Supplemental information

Microfluidics capillary electrophoresis-mass spectrometry for analysis of monosaccharides, oligosaccharides and glycopeptides

Kshitij Khatri[1], Joshua A. Klein[2], John R. Haserick[1], Deborah R. Leon[1], Catherine E. Costello[1], Mark E. McComb[1], Joseph Zaia[1,2,*]

[1]Dept. of Biochemistry, Center for Biomedical Mass Spectrometry

[2]Bioinformatics Program

Boston University

*Email: jzaia@bu.edu

The following information is included:

Methods:

Glycopeptide sample preparation and MS data acquisition

The glycoprotein standards, bovine ribonuclease B (RNase B), human transferrin (transferrin) and human alpha-1-acid glycoprotein (AGP), were purchased from Sigma-Aldrich (St. Louis, MO). Approximately 250µg of each glycoprotein was subjected to reduction and alkylation, using dithiothreitol and iodoacetamide, followed by proteolysis using trypsin (Promega Corp., Madison, WI), as described previously [1]. Tryptic digests were desalted using C18 spin columns (Thermo Pierce) and then dried down using a centrifugal evaporator. Dry peptide/glycopeptide samples were resuspended in 50µl of the background electrolyte (BGE) containing 100 mM ammonium acetate, to improve conductivity and sample stacking via isotachophoresis. MS1 spectra were acquired over the range *m/z* 665 - 2000*,* with 1 microscan per spectrum at a resolution of 70,000. A high starting *m/z* was used to avoid interference from abundant siloxane peaks that were coming from the CE-ESI source. Siloxane peaks corresponding to *m/z* 667.1764 and 741.1952 were consistently present in the spectra and were used as lock masses for internal mass calibration. The automatic-gain-control (AGC) target set at 1e6 and a maximum ion injection time of 110 msec. Data-dependent tandem MS were acquired on the top 10 precursors with charge states ranging between 3 and 6 using (higher energy collisional induced dissociation) collisional-dissociation at a normalized collision energy (NCE) of 27 eV. Tandem mass spectra were acquired with 2 microscans per spectrum at a resolution of 35,000 and an AGC target set at 5e5. Peaks for migrating glycopeptides were between 15-30 seconds wide. A minimum intensity threshold of 1.4 e5 was used for precursor selection and following a tandem MS precursors were dynamically excluded for 1 sec from being selected again. Source capillary temperature and RF level were set at 320^{o}C and 55, respectively.

N-glycan sample preparation and data acquisition

N-linked glycans were released from 100µg of standard glycoprotein tryptic digests using PNGase-F (New England Biolabs, Ipswich, MA). The released glycans were separated from the peptides using a C18 spin column, as the unbound fraction. Glycan samples were dried and then subjected to reducing-end labeling using Aminoxy-TMT^{131} (Thermo Scientific) reagent as per the manufacturer's protocol. Labeled glycans were desalted using PD Minitrap G10 columns (GE Healthcare, Marlborough, MA). Desalted glycan samples were dried *in vacuo* and resuspended in 20µL of background electrolyte containing 100mM ammonium acetate.

Glycan MS1 spectra were acquired over the range *m/z* 600 – 2000*,* with 1 microscan per spectrum at a resolution of 70,000, with the AGC target set at 1e6 and a maximum ion injection time of 110 msec. Siloxane peaks with *m/z* 610.18416, 667.1764 and 741.1952 were used as lock-masses. Data-dependent tandem MS were acquired on the top 5 precursors with charge states of 2 or 3, using collisional-dissociation at a stepped NCE of 25 and 30 eV. While glycan compositions were assigned based on MS1 only, a low minimum intensity threshold of 5.5 e3 was used for precursor selection in data dependent tandem MS to confirm that the migrating peaks were glycans and following a tandem MS precursors were dynamically excluded for 1 sec from being selected again. Tandem mass spectra were acquired with 2 microscans per spectrum at a resolution of 35,000 and an AGC target set at 2e4 to enable tandem MS of low abundance glycoforms. Source capillary temperature and RF level were set at 320^{o}C and 55, respectively.

Monosaccharide sample preparation

Preparation of TMT-labeled Monosaccharide Standards

Monosaccharide standard pentoses (D-xylose, D-ribose, D-arabinose), deoxyhexoses (L-fucose, L-rhamnose), hexoses (D-glucose, D-galactose, D-mannose), *N*-Acetyl-hexoses (*N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, *N*-acetyl-D-mannosamine), and a sialic acid, *N*-acetylneuraminic acid, were purchased from Sigma Aldrich. Aqueous solutions of each monosaccharide were made at a concentration of 20 mg/mL in ddH_2O. One microliter of each was added to three separate tubes, in groups, to be labeled with the

aminoxyTMT^6-127, aminoxyTMT^6-128, and aminoxyTMT^6-130 tags. The monosaccharides were grouped in the following manner: aminoxyTMT^6-127 (Xyl, Fuc, Gal, GlcNAc), aminoxyTMT^6-128 (Rib, Rha, Man, GalNAc), and aminoxyTMT^6-130 (Ara, Glc, ManNAc, Neu5Ac). Prior to labeling, the three mixes were dried under vacuum and then reactions were performed according to the manufacturer's protocol, excluding quenching with acetone. Briefly, the entire contents of a packet of aminoxy-TMT reagent (0.8 mg) was re-suspended in 200 µL of 95% methanol/ 0.1% acetic acid, and added to one of the tubes containing dry monosaccharide samples. The tubes were then vortexed vigorously, and allowed to sit for 10 minutes at ambient temperature. The solvents were removed with a centrifugal evaporator. Then, 200 µL of 95% methanol was added; the samples were vortexed and allowed to sit for 10 min. at room temperature, and dried using the centrifugal evaporator. After drying, the three mixes were stored at -20°C until analyzed.

Just prior to analysis, the three standard monosaccharide mixes were reconstituted with 50:50 methanol:water, and each solution was mixed with an equal volume of 50:50 methanol:water containing 200 mM ammonium acetate to yield a final concentration of 100 mM ammonium acetate for CE-MS analysis.

AGP monosaccharide preparation

A 1 mg quantity of AGP was treated with PNGase-F overnight at 37 ^{o}C. The released *N*-glycans were separated from the protein using C-18 Sep-Pak cartridges (Waters Corporation, Milford, MA), as described previously [2,3]. The released, purified *N*-glycans were placed into a conical vial and subjected to acid hydrolysis using 2 M trifluoroacetic acid at 110 ^{o}C, for 30 min.

The resulting monosaccharide mixtures were dried down in a centrifugal evaporator and labeled with aminoxyTMT^6-127 as described above. The labeled sample was dried and re-suspended for analysis in 50 µL of 50:50 methanol:water. A 1:200 dilution was made in 50:50 MeOH/H_2O containing 100 mM ammonium acetate and this solution was used for CE-MS analysis.

