SUPPORTING INFORMATION

Strategic applications of negative mode LC-MS/MS analyses to expedite confident mass spectrometry-based identification of multiple glycosylated peptides.

Chu-Wei Kuo and Kay-Hooi Khoo*
Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 115, Taiwan

*Corresponding Author: kkhoo@gate.sinica.edu.tw

Contents:

Table S1 is provided as a separate Excel file, containing a compiled list of all identified glycoforms of the first 2 glycopeptides from the N-terminus of recombinant PTPRA by LC-MS/MS analyses.

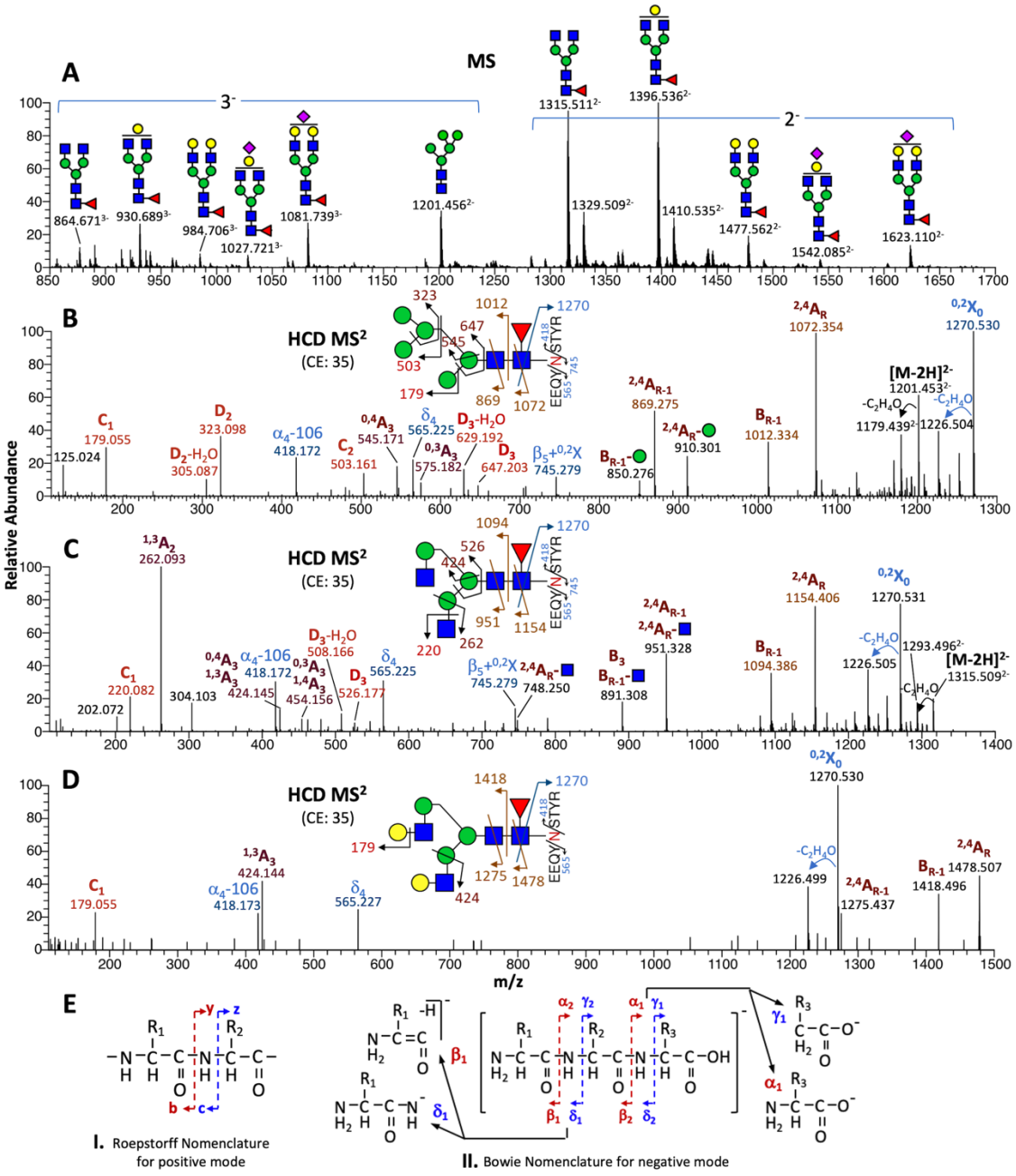

Figure S1. Negative mode LC-MS/MS analyses of IgG tryptic N-glycopeptide EEQY<u>N</u>STYR. (A) MS profile averaged from 431 scans (~ 5 min LC elution time), showing the charge states and *m/z* values of the major glycoforms. The negative mode HCD MS^2 spectra of the doubly charged N-glycopeptides with high mannose type $Man_5GlcNAc_2$ N-glycan at *m/z* 1201.456^{2-} (B), A2G0F biantennary N-glycan at *m/z* 1315.511^{2-} (C), and A2G2F biantennary N-glycan at *m/z* 1477.562^{2-} (D), each afforded characteristic $^{2,4}A_R$, B_{R-1}, $^{2,4}A_{R-1}$ ions, $^{0,2}X_0$ ions, and the D ions indicative of 6-arm extension. The MS^2/MS^3 of A2G1F and monosialylated N-glycopeptides are shown in Fig. 1, along with an illustration on the ion nomenclature adopted for the various fragment ions arising from glycosidic or ring cleavages. In cases when an ion can be alternatively assigned, both the equally probable assignments are given, *e.g.* $^{2,4}A_{R-1}$ or $^{2,4}A_R$-HexNAc, $^{0,4}A_3$ or $^{1,3}A_3$ in (C). The negative mode peptide cleavage ions are designated in accordance with the nomenclature suggested by Bowie [23], in which the N-terminal fragment ions similar to positive mode b and c ions (according to Roepstorff nomenclature, Roepstorff and Fohlman *Biomed Mass Spectrom* **1984,** *11* (11), 601) were named β and δ ions, respectively, while the C-terminal fragment ions similar to y and z ions were called α and γ ions, respectively (E). Loss of 106 u from the α_4 ion corresponds to side chain of Tyr.

