Structural characterization of glycopeptides: Integration of complementary CID and ETD based information.

K.F. Medzihradszky[1,2], A.J. Lynn[1], P. Baker[1], R.J. Chalkley[1], Z. Darula[2] and A.L. Burlingame[1].

[1]Mass Spectrometry Facility, University of California San Francisco
[2]Proteomics Research Group, Biology Research Center of the Hungarian Academy of Sciences, Szeged, Hungary.

O-glycosylation

Two functionally different types:

1. Long-chain modifications of extracellular proteins
2. Single residue regulatory modification of intracellular proteins – O-GlcNAcylation (analogous to phosphorylation)

O-GlcNAcylation is increasingly studied for its regulatory role and interaction with phosphorylation.

Secreted O-glycosylation is a somewhat neglected research area, even though aberrant glycosylation has been linked to cancer and Alzheimer's disease.

The study of both types of O-glycosylation present similar analytical challenges by mass spectrometry.

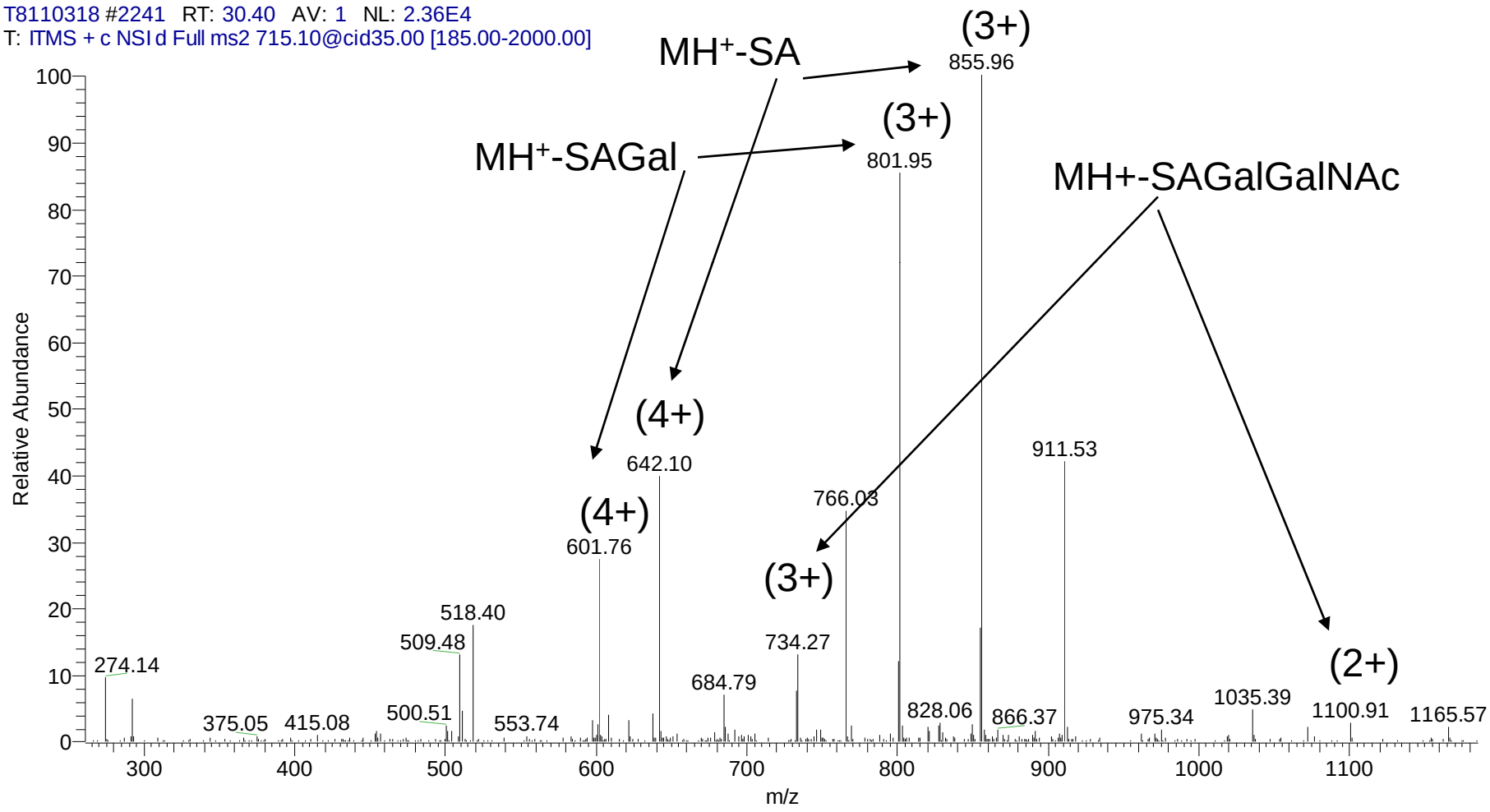
All O-linked glycopeptides display characteristic fragmentation in CID:

