Supporting information for:

Simply extending the EThcD MS/MS range increases the confidence in N-glycopeptide identification.

Tomislav Čaval[§†#], Jing Zhu[§†#], and Albert J.R. Heck[§†*].

[§]*Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science4Life, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands;*

[†]*Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands;*

Corresponding Author

*E-mail: A.J.R.Heck@uu.nl Tel.: +31- 302536797.

This supporting information contains:

Materials and methods.
Figure S1. Annotated EThcD spectra of various N-glycopeptides.
Figure S2. Annotated EThcD fragmentation spectrum of doubly glycosylated glycopeptides.
Figure S3. Comparison of oxonium ion intensities between standard *vs*. extended *m/z* range.
Figure S4. Comparison of the distribution of LogP scores using HCD with a standard *vs.* an extended *m/z* range.
Figure S5. Examples of extended m/z range HCD spectra.

Materials and Methods

Sample preparation

The CHO cell sample (a kind gift from Henrik Clausen, University of Copenhagen) preparation was identical as previously reported[1]. In short, the cell pellet was resuspended in lysis buffer containing 100 mM Tris-HCl (pH 8.5) (Sigma-Aldrich), 7 M urea (Sigma-Aldrich), 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich), 30 mM chloroacetamide (CAA, Sigma-Aldrich), Triton X-100 (1%) (Sigma-Aldrich), 2 mM magnesium sulfate (Sigma-Aldrich), Benzonase (1%) (Merck Millipore, Darmstadt, Germany), phosphoSTOP (Roche) and complete mini EDTA free (Roche, Woerden, the Netherlands). Next, the sample was disrupted by sonication (45 min; 20 s on, 40 s off) with a Bioruptor Plus (Diagenode, Seraing, Belgium). Following sonication the sample was centrifugated at 14,000 rpm for 1h and the supernatant was collected. Impurities were further removed by methanol/chloroform protein precipitation: 1 mL of supernatant was mixed with 4 mL of methanol (Sigma-Aldrich), 1 mL chloroform (Sigma-Aldrich), and 3 mL ultrapure water. This mixture was then centrifuged at 5000 rpm for 10 min; the upper layer was discarded and 3 mL of methanol was added. Another round of sonication and centrifugation was performed after which the solvent was removed. Air dried precipitate was resuspended in the digestion buffer: 100 mM Tris-HCl (pH 8.5), 0.5% Rapigest (Waters, Etten-Leur, the Netherlands), 5 mM TCEP and 30 mM CAA. The proteases trypsin (Sigma-Aldrich) and Lys-C (Wako, Neuss, Germany) were added to a 1:50 and 1:100 ratio (w/w), respectively. Protein digestion was performed overnight at 37 °C. Next day the Rapigest was removed using acid precipitation (0.5% TFA) and the peptides were desalted by using a Sep-Pak C18 vac cartridge. For Fe^{3+}-IMAC 2 mg of tryptic peptides were resuspended in buffer A (30% ACN and 0.07% TFA) whereby the pH was adjusted to 2.3. Samples were injected onto