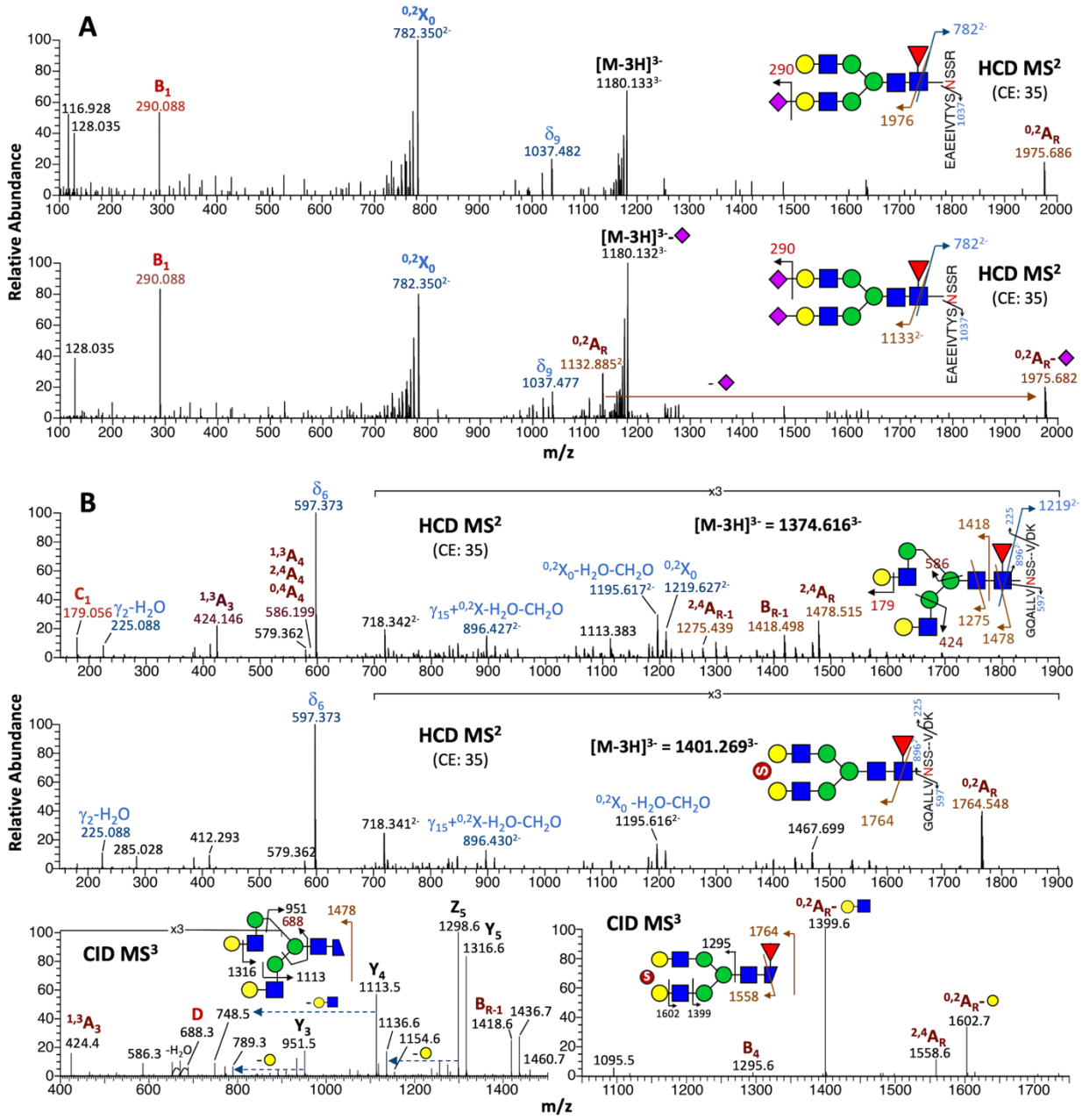

Figure S2. Negative mode HCD MS² of tryptic mono- and disialylated biantennary N-glycopeptides from bovine thyroglobulin (A), and non- and monosulfated biantennary N-glycopeptides from EPO (B). As shown for IgG N-glycopeptides (Fig. 1), the non-sialylated/sulfated N-glycopeptides produced $^{2,4}A_R$ ion, whereas the sialylated and sulfated N-glycopeptides afforded $^{0,2}A_R$ ion. Loss of sialic acid from precursor and the $^{0,2}A_R$ ion is only prominent for disialylated and not monosialylated N-glycopeptides. Multiple neutral losses from amino acid side chains can lead to precursor and $^{0,2}X_0$ ions showing similar pattern of satellite signals. At extreme case for longer peptide, this may compromise the signal intensity of the $^{0,2}X_0$ ion. Either the $^{2,4}A_R$ ion or $^{0,2}A_R$ ion can be selected for CID MS³, as shown in the bottom panels in (B) for the non- and monosulfated biantennary N-glycopeptides from EPO, respectively.