- The glycosydic bond is weaker than the peptide bonds, thus the spectrum is usually dominated by ions present due to carbohydrate losses and non-reducing end oxonium ions.
- Gas-phase deglycosylaton occurs - the sugar is eliminated without leaving any telltale sign on the originally modified amino acid, making site assignments impossible in most instances. (Figure 1).

MS/MS analysis utilizing ETD permits both the identification of the modified sequence and the unambiguous determination of the modification site(s) (Figures 2 & 3) [2,3].

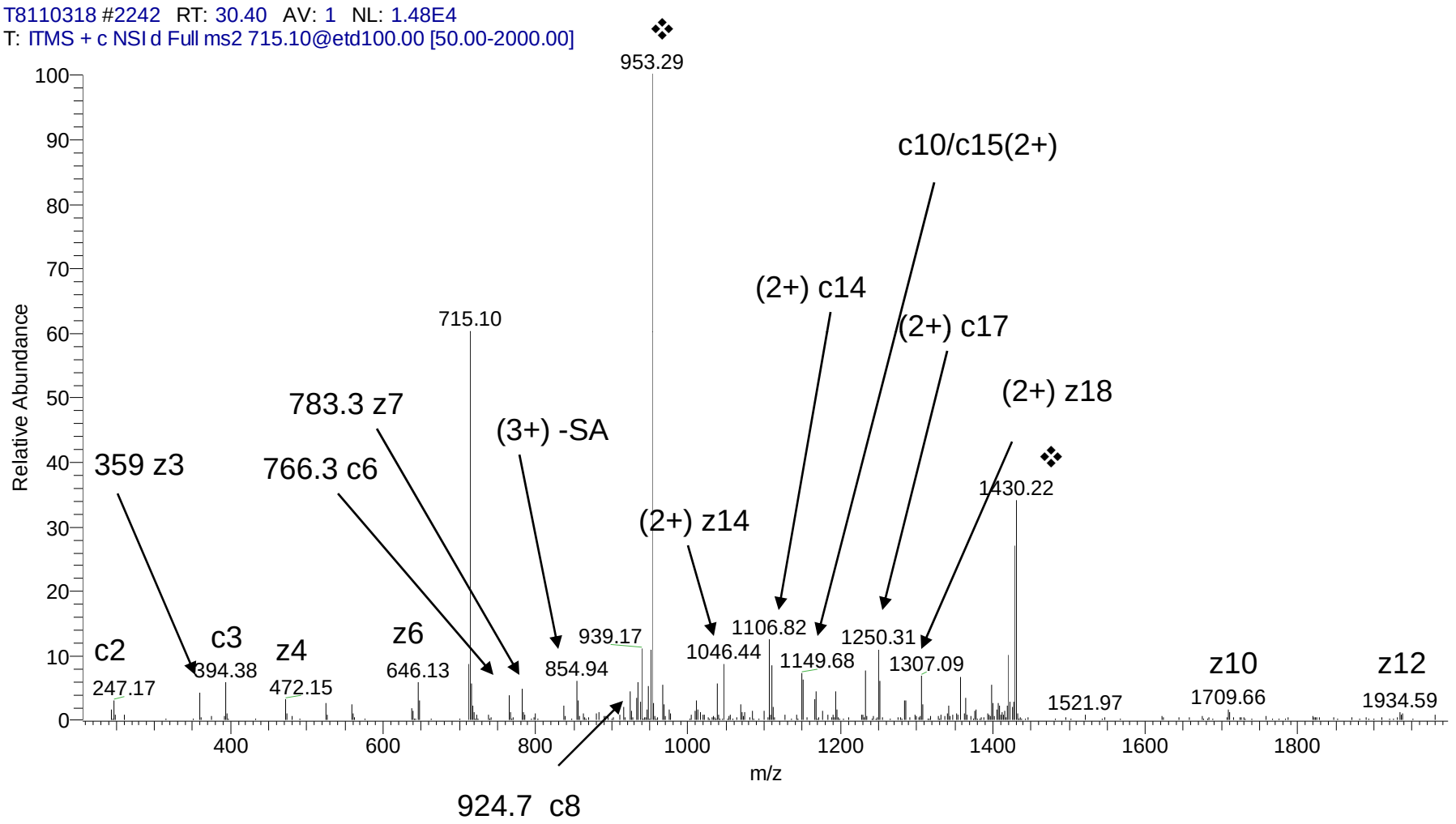
Figure 1: CID of SAGalGalNAc-Modified Peptide