CE-MS of Monosaccharides

MS1 alternating with all-ions fragmentation (AIF) was performed (MS/AIF). For MS, the chosen parameters were resolution 17,500, AGC target 2e5, maximum injection Time 40 ms, scan range *m/z* 440 - 640 For AIF, the chosen parameters were NCE 35 eV, resolution 17,500, AGC target 2e5, maximum injection time 40 ms, scan range *m/z* 440-640. Lock masses used were, *m/z* 445.12003, 519.13888 and 610.18416. Source capillary temperature and RF level were set at 320^{o}C and 55, respectively.

CE-MS data analysis:

Data analysis was performed using in-house software tools, as described below.

Database Construction

The database of theoretical glycan compositions was generated using a combinatorial algorithm considering every combination of Hex between 3 and 9, HexNAc between 2 and 8, Fuc between 0 and 4, and NeuAc between 0 and 5. The number of NeuAc residues was constrained to be ≤ (# HexNAc – 1) and the number of fucose were constrained to be ≤ (# HexNAc) residues. Each composition was considered with and without a tandem mass tag (Mass Shift Formula: C10 C[13]5 H28 N4 O2). For each composition containing NeuAc, each NeuAc was considered in three forms, underivatized, dimethylaminated, or lactonized.

The glycopeptide database was constructed by combining this same list of glycan compositions without derivatization or tandem mass tags with a list of common tryptic peptide sequences around sequons of interest.

Migration Profile Reconstruction

Each MS1 scan for each sample was processed by a deisotoping and charge state deconvolution algorithm using an *N*-glycan-specific averagine[4] formula C7 H11.83 N0.5 O5.17, using the BRAIN [5] algorithm to generate isotopic patterns for the interpolated experimental composition. Extrapolated neutral mass monoisotopic peaks were aggregated into migration profiles by mass using a 10 parts-per-million error

tolerance to collect peaks across scans. Profiles which contained gaps greater 250 milliseconds wide were split into separate migration profiles, corresponding to separate events. Migration profiles were expected to conform to a bi-Gaussian shape after Gaussian smoothing ($\sigma = 1$). Un-gapped multimodal profiles were partitioned at valleys and evaluated independently. The bi-Gaussian fit for each smoothed profile was computed using non-linear least squares and the residual-sum-of-squares (RSS) for this fit was recorded, called the "shape fit RSS". Each migration profile was also compared against a null model which assumed that the distribution of signal over time was random, using a horizontal line at the average intensity across the migration profile to calculate the "null model RSS". Profiles for which $1 - \frac{\text{shape fit RSS}}{\text{null model RSS}} > 0.4$ were considered well resolved.

For glycomics data, reconstructed electropherograms were searched against a database of theoretical glycan compositions by neutral mass, using a 10 parts-per-million error tolerance to assign putative compositions. Under the conditions used, ammonium adducts were expected, and a second round of searches was performed using a mass shift corresponding to one or two ammonium adducts. The results which matched overlapping migration profiles of different adduct types were summed. These summed profiles were also evaluated following the above shape rules. Glycopeptide electropherograms were reconstructed in the same fashion; however, adducts were not considered.

Migration Profile Visualization

All migration profile figures were produced by applying the same Gaussian smoothing described above to the matched raw profile data.

