Concurrent Automated Sequencing of the Glycan and Peptide Portions of *O*-Linked Glycopeptide Anions by Ultraviolet Photodissociation Mass Spectrometry

James A. Madsen,[1] Byoung Joon Ko,[1] Hua Xu,[2] Jeremy A. Iwashkiw,[3] Scott A. Robotham,[1] Jared B. Shaw,[1] Mario F. Feldman,[3] and Jennifer S. Brodbelt*[1]

[1]Department of Chemistry and Biochemistry
The University of Texas at Austin
1 University Station A5300
Austin, TX, USA 78712

[2]Center for Proteomics and Bioinformatics
Case Western Reserve University
10900 Euclid Avenue, BRB 9th Floor
Cleveland, OH, USA 44106

[3]Alberta Glycomics Centre, Department of Biological Sciences
University of Alberta
Edmonton, Alberta T6G 2E9, Canada

SUPPLEMENTAL EXPERIMENTAL

Materials, Reagents, and Protein Preparation

OmpA/MotB was cloned with a decahistidine C-terminal tag into the plasmid pEXT20 using the forward oligonucleotide primer 1193fwd (BamHI) ATATGGATCCATGTCGATAAACCCTATAG and reverse oligonucleotide primer 1193rev10his (XbaI) TTAAtctagattaGTGGTGGTGGTGGTGGTGGTGGTGGTGTTGAGCTACTGAAAACTC . After sequencing confirmed the correct sequence, *ompA/motB* was subcloned into the plasmid pEC and transformed into *A. baumannii* ATCC 17978. Colonies positive for plasmid expression were tested via α-His Western blot (data not shown). A positive colony was inoculated into 4L of LB media, and induced in midlog growth phase with 0.2% arabinose, with an additional induction after 4 hours and incubated overnight at 37^{o}C @ 200 rpm. The overnight culture was pelleted (8000 x g, 10 mins, 4^{o}C), washed with PBS, and resuspended in 200 mL of PBS. Cells were lysed via cell disruptor (Constant Systems Ltd) at a pressure of 35 KPSI, and unbroken cells were pelleted via centrifugation (8000 x g, 10 mins, 4^{o}C). The supernatant was filtered (0.45 µm PVDF low protein binding; Millipore) and membranes were isolated by ultracentrifugation (100,000 x g, 1 hr, 4^{o}C). The pelleted membranes were resuspended in PBS+ 1% Triton X-100 with one Complete ETDA-free protease inhibitor tablet (Roche) and solubilized overnight at 4^{o}C. The sample was subsequently diluted to 0.5% Triton X-100, and soluble membrane proteins were isolated by ultracentrifugation (100,000 x g, 1 hr, 4^{o}C).

OmpA/MotB was purified from the soluble membrane protein extract via Ni^{2+}-NTA chromatography using the ATKA explorer. The membrane proteins were equilibrated with 1/9 vol 10 × loading buffer (200 mM Tris-HCl pH 8.0, 3 M NaCl, 0.5 % Triton X-100, 400mM Imidazole). The column was equilibrated with 10 column volumes of 1 × loading buffer and the

sample was loaded on a HisTrap HP column (Amersham Pharmacia Biosciences) at a flow rate of 1 mL/min. The column was washed with 10 column volumes of wash buffer (0.08 M Imidazole, 0.3 M NaCl, 0.02 M Tris-HCl pH 8.0, 0.5% Triton X-100), and eluted from the column by elution buffer (0.250 M Imidazole, 0.3 M NaCl, 0.02 M Tris-HCl pH 8.0, 0.5% Triton X-100). The purification was analyzed by Western blot and Coomassie to ensure purity of OmpA/MotB.