Precursor ion = m/z 714.8520(4+).



- Sugar structure can be identified, but neither peptide sequence nor modification site can be determined.

Figure 2: ETD of Same SAGalGalNAc-Modified Peptide as Figure 1.



^{636}K**T**F**M**L**Q**A**S**Q**P**A**P**T(GalNAcGalSA)**H**S**S**L**D**IK^{655} from Inter-alpha-trypsin inhibitor heavy chain H1 precursor was confidently identified.

Sample Preparation & Mass Spectrometry

* GlcNAc-modified glycopeptides were enriched on a wheat germ agglutinin column as published earlier [1].
* Secreted glycopeptides bearing mucin core-1 type structures were isolated by Jacalin-affinity chromatography, and were analyzed intact or after digestion to leave only the GalNAc core [2].

MS analyses were performed on an LTQ-Orbitrap mass spectrometer in LC/MS mode. The instrument was operated in a data-dependent fashion: MS acquisitions were followed by CID and ETD analyses of 3 software-selected multiply charged ions. Precursor masses were measured in the Orbitrap, while MS/MS experiments were performed and the fragments were measured in the linear trap.

Data-processing I

- Raw data were converted into peaklists using in-house software, PAVA. Database searching was performed against the UniProt database supplemented with a random sequence for each entry, and species specified as *Bos taurus* (31074 entries searched).
- Trypsin was specified as the enzyme, 1 missed cleavage and non-specific cleavage at one of the peptide termini were permitted.
- Mass accuracy: 15 ppm for precursor ions and 0.6 Da for fragment ions.
- Carbamidomethylation of Cys = fixed modification; variable modifications: acetylation of protein N-termini; Met oxidation; cyclization of N-terminal Gln residues; and HexHexNAc or SAHexHexNAc (for intact glycopeptides) and HexNAc (for partially deglycosylated or GlcNAc-modified peptides) modification on Thr and Ser residues. 3 modifications per peptide.

Data-processing II

- The same peaklists were used by 2 versions of the search engine, Protein Prospector: v.5.3 and v5.4:
- Scoring in v5.3 was optimized for tryptic peptides
- Scoring in v5.4 uses different weighting depending on precursor charge and basic residue location.