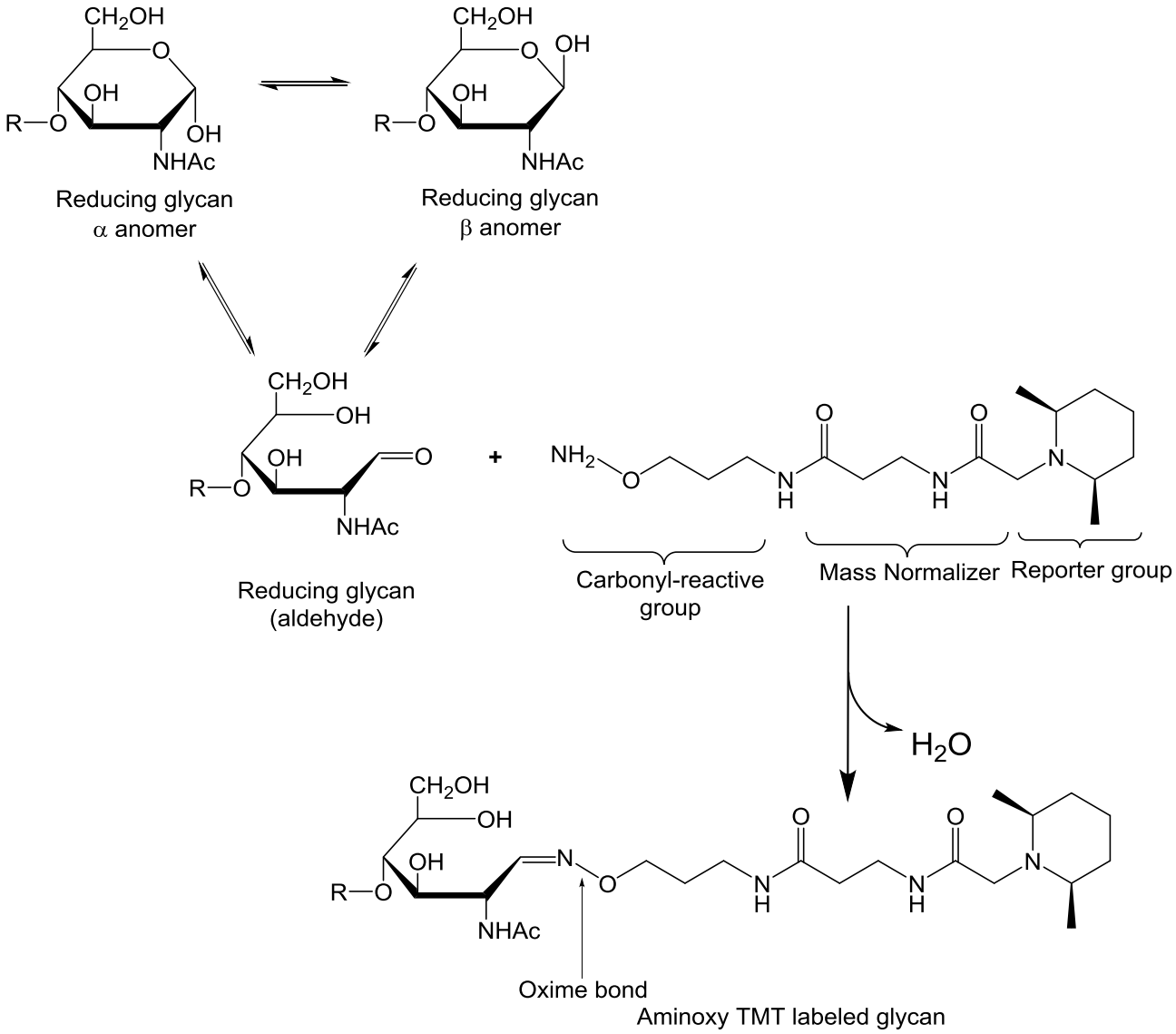


Figure S-1 Glycan derivatization using aminoxy TMT

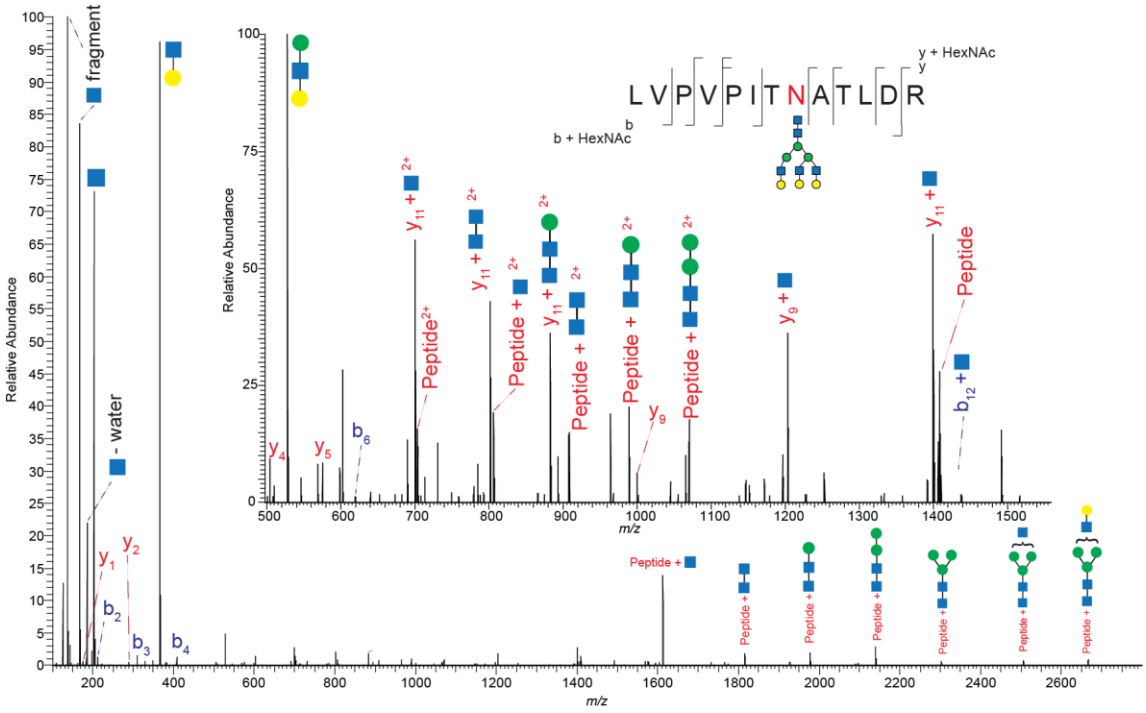


Figure S-2: Tandem MS of an asialo AGP glycopeptide, identified in Figure 5 in main text

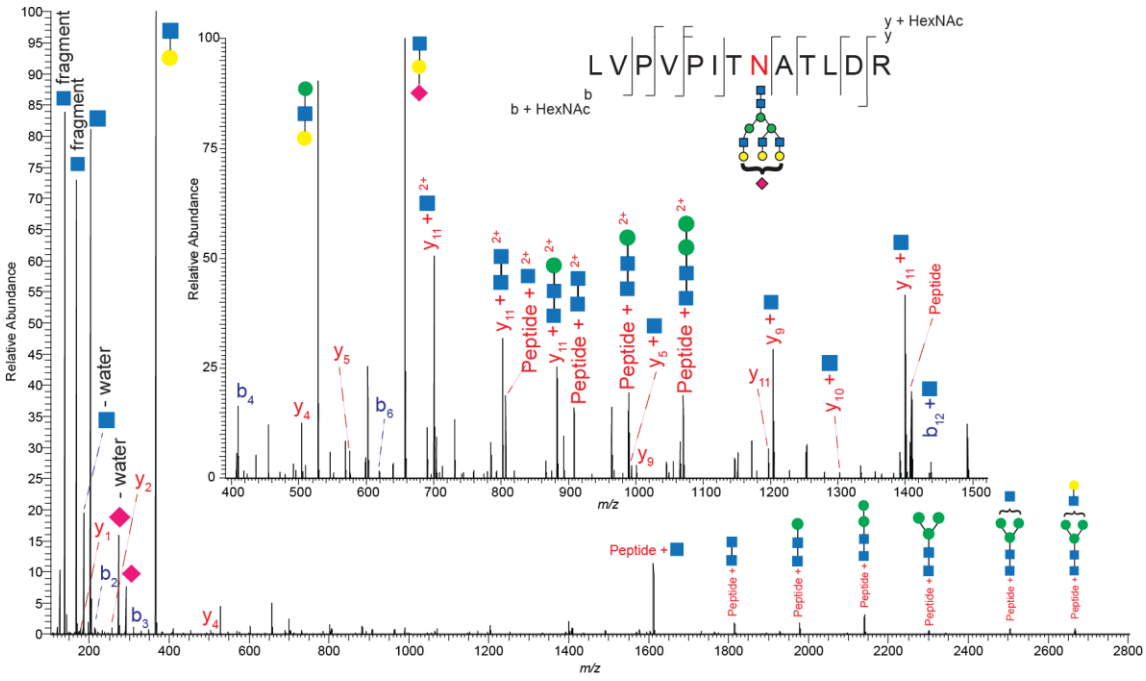


Figure S-3: Tandem MS of a sialylated AGP glycopeptide, identified in Figure 5 in main text