GluC and GluC + Trypsin protease digestions

Equal volumes of protein solution and 0.2% ProteaseMax in 1X PBS (approximately 200 µL total) were combined with dithiothreitol (added to a concentration of 5 mM). This mixture was reacted at 55º C for 30 minutes. After cooling, iodoacetamide was added to a concentration of 15 mM, followed by incubation at room temperature for 30 minutes in the dark. GluC was then added at a 1:20 enzyme:substrate ratio and reacted for 4 hours at 37º C. For the combined protease GluC plus trypsin experiments, the same procedure was followed as just described, but trypsin was included at a 1:20 enzyme:substrate ratio along with the addition of GluC.

Glycopeptide Enrichment via ZIC-HILIC SPE

ZIC-HILIC cartridges were first conditioned with 1 mL of water followed by 2 x 1 mL of 80% acetonitrile/15% water/5% formic acid. Digested glycoproteins were speed-vac dried, reconstituted in 1 mL of 80% acetonitrile/15% water/5% formic acid, and loaded into the SPE cartridge. The loaded sample was then washed with 2 x 1 mL of 80% acetonitrile/15% water/5% formic acid. Bound peptides were eluted with 2 x 250 µL of 100% water (total of 500 µL), speed-vac dried, and reconstituted with HPLC mobile phase.

Liquid Chromatography and Mass Spectrometry

Enriched glycopeptides were first injected onto a Dionex Acclaim PepMap100 C_{18} preconcentration column (Santa Clara, CA) (75 µm × 2 cm, 3 µm particle size) at a volume of 1

μL. Samples were preconcentrated for 3 minutes using 2% acetonitrile/0.05% acetic acid at a flow rate of 5 μL/min. The preconcentration column was then switched in-line with a Dionex Acclaim PepMap RSLC C_{18} analytical column (Santa Clara, CA) (75 μm × 15 cm, 2 μm particle size). Separation was performed with eluent A consisting of 0.05% acetic acid in water and eluent B consisting of 0.05% acetic acid in acetonitrile with a 120 min linear gradient from 5% to 30% eluent B at a flow rate of 300 nL/min.

Mass spectrometric analysis was performed on a Thermo Fisher Scientific LTQ Velos dual linear ion trap mass spectrometer (San Jose, CA) equipped with a Coherent (Santa Clara, CA) ExciStar XS excimer laser operated at 193 nm. The coupling of the laser to a linear ion trap mass spectrometer was described in detail previously.[65, 66] Data-dependent nanoLC-MS/UVPD was performed as follows: the first event was the survey negative mass scan (*m/z* range of 400 – 2000) followed by five UVPD events on the five most abundant ions from the first event. A series of four 2 mJ UV pulses (applied during an 8 ms activation period) was used per MS/MS scan with a *q*-value set to 0.1. Data-dependent nanoLC-MS/CID was performed as follows: the first event was the survey positive mass scan (*m/z* range of 400 – 2000) followed by ten CID events on the ten most abundant ions from the first event. A normalized collision energy of 35%, a *q*-value of 0.25, and an activation period of 10 ms was used for CID. The maximum ion injection time for all experiments was 100 ms for both survey scan and MS/MS events. A dynamic exclusion duration of 50 s was used with a list size of 500 allowed m/z values and a single repeat count. CID and UVPD spectra were the average of three and eight microscans, respectively.