- Acceptance criteria: Protein score: 22, E>0.05; peptide score: 15, E>0.1.

- Protein Prospector (v5.5) is publicly available:

http://prospector.ucsf.edu

New features in Protein Prospector v5.4 presented on Tuesday (Poster 2781)

1) Neutral loss peaks from charge reduced species are removed prior to database searching.
2) Multiply-charged fragment ions are now considered in ETD data (for precursors of charge state 3+ or higher).
3) New scoring was introduced that gives different fragment ion type weighting depending on the precursor ion charge state and presence of basic residues at the peptide termini.

Re-interrogation of Published [2] Datasets

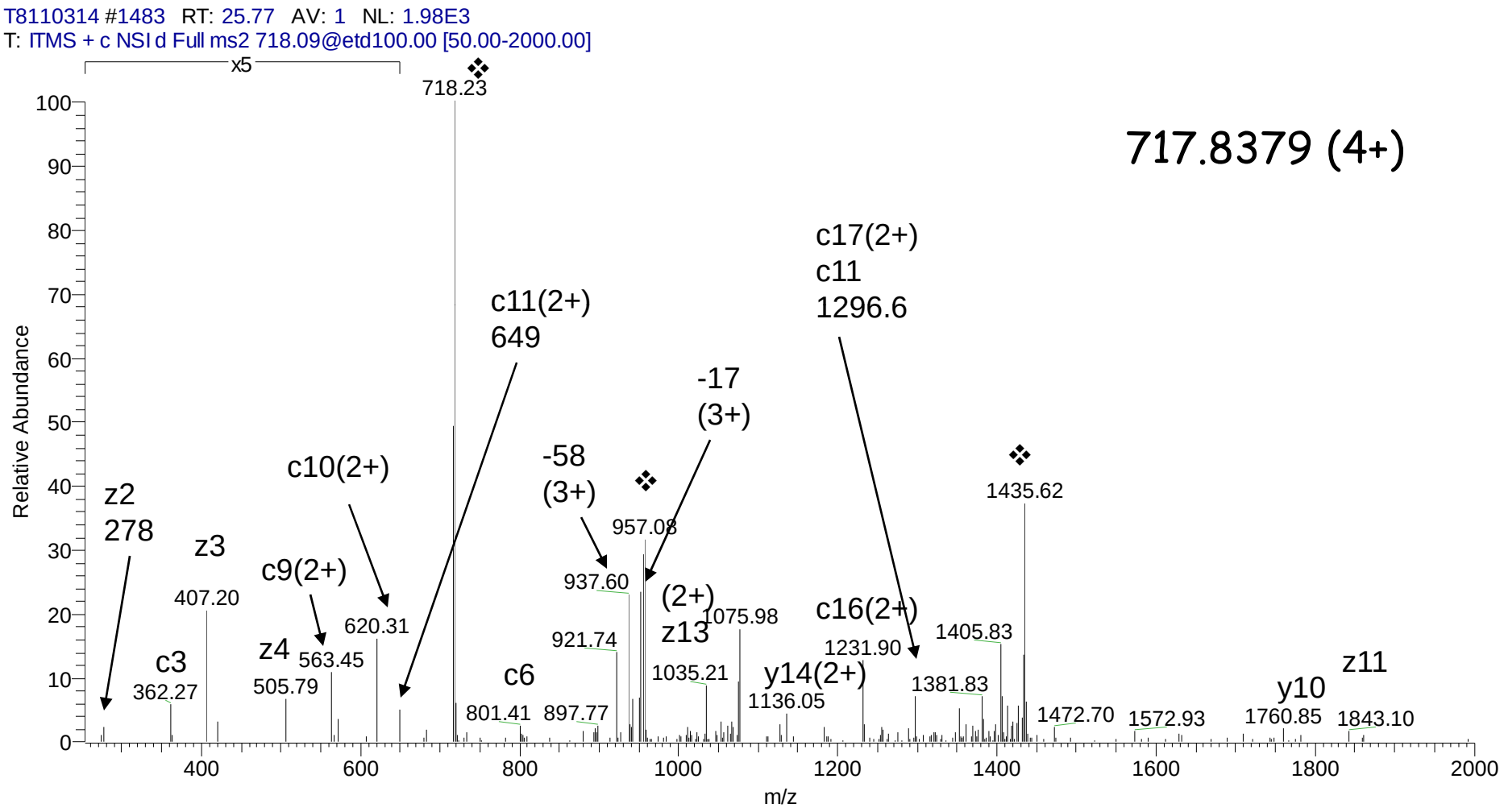
Modified with SA_{1-0}GalGalNAc: 16 LC/MS/MS files

