met)	16-17	22-23
AL**M**NGSESR	Asn268	Carbamid. (Met)	ND**	24-25
ALMNGSESR	Asn268	No modification	19-20	27-28
ITDSED**K**ALM**N**GSESR	Asn268	*N*-glycan (Asn), Missed cleavage (Lys)	26-28	31-33
ITDSED**K**AL**M****N**GSESR	Asn268	Ox. (Met), *N*-glycan (Asn), Missed cleavage (Lys)	22-23	28-29
ITDSED**K**AL**M****N**GSESR	Asn268	Carbamid. (Met), *N*-glycan (Asn), Missed cleavage (Lys)	19-21	26-28
ITDSED**K**AL**M****N**GSESR	Asn268	Ox. (Met), Deamid. (Asn), Missed cleavage (Lys)	23-24	29-30
ITDSED**K**AL**M****N**GSESR	Asn268	Carbamid. (Met), Deamid. (Asn), Missed cleavage (Lys)	21-22	27-28
ITDSED**K**ALM**N**GSESR	Asn268	Deamid. (Asn), Missed cleavage (Lys)	28-29	28-29
ITDSED**K**AL**M**NGSESR	Asn268	Ox. (Met), Missed cleavage (Lys)	23-24	29-30
ITDSED**K**AL**M**NGSESR	Asn268	Carbamid. (Met), Missed cleavage (Lys)	20-21	27-28
ITDSED**K**ALMNGSESR	Asn268	Missed cleavage (Lys)	19-20	32-33
CNGTSSK	Asn302	Carbamid. (Cys), *N*-glycan (Asn)	5-7	< 1

Supplementary Table S-3. Overview of Met267-modified basigin Asn268 glycopeptides posing significant challenges for correct identification due to significant neutral losses of Met267 oxidation (-64.1079 Da, CH_3SOH) and carbamidomethylation (-105.1300 Da, C_3H_7NSO), which were not assigned and scored by Byonic and the fact that their mass difference (oxidation [+15.9949 Da] and carbamidomethylation [+57.0215 Da], Δm = 41.0266 Da) equates to the mass difference between a HexNAc [203.0794 Da] and Hex [162.0528 Da]. The Byonic scores are listed for these peptides when disabling (left) and enabling (right) Met carbamidomethylation as a variable peptide modification in Byonic. The correct assignments as assessed by manual annotation are highlighted in bold. Only the peptide at $[M+H]^+$ 2237.896 Da could be correctly identified after enabling Met carbamidomethylation. This is due to Byonic's limited ability to accurately identify, annotate and score neutral losses arising from Met oxidation and carbamidomethylation, see also Supplementary Fig. S-10 for examples. ND, not detected.