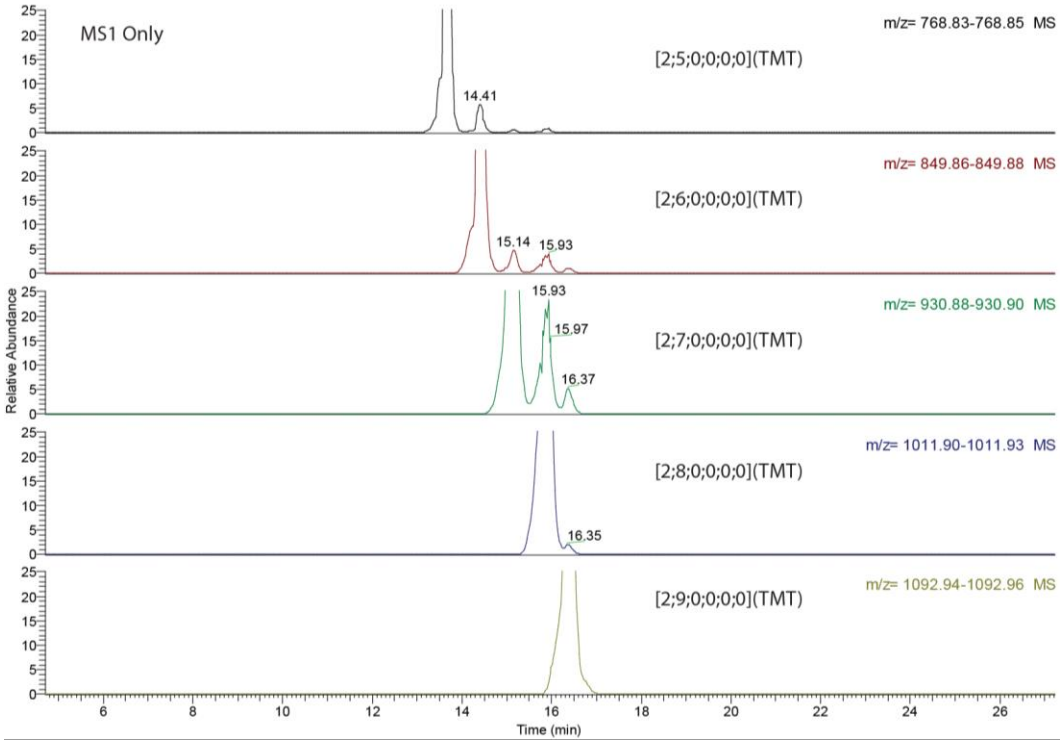


Figure S-4: EIEs from CE-MS of high-mannose type *N*-glycans, showing presence of in-source decay products. Y-axis has been zoomed-n to make visualization easier.

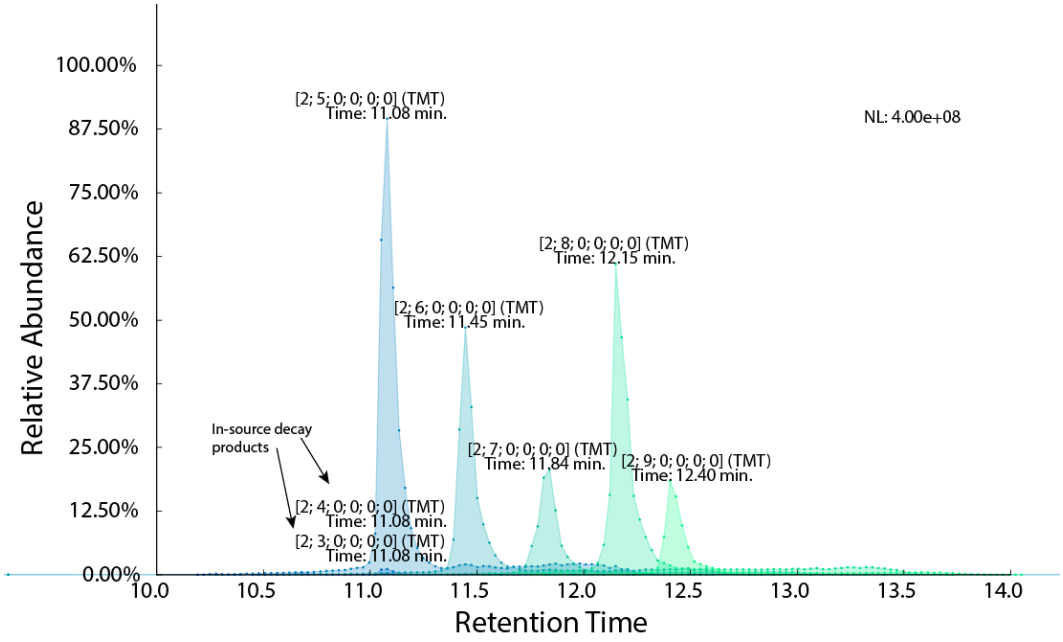


Figure S-5: Overlaid EIEs showing migration of high-mannose type *N*-glycans shown in Figure 2 in main text. Glycan compositions are represented as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; Lactonized NeuAc].

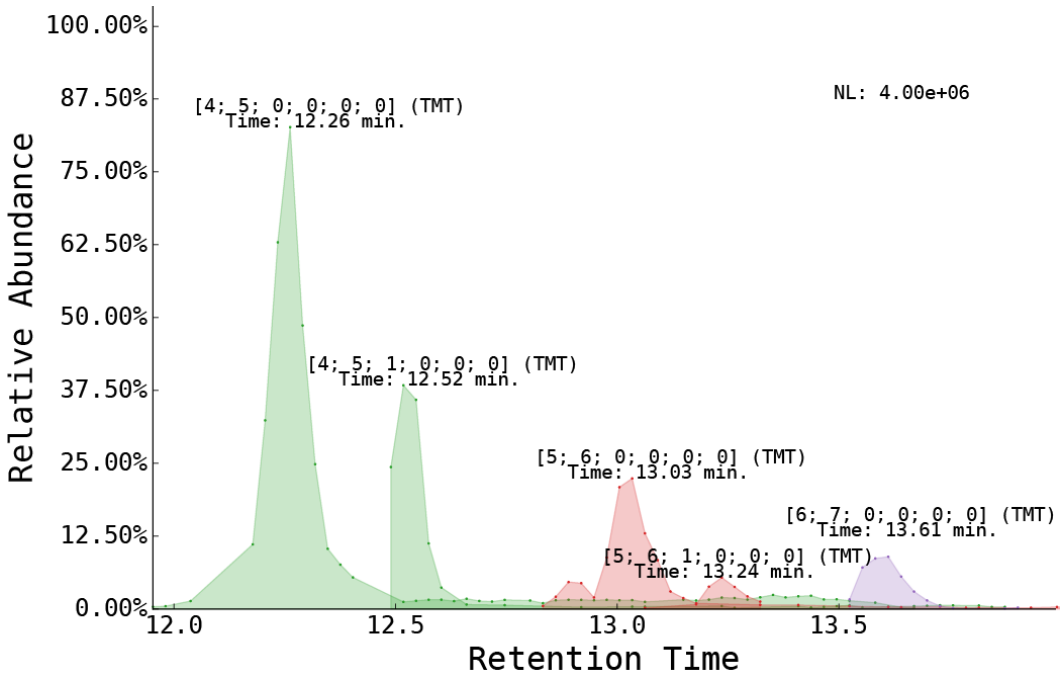


Figure S-6: Overlaid EIEs showing migration of asialo complex type *N*-glycans shown in Figure 2 in main text. Glycan compositions are represented as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; Lactonized NeuAc].