MassMatrix Automated Searching of Glycopeptides

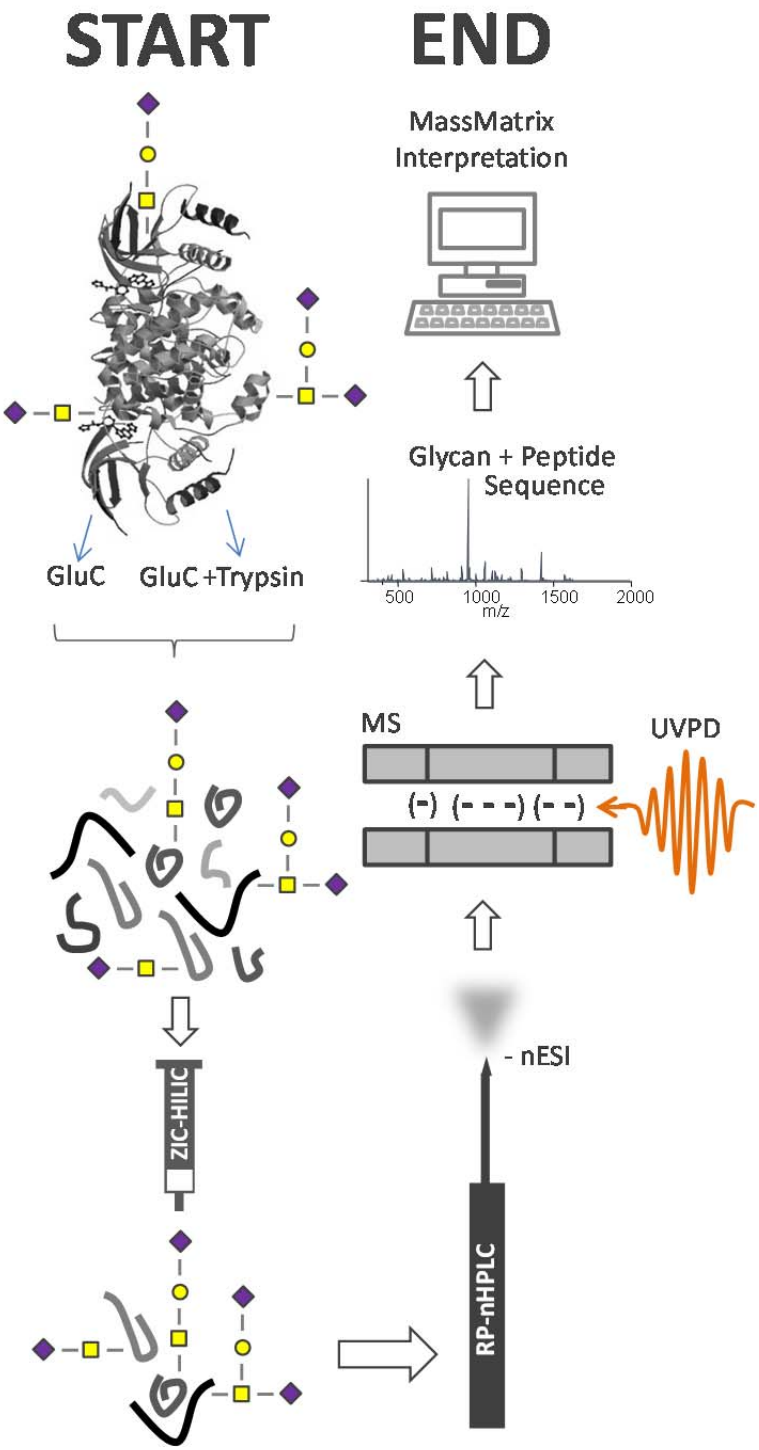
UVPD generates product ions from both the peptide and glycan portions of glycopeptide anions in a single spectrum. A modified version of the MassMatrix[61-64] database searching

algorithm was developed in this study for the simultaneous identification of both types of ions, thus allowing the complete sequencing of glycopeptides. Based on fragmentation behavior previously characterized,[57, 59] MassMatrix was first modified to search peptide-specific product ions associated with UVPD of peptide anions. The algorithm was then adapted further to search glycan-specific product ions and neutral-loss ions from the fragmentation trends observed in this study. The peptide and glycan-specific ions incorporated into MassMatrix are summarized in Supplemental Table 1 and Supplemental Table 2, respectively. The main scoring models used for MassMatrix are the pp and pp2 scores, which are a statistical measure of the number of matched product ions and the total abundance of matched product ions, respectively.[61] A set of pp and pp2 scores are generated for both the peptide and glycan products for each spectrum collected from LC-MS/UVPD runs; these scores can then be combined to create a "glycopeptide pp score" (peptide pp2 x glycan pp).