Peptide mass $[M+H]^+$, Da	**Met carbamidomethylation disabled** [mass of modification, Da] [*N*-glycan monosaccharide composition]	**Byonic score or ND**	**Met carbamidomethylation enabled** [mass of modification, Da] [*N*-glycan monosaccharide composition]	**Byonic score or ND**
2075.844	**ALM[+15.995]N[+1095.397]GSESR [$HexNAc_3Hex_3$]**	49.9-68.5	**ALM[+15.995]N[+1095.397]GSESR [$HexNAc_3Hex_3$]**	ND
	ALM[+57.021]N[+1054.370]GSESR [$HexNAc_2Hex_4$]	ND	ALM[+57.021]N[+1054.370]GSESR [$HexNAc_2Hex_4$]	ND
2237.896	ALM[+15.995]N[+1257.449]GSESR [$HexNAc_3Hex_4$]	102.6-150.7	ALM[+15.995]N[+1257.449]GSESR [$HexNAc_3Hex_4$]	ND
	ALM[+57.021]N[+1216.423]GSESR [$HexNAc_2Hex_5$]	ND	**ALM[+57.021]N[+1216.423]GSESR [$HexNAc_2Hex_5$]**	11.0-38.8
2278.923	ALM[+15.995]N[+1298.476]GSESR [$HexNAc_4Hex_3$]	33.9-175.0	ALM[+15.995]N[+1298.476]GSESR [$HexNAc_4Hex_3$]	20.8-33.5
	ALM[+57.021]N[+1257.449]GSESR [$HexNAc_3Hex_4$]	ND	**ALM[+57.021]N[+1257.449]GSESR [$HexNAc_3Hex_4$]**	ND
2424.980	ALM[+15.995]N[+1444.534]GSESR [$HexNAc_4Hex_3Fuc_1$]	39.2-152.4	ALM[+15.995]N[+1444.534]GSESR [$HexNAc_4Hex_3Fuc_1$]	91.5-124.6
	ALM[+57.021]N[+1403.507]GSESR [$HexNAc_3Hex_4Fuc_1$]	ND	**ALM[+57.021]N[+1403.507]GSESR [$HexNAc_3Hex_4Fuc_1$]**	ND
2482.002	ALM[+15.995]N[+1501.555]GSESR [$HexNAc_5Hex_3$]	2.7-188.6	ALM[+15.995]N[+1501.555]GSESR [$HexNAc_5Hex_3$]	6.0-119.4
	ALM[+57.021]N[+1460.529]GSESR [$HexNAc_4Hex_4$]	ND	**ALM[+57.021]N[+1460.529]GSESR [$HexNAc_4Hex_4$]**	ND
2628.060	ALM[+15.995]N[+1647.613]GSESR [$HexNAc_5Hex_3Fuc_1$]	15.4-58.4	ALM[+15.995]N[+1647.613]GSESR [$HexNAc_5Hex_3Fuc_1$]	124.1
	ALM[+57.021]N[+1606.587]GSESR [$HexNAc_4Hex_4Fuc_1$]	ND	**ALM[+57.021]N[+1606.587]GSESR [$HexNAc_4Hex_4Fuc_1$]**	ND
2685.081	ALM[+15.995]N[+1704.635]GSESR [$HexNAc_6Hex_3$]	55.9-112.3	ALM[+15.995]N[+1704.635]GSESR [$HexNAc_6Hex_3$]	106.9-180.3
	ALM[+57.021]N[+1663.608]GSESR [$HexNAc_5Hex_4$]	ND	**ALM[+57.021]N[+1663.608]GSESR [$HexNAc_5Hex_4$]**	ND
2831.139	ALM[+15.995]N[+1850.693]GSESR [$HexNAc_6Hex_3Fuc_1$]	70.4	ALM[+15.995]N[+1850.693]GSESR [$HexNAc_6Hex_3Fuc_1$]	61.4-146
	ALM[+57.021]N[+1809.666]GSESR[$HexNAc_5Hex_4Fuc_1$]	ND	**ALM[+57.021]N[+1809.666]GSESR[$HexNAc_5Hex_4Fuc_1$]**	ND

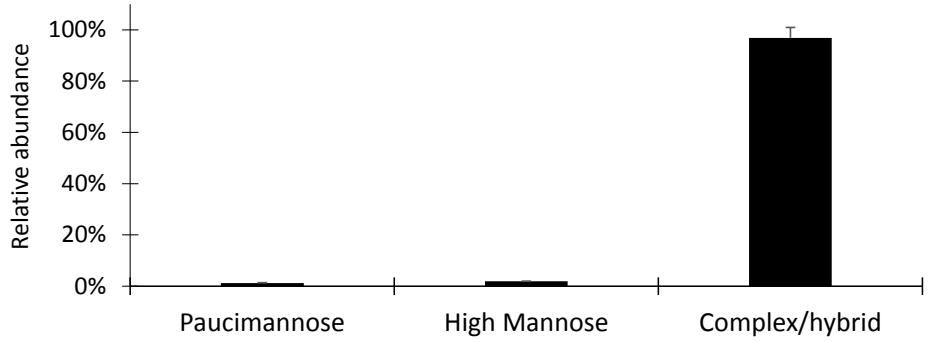
Supplementary Table S-4. The *N*-glycosylation site occupancies of Asn160 and Asn268 determined at Byonic scores threshold of 50, 100 and 150 were quantitatively comparable to the site occupancies determined in the reference profiles i.e. Asn160, 87-89% and Asn268, 20-22%.