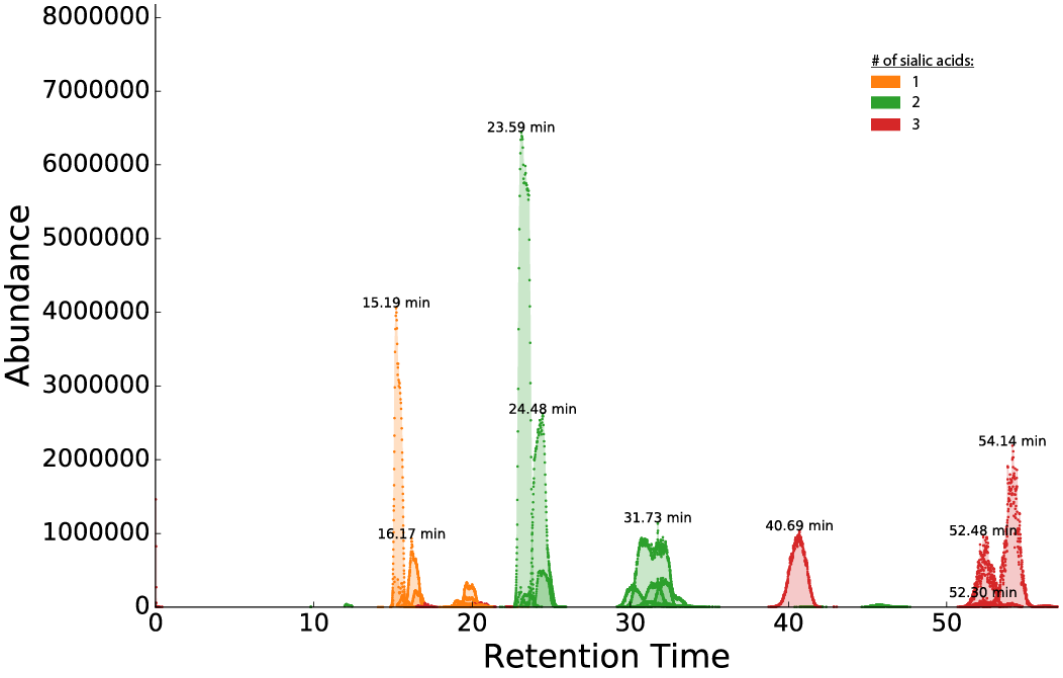


Figure S-7: Overlaid EIEs showing migration of sialylated complex type *N*-glycans shown in Figure 3 in main text. Compositions have been segregated using different colors based on number of sialic acids.

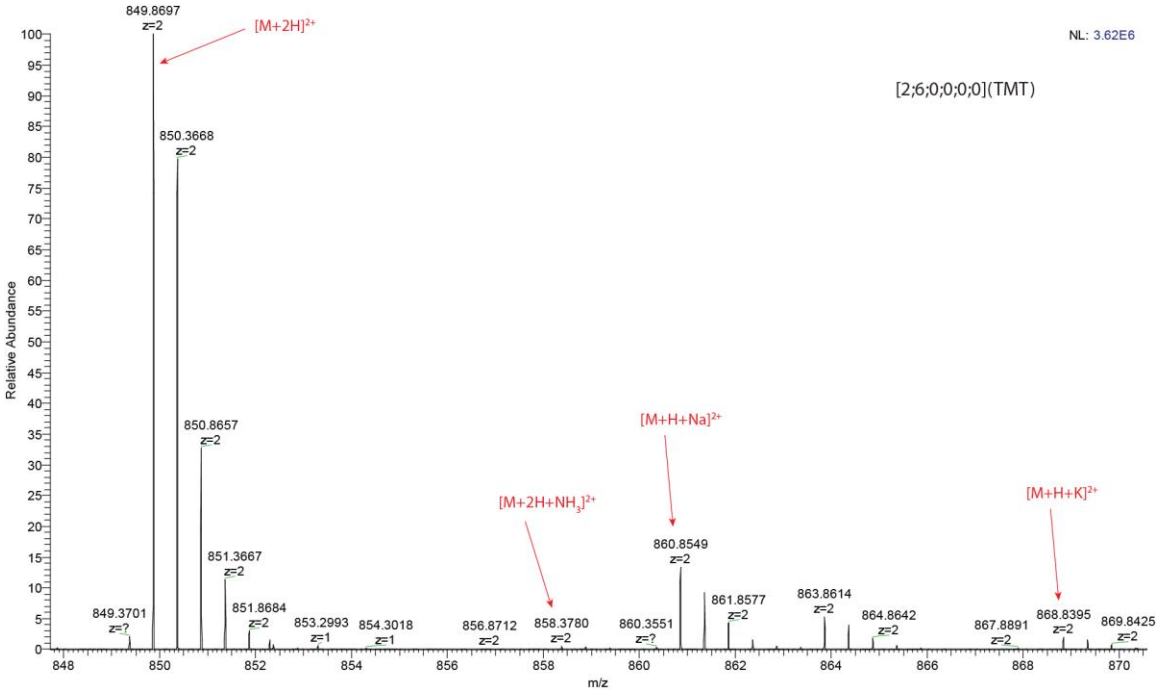


Figure S-8: Example extracted mass spectrum for [2;6;0;0;0;0](TMT), showing presence of adducted ions.

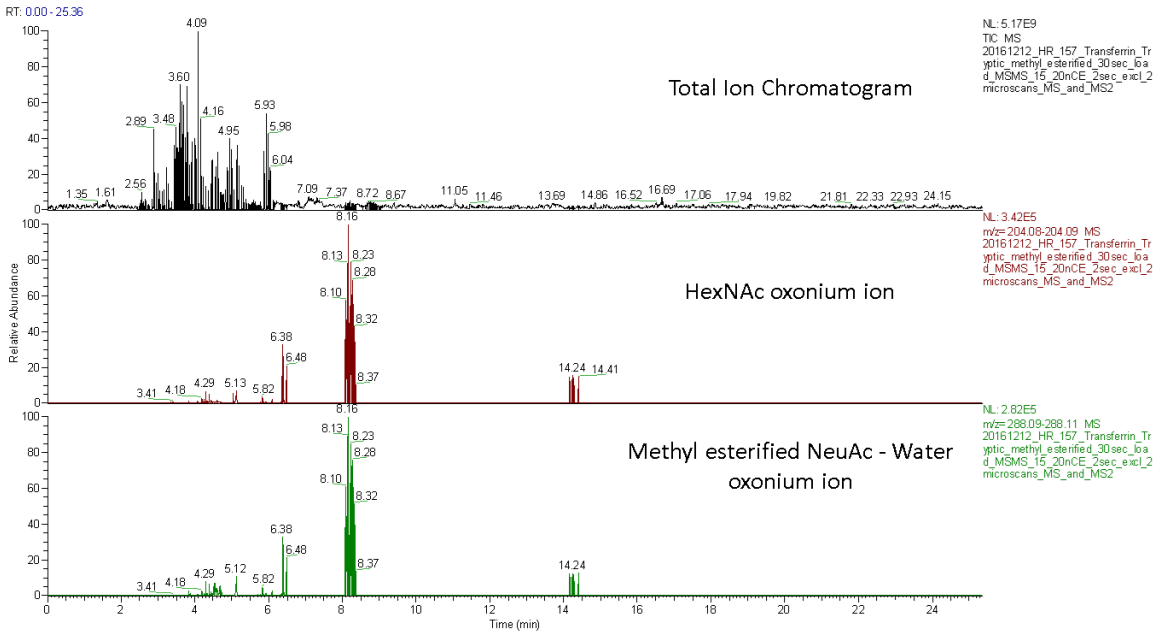


Figure S-9: Overall migration profile of methyl esterified glycopeptides from Human Transferrin tryptic digest