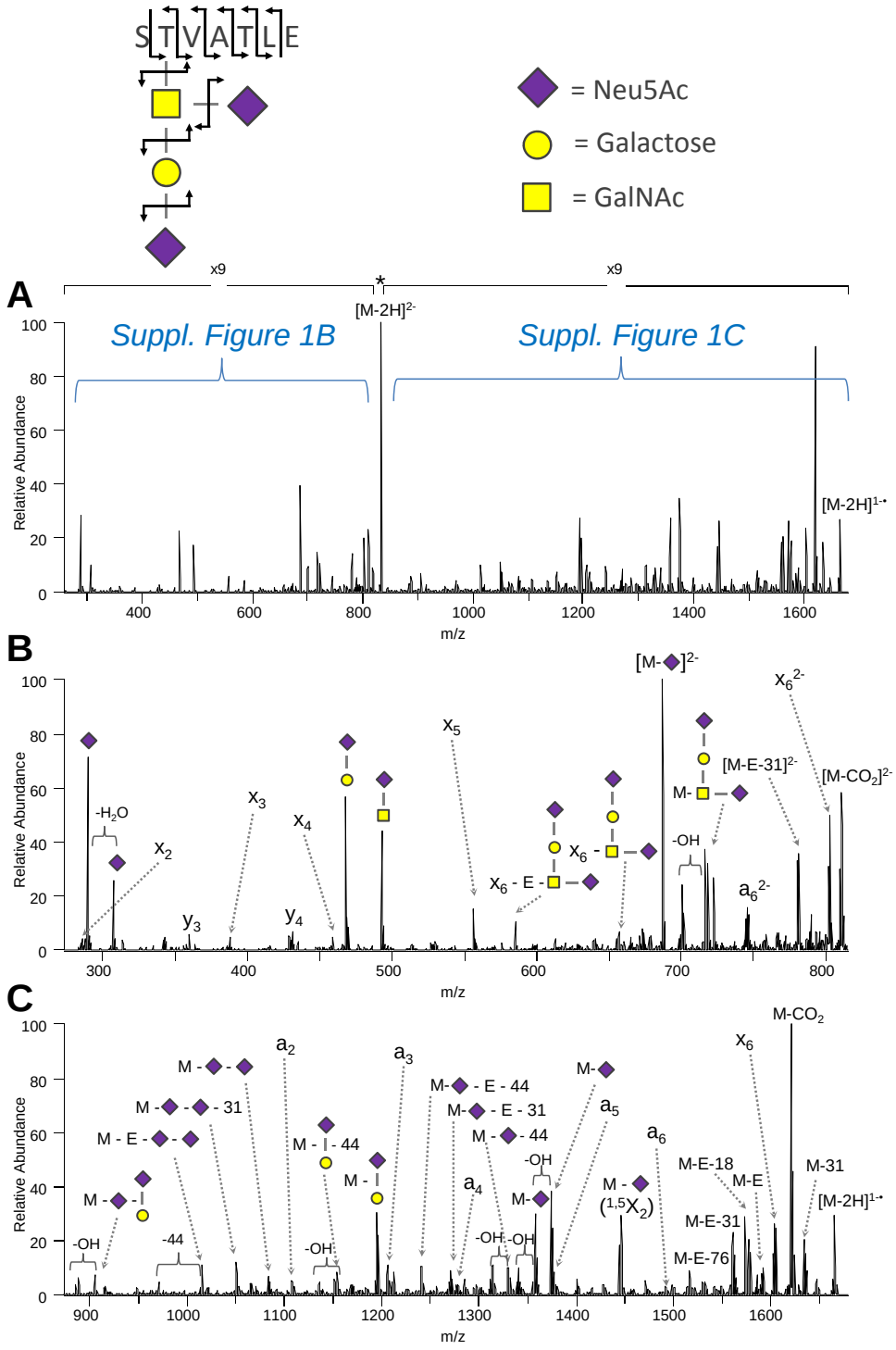
For MassMatrix searching, a precursor mass tolerance of 2.0 Da and a fragment mass tolerance of 2.0 Da were utilized. Glycans of interest were set as variable side-chain modifications to serine and threonine. Carbamidomethyl of cysteine (via reduction of disulfides and iodoacetamidation of cysteine residues) was set as a fixed modification. Experimental UVPD spectra were searched against a database consisting of the proteins of interest. Peptide hits were first filtered based on a minimum peptide pp and pp2 score of 2 and a max peptide ranking of one. The results were then further filtered by a precursor tolerance of 1.0 Da, and a glycan pp score of "0" or "NA". The final peptide list was ranked based on the glycopeptide pp. The trisaccharide structural isomer "A" (HexNAc bound to Neu5Ac) and trisaccharide structural isomer "B" (Hex bound to Neu5Ac) were differentiated by running separate MassMatrix searches to allow multiple peptide hits per spectrum, then filtering the two species based on the glycan pp score, thus allow rejection of one of the two. LC-MS/CID runs were also processed