the Fe^{3+-}IMAC column (Propac IMAC-10 4 × 50 mm column, ThermoFisher Scientific, Landsmeer, the Netherlands). Loading was performed at a flow rate of 0.1 mL/min for 7 min with 0% solvent B (0.3% NH4OH), next non-phospho and non-M6P containing peptides were washed off at a flow rate of 1mL/min for 5 min with 0% B whereby the flow through was collected. Mannose-6-phosphate glycopeptides (M6P) and phosphopeptides were eluted with 50% B 1mL/min for 1.5 min, followed by 50% at 0.5 mL/min for 2.5 min, and finally held at 0% B 1mL/min for 9 min. Collected peptides were then lyophilized to dryness. The Fe^{3+}-IMAC column flow-through was loaded onto a Hypersep SAX cartridge (ThermoFisher Scientific) to enrich for glycopeptides missed by the initial Fe^{3+}-IMAC enrichment. The SAX cartridge was preconditioned with 3x1 mL of ACN, 100 mM triethylammonium acetate in water, 1% TFA, and 95%ACN with 1% TFA. The sample was loaded in 95% with 1% TFA and washed with 6x1mL of 95% ACN with 1% TFA. The glycopeptides were eluted with 1 mL 50% ACN with 0.1% TFA and 0.5 mL of 5% ACN with 0.1% TFA and dried in a Speedvac. The human milk sample was collected, with written consent, from one donor in week 6 post-delivery. After centrifugation at 1000 g for 30min to remove fat and ultracentrifugation at 10000g for 60 min, the upper milk serum was separated. Up to a final concentration of 1% w/v sodium deoxycholate (SDC), 100mM Tris, 5mM TCEP and 30 mM CAA were added to the milk serum. Trypsin and Lys-C were added to a 1:50 and 1:100 ratio (w/w), respectively. Digestion was performed overnight at 37 °C. Next day SDC was removed with acid precipitation (0.5% TFA) and the peptides were desalted by Oasis PRiME HLB plate (Waters). For hydrophilic interaction liquid chromatography (HILIC) enrichment, 150μg of tryptic peptides were resuspended in 200μL 80% ACN/0.5% TFA and applied on GlykoPrep cartridge (ProZyme, Ballerup, Denmark). After washing with 50 μL 80% ACN/0.5%TFA, the glycopeptides were stepped-eluted by 50 μL 75% ACN/0.5%TFA, 50 μL 70% ACN/0.5%TFA, 50 μL 65% ACN/0.5%TFA and 50 μL 0.5%TFA and all elution were combined and dried in a Speedvac.

Mass Spectrometry

Peptides were resuspended in 0.1% TFA and loaded onto an Agilent 1290 (Agilent Technologies, Middelburg, Netherlands) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific, Bremen, Germany). Peptides were separated on a 100 μm

inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 μm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% FA in water, and mobile-phase solvent B consisted of 0.1% FA in ACN. Trapping was performed at a flow rate of 5 μl/min for 5 min with 0% B and peptides were eluted using a passively split flow of 300 nl/min for 85 min with 8% to 40% B over 75 min, 40% to 100% B over 3 min, 100% B for 1 min, 100% to 0% B over 1 min, and finally held at 0% B for 10 min. Peptides were ionized using 2.0 kV spray voltage and a capillary temperature of 320 °C. The mass spectrometer was set to acquire full-scan MS spectra (375–2000 m/z) for a maximum injection time of 50 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 4e5. The dynamic exclusion was set to 20s at exclusion window of 10 ppm with a cycle time of 3s. Charge-state screening was enabled, and precursors with +2 to +6 charge states and intensities >1e5 were selected for tandem mass spectrometry (MS/MS). HCD MS/MS (120–2000 m/z) or (120-4000) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.6 Th) and an AGC target value of 5e4 or a maximum injection time of 75 ms with a normalized collision energy of 30%. When a signature MS/MS oxonium ion of glycans was observed, EThcD MS/MS on the same precursor was triggered (isolation window of 1.6 Th) and fragment ions (120–2000 m/z) or (120-4000) were analyzed in the Orbitrap mass analyzer at a resolution of 30,000, AGC target value of 2e5 or a maximum injection time of 200 ms with activation of ETD and supplemental activation with a normalized collision energy of 27%. In both cases the instrument was operated in the “normal” mass setting.

Data Analysis

Raw files of LS MS runs of peptides extracted from the CHO cell lysate were processed with Byonic (ver 2.15.10) (Protein Metrics Inc.) using the following search parameters: trypsin digestion with a maximum of 2 missed cleavages, unreviewed CHO database + human alpha-galactosidase (Uniprot, 34 962 entries, March 2018). To enable comparison of HCD and EThcD fragmentation files were searched with fragmentation type HCD or EThcD separately with 10 ppm tolerance for the precursor mass and 20 ppm for product mass tolerance.