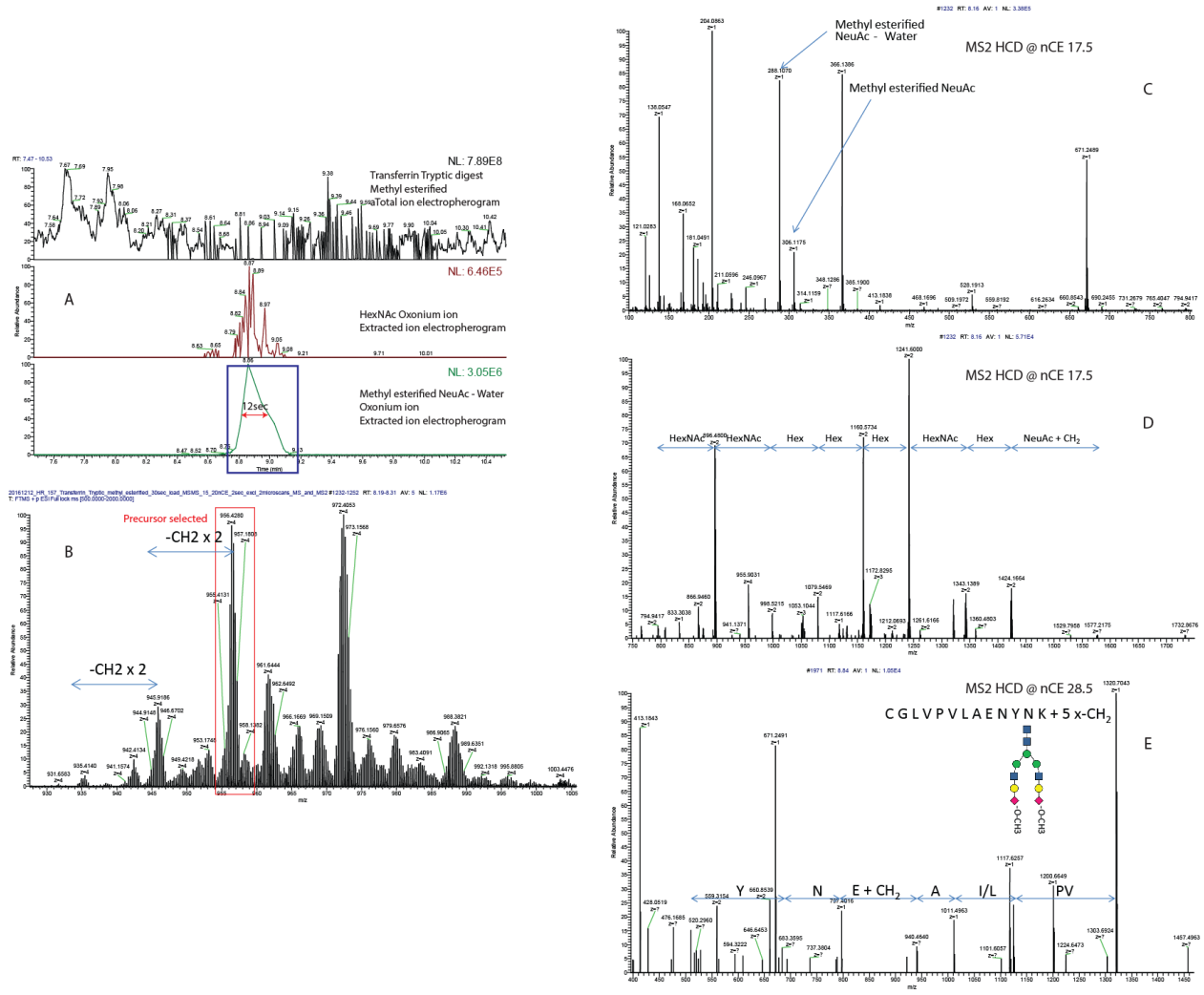


Figure S-10: CE-MS of tryptic transferrin glycopeptides after methyl esterification. A. Total-ion electropherogram overlaid with oxonium ion traces for HexNAc and methy-esterified NeuAc, showing peak shape for migrating glycopeptides; B. MS1 corresponding to the migrating peak in blue box, showing the complexity from numerous methyl esterification events; C and D. Tandem MS for the precursor shown in panel B at nCE 17.5; E. Tandem MS for the same precursor at higher nCE (28.5), from a different run.