with MassMatrix, but typical CID *b*/*y* product ions were used for peptide searching. The raw data was made publically available through Tranche - hashes wWA0fHeYM0/dK08OdkdpK9BknfyAnx5CLMfMHWm0pWiVCoxuHAU6ATBH2h7TYikEh HWPQ9V7Fpj36CcBH32RmiYwzRUAAAAAAAABlg== , MXpNB3lZeYcbYX4vWw+8bhV+jRORX/h4Oc6CJ6Z/sShLbipfRwbTkyH3LQDbS0nx+Mvp9 /Km3OHyQ6WFvLLdoQ8ICKQAAAAAAAAB8A== , 8Em6skgzokNQp09NG0cSrBSIty8kQf4E0wE695rAL+8+F+3217ndedmIIi3KkFp/NNYdzX2m x3viSjO1DiA68evjvGQAAAAAAABoA== , m7XImAeIuMkDE9YGujI1g29e687jDA9buSbVcRymypWkQ+6zCyacUQtIN7fl4MWr8amaLy /IoY2PN5Qe4laAg6J3CVgAAAAAAAB1g== , and wtxFJbde0mG3PLwbnIm7nv/PH9PM2/FM8uatnAIZ71bKZyiugdjtBBn0L9JI9JyUGqEY1ctLe1 I4NcaUp8OSUuL6RsUAAAAAAAAB4A== All raw MassMatrix output files can be found in the supplemental information.