Carbamidomethylation of cysteines as a fixed modification; variable modifications: methionine oxidation, phosphorylation on serine, threonine and tyrosine residues. For glycan analysis we used the Byonic database of 182 glycans, whereby we added manually several M6P glycan compositions as previously described[1]. Raw files of human milk glycopeptide runs were processed using the same parameters as above, except a focused human milk protein database (1259 entries), and a Byonic database of 309 known human glycans. Result files from Byonic search were further processed to remove nonmodified peptides and reversed hits. Mass spectrometry data files have been deposited to MassIVE with the identifier: MSV000083710. Glycan databases used in this work can be found in File S1.

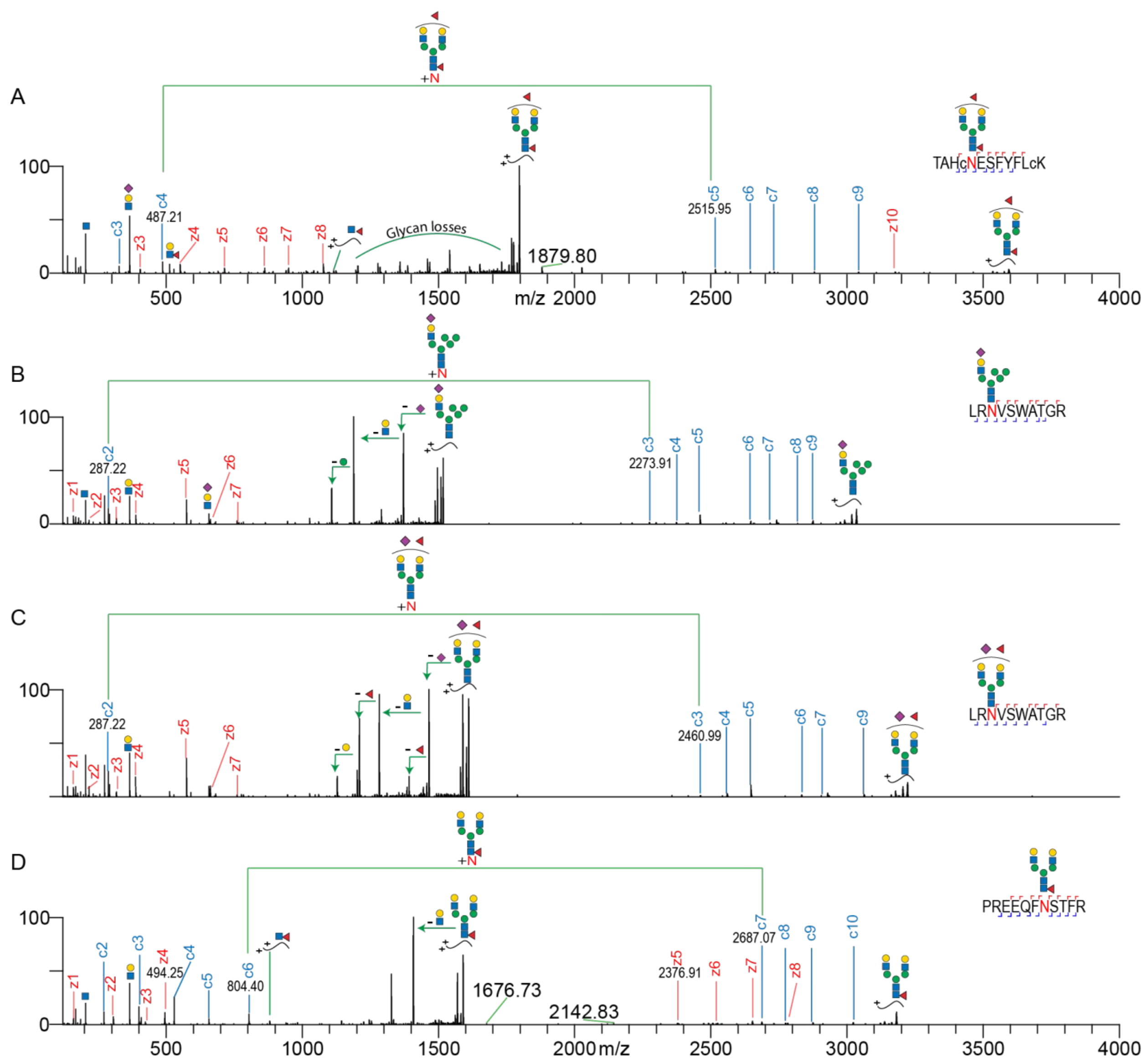


Figure S1. Annotated EThcD spectra of various N-glycopeptides. EThcD spectrum of a glycopeptide originating from Macrophage mannose receptor 1 (A), CD14 (B and C), and Ig gamma 2 chain C region (D). The c and z ion series are annotated in blue and red colors, respectively. Oxonium ions are annotated with their corresponding glycan structures. Green lines connect c ions fragment pairs containing asparagine + intact glycan mass increment. Lower case c in the peptide sequence indicates a carbamidomethylated Cysteine.

