Extended sialylated O-glycan repertoire of human urinary glycoproteins discovered and characterized using EThcD

AUTHOR NAMES

Zsuzsanna Darula[1], Ádám Pap[1,2], Katalin F. Medzihradszky[1]**

CORRESPONDING AUTHORS

Zsuzsanna Darula

e-mail: darula.zsuzsanna@brc.mta.hu

Katalin F. Medzihradszky

e-mail: medzihradszky.katalin@brc.mta.hu

ADDRESS

[1]Biological Research Centre of the Hungarian Academy of Sciences, Temesvari krt. 62., H-6726 Szeged, Hungary

[2]Doctoral School in Biology, Faculty of Science and Informatics, University of Szeged, Kozep fasor 52., H-6726 Szeged, Hungary

ABSTRACT

A relatively novel activation technique, EThcD was used in the LC-MS/MS analysis of tryptic glycopeptides enriched with wheat germ agglutinin from human urine samples. We focused on the characterization of mucin-type O-glycopeptides. EThcD in a single spectrum provided information on both the peptide modified and the glycan carried. Unexpectedly, glycan oxonium ions indicated the presence of O-acetyl, and even O-diacetyl-sialic acids. B and Y fragment ions revealed that i) in core 1 structures the Gal residue featured the O-acetyl-sialic acid, when there was only one in the glycan; ii) several glycopeptides featured core 1 glycans with disialic acids, in certain instances O-acetylated; iii) the disialic acid was linked to the GalNAc residue whatever was the degree of O-acetylation; iv) core 2 isomers with a single O-acetyl-sialic acid were chromatographically resolved. Glycan fragmentation also helped to decipher additional core 2 oligosaccharides: a LacdiNAc-like structure, glycans carrying sialyl Lewis$^{X/A}$ at different stages of O-acetylation, and blood antigens. A sialo core 3 structure was also identified. We believe this is the first study when such structures were characterized from a very complex mixture and were linked not only to a specific protein, but also the sites of modifications have been determined.

KEYWORDS

affinity chromatography, blood-type antigen, disialic acid, EThcD, mass spectrometry, O-acetyl-sialic acid, O-glycopeptides, urinary glycoproteins, wheat germ agglutinin

INTRODUCTION

Protein glycosylation is a common post-translational modification that is responsible for a wide range of biochemical interactions. Protein glycosylation can be C-, N-, O- and S-linked and hence a range of amino acid side-chains harbor these glycan linkages[1]. O-glycosylation may alter Ser and Thr side-chains within the cytosol as well as within the Golgi[1,2]. While the intracellular proteins feature a single GlcNAc as modification, the Golgi-derived glycans are much more complex, with alternate cores that are usually elongated[1,2]. In mammals the GalNAc-core (mucin-type) glycosylation is the most common[1,2]. Mucin-type O-glycosylation of Tyr residues has also been reported recently[3]. Early on it has also been observed that the same sequence positions may or may not be always glycosylated (macroheterogeneity) and may feature several different glycans (microheterogeneity)[1]. For technical reasons glycoprotein decorating glycans are frequently studied after being released[4-7]. This permits the fractionation and in-depth characterization of these oligosaccharides, however, the site-specific information is lost in the process. With the advent of very sensitive, high resolution mass spectrometers and new activation techniques, such as ETD (electron transfer dissociation[8]) and EThcD (electron-transfer/higher-energy collision dissociation[9]), the focus is shifting towards intact glycopeptide or even intact glycoprotein (or at least middle-down) analysis[10]. While mass spectrometry is suitable for identifying most PTM-bearing sequences in an unbiased manner, collision-induced dissociation (CID), the most commonly used MS/MS activation method for obtaining structural information on peptides, performs poorly for O-glycopeptide characterization. The culprit is the glycosidic bond that is more prone to fragmentation than the peptide bond and thus, fragment ions resulting from sugar cleavages dominate the spectra. The glycan is usually completely released from the peptide, without leaving any clue behind on the originally modified amino acid. As a result, little to no information is available to assign the glycosylation site(s).[10] Hence ion trap CID, i.e. resonance activation only delivers information about the glycan size and composition, and the peptide size, while beam-type CID (HCD) may produce peptide sequence information, but information on the glycan(s) may be more limited than in ion trap CID because of its extensive fragmentation[10]. Electron transfer dissociation usually results solely in the peptide backbone fragmentation, thus, both the modified amino acid sequence and the modification site(s) can be

assigned[8]. However, ETD spectra frequently yield incomplete sequence ion series, and thus, limited information about the glycans present (i.e. it cannot be decided whether the glycopeptide features two small glycans or a larger oligosaccharide of the same combined sugar composition) and their location is available[10]. In the newest activation method, EThcD, electron-transfer is performed first, then newly generated fragments and intact precursor ions (activated) are subjected to beam-type collisional activation[9]. In this process only some of the ETD-generated fragments undergo further breakdown (such as w ion formation[11]), but the precursor ions are definitely cleaved[9]. Most of this fragmentation results from glycosidic bond cleavages, however, the activation is gentler than in normal HCD and the resulting glycan fragments are more abundant, while larger structures also survive[12]. Thus, we gain information both about the peptides modified and the oligosaccharides featured.

In this manuscript we demonstrate that the information content of EThcD spectra enabled the characterization of unexpected/unusual sialylated structures on intact glycopeptides that were enriched from the urine of healthy individuals. As far as we know this is the first time that such structures have been identified from a complex biological matrix, in a high throughput manner, with protein site specificity. The presented method opens new possibilities for the discovery of novel glycan structures with site specificity on proteins to reveal their biological function and potential as diagnostic biomarkers.

EXPERIMENTAL

Here we utilized a glycopeptide enrichment strategy as previously described [12]. We would like to emphasize that no acetate buffer or acetic acid containing solution was used in any part of the protocol. Briefly, tryptic digests of urine of 4 healthy donors (collected with appropriate consents approved by the Hungarian Scientific and Research Ethics Committee (approval number: 1011/16)) were subjected to a 2-round glycopeptide enrichment using a wheat germ agglutinin affinity column collecting 2 glycopeptide fractions representing the shoulder of the flow through peak and a fraction eluted by GlcNAc. Fractions were analyzed separately by LC-MS/MS using a Waters M-Class nanoUPLC on-line coupled to an Orbitrap Fusion Lumos mass spectrometer operated in positive ion mode. Peptides loaded onto a trap column (Waters Acquity UPLC MClass Symmetry C18 180 µm× 20 mm column, 5-µm particle size, 100-Å pore size; flow rate 10 µl/min) were separated by a linear gradient of 10-30% B in 60 min (Waters Acquity UPLC M-Class BEH C18 75 µm× 250 mm column, 1.7-µm particle size, 130-Å pore size; solvent A: 0.1% formic acid/water; solvent B: 0.1% formic acid/ACN; flow rate: 300 nl/ min), MS/MS data were acquired using HCD product-ion dependent EThcD data acquisition mode. The HexNAc-specific oxonium ion, m/z 204.0867 among the 20 most abundant HCD fragments triggered EThcD acquisition. HCD spectra were acquired at 28% NCE, while supplemental activation in EThcD was set at 15% NCE. Precursors with z=3-5 were considered (intensity threshold: 10^6) according to a decision tree with decreasing charge state and increasing precursor m/z, in a total cycle time of 3 s. Some samples were also analyzed using direct injection (i. e. without trapping) and allowing z=2 precursors. All measurements were performed in the Orbitrap, with a resolution of 60000 and 15000 for MS1 and MS/MS, respectively.

Separate HCD and EThcD peaklists were generated using Proteome Discoverer (Thermo Scientific, v2.2.0.388) requiring at least 40 peaks per spectrum. EThcD data were filtered for the presence of m/z 292.0927 within the top 20 most abundant fragment ions (required mass accuracy: 10 ppm) using the MS-Filter software of Protein Prospector[13] and retained spectra were searched using the Protein Prospector search engine with the following settings: enzyme: semitrypsin with maximum 1 missed cleavage site; database: human subset of the Swissprot database (2017.9.19.version, 20219 sequences) concatenated with

the same number of random sequences; mass accuracy: 5 ppm for precursor ions, 10 ppm for fragment ions; fixed modification: carbamidomethylation (Cys), variable modifications: acetylation (protein N-terminus), cyclization (N-terminal Gln), oxidation (Met), and a set of 28 glycan structures representing core 1 and core 2 O-glycans with up to 4 sialic acids/glycan (Supporting Information 1, Table S-1) on Ser/Thr. All variable modifications were defined as "common" allowing 2 variable modifications per peptide. The 80 most abundant peaks were considered from each MS/MS spectrum. Glycopeptide identifications were accepted applying 5% and 1% FDR on the protein- and peptide-level, respectively, and setting a SLIP-score[14] threshold of 6. All identifications presented were validated manually.

The raw data files have been uploaded to MassIVE (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp), the project's identifier is MSV000083070.

RESULTS

The workflow

We intend to gain information on the site-specific O-glycosylation of human proteins, and for this purpose the analysis of intact glycopeptides from different sources is necessary. In this study tryptic glycopeptides were enriched from urine of healthy donors. Previous studies on urinary glycoproteins, applying the sialic acid release and capture method[15] or the glycoblotting-assisted O-glycomics approach[16,17] indicated that core 2 and disialylated core 1 O-glycans are present at considerable amounts. As glycopeptides bearing these structures would not be enriched by lectins frequently used for O-glycopeptide analysis including Jacalin and peanut agglutinin, we used wheat germ agglutinin (WGA) that binds a wide array of glycan structures[18-20] including both N- and O-glycopeptides. The glycopeptide mixtures were analyzed applying HCD product ion-dependent EThcD with mild supplemental activation (15% NCE), and EThcD data were interpreted using Protein Prospector[12]. As deduced from the number of MS/MS spectra acquired, the two-round lectin enrichment was highly efficient: ≥ 94% of the acquired HCD spectra featured the diagnostic glycan oxonium ion m/z 204 in all collected WGA fractions[12]. N- and O-glycopeptides were identified in separate searches (for details see [12] and Online Resources 3-4 therein). As the vast majority of the glycopeptides represented sialylated glycans[12] we used peak lists filtered for the presence of the diagnostic m/z 292. For O-

glycopeptide identification a-, mono- and disialo core 1 and the disialo core 2 O-glycan structures were considered. Approximately 2/3 of the O-glycosylated sequences reported earlier for urinary proteins[15] were identified along with several new, mostly multiply glycosylated peptides indicating that WGA affords efficient enrichment of O-glycopeptides. However, the overall identification rate was rather low (~5%)[12]. One of the obvious reasons is that in order to keep false discovery rates at acceptable levels, only 2 variable modifications per peptide was permitted in these searches. Thus, multiply modified peptides could be (mis)assigned only if the combined sugar composition of their glycans equaled to a permitted combination. Obviously, glycopeptides with glycans other than the specified structures were also missed.

Iterative search for unexpected/unusual structures

In order to learn what additional structures occur on the glycopeptides enriched, a non-specified modification search[18,19] was performed on proteins whose presence was already ascertained: any modification up to 3000 Da was allowed on Asn, Ser and Thr residues. From the recurring modification masses and manual inspection of some corresponding MS/MS spectra we concluded that structures containing disialic acids or O,N-diacetylneuraminic acid(s) also were present and needed to be considered. Thus, a more complete glycan database was designed (Table S-1) and a new database search was performed with these accurately defined structures as variable modifications. Then the assigned spectra were carefully inspected. We observed earlier with O-glycopeptides bearing sialyl core 1 and core 2 structures, that the mild supplemental activation during EThcD enables the detection of higher-mass oxonium ions, sometimes even the intact glycan[12]. We hypothesized that such ions together with Y-type fragments (for nomenclature see [21]) could be used to decipher the glycan structures, for example, the linkage position of the unusual disialic acid moiety and the O,N-diacetylneuraminic acids. In addition, our observations could be used to establish fragmentation rules for the identification of novel structures.

The final list of unexpected sialic acid-containing glycoforms derived from the above searches is presented in Table 1. Table S-2 and Figures S-1-29 contain additional information about the identifications and display the supporting MS/MS data. As the focus of this manuscript is to demonstrate how EThcD data can be used for glycan structure elucidation, only fragmentation relevant to the glycan structures is detailed below

(although we would like to point out that peptide side-chain w ions were observed in several spectra that may aid peptide identification). Here we would like to emphasize that our data does not distinguish the isomeric sugar units, resolve position or the stereochemistry of the glycan linkages. These assignments (if presented) are based on assumptions and on our general knowledge on mucin-type glycosylation. However, as described below, the MS/MS spectra usually provided sufficient information on the direct connection of the sugar units or on the branching of the glycans.

Table 1. Glycopeptides carrying O-glycans with O-acetyl- and/or disialic acid(s)[1]

Uniprot Acc #	Protein Name	Peptide[2]	Tetra[3]		Penta[3]				Hexa[3]	
			1	2	0	1	2	3	1	2
Q9GZM5	Protein YIPF3	^{342}AVAVTLQSH^{350}	+	+	+	+	+	+	+	+
P01344	Insulin-like growth factor II	^{93}DVSTPPTVLPDNFPR^{107}			+	+				
P78423	Fractalkine	^{175}AQDGGPVGTELF^{186}				+				
		^{175}AQDGGPVGTELFR^{187}				+				
		^{247}VWGQGQSPRPE^{257}				+				
		^{247}VWGQGQSPRPENSLER^{262}			+	+				
		^{263}EEMGPVPAHTDAF^{275}			+	+				
Q9BZL3	Small integral membrane protein 3	^{1}MDAVSQVPMEVVLPK^{15}	+			+				
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	^{33}SLPEDIETVTASQMR^{47}			+					
		^{48}WTHSYL^{53}				+				
Q13508	Ecto-ADP-ribosyltransferase 3	^{316}ILEPTQIPGMK^{326}	+							
Q9H2B2	Synaptotagmin-4	^{9}EEFDEIPTVVGIF^{21}				+				
Q8TBP5	Membrane protein FAM174A	^{50}PRTLPPLPPGPTPAQQPGR^{68}	+							
P09603	Macrophage colony-stimulating factor 1	^{370}VGPVRPTGQDWNHTPQK^{386}			+					
P14209	CD99 antigen	^{29}SDALPDNENKKPTAIPK^{45}			+					
P60022	Beta-defensin 1	^{22}GNFLTGLGHR^{31}			+					

[1] identified by Protein Prospector and validated manually.