Basigin *N*-glycosylation site	**Byonic score**		
	50	**100**	**150**
Asn160	89.5%±1.0%	89.3%±1.0%	88.9%±1.0%
Asn268	19.4% ±1.5%	16.7% ±1.4%	15.7% ±1.3%

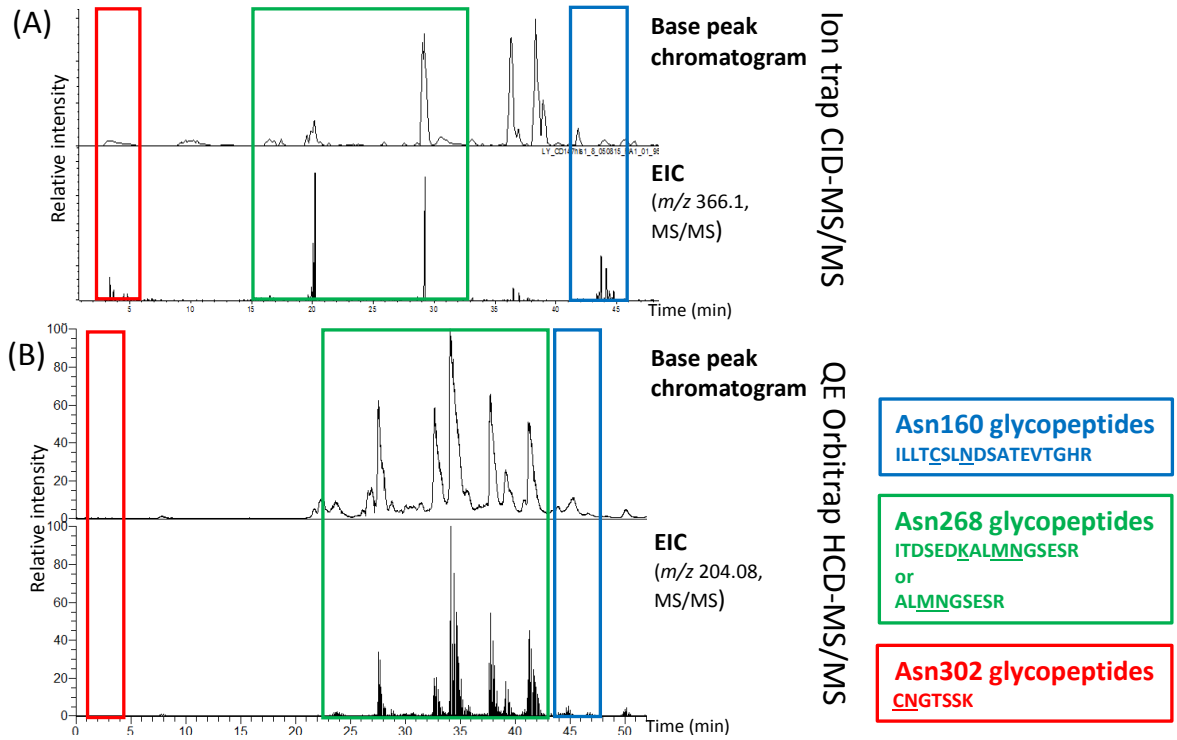
Supplementary Figure S-1

[1]MAAALFVLLGFALLGTHGASGAAGFVQAPLSQQRWVGGSVELHCEAVGSPVPEIQWWFEG
QGPNDTCSQLWDGARLDRVHIHATYHQHAASTISIDTLVEEDTGTYECRASNDPDRNHLTRA
PRVKWVRAQAVVLVLEPGTVFTTVEDLGSKILLTCSL**NDS**ATEVTGHRWLKGGVVLKEDAL
PGQKTEFKVDSDDQWGEYSCVFLPEPMGTANIQLHGPPRVKAVKSSEHINEGETAMLVCKSE
SVPPVTDWAWYKITDSEDKALM**NGS**ESRFFVSSSQGRSELHIENLNMEADPGQYRC**NGT**SSK
GSDQAIITLRVRSHLAALWPFLGIVAEVLVLVTIIFIYEKRRKPEDVLDDDDAGSAPLKSSGQH
QNDKGKNVRQRNSS[385]