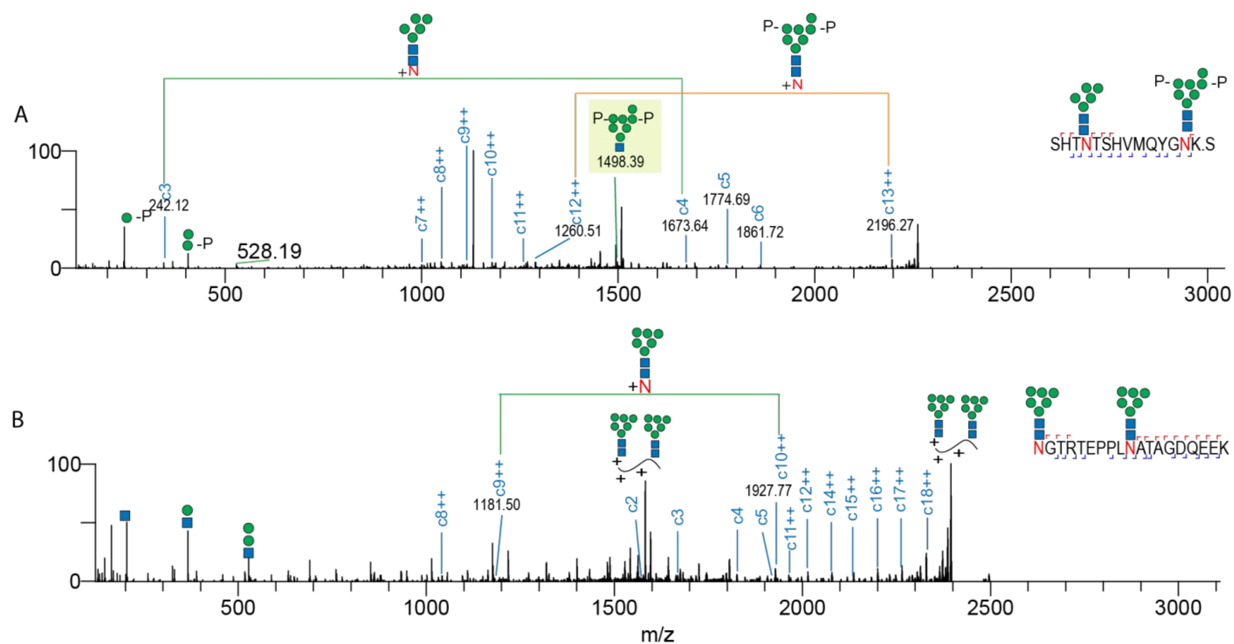
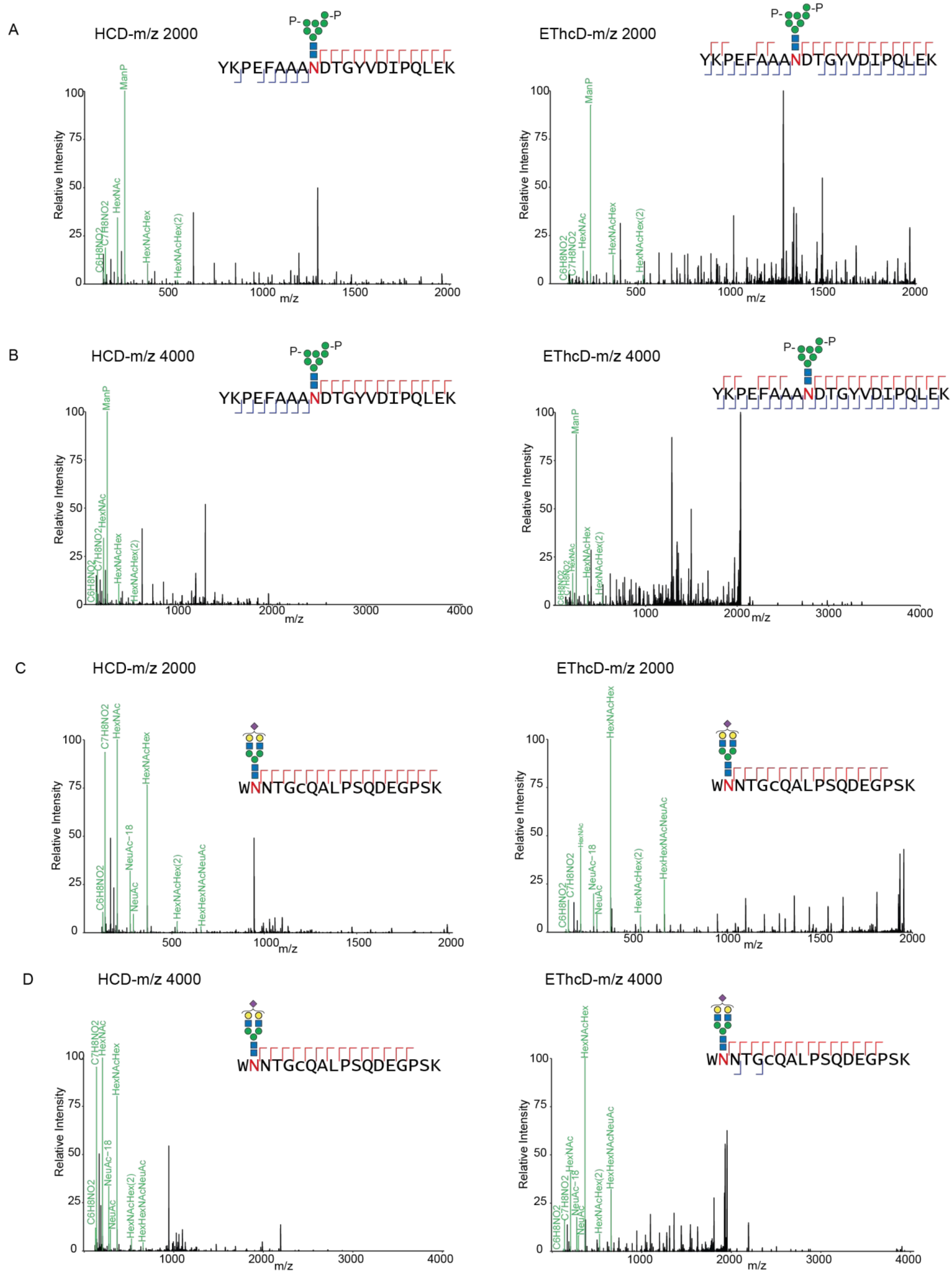