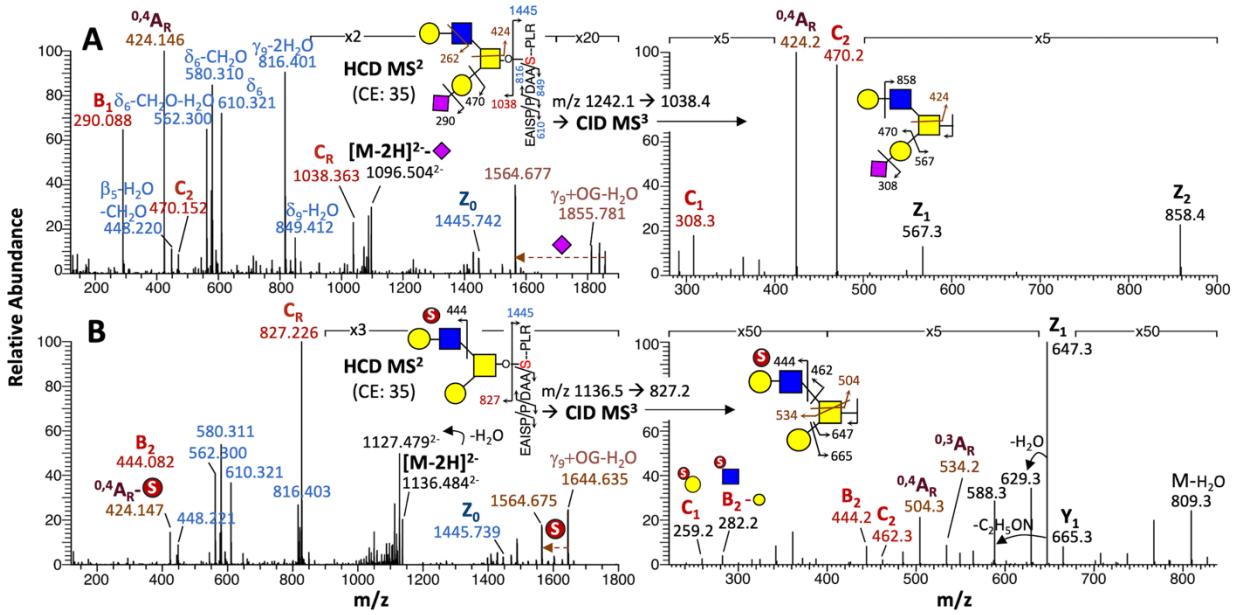

Figure S3. Negative mode HCD MS² and CID MS³ analyses of *O*-glycopeptides from EPO expressed in BHK21 cells. HCD MS² of the tryptic O-glycopeptide carrying a sialylated (A) or sulfated (B) core 2 O-glycan afforded prominent C_R ion (annotated in red), which defines the entire O-glycan moiety and can be further targeted for CID MS³ (shown in right panels). The commonly produced peptide-specific ions were annotated in blue, including the δ_6 ions (*m/z* 610 and *m/z* 580/562 after further losses of 30 and 18 u) and γ_9 ion after elimination of the O-glycan (*m/z* 816). The corresponding γ ions retaining the O-glycan but losing a H_2O moiety from Asp were also commonly detected.