[2] Ser/Thr residues printed in red denote the site of glycosylation.

[3]Tetra, penta and hexa stands for the total number of the monosaccharides within the glycan structure. Numbers in the next row specify the number of NeuAcAc residues in the glycan. Proposed structures are shown in Table S-1, and in Figures S1-29.

EThcD fragmentation 'rules' at low normalized collision energy

As a general observation, gentle supplemental energy (NCE: 15%) applied in EThcD typically seemed to favor single bond cleavages. Glycan fragmentation yielded non-reducing end B ions, frequently even larger (up to 6 residues) oligosaccharides were detected[12], and less prominent internal fragments could be observed compared to HCD (acquired with NCE: 28%) (Figures S-1-29 display the corresponding HCD and EThcD spectra for each assigned glycopeptide, in certain instances we highlighted the fragmentation differences). These ions together with reducing end Y fragments usually provided insight into the glycan structure as detailed below.

O-acetyl-sialic acid-containing glycans

Fragments at m/z 334.113 and 316.103 indicated the presence of 9-O-acetyl,N-acetylneuraminic acid ($Neu5,9Ac_2$ according to convention, but defined as NeuAcAc in the database searches) within some glycan structures (m/z 316 represents water-loss from $Neu5,9Ac_2$). Obviously, from these data it cannot be determined which position was acetylated, however, this structure is the most likely[1]. Screening for these fragments with MS-Filter of Protein Prospector[13] showed that ~5% of all EThcD spectra featured at least one of these oxonium ions. Database searches allowing $Neu5,9Ac_2$ (NeuAcAc) containing glycans yielded confident identifications (Table 1, Table S-2, Figures S-1-29).

$Neu5,9Ac_2$ was always found in terminal position as evidenced by the Y ions formed via $Neu5,9Ac_2$-loss from the charge-reduced precursors. Confidently assigned glycopeptides featured tetra-, penta- and hexasaccharides with up to three $Neu5,9Ac_2$ residues/glycan (Table 1). However, manual evaluation altered the assignment that indicated the presence of an 'internal' $Neu5,9Ac_2$ (see below).

In the core 1 tetrasaccharides containing one $Neu5,9Ac_2$, this sugar was Gal-linked as evidenced by the B_2 ion representing $GalNeu5,9Ac_2$ (at m/z 496.166) and the corresponding $Y_{1,\alpha}$ ion, at m/z 710.345(2+) (Figure

1). This B_2 ion in EThcD is usually of higher abundance than the GalNAcNeuAc internal fragment (at m/z 495.182) that is more prominent in HCD (see Figure S-1). Actually, the GalNeu5,9Ac_2 B_2 ion may be overlooked in HCD as the isotopic peak of the internal fragment. The ion at m/z 990.34 represents the intact oligosaccharide.

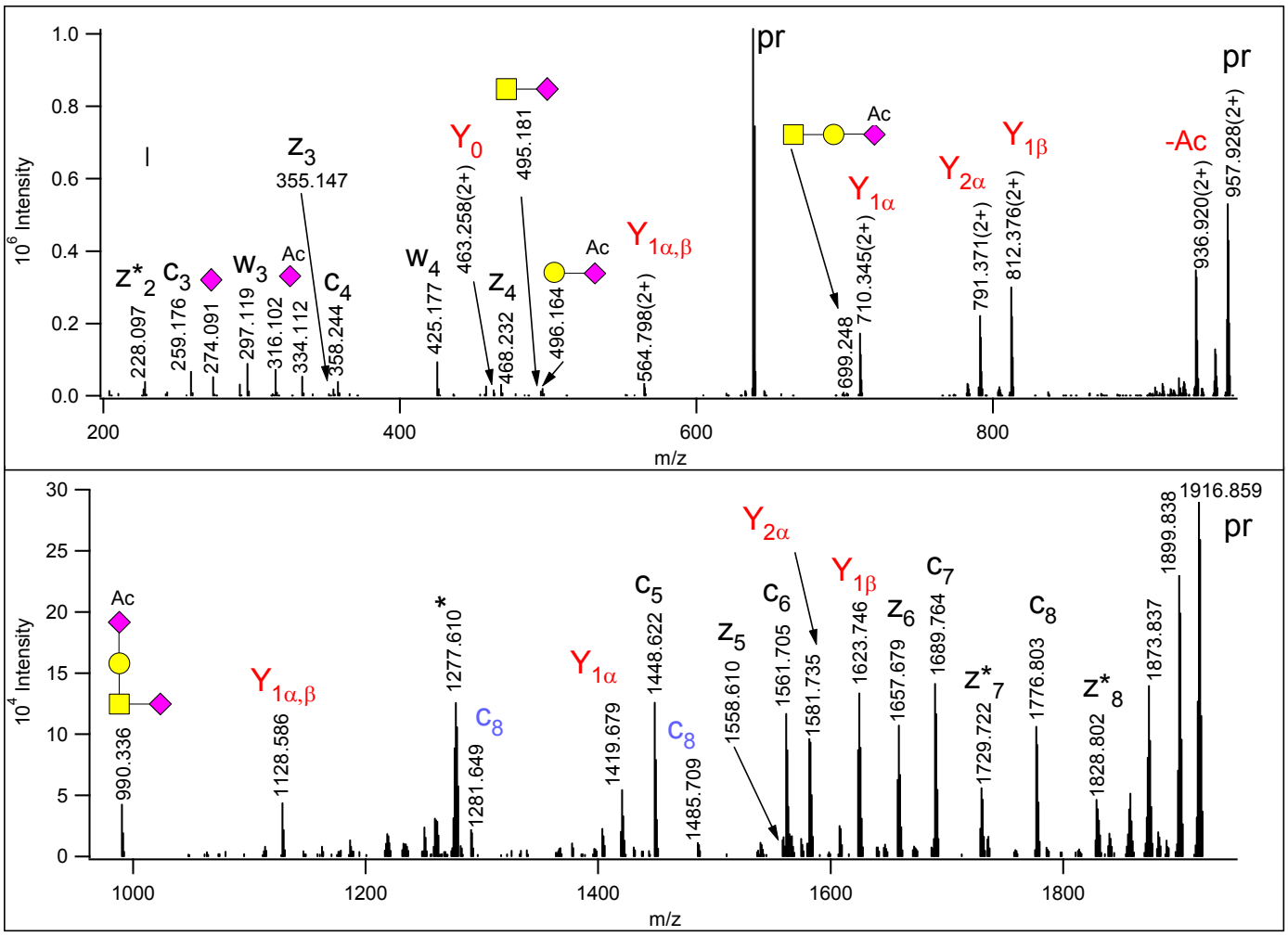


Figure 1. EThcD spectrum of m/z 638.954(3+), identified as ^{342}AVAVT(HexNAcHexNeuAcNeuAcAc)LQSH^{350} of Protein YIPF3. The Gal of the mucin-type core 1 structure is capped with a Neu5,9Ac_2 moiety. (This sugar is listed as NeuAcAc in the glycan library for database searches). Oxonium and related ions are labeled according to the CFG recommendations, for the reducing end fragments – printed in red - the Domon-Costello nomenclature is followed[21]. The sialic acid loss from the core GalNAc yields the $Y_{1\beta}$ fragment. z+1 peptide ions are distinguished with asterisks. Peptide fragments in blue indicate that some glycan fragmentation occurred in EThcD, the loss of the

NeuAc and GalNeu5,9Ac_2 from the c_8 fragment, respectively. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.

In a disialo core 2 hexasaccharide with one Neu5,9Ac_2, this capping residue can be linked to either of the galactoses. Protein Prospector delivered the same identification for EThcD data acquired at different time points from precursor ion m/z 760.664(3+), as ^{342}AVAVT(HexNAc_2Hex$_2$NeuAcNeuAcAc)LQSH^{250} of Protein YIPF3 (Q9GZM5) (Figures S-7 & S-8). Both spectra display the intact glycan fragment (at m/z 1355.46) to confirm the presence of a hexasaccharide as opposed to two smaller structures, and Neu5,9Ac_2 is Gal-linked (see m/z 496.166). However, as Figure 2 shows in detail, two isomeric glycoforms were present in this sample. In the earlier eluting, more abundant glycoform, the Neu5,9Ac_2 is capping the GlcNAc-Gal arm (designated as 'β-chain', Figure 2, lower panel). This arm can readily be removed by a single cleavage yielding an oxonium ion of HexNAcHexNeuAcAc composition at m/z 699.245 ($B_{3\beta}$), and the corresponding $Y_{1\beta}$ fragment at m/z 791.372(2+). These ions are missing from the other isoform, where the Neu5,9Ac_2 is capping the core GalNAc-linked Gal (designated as 'α-chain', Figure 2, upper panel). Thus, when the bond between the core GalNAc and its modifying GlcNAc is cleaved, an abundant $B_{3\beta}$ can be observed at m/z 657.234, and the corresponding $Y_{1\beta}$ was recorded at m/z 812.378(2+). The other fragments are of identical composition, but they are the products of different bond cleavages as indicated in their assignments. In addition, the relative intensities of $Y_{1\alpha}$ and $Y_{2\beta}$ are different, it seems that the cleavage between the core GalNAc and Gal is preferred to that between the GlcNAc and the Gal linked to it. Here again, we would like to point out that the internal fragmentation of the glycan is less extensive than in HCD: m/z 366.139 representing HexNAcHex is more abundant in the HCD spectra, and the HexNAc_2 fragment (m/z 407.166) was not detected in EThcD (See Figures S 7-9).

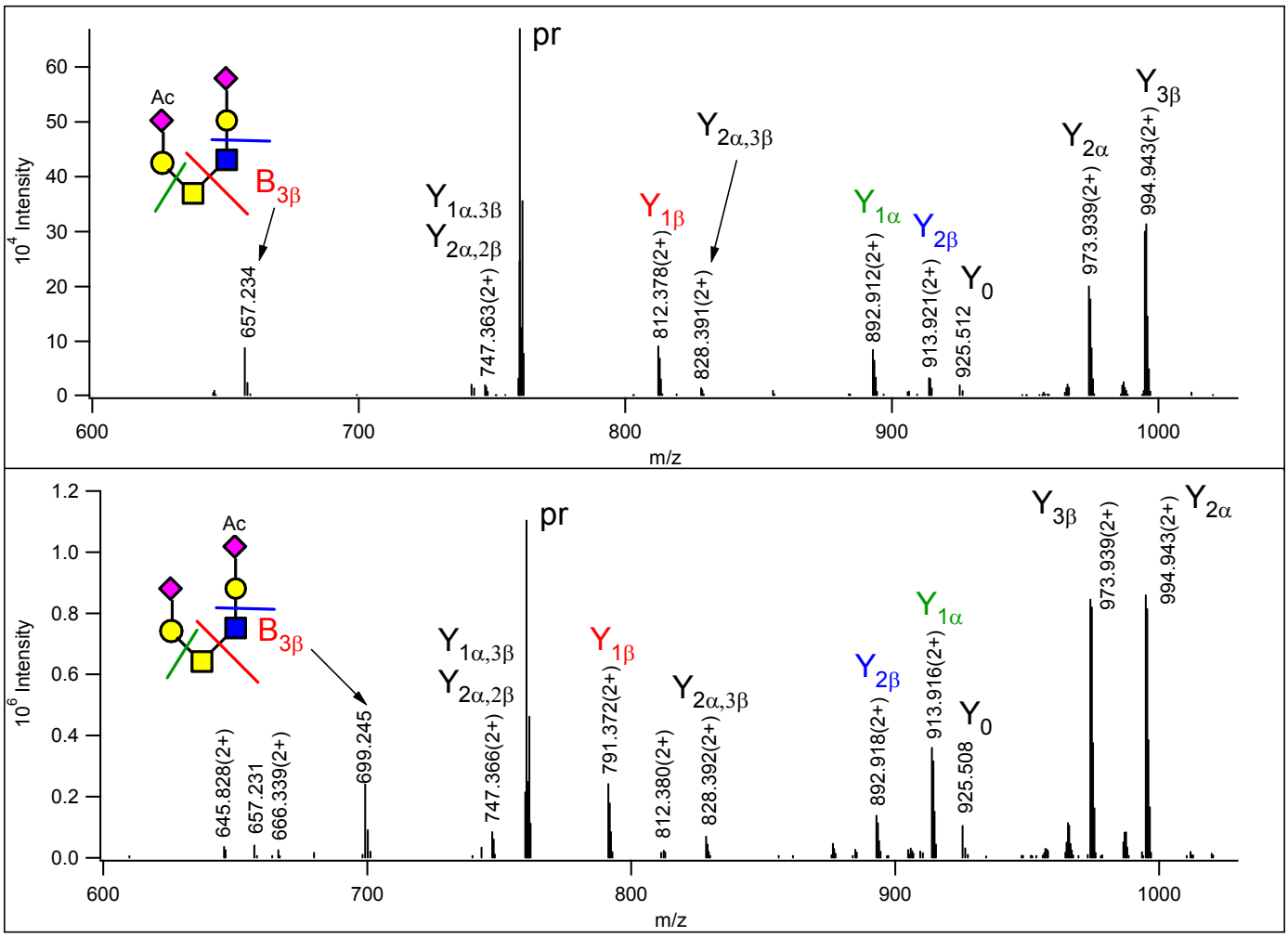


Figure 2. Part of the EThcD spectra acquired from m/z 760.664(3+) at RT=14.95min (Lower panel) and at RT=18.29min (Upper panel). Both spectra were assigned as ^{342}AVAVT(HexNAc2Hex2NeuAcNeuAcAc)LQSH^{350} of Protein YIPF3 (for full spectra see Figures S-7 & S-8). (NeuAcAc stands for Neu5,9Ac$_2$ in the glycan library for database searches.) However, the glycan fragments revealed isomeric glycoforms. The diagnostic fragment ions and the underlying cleavages are indicated in matching colors.