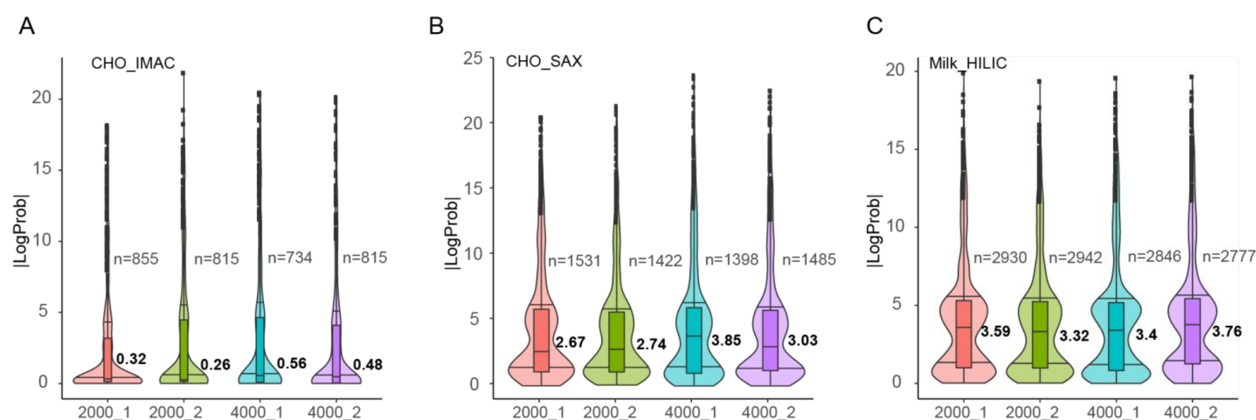