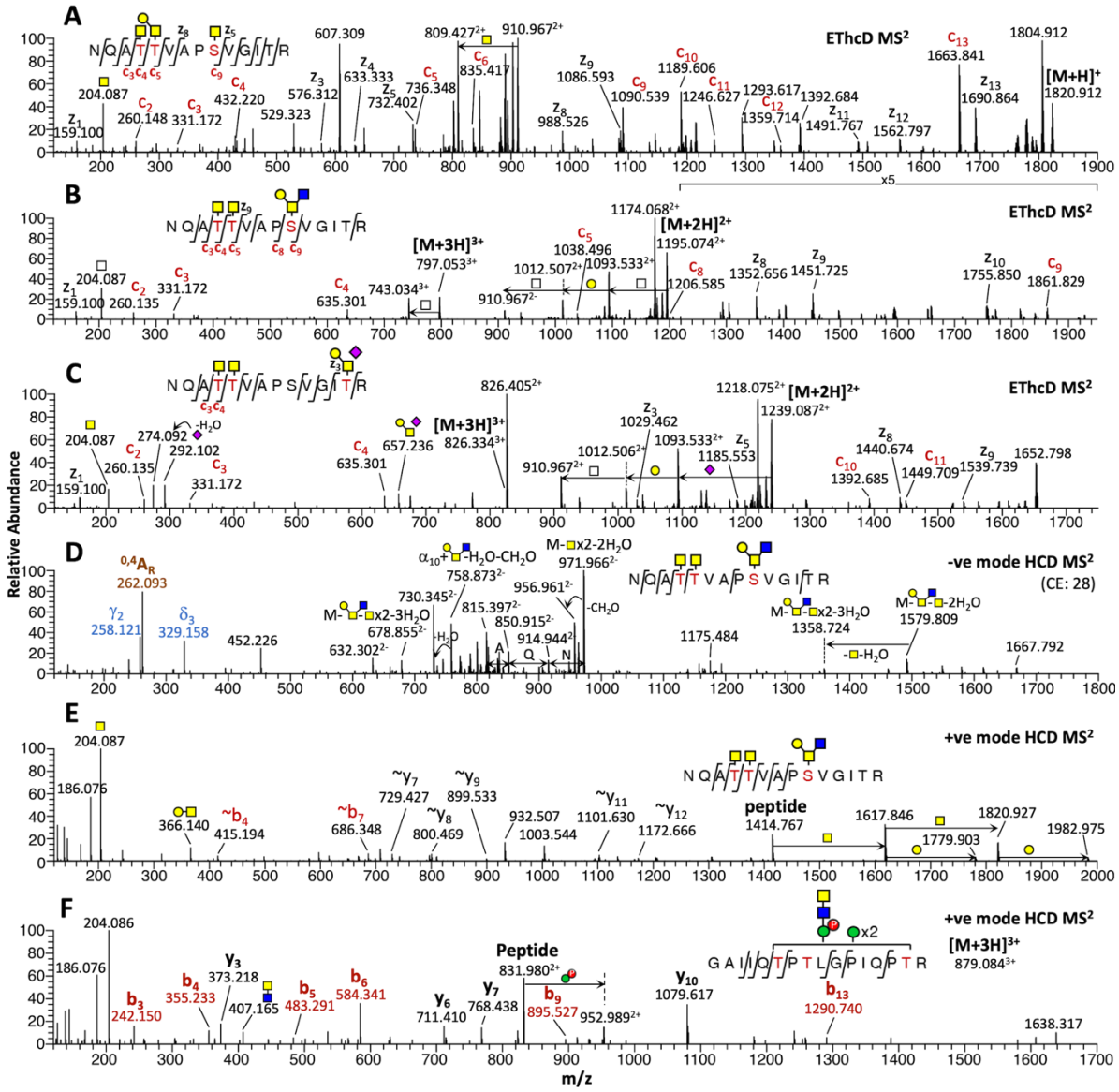

Figure S4. Supporting MS² spectra for the multiply O-glycosylated peptides. The respective positive mode EThcD MS² spectra (A-C) shown here for the three PTPRA O-glycopeptides analyzed by negative mode CID MS² (shown in Fig. 3A-C) allowed unambiguous assignment of the distribution of O-glycans at 3 different sites (as annotated on the spectra). These were used as supporting evidence for the observed elimination of individual O-glycans in negative mode. In cases when good quality EThcD MS² for the multiply O-glycosylated peptides cannot be obtained, the negative mode MS² analysis is useful to identify the presence of additional single HexNAc at different sites, which would otherwise not be discriminated from a single larger O-glycan carrying the extra HexNAc. Also shown here are exemplary negative and positive mode HCD MS² for the PTPRA O-glycopeptide with 2 single HexNAc and a $Hex_1HexNAc_2$. The negative mode HCD MS² (D) is similar to that of negative mode CID MS² shown in Fig. 3B, but with additional detection of low mass $^{0,4}A_R$ ion. However, more extensive elimination of at least 2 O-glycan moieties was observed. As expected, the positive mode HCD MS² (E) afforded sufficient numbers of peptide b and y ions to allow unambiguous identification of the peptide core but yielded little insight with respect to site-specific distribution of O-glycans. Likewise, the corresponding positive mode HCD MS² analysis (F) for the O-mannosylated peptide from recombinant human α-dystroglycan confidently identified its peptide core by the peptide b and y ions as annotated but could only spot a peptide core $+P_1Hex_1$ ion at *m/z* 952.989²⁺, not informative of the distribution of individual O-glycans at one or more sites. The positive mode EThcD and HCD MS² data were first searched against the protein sequence by using Byonic for initial glycopeptide identification and then manually examined to verify the automated assignment.