Finally, an identification indicated the presence of a pentasaccharide structure with three NeuAcAc residues (Figure 3). Although the additive mass of the glycan (1364.45 Da) is correct, we found this identification highly unlikely as it suggests the presence of a non-terminal Neu5,9Ac$_2$ considering the structure of the pentasaccharide glycans (see below). Careful investigation of the EThcD spectra revealed that three different sialic acids modified the core 1 glycan: the Gal is capped with a Neu5,9Ac$_2$, while the GalNAc is linked directly to a NeuAc that is further elongated with a sialic acid bearing 2 'extra' acetyl groups. This is likely

an O-4,9-diacetyl-N-acetylneuraminic acid[1]. The presence of such a neuraminic acid derivative, perhaps Neu4,5,9Ac_3, was confirmed by its proper oxonium ion and the corresponding water-loss fragment at m/z 376.124 and 358.112, respectively (both within 4 ppm of the theoretical m/z values). The GalNeu5,9Ac_2 linkage was validated by $Y_{1\alpha}$ representing the loss of the hexose along with the terminal sialic acid, at m/z 897.899(2+). The GalNAc-linked NeuAcNeu4,5,9Ac_3 structure is supported by the observation of the above described $Y_{1\alpha}$ fragment, and by the consecutive losses of Neu4,5,9Ac_3 and NeuAcNeu4,5,9Ac_3 at m/z 957.924(2+) and 1623.747, respectively.

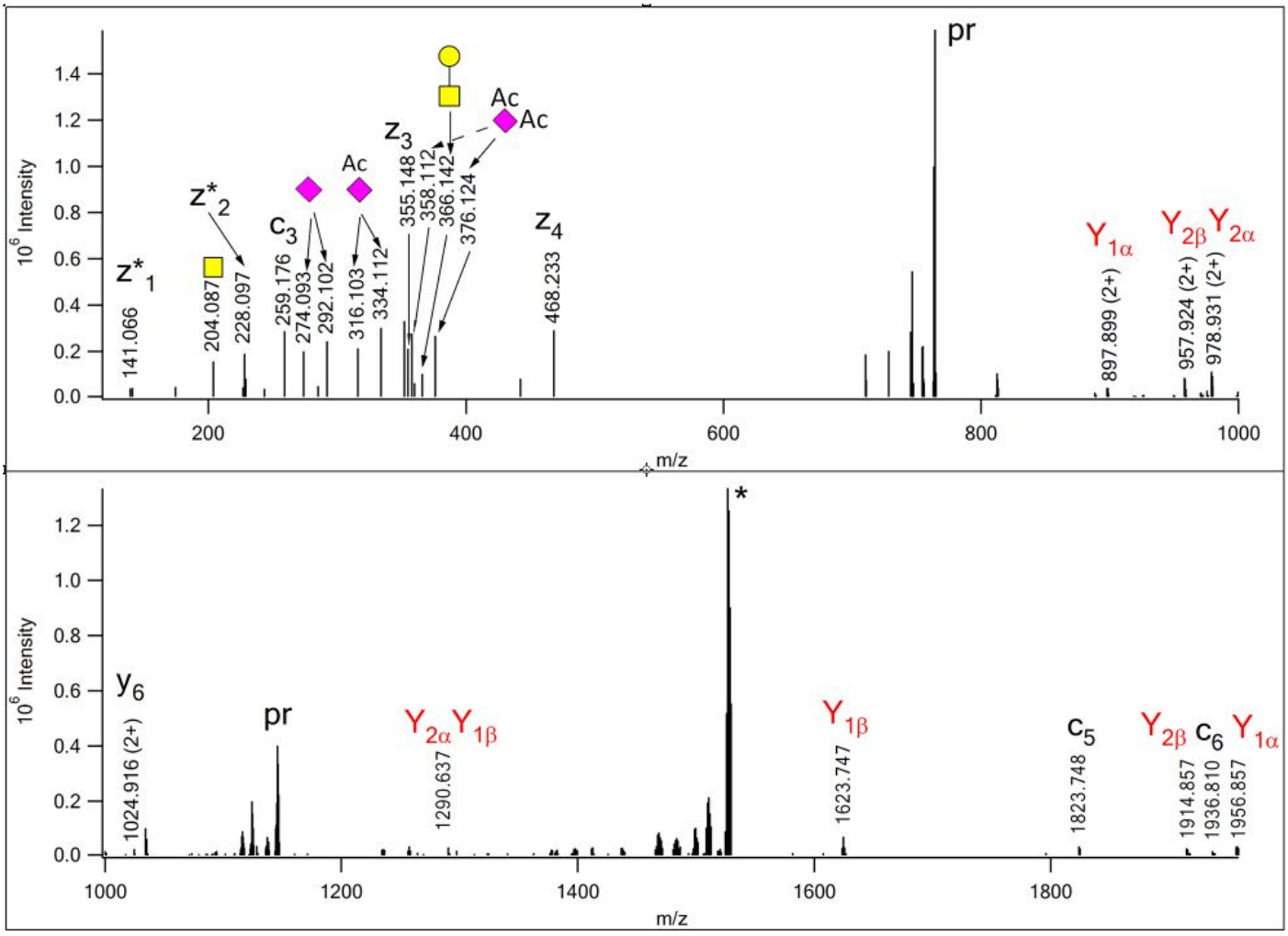
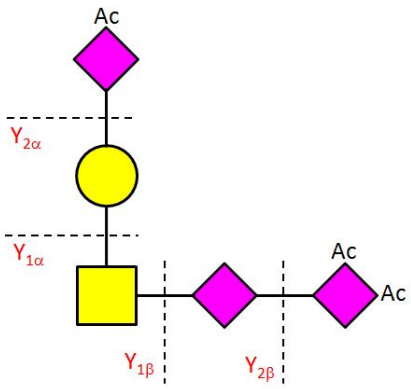


Figure 3. EThcD spectrum of m/z 763.992(3+) manually assigned as ^{342}AVAVTLQSH^{350}, modified at Thr-346 with a core 1 glycan that features a Neu5,9Ac_2 on the Gal residue, and a NeuAc-Neu4,5,9Ac_3 disialo-unit on the core GalNAc. 'Pr' indicates the different forms of the precursor ion. z+1 peptide fragments are distinguished with asterisks. The oxonium ions identifying the three different sialic acids are labeled with the CFG symbols, for the reducing end fragments – printed in red - the Domon-Costello nomenclature is

followed[21]. The GalNAc modifications were considered the β-arm. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.



Scheme 1. Fragmentation of the glycan structure assigned from the spectrum in Figure 3.

Disialic acid-containing glycans

Obviously, there is an overlap between the unusual glycopeptides featuring O-acetyl-sialic acids and the ones bearing core 1 structures modified with a disialic acid unit (Table 1). The terminal residue of this disialic unit may be NeuAc or may bear additional O-acetyl groups, i.e. it may be Neu5,9Ac_2 or Neu4,5,9Ac_3 (as described above). The disialic acid unit usually produces a relatively abundant oxonium ion at m/z 583.197 or 625.210 for $(NeuAc)_2$ or NeuAcNeu5,9Ac_2, respectively. Filtering for the presence of these ions with MS-Filter[13] showed that ~2% of all EThcD spectra represented glycopeptides with disialic acid within the glycan structure. The composition of the modifying glycan was frequently confirmed by an oxonium ion representing the intact glycan, at m/z 1239.43 or 1281.44.

EThcD fragmentation of a glycopeptide with a pentasaccharide glycan is illustrated in Figure 4. The glycan carries three "simple" NeuAc units, Y ions representing neutral losses of $(NeuAc)_2$ at m/z 1156.04(2+), and HexNeuAc at m/z 1220.56(2+) prove that the glycan is branched, and the disialic acid is linked to the core GalNAc. The inter-sialic acid bond is confirmed by the $(NeuAc)_2$ oxonium ion ($B_{2\beta}$) at m/z 583.197. The oxonium ion representing the intact glycan at m/z 1239.41 confirms that a single glycan modifies the peptide. In a similar fashion we could decipher that (whenever there was conclusive evidence) the disialic acid was always GalNAc-linked. Limited internal fragmentation of the glycan was observed for this class of glycopeptides as well. While in HCD the internal fragment representing GalNAcNeuAc (m/z 495.182) was

usually abundant, in EThcD the B_2 ion representing GalNeuAc (m/z 454.155) was dominant (for example, see Figures S-10-12).

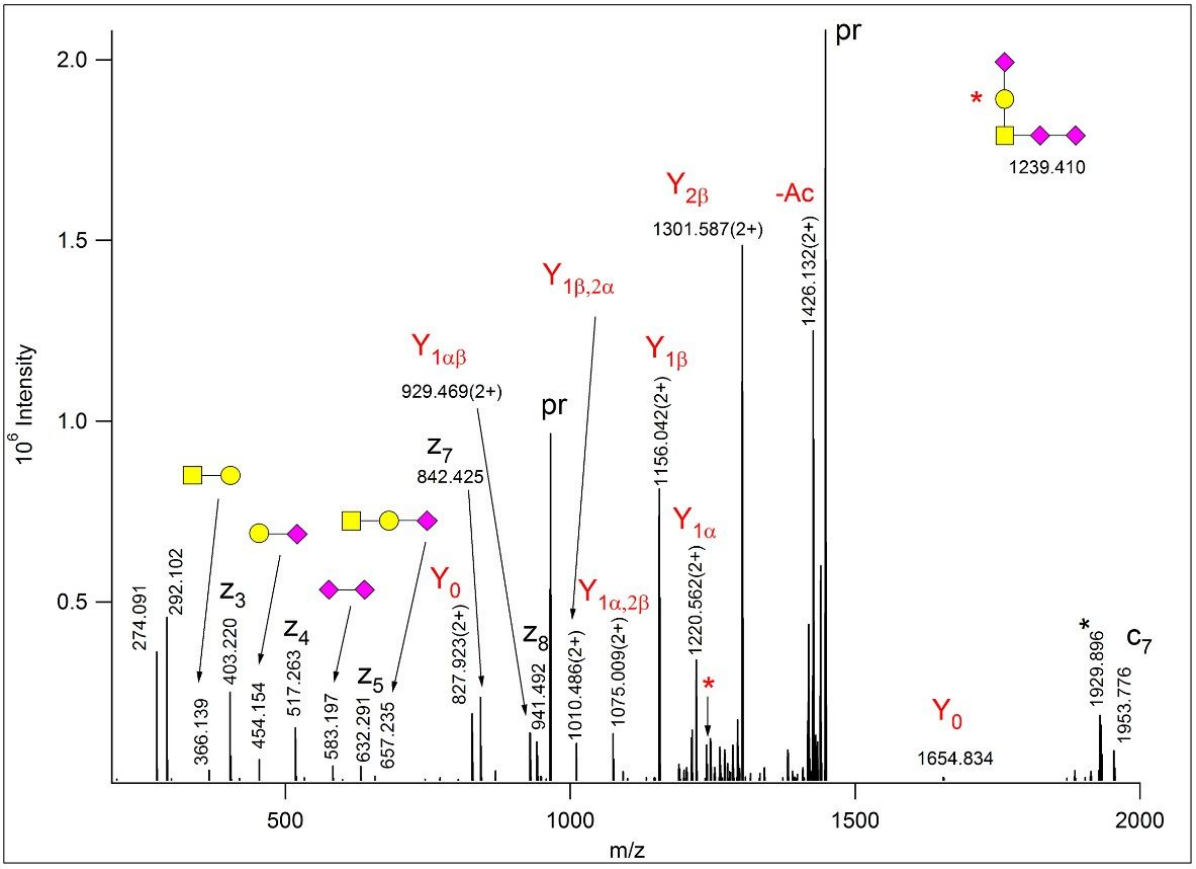


Figure 4. EThcD spectrum of m/z 965.092(3+), identified as ^{93}DVSTPPT(HexNAcHexNeuAc$_3$)VLPDNFPR^{107} of Insulin-like growth factor II. The precursor ion and its charge-reduced form are labeled with 'pr'. The nonreducing end fragments are labeled with cartoons according to the CFG recommendations, the reducing end fragments, printed in red, follow the nomenclature[21]. The inset shows the intact glycan and its oxonium ion is labeled by a red asterisk. The other asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule. For full peaklist see Figure S-10.

Interestingly, if only one Neu5,9Ac$_2$ was present in the glycan structure, it was always Gal-linked in the tetrasaccharides, while in pentasaccharides the GalNAc-linked disialic acid was terminated with it.

Additional unusual glycans

Finally, the database search permitting unspecified modifications delivered two intriguing suggestions. First, it identified a glycopeptide with a glycan mass of 1062.39 Da (Figure 5). This mass corresponds to a $HexNAc_3HexNeuAc$ glycan composition, that is most unusual for an O-glycan (the calculated mass is 1062.386, within 2 ppm). The observed non-reducing end B and reducing end Y carbohydrate fragments suggest that this unusual glycan might be a Galβ3(GalNAcβ4GlcNAcβ6)GalNAc, the LacdiNAc-like structure described by Jin at al.[22], from a human gastric mucin sample. We detected this structure modified with a sialic acid on the galactose (Figure 5, Scheme 2). More specifically, the fragment ion at m/z 407.165 confirms the inter-HexNAc linkage and also indicates that this is a terminal HexNAc-HexNAc structure. Although m/z 407 can, in principle, also be formed as an internal fragment of core 2 glycans, it was not detected in their respective EThcD spectra (see Figures S-7-9). Similarly, m/z 454.154 indicates a terminal HexNeuAc group. These ions together with reducing end Y ions unambiguously show that the glycan is branched and a HexNeuAc and a $HexNAc_2$ are linked to the core GalNAc. The B_3 fragment representing the intact glycan was detected at m/z 1063.398.