Comparison of Number of Glycopeptides Identified at Different Acceptance Thresholds

		Version 5.3.3	Version 5.4.2	improvement
Intact glycopeptides	E<0.10	21	61	2.9
	E<0.05	16	42	2.63
	E<0.01	10	28	2.8
Partially deglycosylated peptides	E<0.10	23	35	1.52
	E<0.05	21	30	1.42
	E<0.01	17	23	1.35

Only the core GalNAc retained: 1 LC/MS/MS file

Figure 3: One of the new glycosylation sites identified: Thr-605 of Kininogen-1



C(Carbamidomethyl)PSRPWKPVNGVNPT(HexNAcHexSA)VEM(Oxidation)K

Filtering for glycopeptides

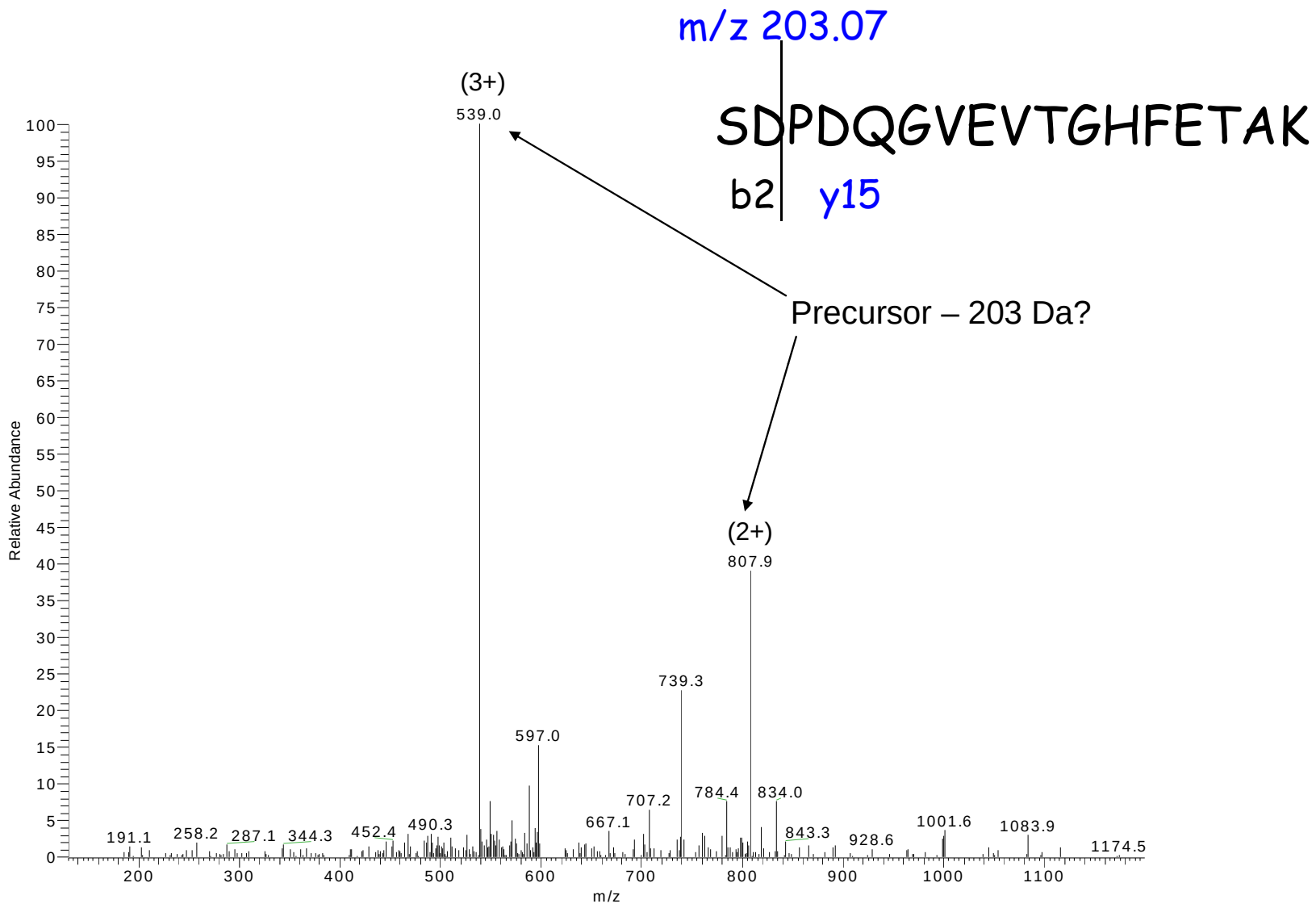
- CID data provides useful information to support ETD identification:
  1. Characteristic sugar losses indicate the presence of the carbohydrate
  2. Limited peptide fragmentation may aid ID confirmation.