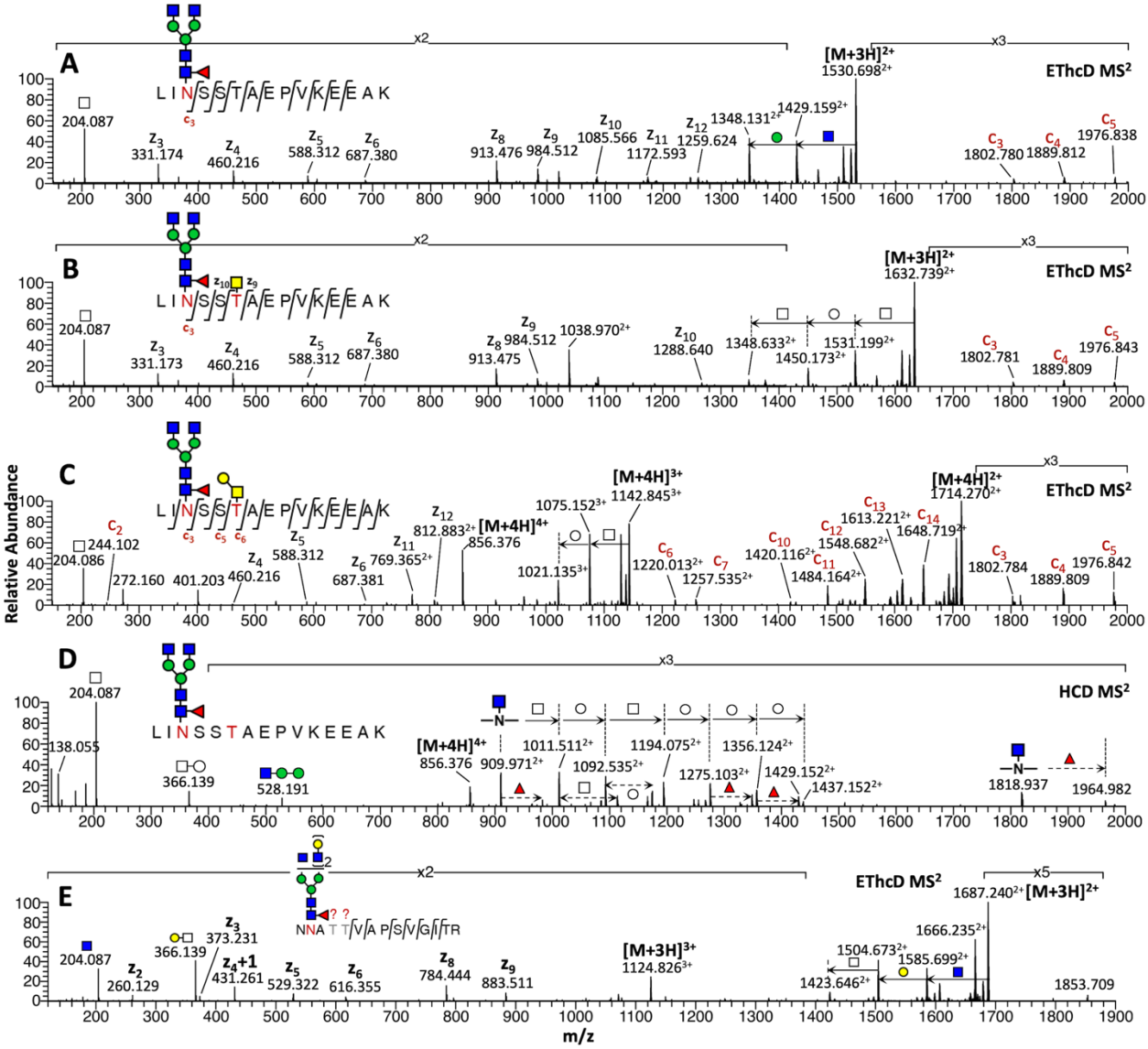

Figure S5. Supporting positive mode MS² spectra for the O-glycosylated N-glycopeptides of recombinant PTPRA. The respective positive mode EThcD MS² spectra (A-C) for a set of three related PTPRA glycopeptides allowed unambiguous assignment of the same N-glycopeptides with no additional O-glycan (A), or site-specifically O-glycosylated by a HexNAc (B) or $Hex_1HexNAc_1$ (C), as annotated. The common c_3, c_4 and c_5 ions detected not only defined the same N-glycan moiety but also ruled out the presence of O-glycans on the adjacent two Ser residues. On the other hand, the common z ion series up to z_9 indicated a non-glycosylated stretch from C-terminus up to the Thr residue which can be additionally O-glycosylated. In contrast, the common loss of a HexNAc and then a Hex from the precursor ions cannot be attributed specifically to losses from N- or O-glycans. Similarly, the positive mode HCD MS² (D) would yield the peptide core Y1 ion accompanied by HexNAc and Hex increments that could be remnants from