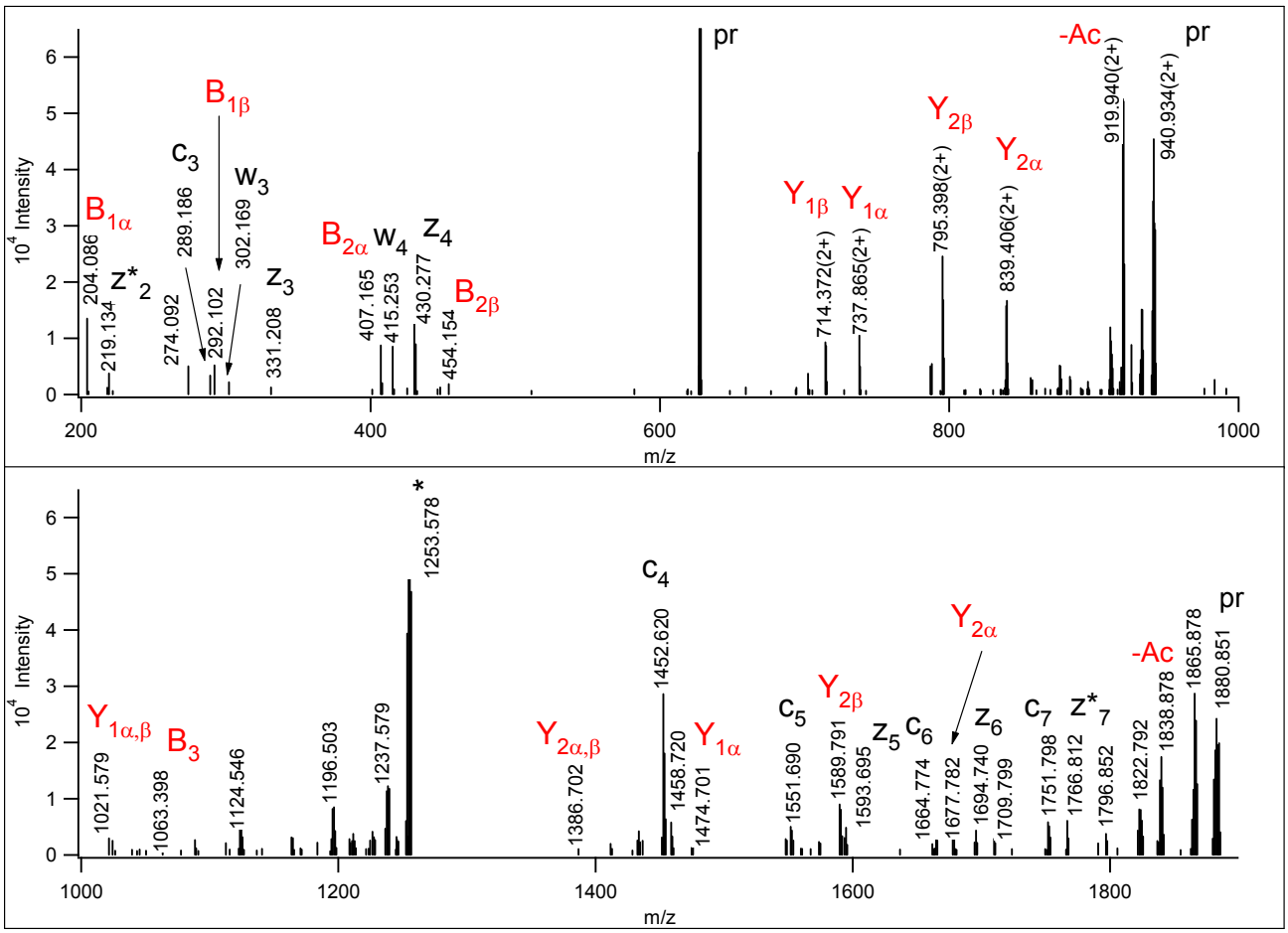
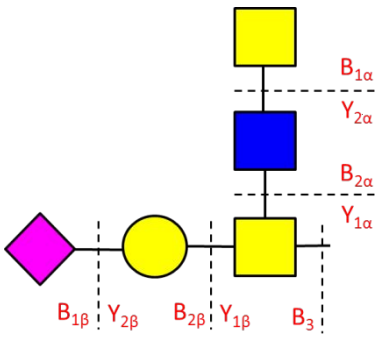


Figure 5. EThcD spectrum of m/z 627.634(3+). The database search permitting unspecified modifications identified as ^{45}VATT(1062.388)VISK^{52} of Plasma protease C1 inhibitor. z* indicates a z+1 peptide fragment. The intact glycan oxonium ion confirms that the peptide is modified by a single oligosaccharide of $HexNAc_3HexNeuAc$ composition. The glycan fragments helped to decipher the branching and linkages within the oligosaccharide, a core 2 based LacdiNAc-like structure[22]. Scheme 2 illustrates which bonds are cleaved. z* indicates a z+1 peptide fragment. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.



Scheme 2. Fragmentation of the glycan structure assigned from the spectrum in Figure 5.

The second fascinating identification was ^{345}SLTVS(1370.497)LGPVSKT(HexNAcHexNeuAc)EGFPK^{361} from Protein HEG homolog 1 (Q9ULI3), where the modification mass, suggested by the search engine, corresponded to a $HexNAc_3Hex_2FucNeuAc$ composition, that could represent a core 2 structure bearing an A-type antigen (see later). However, an abundant oxonium ion at m/z 674.252 ($HexNAcHex_2Fuc$) and the similarly abundant corresponding Y fragment, at m/z 1034.16 (3+) strongly suggested the presence of a core 2 glycan carrying a blood-type B antigen. The assigned modifications and the fragments observed contradicted each other. Our investigation revealed that a series of unassigned doubly charged fragment ions could be identified as z ions (z_7-z_{12}) from the mass differences between them. These fragments' mass shift (697 Da) clearly indicated that the Thr carries a $HexNAc_2NeuAc$ oligosaccharide, most likely a sialylated core 3 structure , i.e. a GlcNAcβ3(NeuAcα6)GalNAc instead of the core 1 trisaccharide. The oxonium ion of this glycan was subsequently identified at m/z 698.266. Thus, the 'leftover' glycan on Ser-349 features a $HexNAc_2Hex_3FucNeuAc$ composition that is indeed a core 2 structure decorated with the B blood-type antigen (Figure 6, Scheme 3). The donor's blood type was AB.

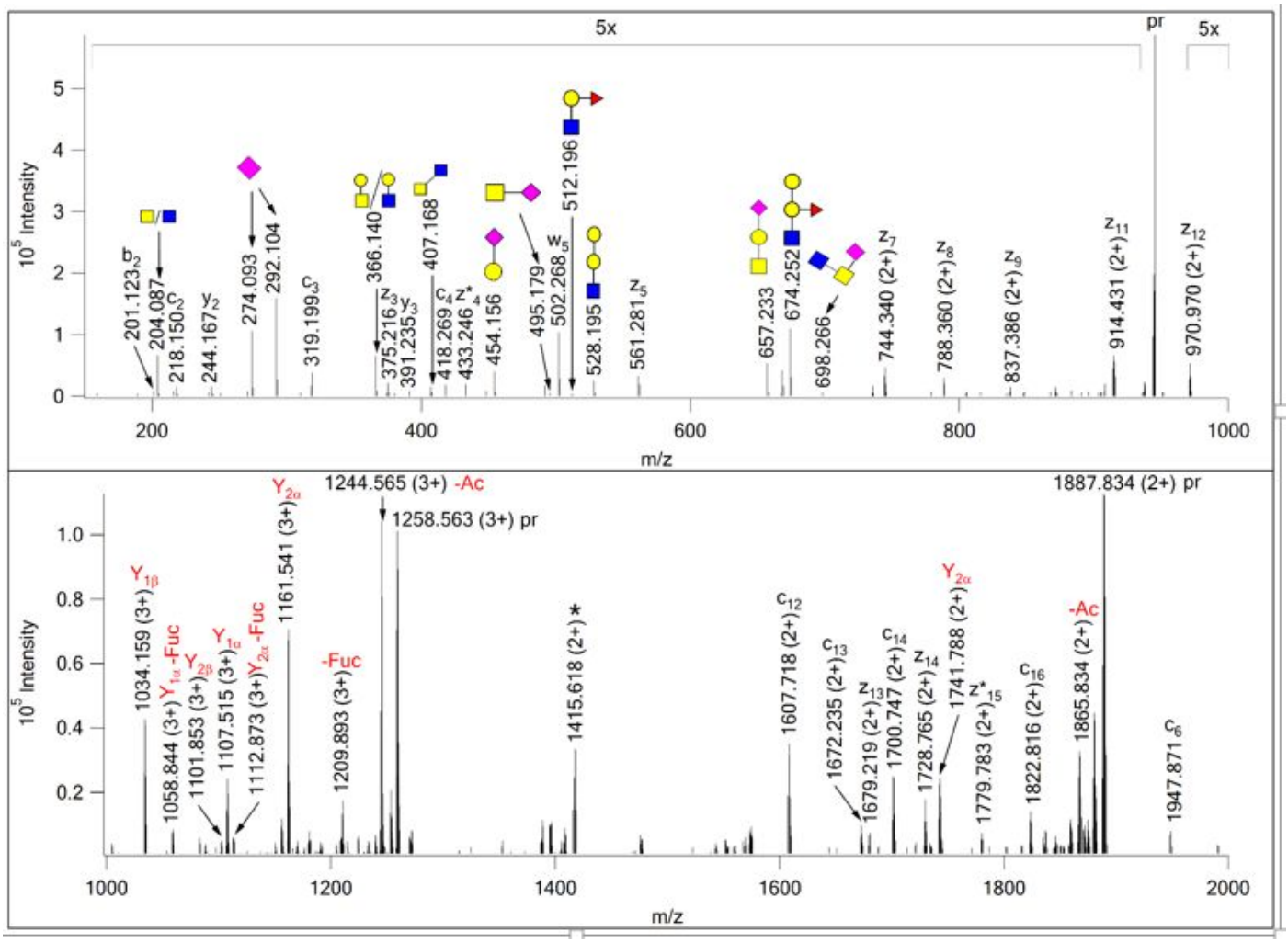
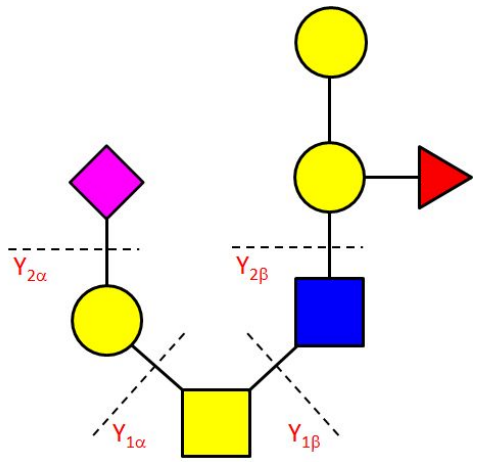


Figure 6. EThcD spectrum of m/z 944.178 (4+) manually assigned as ^{345}SLTVS*LGPVSKT*EGFPK^{361} from Protein HEG homolog 1 (Q9ULI3), with a core 2 structure carrying a B-antigen on Ser-349 and a sialyl core 3, GlcNAcβ3(NeuAcα6)GalNAc on Thr-356. The intact sugar oxonium of this glycan was detected at m/z 698.266. Its modification site was determined from the series of doubly charged z_7-z_{12} fragments. Scheme 3 illustrates bond cleavages within the larger glycan structure.The precursor ion and its charge-reduced form are labeled with 'pr'; the charge-reduced form of a coeluting doubly charged component is indicated with an asterisk. Y ions reflect the fragmentation of the core 2 glycan carrying the B antigen.



Scheme 3. Fragmentation of the mucin core 2 glycan carrying blood-type antigen B structure assigned from the spectrum in Figure 6.

Targeted glycoform identification for a specific peptide

These unusual structures prompted us to further investigate the presence of additional glycoforms for the peptide that already featured the highest microheterogeneity: ^{342}AVAVTLQSH^{250} of Protein YIPF3. Using the MS-filter option of Protein Prospector[13] with one of the datafiles representing the earlier eluting glycopeptides, we identified all HCD spectra that displayed the gas-phase deglycosylated Y_0 fragment (m/z 925.510) among the 5 most abundant ions and with a mass measurement error below 10 ppm. Then the potential glycan masses were calculated for each such component and 4 new glycan structures emerged with compositions of $HexNAc_3Hex_2FucNeuAc$, $HexNAc_2Hex_2FucNeuAc_2$, $HexNAc_2Hex_2FucNeuAcNeuAcAc$ and $HexNAc_2Hex_2Fuc(NeuAcAc)_2$, represented by incremental monoisotopic masses of 1370.499, 1458.514, 1500.516 and 1542.537 Da, respectively (Table S-3). Placing these numbers as modifications within the sequence, and comparing the MS-product predictions (Protein Prospector) with the EThcD data it became clear that Thr-346 was always the glycosylation site.

The first structure is most likely, a core-2 glycan carrying the A blood antigen as shown in Figure 7, Scheme 4. While the identity of the sugar units and their linkage positions obviously could not be determined from these data, the connections between the units could be deciphered from the EThcD spectrum (Figure 7). An oxonium ion at m/z 512.202 and the corresponding Y fragment indicated a terminal HexNAcHexFuc structure. Other reducing end fragments identified a hexose-linked sialic acid as well as a terminal HexNAc.

Since single bond cleavages are preferred in EThcD, the presence of an abundant B-fragment at m/z 657.237 and its reducing end counterpart at m/z 820.399(2+) suggest that the core GalNAc is modified with a HexNAcHexNeuAc structure and the other extension might be the A-antigen because of its composition and terminal HexNAc. These conclusions were further strengthened by the $Y_{2\beta,1\alpha}$ fragment detected in HCD at m/z 1437.715 (Figure S-30). This A-antigen featuring peptide was identified from the same donor as the B-antigen carrying molecule described above.

The other sugar compositions suggest the presence of Sialyl Lewis$^{X/A}$ in different stages of O-acetylation (See Table S-3; Figures S-31-34). The isomeric structures containing one O-acetyl-sialic acid were chromatographically resolved, and the earlier eluting isoform carried this modification on the GlcNAc-containing arm, just like in the hexasaccharide isomers presented in Figure 2. Interestingly, all fucose-containing glycoforms produced a few ions that indicated fucose transfer between the extensions (Figures 6-7, Figures S-31-35). Fucose 'migration' has been reported in glycan analyses[23] as well as for N-linked glycopeptides[24,25].