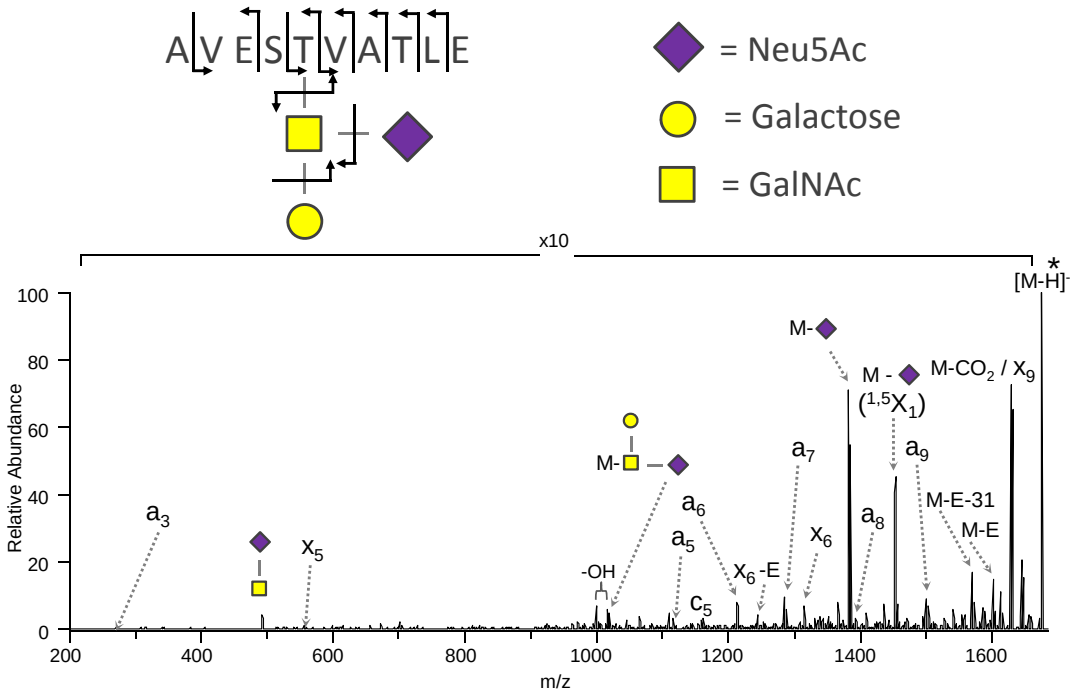
Supplemental Scheme 1. Scheme of the procedure described in this study for glycoprotein digestion, glycopeptide enrichment, LC-MS/UVPD, and automated data interpretation.



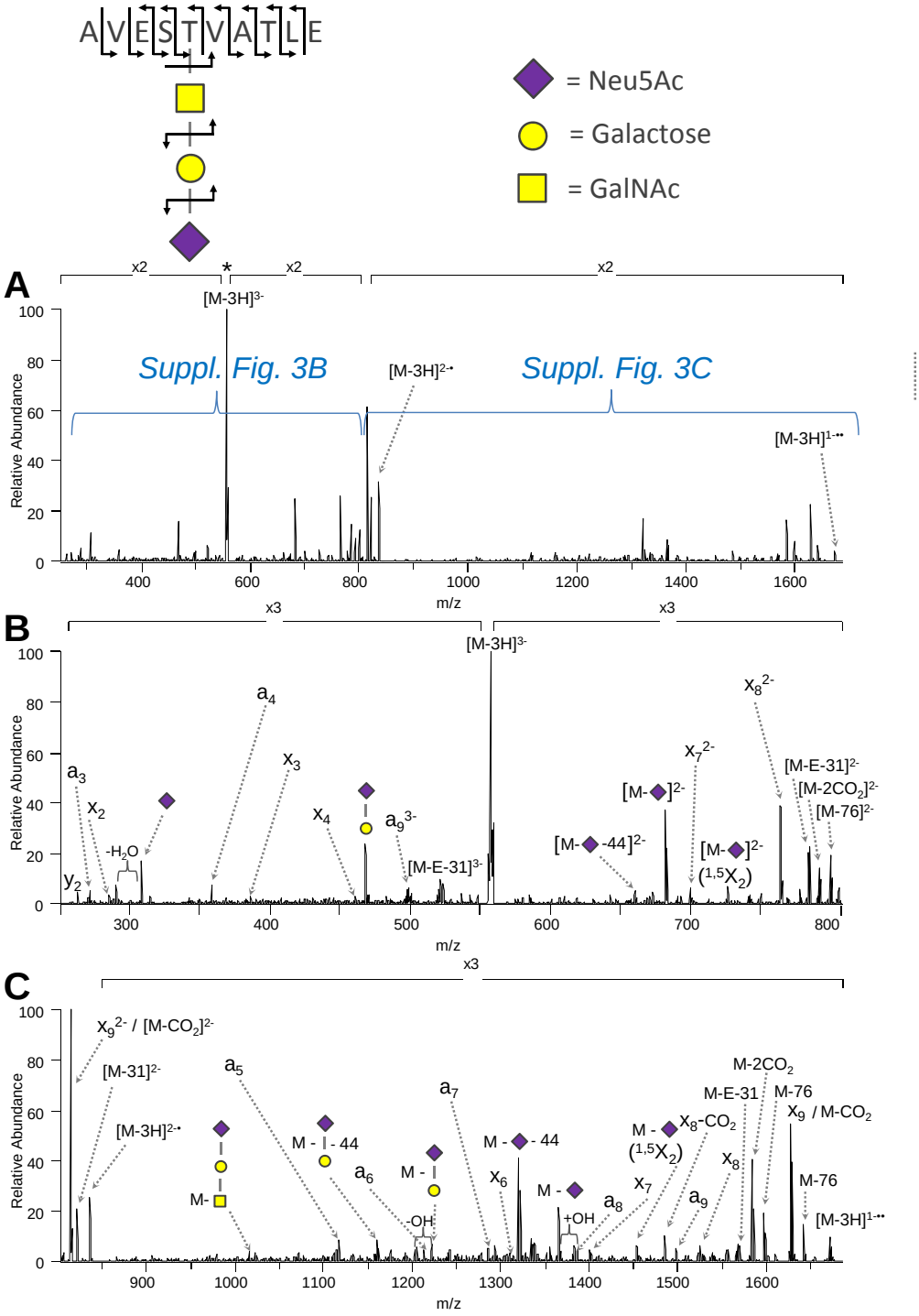
Supplemental Figure 1. UVPD of the doubly deprotonated kappa casein glycopeptide ST(glycan)VATLE (t_r = 43.4 min): (A) UVPD spectrum, (B) expansion of the low *m/z* region, and (C) expansion of the high *m/z* region.