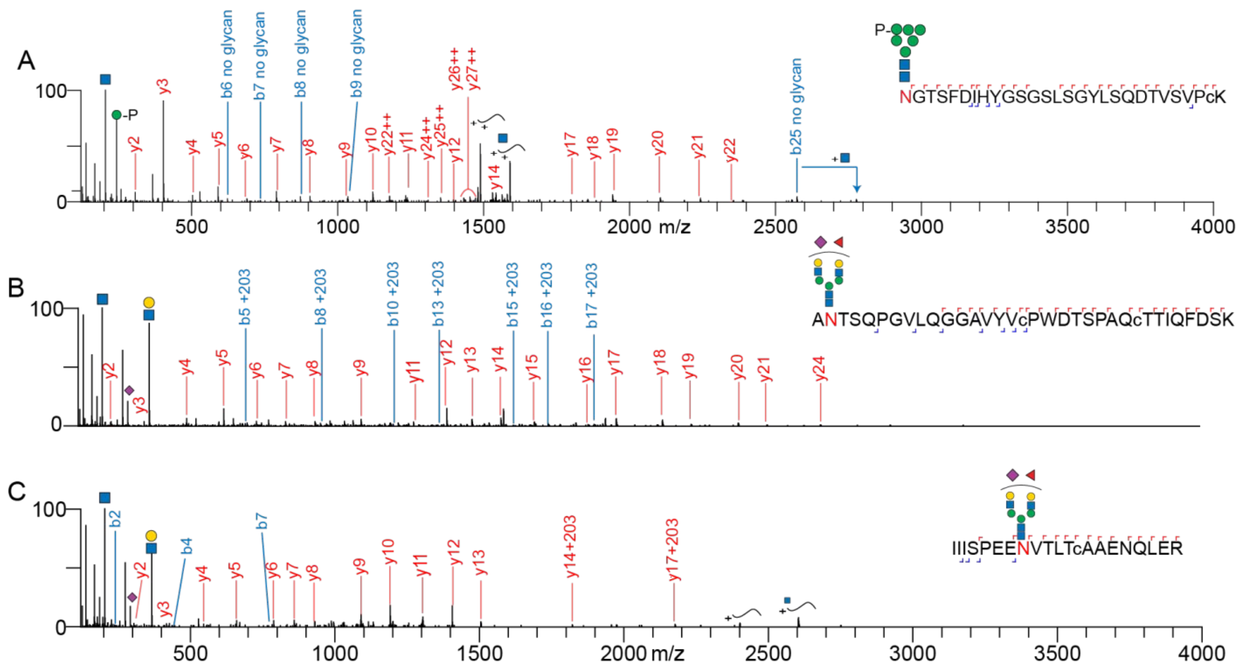
Figure S2. Annotated EThcD fragmentation spectra of doubly glycosylated peptides. Doubly glycosylated peptides originating from Legumain (A) and Hypoxia up-regulated protein 1 (B) are depicted. For simplicity sake only observed c ions are annotated in the spectrum while observed z ions are denoted on the peptide sequence itself. Green lines connect fragment ions corroborating the glycan composition of the first N-glycosylation site, while orange lines connect c ions confirming the glycan composition of the second N-glycosylation site. The annotated glycan composition of Legumain is further validated by the presence of a signature M6P compositional peak, depicted as a shaded fragment at 1498 m/z.