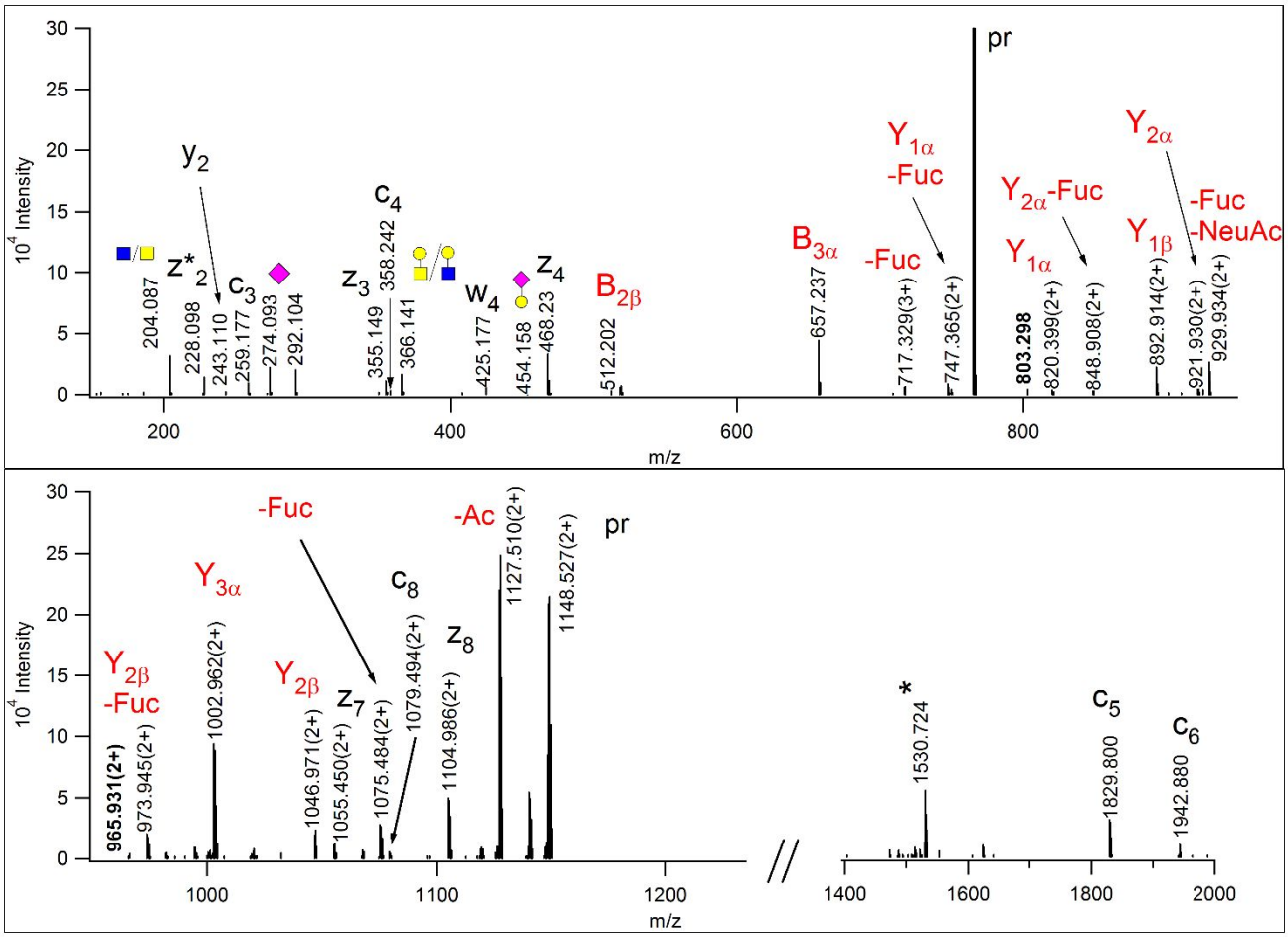
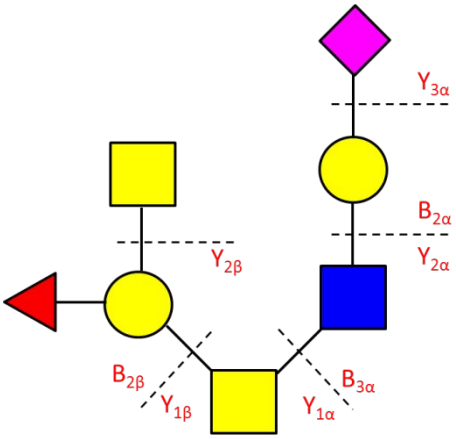


Figure 7. EThcD spectrum of m/z 766.008(3+) that was manually assigned as ^{342}AVAVTLQSH^{350} of Protein YIPF3 bearing a $HexNAc_3Hex_2FucNeuAc$ glycan on Thr-346. The precursor ion and its charge-reduced form are labeled with 'pr'; the charge-reduced form of a coeluting doubly charged component is indicated with an asterisk. The glycan was assigned as a core 2 glycan featuring the A antigen linked to the GalNAc. A similar core 2 A antigen has been described on gastric mucin by Rossez *et al*.[26] Scheme 4 illustrates which bonds are cleaved. m/z values printed in bold label fragments that indicate fucose transfer between the antennae: m/z 803.298 represents a HexNAcHexFucNeuAc structure, while m/z 965.931(2+) was formed *via* HexNAcHex loss from the precursor ion.



Scheme 4. Fragmentation of the glycan structure assigned from the spectrum in Figure 7.

In a similar attempt as described above, an additional glycoform of this peptide bearing a more extended sugar structure ($HexNAc_3Hex_3FucNeuAc_2$, mass increment: 1823.65 Da), representing an extended core 2 glycan decorated with Sialyl Lewis$^{X/A}$ was identified in a later eluting glycopeptide fraction (Figure S-35, Table S-3).

DISCUSSION

In order to understand the biological role(s) of O-glycosylation addressing the site-specificity of the modifications is necessary. However, the enrichment and characterization of O-glycopeptides is an even more daunting task than deciphering N-glycosylation[10].

One tends to believe that mucin-type O-glycosylation operates with only a handful of structures. For example, serum proteins feature primarily mono- and disialo core 1 glycans[27]. However, the glycan repertoire is much more diverse for gastric mucins[22,26], and as presented here the urinary glycoproteome is much more complex as well. Our results also indicate that removing the sialic acids indeed may make the investigated mixtures simpler, but at a significant price i.e. losing valuable information on microheterogeneity. Peptides featuring multiple potential glycosylation sites may display an overwhelming degree of macro- and microheterogeneity. For example, the complexity generated by macroheterogeneity could be illustrated by the glycoforms reported for the ^{154}Ser-Lys^{175} region of bovine insulin-like growth factor II[28] , while for microheterogeneity we have just presented the ^{342}Ala-His^{350} stretch of Protein YIPF3 (its 'normal' disialo core 1, and core 2 glycoforms were described in our earlier report[12]). Obviously, this

heterogeneity makes the chromatographic separation of the different glycoforms harder, and may also prevent the detection of minor components. Differentiation between the isomeric structures as well as positional isomers presents an even greater challenge.

As briefly summarized in the introduction, the different MS/MS activation techniques deliver different clues to solve the glycopeptide puzzle. Unfortunately, these data are rarely used in concert, and primarily via manual data interpretation. Perhaps eventually a software tool will be developed that can utilize orthogonal data sources, but until then EThcD, yielding both glycan and peptide fragments is the best approach for identification. For example, abundant oxonium ions indicate the presence of the different neuraminic acids. The diagnostic value of such Neu5,9Ac_2 fragments has been reported for beam-type CID data[13], here similar ions revealed the presence of a Neu4,5,9Ac_3 moiety. Oxonium ions of sialic acids are very abundant in the EThcD spectra[12], and in complex mixtures numerous spectra may display them due to precursor ion interference. Thus, they have to be considered with caution as unambiguous evidence for sialylation. However, the lack of these diagnostic oxonium ions clearly signals that no glycan featuring the corresponding sialic acid might be present. Thus, screening the peaklists for such diagnostic fragments could be recommended prior to database searches, or PSMs assigned to sialic acid-containing structures should be considered false whenever the diagnostic oxonium ions were not observed.

In an earlier publication beam-type CID (HCD) spectra of murine nucleobindin-1 glycopeptides illustrated that internal glycan fragments representing the core GalNAc modified with N-acetyl, N-glycolyl, or N,O-diacetylneuraminic acid helped assign the different versions of the disialo-core 1 tetrasaccharides[13]. The HCD data acquired in this study (Figures S-1-35) are in good agreement with these observations. However, while the internal oxonium ion is the dominant species in HCD, the terminal B_2 fragment produced by a single bond cleavage is significantly more abundant in the EThcD spectra. With the mild EThcD activation the products of single bond cleavages are detected preferentially, even if internal fragments may be observed. As described, the glycan fragments help to determine the direct linkages between the sugar units. In fact, the occurrence of B_2 ions representing disialic acids is so high that they might be used as reporter ions for the presence of these disaccharide units, and can be used for prefiltering the peaklists when

searching for glycopeptides bearing such glycans. While most such glycoforms will be identified in general database searches as well, the FDR rate is higher, and thus, lower scoring true identifications may 'sink' into the background[12]. Obviously, these fragments also should be considered as strong evidence in glycopeptide fragmentation scoring. Glycan fragmentation in EThcD also may permit the correct assignment of isomeric structures as presented in this manuscript. It would be beneficial if these observations were incorporated into the automated evaluation of glycopeptide EThcD spectra. The gentler activation also enables, not infrequently, the survival of the oxonium ions representing even penta- or hexasaccharides. The presence of these fragments can help to differentiate between the different glycans or glycan combinations that represent identical mass additions. As these fragment ions are often rather weak, it is not very likely that precursor ion interference could 'inject' them into the spectra. Considering them as positive evidence for the presence of larger glycans is recommended.

Manual investigation of the spectra also drew our attention to a peak picking issue. Here we do not refer to the misidentification of the monoisotopic mass in the precursor ion cluster – although this is still a persistent problem. When evaluating the raw spectra there were often additional peaks that could provide glycan structural or site localization information but were not considered by the search engine. Some of these were due to incorrect determination of a fragment monoisotopic peak or removal of a peak during deisotoping, and others were removed by an intensity-based filter. This is likely a result of search engines being optimized for peptide identification; as adding in lower intensity peaks, several of which are background ions, can reduce statistical significance of the peptide assignment.

As for the biological significance of our findings, presently we cannot offer any convincing hypothesis. Polysialic acid has been found both in N- and O-glycans, and depending on the degree of polymerization can be categorized as di-, oligo- and polysialic acids[29]. The presence of such structures is usually determined using immunochemistry. While polysialylation clearly has a role in neural development[30], it has also been implied in cancer metastasis[31]. On the other hand, although disialic acid units have been detected on multiple glycoproteins[32-36], no clear biological role has been linked to this structure yet. The proteins identified in the present report also participate in diverse biological processes including cell adhesion (Macrophage colony-

stimulating factor 1, fractalkine, CD 99), glucose metabolism (insulin like growth factor II) or maintaining the structure of the cell membrane (basement membrane-specific heparan sulfate proteoglycan core protein) or the Golgi (YIPF3).

Nonetheless, the sialyltransferases catalyzing the formation of disialic acids are known. ST8Sia VI is responsible for disialic acid formation in O-glycans catalyzing the formation of both NeuAcα2,8NeuAcα2,3Gal and NeuAcα2,8NeuAcα2,6GalNAc structures[37,38] with higher preference towards the NeuAcα2,6GalNAc substrate. Our findings are in good agreement with the above observation, disialic acid was linked to the core GalNAc in the glycopeptides characterized. In addition, ST8Sia VI is expressed at higher levels in kidney[37], thus, perhaps urinary proteins may feature more of these structures. As far as we know this is the only report where site-specific information was obtained on multiple proteins for the disialic acid carrying O-glycopeptides. The only other study describing a similar pentasaccharide structure was reporting on amyloid peptides isolated from human cerebrospinal fluid. However, in that case the pentasaccharide modified a Tyr residue, and MS/MS data indicated that the disialic acid unit was linked to the galactose[3].

O-acetylation of sialic acid plays a role in cell-cell interactions and non-immune protection of mucosa[39]. It was suggested that O-acetylation of neuraminic acids on cell surfaces "may protect against pathogen infection by preventing degradation of membrane-associated mucins"[40]. At the same time another report links the expression of O-acetyl sialoglycoproteins on erythrocytes to leukemia[41]. None of these studies characterized the O-acetylneuraminic acid-containing glycan structures, or pinpointed their location within any particular protein or residue.

Finally, the presence of blood-type antigens has been documented on gastric mucins[26], but as far as we are aware this is the first time that a glycopeptide is presented bearing such a structure.

CONCLUSION

We demonstrated that EThcD with gentle supplemental activation (NCE 15%) may deliver quite comprehensive information for O-glycopeptide structure elucidation. In this process we extended the sialoglycan repertoire of urinary glycoproteins. We unambiguously established the presence of a variety of,

never reported, mucin-type structures with O-acetyl-sialic acid(s), disialic acids or both. In addition, we also identified a trisialo core 1 glycan with O-diacetylsialic acid, and characterized a series of unusual core-2 structures, including ones displaying a blood-type antigen.

We observed that the detected glycan oxonium ions were mostly B fragments, and the Y fragments also indicated some preference for single bond cleavages. We identified some diagnostic oxonium ions that may not unambiguously prove the presence of certain structures because of the precursor ion interference in such complex mixtures, but their absence is determinative. Larger fragments representing intact penta- and hexasaccharides were also detected. We believe these should be more positively considered during scoring. These observations should lead to improved automated glycopeptide assignments using EThcD data and the resulting methods will provide tools aimed at elucidating this complex yet understudied class of protein post-translational modifications.

ASSOCIATED CONTENT

Supporting Information

The following files are available free of charge:

Table S-1: Glycan compositions considered in database search; Table S-2: List of glycopeptides with unusual sialyl structures, identified from human urinary glycoproteins; Table S-3: Additional glycan structures deciphered from EThcD and HCD data manually; Figures S-1-29: EThcD data and the corresponding HCD spectra for the glycopeptides listed in Table 1 and Table S-2; Figures S-30-35: Supporting MS/MS data for the fucosyl glycoforms of Thr-346 of Protein YIPF3.

AUTHOR INFORMATION

Corresponding Authors

*Email: darula.zsuzsanna@brc.mta.hu, Tel/Fax: +36 62 599 773

*Email: medzihradszky.katalin@brc.mta.hu, Tel/Fax: +36 62 599 773

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

The authors wish to thank Robert Chalkley for his constructive comments on the discussion section, and Kirk Hansen for correcting our Hunglish mistakes. We thank the MTA Cloud (https://cloud.mta.hu/) for housing our Protein Prospector server. This work was supported by the following grants: New Szechenyi Plan GOP-1.1.1-11-2012-0452, and the Economic Development and Innovation Operative Programmes GINOP-2.3.2–15-2016–00001 and GINOP-2.3.2–15-2016–00020 from the Ministry for National Economy.

REFERENCES

(1) Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H.; editors. *Essentials of Glycobiology*. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.

(2) Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol*. **1998**, *33*, 151- 8.

(3) Halim, A.; Brinkmalm, G.; Rüetschi, U.; Westman-Brinkmalm, A.; Portelius, E.; Zetterberg, H.; Blennow, K.; Larson, G.; Nilsson, J. Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc. Natl. Acad. Sci. U S A*. **2011**, *108*, 11848-11853.

(4) Haslam, S. M.; North, S. J.; Dell, A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr. Opin. Struct. Biol*. **2006**, *16*, 584-591.