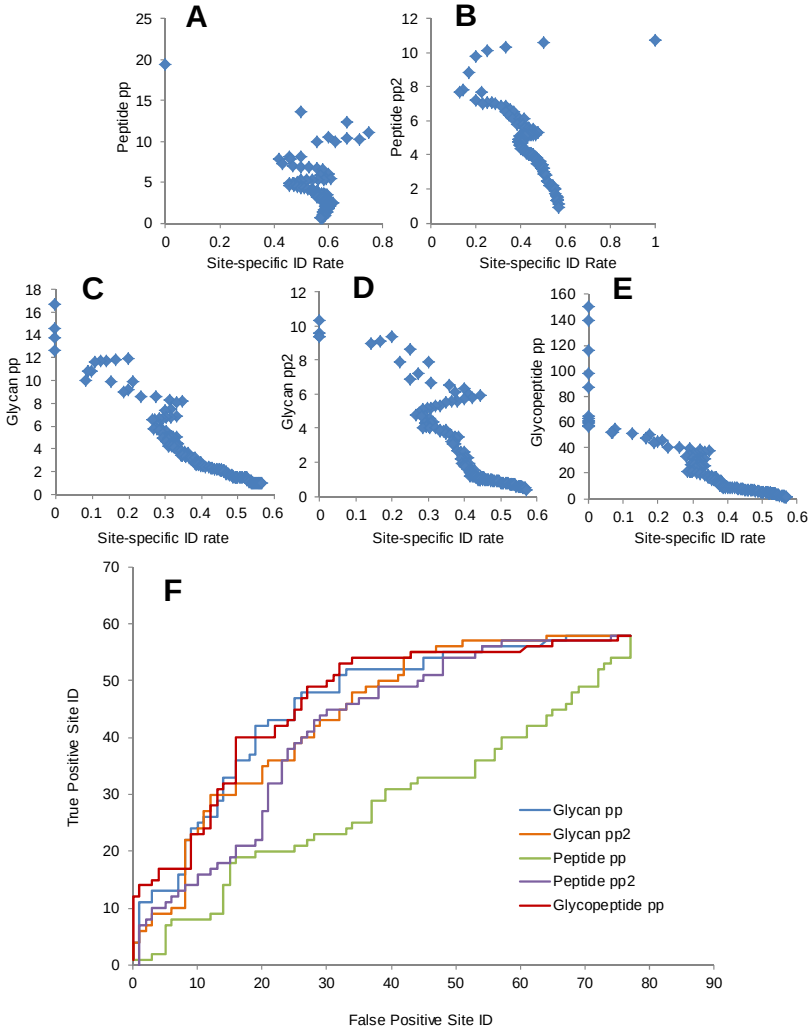
Supplemental Figure 2. UVPD of the singly deprotonated kappa casein glycopeptide AVEST(glycan)VATLE.