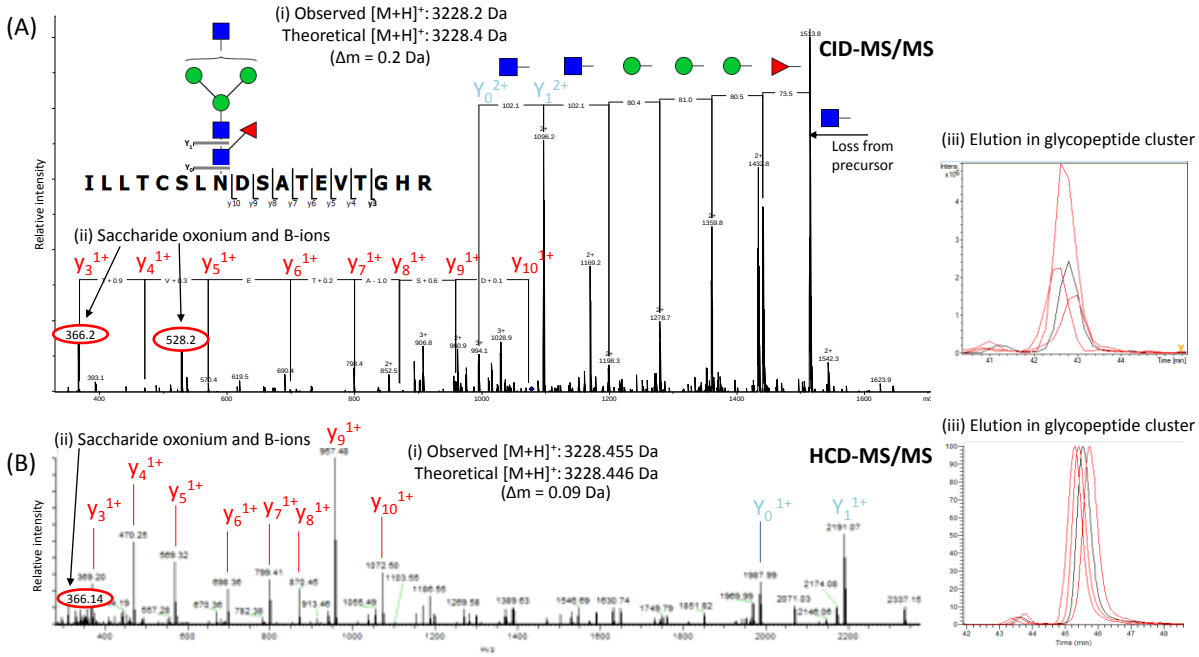
Supplementary Figure S-2



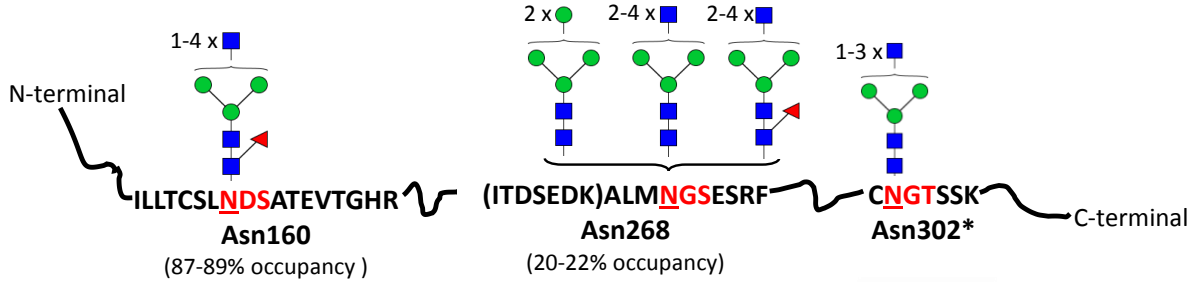
Supplementary Figure S-3



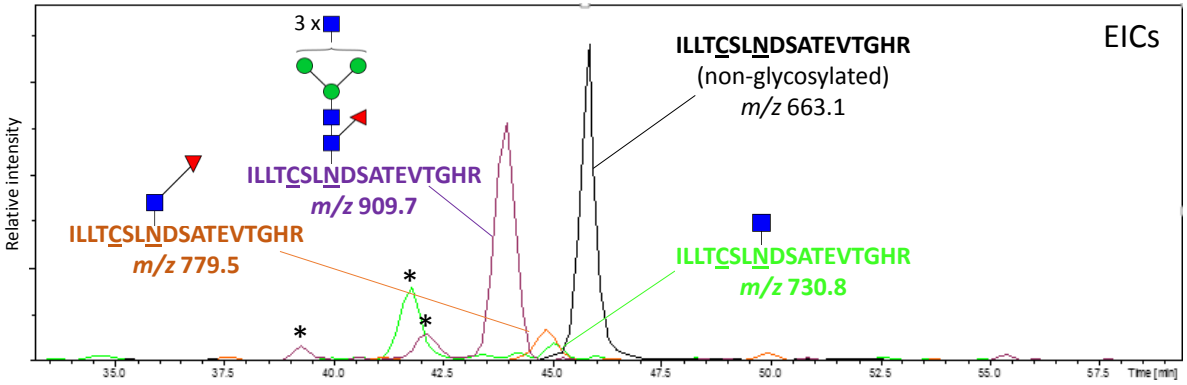