(5) Karlsson, H.; Larsson, J. M.; Thomsson, K. A.; Härd, I.; Bäckström, M.; Hansson G, C. High-throughput and high-sensitivity nano-LC/MS and MS/MS for O-glycan profiling. *Methods Mol. Biol.* **2009**, *534*, 117-131.

(6) Wada, Y.; Dell, A.; Haslam, S. M.; Tissot, B.; Canis, K.; Azadi, P.; Bäckström, M.; Costello, C. E.; Hansson, G. C.; Hiki, Y.; Ishihara, M.; Ito, H.; Kakehi, K.; Karlsson, N.; Hayes, C. E.; Kato, K.; Kawasaki, N.; Khoo, K. H.; Kobayashi, K.; Kolarich, D.; Kondo, A.; Lebrilla, C.; Nakano, M.; Narimatsu, H.; Novak, J.; Novotny, M. V.; Ohno, E.; Packer, N. H.; Palaima, E.; Renfrow, M. B.; Tajiri, M.; Thomsson, K. A.; Yagi, H.; Yu, S. Y.; Taniguchi, N. Comparison of methods for profiling

O-glycosylation: Human Proteome Organisation Human Disease Glycomics/Proteome Initiative multi-institutional study of IgA1. *Mol. Cell. Proteomics* **2010**, *9*, 719-727.

(7) Zauner, G.; Kozak, R. P.; Gardner, R. A.; Fernandes, D. L.; Deelder, A. M.; Wuhrer, M. Protein O-glycosylation analysis. *Biol. Chem*. **2012**, *393*, 687-708.

(8) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U S A.* **2004**, *101*, 9528-9533.

(9) Frese, C. K.; Altelaar, A. F.; van den Toorn, H,; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem*. **2012**, *84*, 9668-9673.

(10) Darula, Z.; Medzihradszky, K. F. Analysis of Mammalian O-Glycopeptides-We Have Made a Good Start, but There is a Long Way to Go. *Mol. Cell. Proteomics* **2018**, *17*, 2-17.

(11) Zhokhov, S. S.; Kovalyov, S. V.; Samgina, T. Y.; Lebedev, A. T. An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1600-1611.

(12) Pap, A.; Klement, E.; Hunyadi-Gulyas, E.; Darula, Z.; Medzihradszky, K. F. Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. *J. Am. Soc. Mass Spectrom*. **2018**, doi: 10.1007/s13361-018-1945-7.

(13) Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J. Characterizing sialic acid variants at the glycopeptide level. *Anal. Chem*. **2015**, *87*, 3064-3071.

(14) Baker, P. R.; Trinidad, J. C.; Chalkley, R. J. Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics* **2011**, doi: 10.1074/mcp.M111.008078.

(15) Halim, A.; Nilsson, J.; Rüetschi, U.; Hesse, C.; Larson, G. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol. Cell. Proteomics* **2012**, 11:M111.013649.

(16) Miura ,Y.; Kato, K.; Takegawa, Y.; Kurogochi, M.; Furukawa, J.; Shinohara, Y.; Nagahori, N.; Amano, M.;, Hinou, H.; Nishimura, S. Glycoblotting-assisted O-glycomics: ammonium carbamate allows for highly efficient o-glycan release from glycoproteins. *Anal. Chem*. **2010**, *82,* 10021-10029.

(17) Takeuchi, M.; Amano,M.; Kitamura,H.; Tsukamoto, T.; Masumori, N.; Hirose, K.; Ohashi, T.; Nishimura, S.I. N- and O-glycome analysis of serum and urine from bladder cancer patients using a high-throughput glycoblotting method. *J. Glycomics Lipidomics* **2013**, *3,* 1-8.

(18) Trinidad, J. C.; Schoepfer, R.; Burlingame, A. L.; Medzihradszky, K. F. N- and O-glycosylation in the murine synaptosome. *Mol. Cell. Proteomics* **2013**, *12*, 3474-3488.

(19) Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J. Tissue-Specific Glycosylation at the Glycopeptide Level. *Mol. Cell. Proteomics* **2015**, *14*, 2103-2110.

(20) Chang, C. F.; Pan, J. F.; Lin, C. N.; Wu, I. L.; Wong, C. H.; Lin, C. H. Rapid characterization of sugar-binding specificity by in-solution proximity binding with photosensitizers. *Glycobiology* **2011**, *21,* 895–902.

(21) Domon, B.; Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988**, *5*, 397-409.

(22) Jin, C.; Kenny, D. T.; Skoog, E. C.; Padra, M.; Adamczyk, B.; Vitizeva, V.; Thorell, A.; Venkatakrishnan, V.; Lindén, S. K.; Karlsson, N. G. Structural Diversity of Human Gastric Mucin Glycans. *Mol. Cell. Proteomics* **2017**, *16*, 743-758.

(23) Wuhrer, M.; Deelder, A. M.; van der Burgt, Y. E. Mass spectrometric glycan rearrangements. *Mass Spectrom Rev.* **2011**, *30*, 664-680.

(24) Wuhrer, M.; Balog, C. I.; Catalina, M. I.; Jones, F. M.; Schramm, G.; Haas, H.; Doenhoff, M. J.; Dunne D. W.; Deelder, A. M.; Hokke, C. H. IPSE/alpha-1, a major secretory glycoprotein antigen from schistosome eggs, expresses the Lewis X motif on core-difucosylated N-glycans. *FEBS J.* **2006**, *273*, 2276-2292.

(25) Hashii, N.; Kawasaki, N.; Itoh, S.; Nakajima, Y.; Harazono, A.; Kawanishi, T; Yamaguchi, T. Identification of glycoproteins carrying a target glycan-motif by liquid chromatography/multiple-stage mass spectrometry: identification of Lewis x-conjugated glycoproteins in mouse kidney. *J Proteome Res.* **2009**, *8*, 3415-3429.

(26) Rossez, Y.; Maes, E.; Lefebvre Darroman, T.; Gosset, P.; Ecobichon, C.; Joncquel Chevalier Curt, M.; Boneca, I. G.; Michalski, J. C.; Robbe-Masselot, C. Almost all human gastric mucin O-glycans

harbor blood group A, B or H antigens and are potential binding sites for Helicobacter pylori. *Glycobiology*. **2012**, *22*, 1193-1206.

(27) Yabu, M.; Korekane, H.; Miyamoto, Y. Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology* **2014**, *24*, 542–553.

(28) Darula, Z.; Sherman, J.; Medzihradszky, K.F. How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* **2012**, doi: 10.1074/mcp.O111.016774.

(29) Sato, C.; Kitajima, K. Disialic, oligosialic and polysialic acids: distribution, functions and related disease. *J. Biochem*. **2013**, *154*, 115-136.

(30) Hildebrandt, H; Dityatev, A. Polysialic Acid in Brain Development and Synaptic Plasticity. *Top. Curr. Chem.* **2015**, *366*, 55-96.

(31) Martersteck, C. M.; Kedersha, N. L.; Drapp, D. A.; Tsui, T. G.; Colley, K. J. Unique alpha 2, 8-polysialylated glycoproteins in breast cancer and leukemia cells. *Glycobiology* **1996**, *6*, 289-301.

(32) Kiang, W. L.; Krusius, T.; Finne, J.; Margolis, R. U.; Margolis, R. K. Glycoproteins and proteoglycans of the chromaffin granule matrix. *J. Biol. Chem*. **1982** 257, 1651-1659.

(33) Fukuda, M.; Lauffenburger, M.; Sasaki, H; Rogers, M. E.; Dell, A. Structures of novel sialylated O-linked oligosaccharides isolated from human erythrocyte glycophorins. *J. Biol. Chem.* **1987**, *262*, 11952-11957.

(34) Kitajima, K.; Sato, C.; Honda, N.; Matsuda, T.; Hara-Yokoyama, M.; Close, B.E.; and Colley, K.J. Occurrence of alpha2,8-linked oligosialic acid residues in mammalian glycoproteins. In Inoue, Y., Lee, Y.C., and Troy, F.A. (eds), Sialobiology and Other Novel Form of Glycosylation, A Monograph. Gakushin Syuppan, Osaka, Japan, **1999**, 69–76.

(35) Sato, C.; Yasukawa, Z.; Honda, N.; Matsuda, T.; Kitajima, K. Identification and adipocyte differentiation-dependent expression of the unique disialic acid residue in an adipose tissue-specific glycoprotein, adipo Q. *J. Biol. Chem.* **2001**, *276*, 28849-28856.

(36) Yasukawa, Z.; Sato, C.; Sano, K.; Ogawa, H.; Kitajima, K. Identification of disialic acid-containing glycoproteins in mouse serum: a novel modification of immunoglobulin light chains, vitronectin, and plasminogen. *Glycobiology* **2006**, *16*, 651-665.

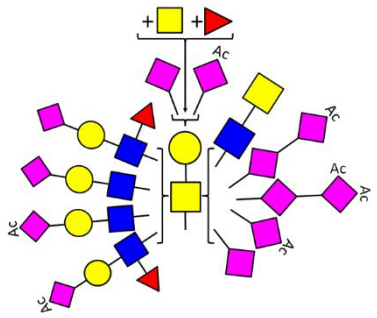
(37) Takashima, S.; Ishida, H. K.; Inazu, T.; Ando, T.; Ishida, H.; Kiso, M.; Tsuji, S.; Tsujimoto, M. Molecular cloning and expression of a sixth type of alpha 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans. *J. Biol. Chem*. **2002**, *277*, 24030-24080.

(38) Teintenier-Lelièvre, M.; Julien, S.; Juliant, S.; Guerardel, Y.; Duonor-Cérutti, M.; Delannoy, P.; Harduin-Lepers, A. Molecular cloning and expression of a human hST8Sia VI (alpha2,8-sialyltransferase) responsible for the synthesis of the diSia motif on O-glycosylproteins. *Biochem. J.* **2005**, *392*, 665-674.

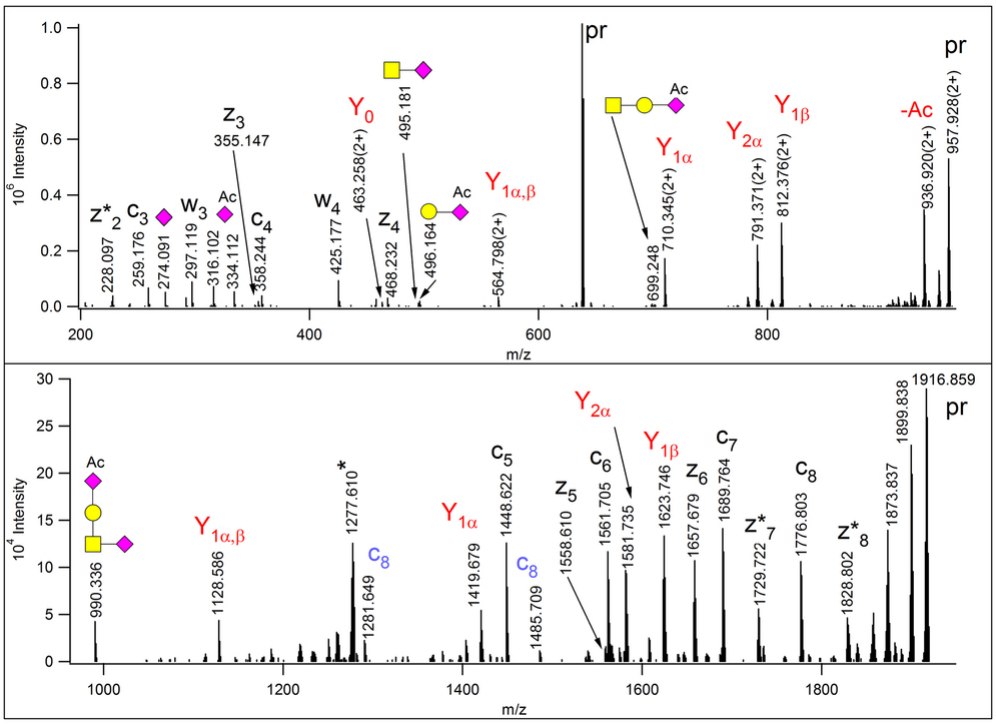
(39) Klein, A.; Roussel, P. O-acetylation of sialic acids. *Biochimie* **1998**, *80*, 49-57.

(40) Argüeso, P.; Sumiyoshi, M. Characterization of a carbohydrate epitope defined by the monoclonal antibody H185: sialic acid O-acetylation on epithelial cell-surface mucins. *Glycobiology* **2006**, *16*, 1219-1228.

(41) Sen, G.; Chowdhury, M.; Mandal, C. O-acetylated sialic acid as a distinct marker for differentiation between several leukemia erythrocytes. *Mol. Cell. Biochem*. **1994**, *136*, 65-70.