A
HCD-m/z 2000
YKPEFAAANDTGYVDIPQLEK
EThcD-m/z 2000
B
HCD-m/z 4000
EThcD-m/z 4000
C
HCD-m/z 2000
WNNTGcQALPSQDEGPSK
EThcD-m/z 2000
D
HCD-m/z 4000
EThcD-m/z 4000
Relative Intensity
m/z
ManP
HexNAc
HexNAcHex
HexNAcHex(2)
C6H8NO2
C7H8NO2
NeuAc-18
NeuAc
HexHexNAcNeuAc

Supplementary Figure S3. Comparison of oxonium ion intensities between standard *vs.* an extended *m/z* range. Only oxonium ions are annotated and highlighted in green. Peptide sequence coverage is denoted for each range and fragmentation mode in the corresponding inset. Comparison of HCD (left) *vs.* EThcD (right) for Cathepsin L glycopeptide is shown for standard range in (A) and extended range (B). Comparison of HCD (left) *vs.* EThcD (right) for polymeric immunoglobulin receptor glycopeptides is shown for standard range in (C) and extended range (D).



Supplementary Figure S4. Comparison of the distribution of Byonic LogP scores using HCD with a standard *vs.* an extended *m/z* range. (A) Distribution of Log P scores for M6P containing glycopeptides enriched from CHO cells by Fe^{3+}-IMAC. (B) Distribution of Log P scores for glycopeptides originating from human milk. (C) Distribution of Log P scores for glycopeptides obtained through SAX enrichment of CHO Fe^{+3}-IMAC flowthrough. Numbers below the graphs denote ranges (2000 m/z *vs.* 4000 m/z). Two technical replicates are shown for each range and all forward glycopeptide PSMs, number of which is represented by n, identified by Byonic were included.



Supplementary Figure S5. Examples of extended range HCD spectra. Glycopeptides originating from Cathepsin D (A), integrin Alpha-5 (B), and CD166 (C) are depicted. Series of y ions are observed and annotated, but no complementary ions retaining the N-glycan can be observed. Sporadic b ions observed in the spectrum have lost the N-glycan moiety during the gas phase fragmentation. Lower case c in the peptide sequence indicates a carbamidomethylated Cysteine.

Reference

(1) Čaval, T.; Zhu, J.; Tian, W.; Remmelzwaal, S.; Yang, Z.; Clausen, H.; Heck, A. J. R. *Mol. Cell. Proteomics* **2019**, *18* (1), 16–27. https://doi.org/10.1074/mcp.RA118.000967.