either N- or O-glycans. In another example, even EThcD MS² (E) failed to identify the presence of additional O-glycan due to lack of c or z ions derived from cleavages at the -NATT- stretch but its presence could be inferred from negative mode MS² data (see Fig. 4C).

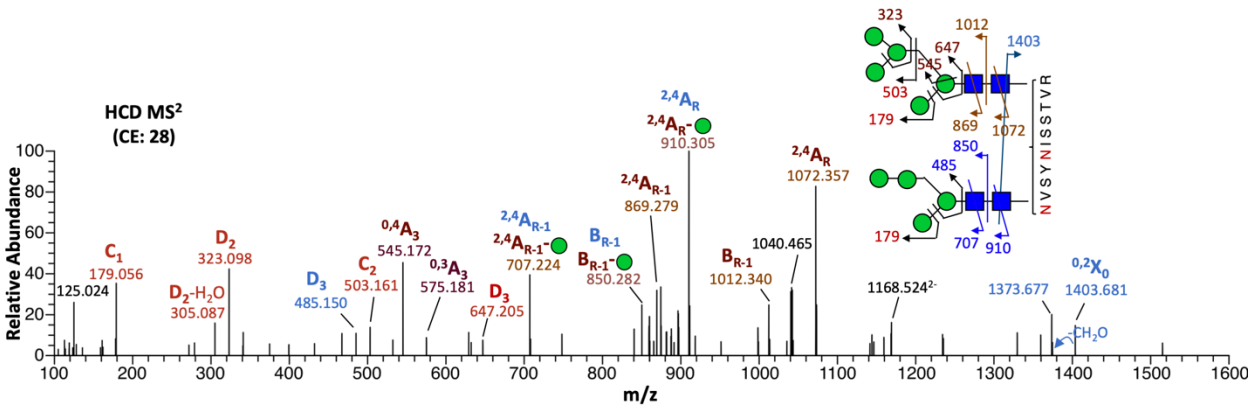

Figure S6. Negative mode HCD MS² spectrum of an N-glycopeptide carrying two N-glycans. A tryptic glycopeptide NVSYNISSTVR from mouse tyrosine-protein phosphatase non-receptor type substrate 1 was identified as carrying two different high mannose type N-glycans, Man_5 and Man_4, by virtue of producing two distinct sets of $^{2,4}A_R$, B_{R-1}, $^{2,4}A_{R-1}$, and D ions in negative mode HCD MS^2 (annotated in different colors). Losing both N-glycans led to a peptide backbone containing not one but two $^{0,2}X$ ion remnants of the innermost GlcNAcs on this peptide. In positive mode, losing both N-glycans would leave instead 2 full GlcNAc as the Y1 ions, which cannot be distinguished from a Y2 ion produced by a singly N-glycosylated peptide.