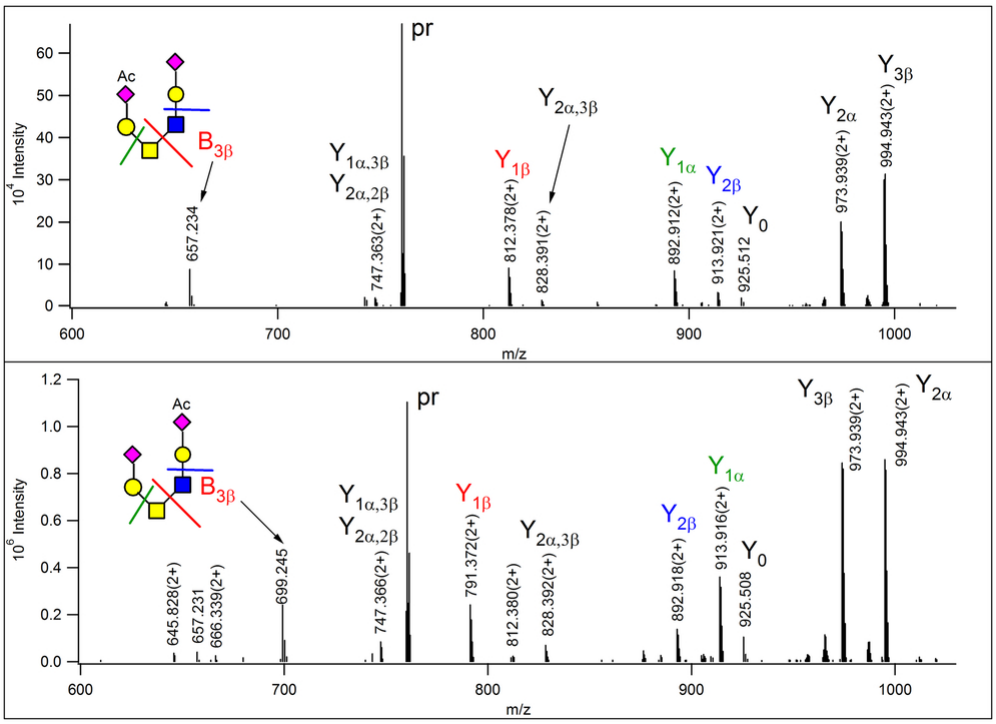


For TOC only



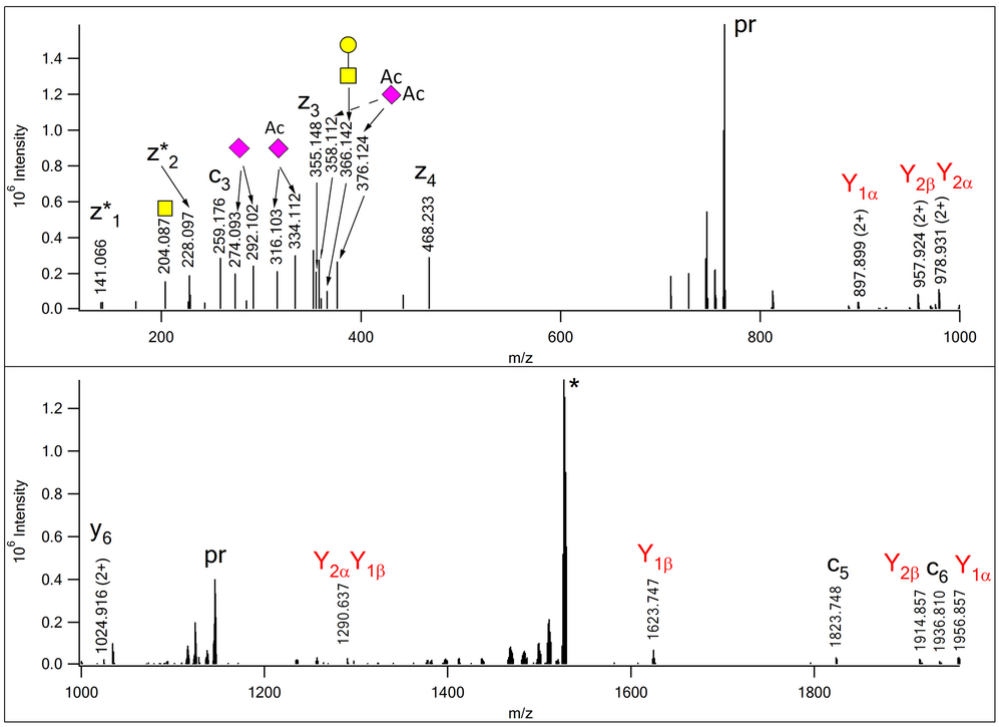
EThcD spectrum of m/z 638.954(3+), identified as 342AVAVT(HexNAcHexNeuAcNeuAcAc)LQSH350 of Protein YIPF3. The Gal of the mucin-type core 1 structure is capped with a Neu5,9Ac2 moiety. (This sugar is listed as NeuAcAc in the glycan library for database searches). Oxonium and related ions are labeled according to the CFG recommendations, for the reducing end fragments – printed in red - the Domon-Costello nomenclature is followed21. The sialic acid loss from the core GalNAc yields the Y1β fragment. z+1 peptide ions are distinguished with asterisks. Peptide fragments in blue indicate that some glycan fragmentation occurred in EThcD, the loss of the NeuAc and GalNeu5,9Ac2 from the c8 fragment, respectively. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.

84x61mm (300 x 300 DPI)



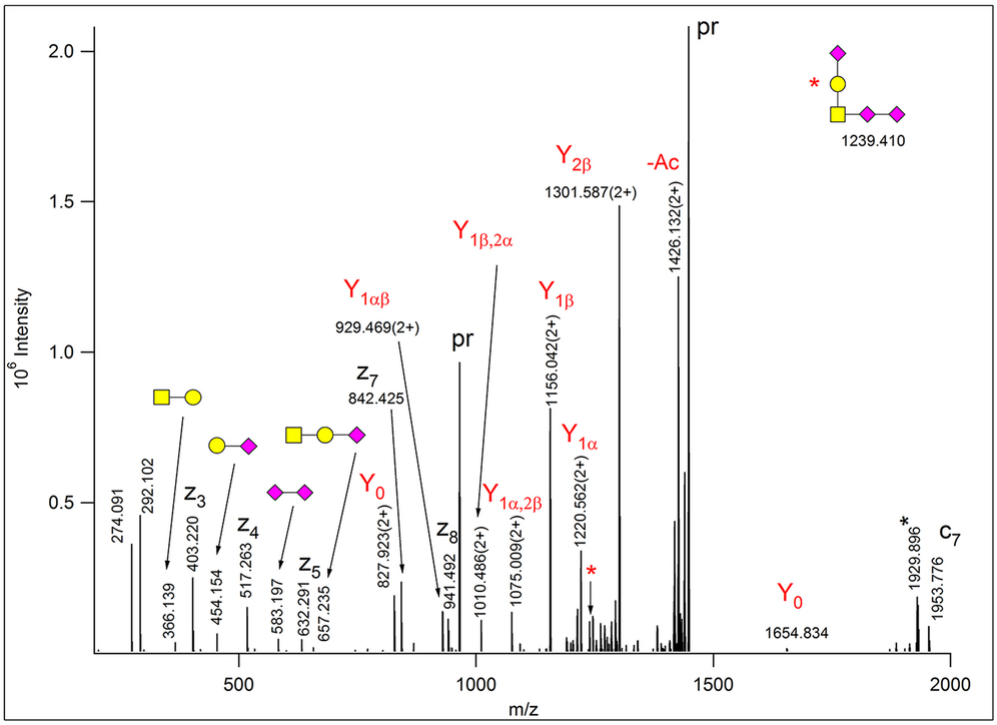
Part of the EThcD spectra acquired from m/z 760.664(3+) at RT=14.95min (Lower panel) and at RT=18.29min (Upper panel). Both spectra were assigned as 342AVAVT(HexNAc2Hex2NeuAcNeuAcAc)LQSH350 of Protein YIPF3 (for full spectra see Figures S-7 & S-8). (NeuAcAc stands for Neu5,9Ac2 in the glycan library for database searches.) However, the glycan fragments revealed isomeric glycoforms. The diagnostic fragment ions and the underlying cleavages are indicated in matching colors.

84x61mm (300 x 300 DPI)



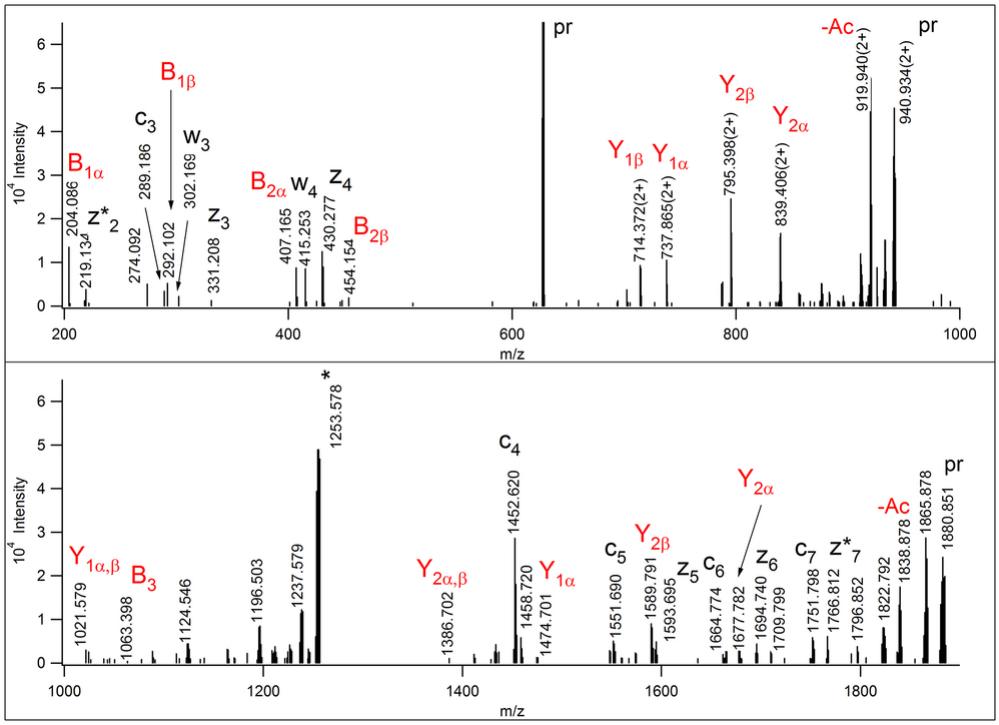
EThcD spectrum of m/z 763.992(3+) manually assigned as 342AVAVTLQSH350, modified at Thr-346 with a core 1 glycan that features a Neu5,9Ac2 on the Gal residue, and a NeuAc-Neu4,5,9Ac3 disialo-unit on the core GalNAc. 'Pr' indicates the different forms of the precursor ion. z+1 peptide fragments are distinguished with asterisks. The oxonium ions identifying the three different sialic acids are labeled with the CFG symbols, for the reducing end fragments – printed in red - the Domon-Costello nomenclature is followed21. The GalNAc modifications were considered the β-arm. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.

84x61mm (300 x 300 DPI)



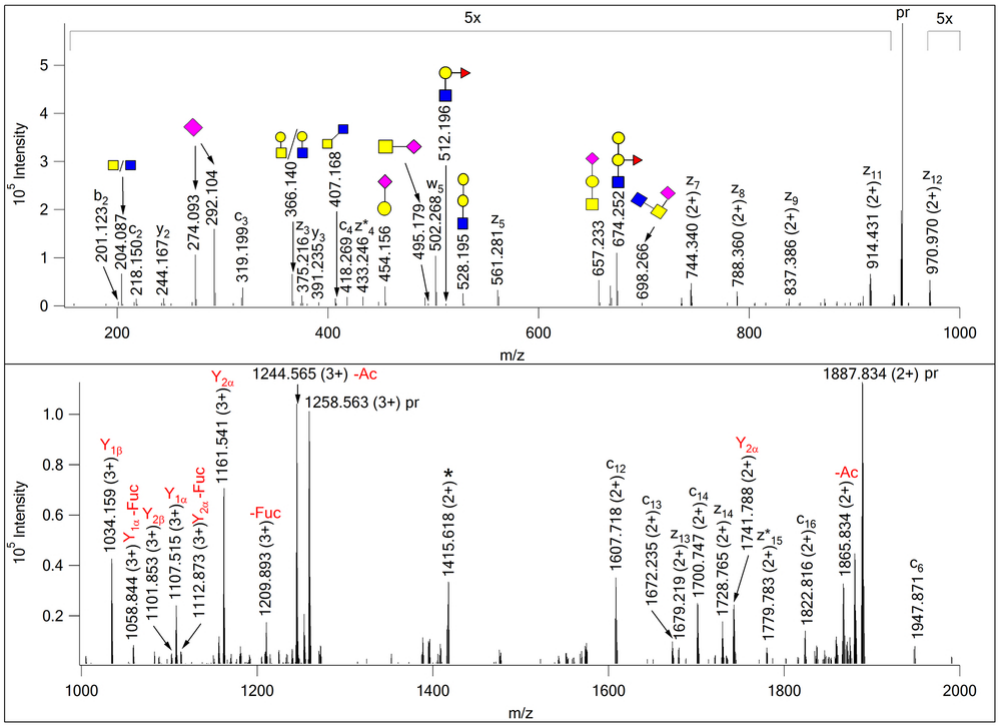
EThcD spectrum of m/z 965.092(3+), identified as 93DVSTPPT(HexNAcHexNeuAc3)VLPDNFPR107 of Insulin-like growth factor II. The precursor ion and its charge-reduced form are labeled with 'pr'. The nonreducing end fragments are labeled with cartoons according to the CFG recommendations, the reducing end fragments, printed in red, follow the nomenclature21. The inset shows the intact glycan and its oxonium ion is labeled by a red asterisk. The other asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule. For full peaklist see Figure S-10.

84x61mm (300 x 300 DPI)



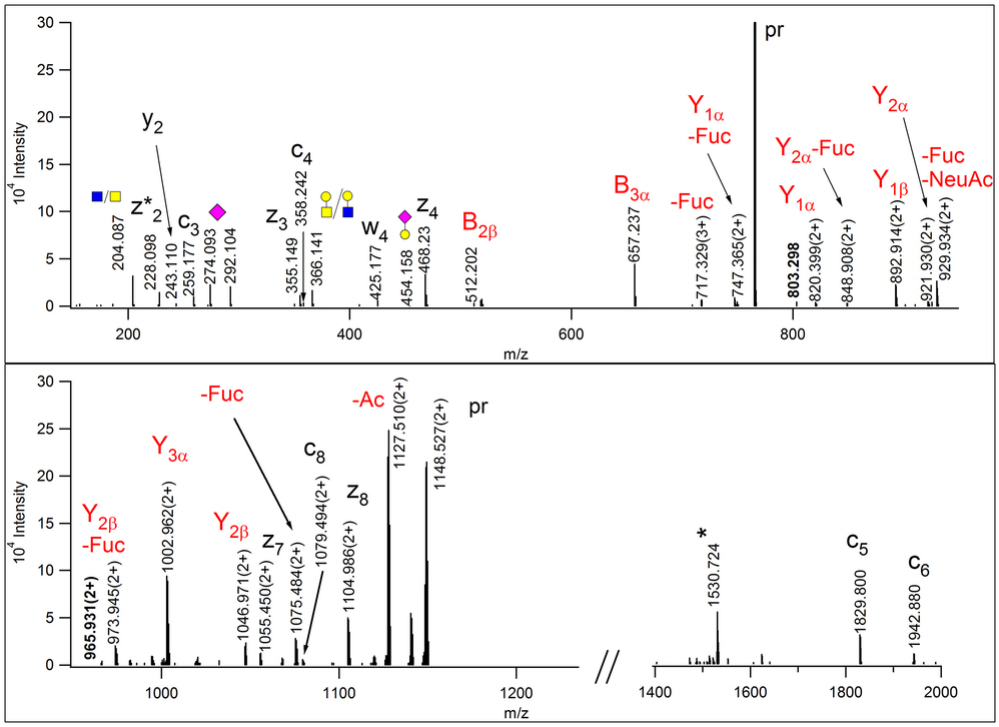
EThcD spectrum of m/z 627.634(3+). The database search permitting unspecified modifications identified as 45VATT(1062.388)VISK52 of Plasma protease C1 inhibitor. z* indicates a z+1 peptide fragment. The intact glycan oxonium ion confirms that the peptide is modified by a single oligosaccharide of HexNAc3HexNeuAc composition. The glycan fragments helped to decipher the branching and linkages within the oligosaccharide, a core 2 based LacdiNAc-like structure22. Scheme 2 illustrates which bonds are cleaved. z* indicates a z+1 peptide fragment. The asterisk-labeled ion is the charge-reduced form of a doubly charged coeluting molecule.