- Identifying the characteristic neutral losses could be automated:
  - Input: separated CID and ETD peaklists in mgf format.
  - Perl script "find_neutral_loss.pl" (version 0.2.1):
  1. Identifies CID spectra in which peaks corresponding to neutral losses of sugar moieties (for example, -203.1 or -291.1 Da, corresponding to HexNAc or sialic acid, respectively) from the precursor ion are observed. The script looks for 1 or 2 losses at all the possible charge states (1+ to z^{n-1}, where n is the precursor charge).
  2. Identifies corresponding ETD spectra of the same precursor mass in the ETD MGF file to create a subset MGF file that only contains ETD spectra of potential glycopeptides.
  - Mass of the neutral loss, the threshold for how many of the most intense peaks are searched for the neutral loss peak and the precursor retention time window for correlating ETD and CID spectra can all be altered.

Testing the filtering script

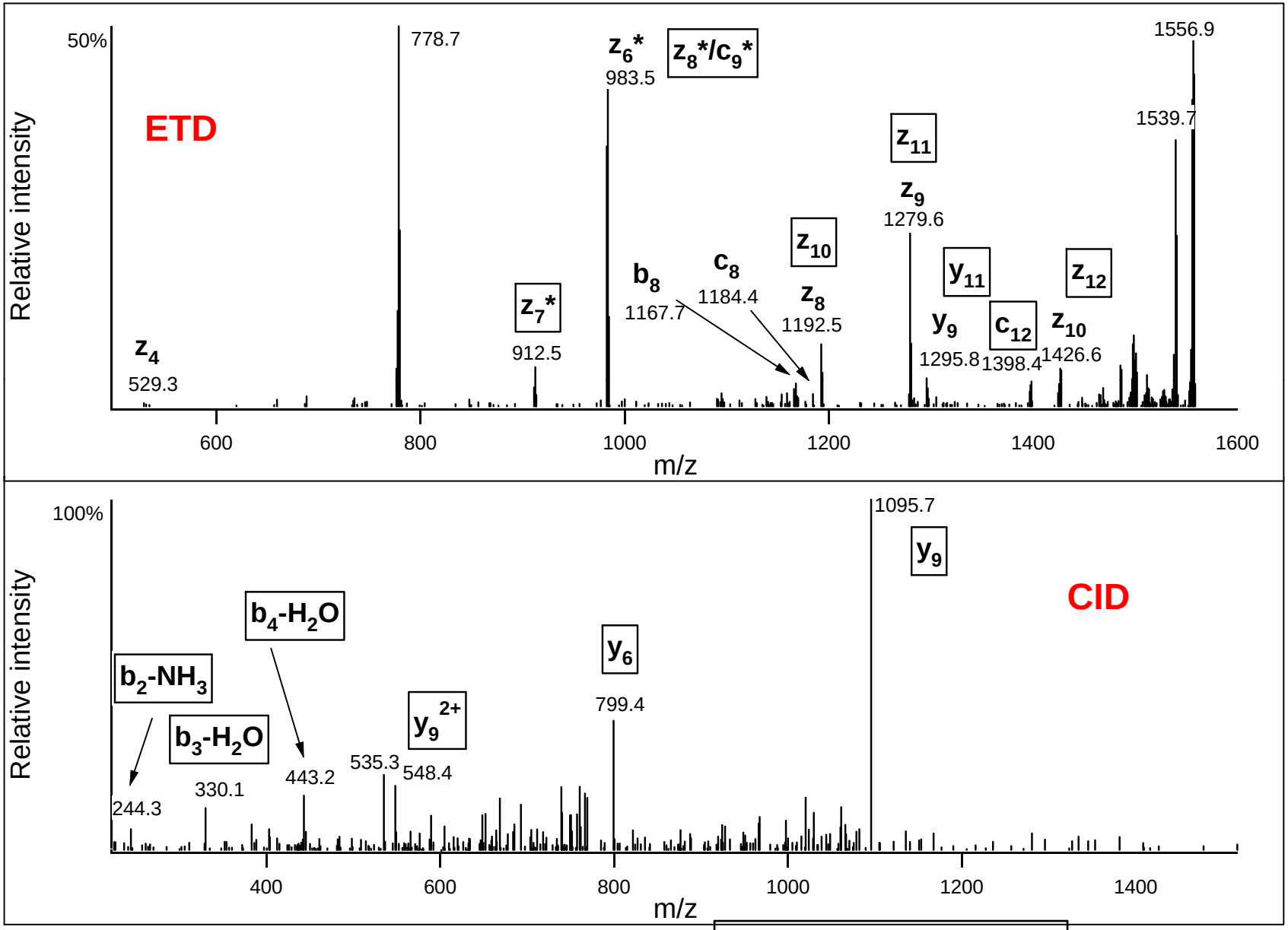
- CID data from LC/MS/MS experiments of fractions enriched in GalNAc- or GlcNAc-modified peptides were screened for glycopeptides
- ~90% of the unmodified peptides were eliminated
- ~5% of the glycopeptides may be affected: either eliminated or identified from a different charge state

Figure 4: False-Positive Glycopeptide "IDd" by the Script



Unmodified peptides may produce glycopeptide-like CID fragmentation !

Figure 5: False-Positive Glycopeptide "ousted" by the script



Summary

- The altered ETD searches with Protein Prospector yielded superior results from the very same data
- "Incorporating" CID data with the help of a glycopeptide-identifying script may boost the confidence level of glycopeptide identifications
- Linking different charge states of the same component will be beneficial, since different charge states may produce the best CID and ETD data

References

1. O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA, Schoepfer R, Burlingame AL. *Mol Cell Proteomics.* **5,** 923-34 (2006).

2. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. Darula Z, Medzihradszky KF. *Mol Cell Proteomics.* **8,** 2515-26 (2009).

3. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. Chalkley RJ, Thalhammer A, Schoepfer R, Burlingame AL. *Proc Natl Acad Sci U S A.* **106,** 8894-9 (2009).

Acknowledgement

This work was supported by NIH NCRR grant P41RR001614 (to ALB) and Hungarian Science Foundation grants OTKA T60283 (to KFM)