Supplemental Figure 3. UVPD of the triply deprotonated kappa casein glycopeptide AVEST(glycan)VATLE: (A) UVPD spectrum, (B) expansion of low *m/z* region, and (C) expansion of high *m/z* region.



Supplemental Figure 4. Site-specific identification false discovery rate (false positive site ID/(true positive site ID + false positive site ID)) for the identification of glycopeptides from bovine kappa casein (GluC digested and HILIC enriched) using various MassMatrix spectral scoring methods: (A) Peptide pp, (B) Peptide pp2, (C) Glycan pp, (D) Glycan pp2, and (E) Glycopeptide pp (Peptide pp2 x Glycan pp). (F) ROC curves for the accurate detection of the site of glycosylation for each scoring method. For panels A-E, a greater number of data points (in blue) towards the upper left of each graph indicates a better scoring system, as evidenced by panel E. For the receiver operating characteristic (ROC) curves in F, trend lines with higher numbers of true positives and lower numbers of false positives (i.e. those favoring the upper left of the graph) represent scoring systems with the most confident assignments.



Supplemental Table 1. UVPD product ions (with amino acid and charge state selectivity) that were incorporated into the MassMatrix algorithm for peptide anion searching. Sequence ions are sorted from most prevalent to least prevalent (top to bottom), and common neutral losses from intact species are listed last.

Product Ion	Amino Acid Preference	Charge State Observed
a/*x*	No preference	All
Y	N-terminal to P	All
$y - H_2O - NH_3$	N-terminal to D	All (1- is best)
$y - H_2O$	N-terminal to E	All (1- is best)
c/*z*	No preference	All
y	No preference	1-
Neutral loss of CO_2	N/A (from intact species)	All
Neutral loss of Y	N/A (from intact species)	All
Neutral loss of W	N/A (from intact species)	All

Supplemental Table 2. Variable glycan modifications and their associated product ions and neutral losses that were incorporated into the MassMatrix algorithm. HexNAc+Hex+Neu5Ac(A) and HexNAc+Hex+Neu5Ac(B) denote two different structural isomers.

	Composition	**Monoisotopic Mass Addition**	**Product Ions (B, C)**	**Neutral Losses from $[M-H]^-$ (Y, Z)**
HexNAc	$C_8H_{13}NO_5$	203.0794	NA	203, 221, 247
HexNac+Hex	$C_{14}H_{23}NO_{10}$	365.1322	NA	162, 180
HexNac+Neu5Ac	$C_{19}H_{30}N_2O_{13}$	494.1748	290, 308	291, 308
HexNAc+Hex+Neu5Ac(A)	$C_{25}H_{40}N_2O_{18}$	656.2276	290, 308, 493	291, 308, 656
HexNAc+Hex+Neu5Ac(B)	$C_{25}H_{40}N_2O_{18}$	656.2276	290, 308, 468 656	291, 308, 453, 470,
HexNAc+Hex+$Neu5Ac_2$	$C_{36}H_{57}N_3O_{26}$	947.3230	290, 308, 468, 493 582, 947	291, 308, 453, 470,
A. baumannii pentasaccharide	$C_{40}H_{62}N_4O_{27}$	1030.3601	300, 316, 679, 842.5, 1045.7	300, 318, 344, 637, 665, 827, 1030, 1048