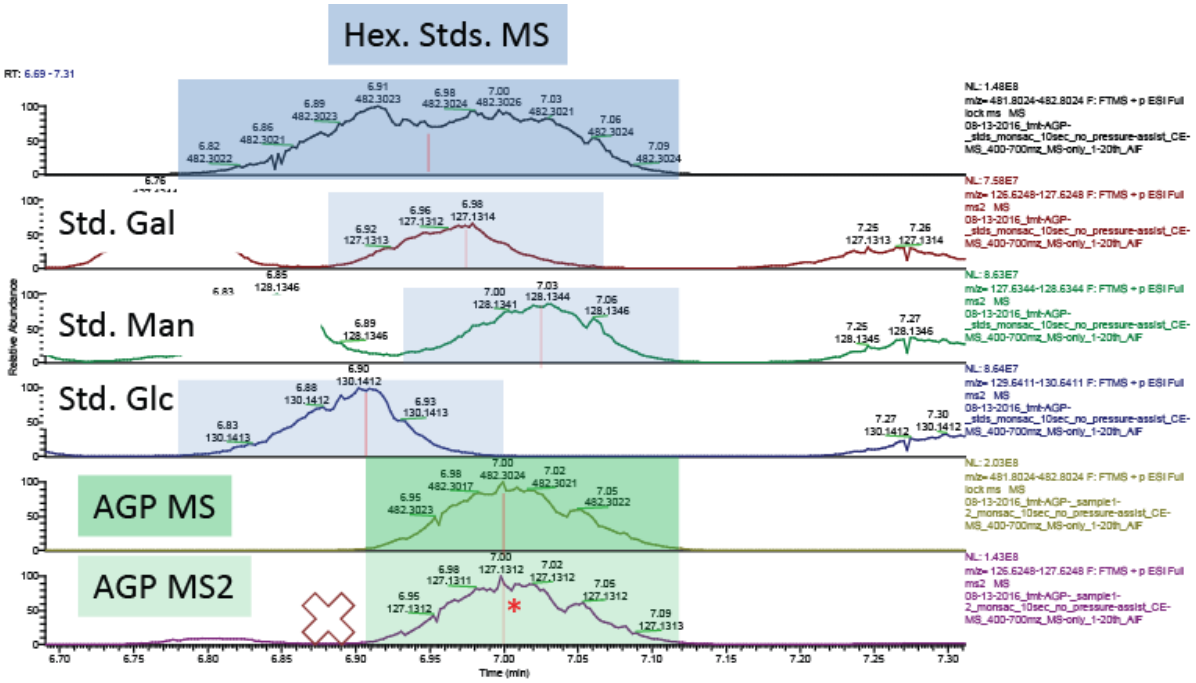
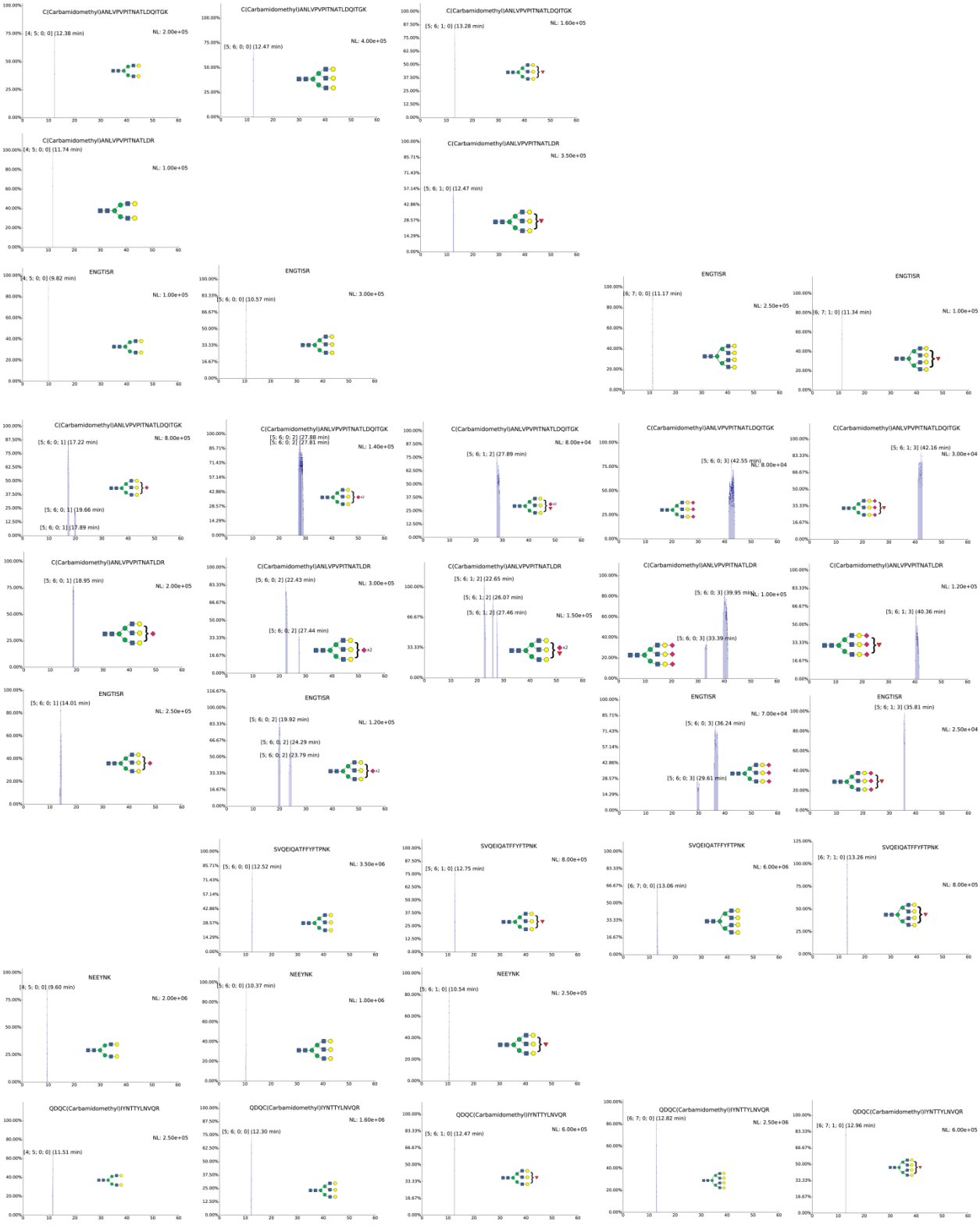


Figure S-11: Overlaid traces for hexose standards and monosaccharides derived from AGP showing that AGP contains a mixture of mannose and galactose but no glucose.


C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[4; 5; 0; 0] (12.38 min)
NL: 2.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 0; 0] (12.47 min)
NL: 4.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 1; 0] (13.28 min)
NL: 1.60e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[4; 5; 0; 0] (11.74 min)
NL: 1.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 1; 0] (12.47 min)
NL: 3.50e+05
ENGTISR
[4; 5; 0; 0] (9.82 min)
NL: 1.00e+05
ENGTISR
[5; 6; 0; 0] (10.57 min)
NL: 3.00e+05
ENGTISR
[6; 7; 0; 0] (11.17 min)
NL: 2.50e+05
ENGTISR
[6; 7; 1; 0] (11.34 min)
NL: 1.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 0; 1] (17.22 min)
[5; 6; 0; 1] (19.66 min)
[5; 6; 0; 1] (17.89 min)
NL: 8.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 0; 2] (27.88 min)
[5; 6; 0; 2] (27.81 min)
NL: 1.40e+05
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 1; 2] (27.89 min)
NL: 8.00e+04
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 0; 3] (42.55 min)
NL: 8.00e+04
C(Carbamidomethyl)ANLVPVPITNATLDQITGK
[5; 6; 1; 3] (42.16 min)
NL: 3.00e+04
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 0; 1] (18.95 min)
NL: 2.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 0; 2] (22.43 min)
[5; 6; 0; 2] (27.44 min)
NL: 3.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 1; 2] (22.65 min)
[5; 6; 1; 2] (26.07 min)
[5; 6; 1; 2] (27.46 min)
NL: 1.50e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 0; 3] (39.95 min)
[5; 6; 0; 3] (33.39 min)
NL: 1.00e+05
C(Carbamidomethyl)ANLVPVPITNATLDR
[5; 6; 1; 3] (40.36 min)
NL: 1.20e+05
ENGTISR
[5; 6; 0; 1] (14.01 min)
NL: 2.50e+05
ENGTISR
[5; 6; 0; 2] (19.92 min)
[5; 6; 0; 2] (24.29 min)
[5; 6; 0; 2] (23.79 min)
NL: 1.20e+05
ENGTISR
[5; 6; 0; 3] (36.24 min)
[5; 6; 0; 3] (29.61 min)
NL: 7.00e+04
ENGTISR
[5; 6; 1; 3] (35.81 min)
NL: 2.50e+04
SVQEIQATFFYFTPNK
[5; 6; 0; 0] (12.52 min)
NL: 3.50e+06
SVQEIQATFFYFTPNK
[5; 6; 1; 0] (12.75 min)
NL: 8.00e+05
SVQEIQATFFYFTPNK
[6; 7; 0; 0] (13.06 min)
NL: 6.00e+06
SVQEIQATFFYFTPNK
[6; 7; 1; 0] (13.26 min)
NL: 8.00e+05
NEEYNK
[4; 5; 0; 0] (9.60 min)
NL: 2.00e+06
NEEYNK
[5; 6; 0; 0] (10.37 min)
NL: 1.00e+06
NEEYNK
[5; 6; 1; 0] (10.54 min)
NL: 2.50e+05
QDQC(Carbamidomethyl)IYNTTYLNVQR
[4; 5; 0; 0] (11.51 min)
NL: 2.50e+05
QDQC(Carbamidomethyl)IYNTTYLNVQR
[5; 6; 0; 0] (12.30 min)
NL: 1.60e+06
QDQC(Carbamidomethyl)IYNTTYLNVQR
[5; 6; 1; 0] (12.47 min)
NL: 6.00e+05
QDQC(Carbamidomethyl)IYNTTYLNVQR
[6; 7; 0; 0] (12.82 min)
NL: 2.50e+06
QDQC(Carbamidomethyl)IYNTTYLNVQR
[6; 7; 1; 0] (12.96 min)
NL: 6.00e+05