Supplementary Figure S-4



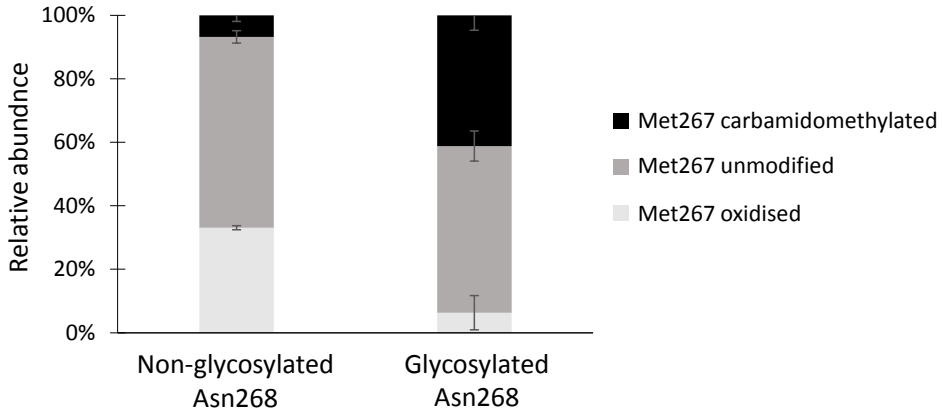
Supplementary Figure S-5



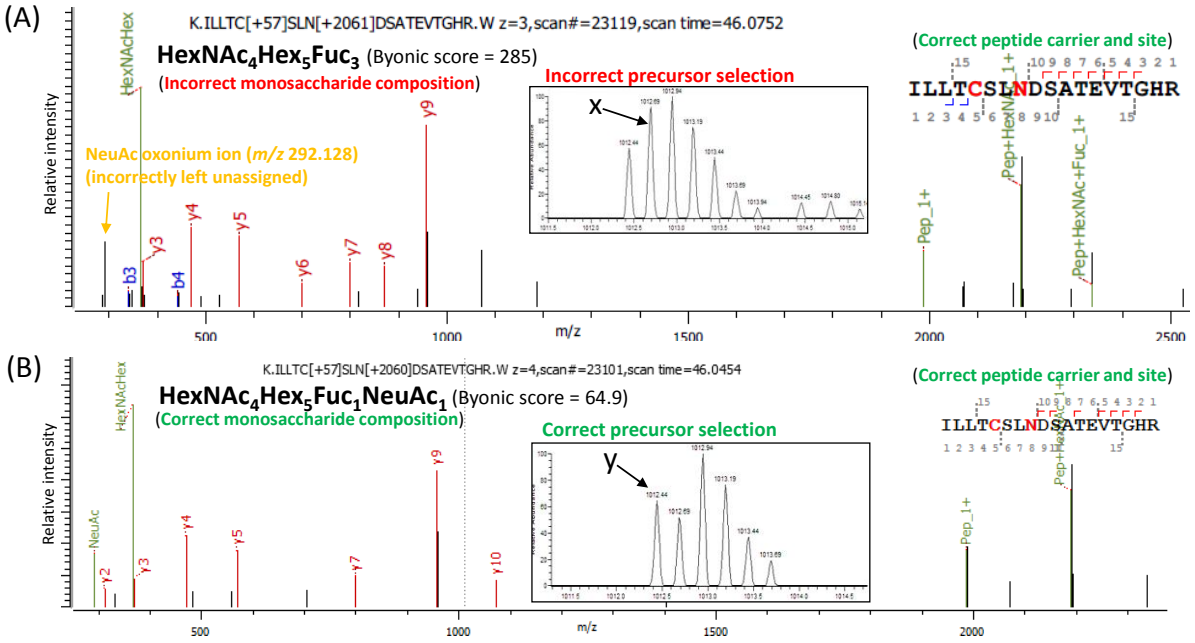