84x61mm (300 x 300 DPI)



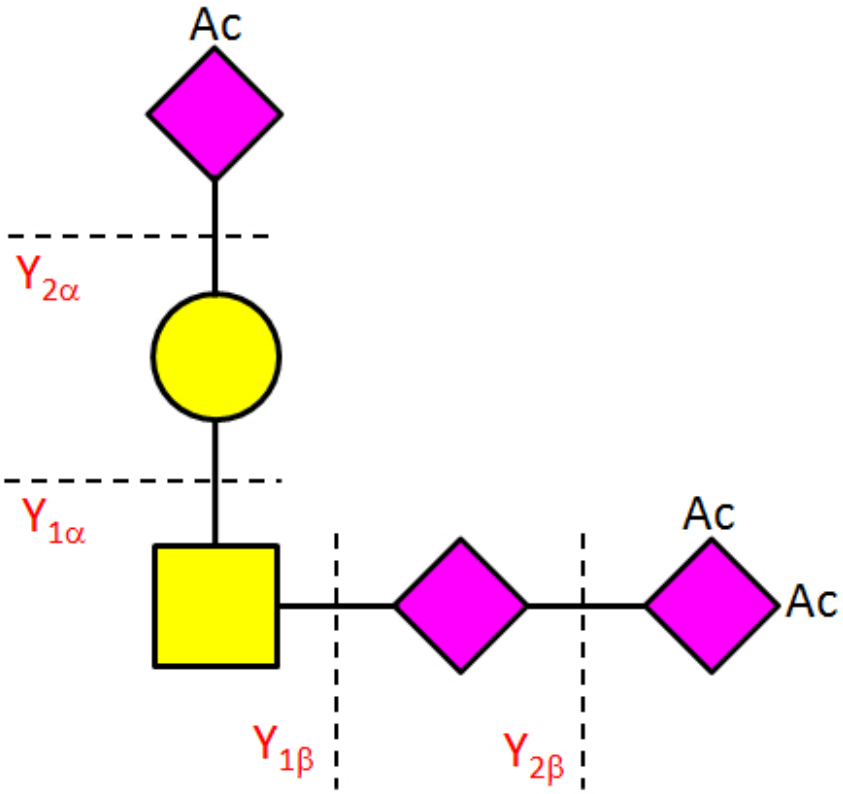
EThcD spectrum of m/z 944.178 (4+) manually assigned as 345SLTVS*LGPVSKT*EGFPK361 from Protein HEG homolog 1 (Q9ULI3), with a core 2 structure carrying a B-antigen on Ser-349 and a sialyl core 3, GlcNAc□3(NeuAc□□)GalNAc on Thr-356. The intact sugar oxonium of this glycan was detected at m/z 698.266. Its modification site was determined from the series of doubly charged z7-z12 fragments. Scheme 3 illustrates bond cleavages within the larger glycan structure.The precursor ion and its charge-reduced form are labeled with 'pr'; the charge-reduced form of a coeluting doubly charged component is indicated with an asterisk. Y ions reflect the fragmentation of the core 2 glycan carrying the B antigen.

84x61mm (300 x 300 DPI)



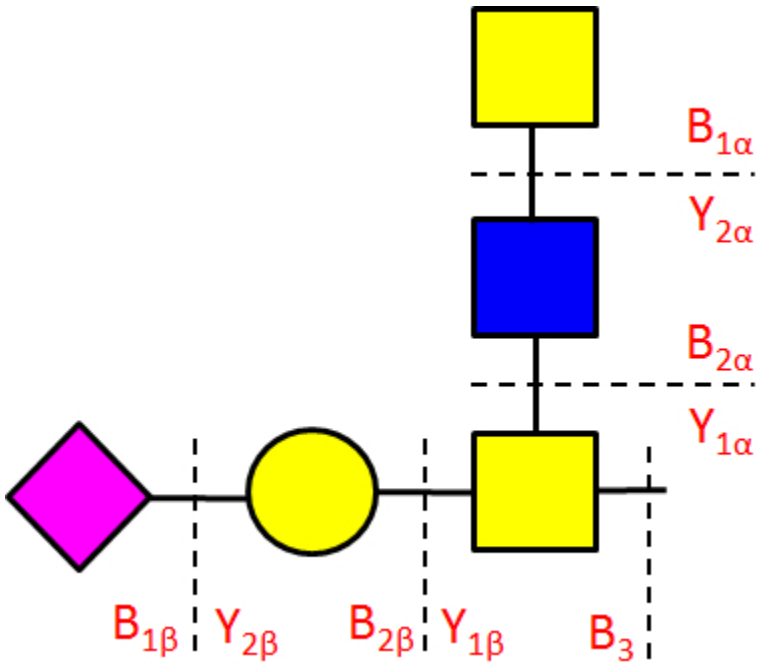
EThcD spectrum of m/z 766.008(3+) that was manually assigned as 342AVAVTLQSH350 of Protein YIPF3 bearing a HexNAc3Hex2FucNeuAc glycan on Thr-346. The precursor ion and its charge-reduced form are labeled with 'pr'; the charge-reduced form of a coeluting doubly charged component is indicated with an asterisk. The glycan was assigned as a core 2 glycan featuring the A antigen linked to the GalNAc. A similar core 2 A antigen has been described on gastric mucin by Rossez et al.26 Scheme 4 illustrates which bonds are cleaved. m/z values printed in bold label fragments that indicate fucose transfer between the antennae: m/z 803.298 represents a HexNAcHexFucNeuAc structure, while m/z 965.931(2+) was formed via HexNAcHex loss from the precursor ion.

84x61mm (300 x 300 DPI)



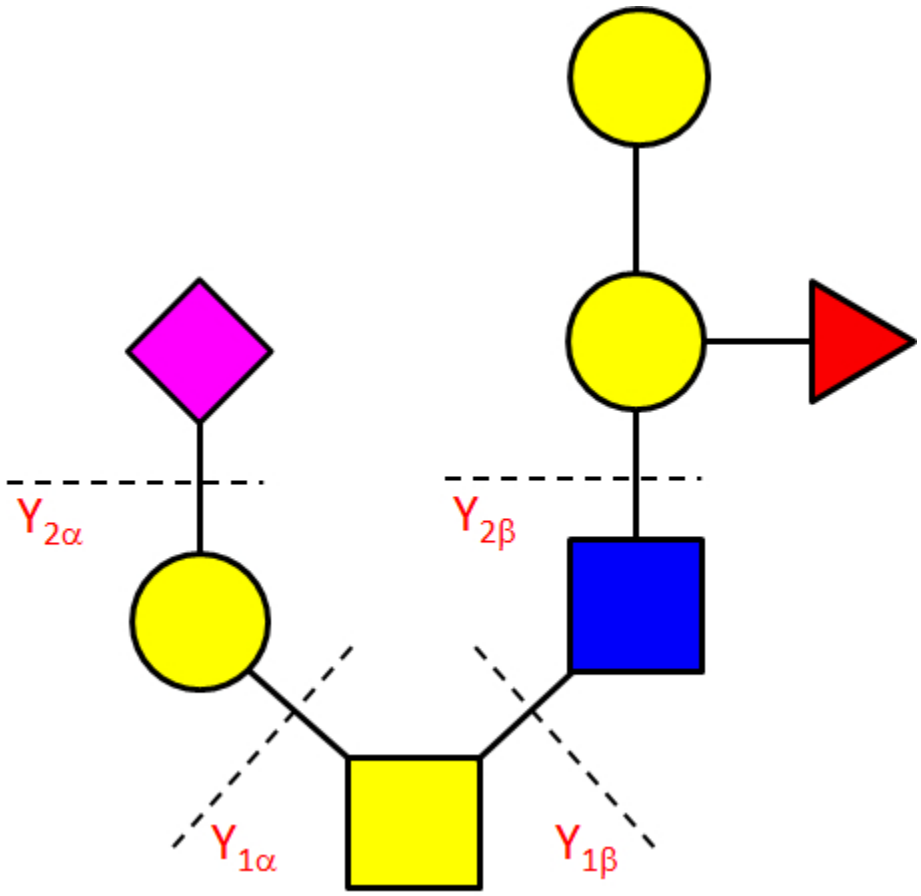
Fragmentation of the glycan structure assigned from the spectrum in Figure 3.

136x114mm (96 x 96 DPI)



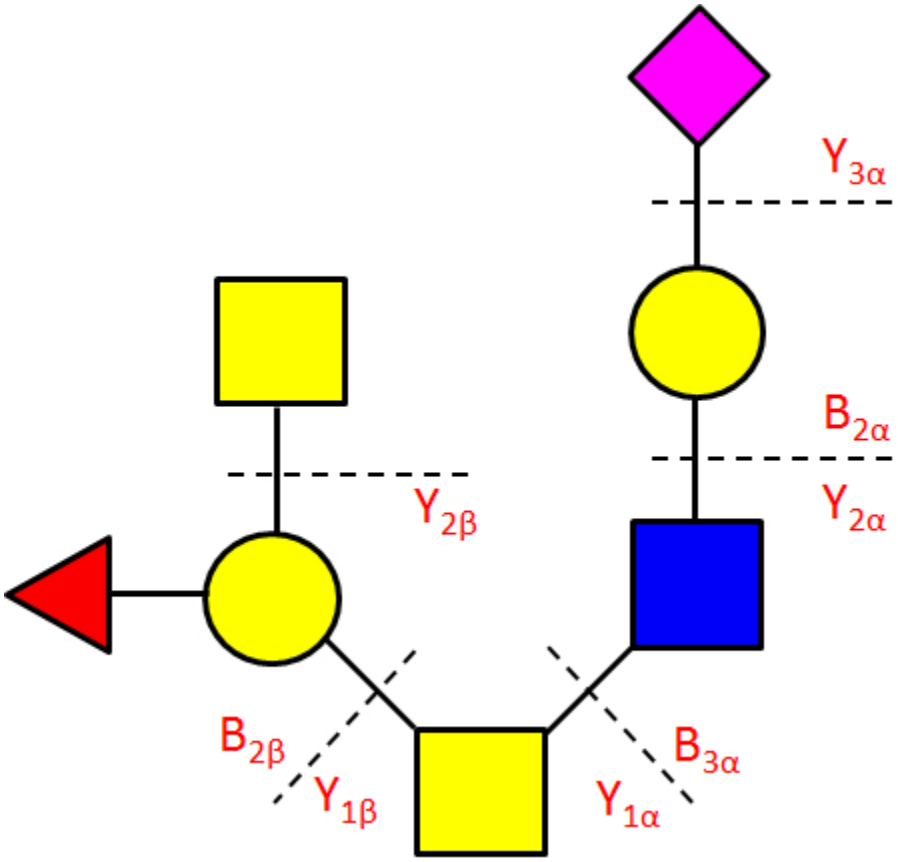
Fragmentation of the glycan structure assigned from the spectrum in Figure 5.

103x91mm (96 x 96 DPI)



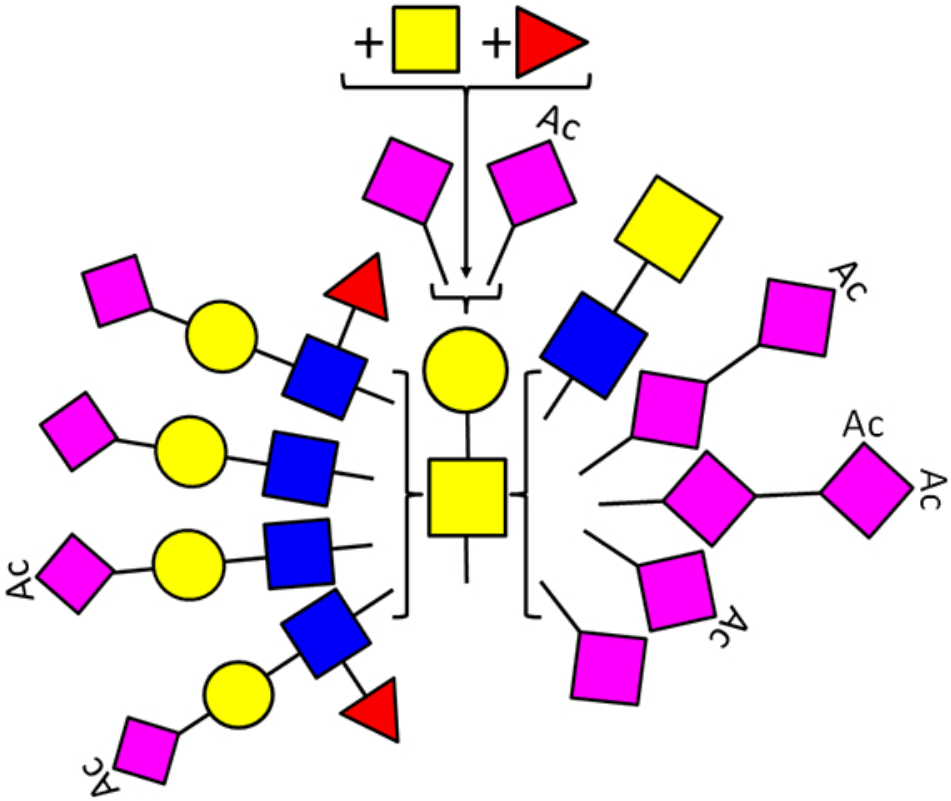
Fragmentation of the mucin core 2 glycan carrying blood-type antigen B structure assigned from the spectrum in Figure 6.

124x118mm (96 x 96 DPI)



Fragmentation of the glycan structure assigned from the spectrum in Figure 7.

125x114mm (96 x 96 DPI)



50x43mm (300 x 300 DPI)