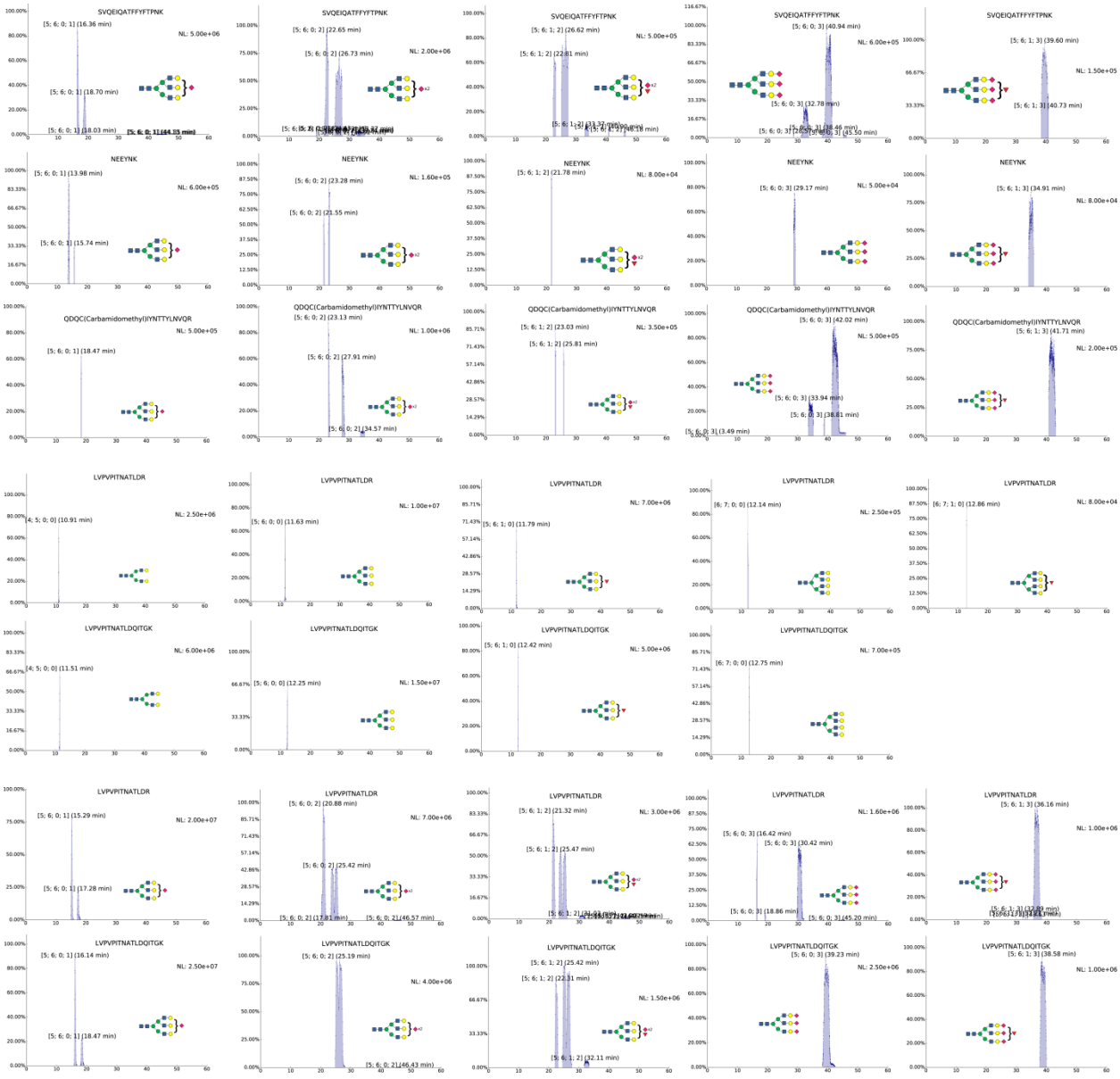


Figure S-12: Extracted ion electropherograms for individual glycopeptides plotted in Figures 6 and 7 in the main text. Glycan compositions are represented as [HexNAc; Hex; dHex; NeuAc].

References cited:

(1) Khatri, K.; Staples, G. O.; Leymarie, N.; Leon, D. R.; Turiák, L.; Huang, Y.; Yip, S.; Hu, H.; Heckendorf, C. F.; Zaia, J. *J. Proteome Res.* **2014**, *13* (10), 4347–4355.

(2) Haserick, J. R.; Leon, D. R.; Samuelson, J.; Costello, C. E. *Mol. Cell. Proteomics MCP* **2017**.

(3) Khatri, K.; Klein, J. A.; White, M. R.; Grant, O. C.; Leymarie, N.; Woods, R. J.; Hartshorn, K. L.; Zaia, J. *Mol. Cell. Proteomics* **2016**, mcp.M116.058016.

(4) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (4), 229–233.

(5) Dittwald, P.; Claesen, J.; Burzykowski, T.; Valkenborg, D.; Gambin, A. *Anal Chem* **2013**, *85*, 1991–1994.