Supplementary Figure S-6



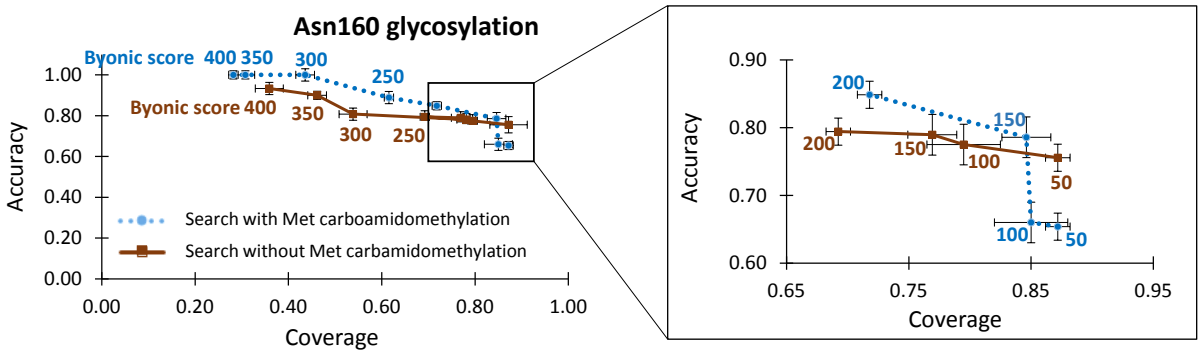
Supplementary Figure S-7



Supplementary Figure S-8



Supplementary Figure S-9



Supplementary Figure S-10