Carbamidomethylation side-reactions may lead to glycan misassignments in glycopeptide analysis

Zsuzsanna Darula[1] & Katalin F. Medzihradszky[1,2]

[1] Laboratory of Proteomics Research, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary; [2] Department of Pharmaceutical Chemistry, University of California San Francisco, USA

* Corresponding author

Katalin F. Medzihradszky

Department of Pharmaceutical Chemistry

University of California, San Francisco

600 16th Street

Genentech Hall, N474A, Box 2240

San Francisco, CA 94158-2517

Phone: 415-476-5160 Fax: 415-502-1655

e-mail: folkl@cgl.ucsf.edu

Keywords: carbamidomethylation; collision-induced dissociation; electron-transfer dissociation; glycopeptide; glycosylation; overalkylation

Abstract

Iodoacetamide is perhaps the most widely used reagent for the alkylation of free sulfhydryls in proteomic experiments. Here we report that both incomplete derivatization of Cys side-chains and overalkylation of the peptides may lead to the misassignment of glycoforms when LC-MS/MS with electron-transfer dissociation (ETD) alone is used for the structural characterization of glycopeptides. Accurate mass measurements do not help, because the elemental compositions of the misidentified and correct modifications are identical. Incorporation of "higher-energy C-trap dissociation" (HCD), i.e. beam-type collision-induced dissociation data into the database searches with ETD data may prove decisive in most cases. However, the carbamidomethylation of Met residues leads to sulfonium ether formation, and the resulting fixed positive charge triggers a characteristic fragmentation, that eliminates the normal Y_1 fragment from the HCD spectra of N-linked glycopeptides, producing an abundant Y_1-48 Da ion instead (the nominal mass diference is given relative to the unmodified amino acid sequence), that easily can be mistaken for the side-chain loss from Met sulfoxide. In such cases, good quality ETD data may indicate the discrepancy, and will also display abundant fragments due to CH_3-S-CH_2CONH_2 elimination from the charge-reduced precursor ions. Our observations also draw attention to the underreported interference of different unanticipated covalent modifications.

Introduction

In most proteomic sample preparation protocols the denaturation of proteins is followed by reduction of the disulfide-bridges and by blocking the newly formed sulfhydryl groups in order to prevent reoxidation. Perhaps the most frequently used alkylating agent is iodoacetamide (approximately 70% of the data deposited at PRIDE as of May 2015, were acquired from carbamidomethylated samples). The length of the alkylation reaction varies. Some protocols then will remove the reagent excess [1, 2]; some try to quench the alkylation reaction with the addition of DTT [3, 4]; others perform the in-solution digestion in the presence of the derivatizing agents [5, 6]. As all chemical reactions, alkylation of Cys side-chains is not perfect either. If the exposure was not long enough, or the protein was not properly denatured underalkylation does occur [7, 8]. At the same time, overalkylation also frequently happens [9-13]. Alkylation of the newly formed N-termini, Lys, His, Asp, and Met side-chains have been reported. Obviously, both under- or overalkylation will cause problems in MS/MS data interpretation in large scale proteomic experiments. Primarily, because it will prevent the identification of the affected sequences. However, it also may result in misassignment of MS/MS data.

Mass spectrometry is the method of choice for the analysis of post-translational modifications both in the characterization of individual proteins and in large scale, systems biology experiments [14]. Mass spectrometry has been used in the characterization of extracellular glycosylation for decades, but usually the released glycan pool was characterized [15] or the formerly glycosylated peptides were identified [16]. There is growing evidence that glycosylation may influence biological processes in a site-specific manner [17-19], and thus, there is a growing need for the analysis of intact glycopeptides. Recent technical developments, such as the introduction of electron-transfer dissociation (ETD) [20] and new, high sensitivity instrumentation equipped with ETD have made large scale glycopeptide analysis a reality [21, 22].

Our group has been involved in glycosylation analysis for almost a decade. Recently we have been using wheat germ agglutinin-based lectin-affinity chromatography for glycopeptide enrichment from

various tryptic digests. The isolated mixtures have then been subjected to LC-MS/MS analysis, where both collisional and electron-transfer activation (beam-type CID (HCD), and ETD, respectively) have been used to reveal the identity of the glycopeptides. Collisional activation of glycopeptides leads to extensive glycan and limited peptide fragmentation. Ion trap CID spectra almost exclusively display fragments formed by glycosidic bond cleavages, while beam-type CID (HCD) data may feature sufficient information for the identification of the sequence modified [23]. The activation of multiply charged glycopeptide ions by electron-transfer leads to peptide backbone cleavages leaving the side-chains intact in principal. This is an advantage and a significant drawback at the same time, since a wide array of glycoforms may be present in each mixture analyzed [21, 24, 25]. In order to interpret such results, all potential glycans have to be considered as variable modifications, and thus, not only a protein, but also a glycan database has to be interrogated. Among the available search engines Byonic is uniquely 'specialized' in glycopeptide analysis, and it contains default glycan databases that make this task relatively straightforward [26, 27]. Nevertheless since large-scale intact glycopeptide analysis is still in its infancy careful evaluation of the search results cannot be skipped.

Here we present examples, when under- or overalkylation led to the misidentification of the glycan structures of glycopeptides in database searches with ETD data. These examples again underscore that ETD data alone may not be sufficient for unambiguous glycopeptide identifications.

Methods

Mouse synaptosome sample

"Thirty milligrams of synaptosome was resuspended in 1 ml buffer containing 50 mm ammonium bicarbonate, 6 M guanidine hydrochloride, 6× Phosphatase Inhibitor Cocktails I and II (Roche), and 20 μM PUGNAc (Tocris). The mixture was incubated for one hour at 57 °C with 2 mM Tris(2-carboxyethyl)phosphine hydrochloride to reduce cysteine side chains, these side chains were then alkylated with 4.2 mM iodoacetamide in the dark for 45 min at 21 °C. The mixture was diluted sixfold with ammonium bicarbonate to a final ammonium bicarbonate concentration of 100 mm and 1:50

(w/w) modified trypsin (Promega, Madison, WI) was added. The *p*H was adjusted to 8.0 and the mixture was digested for 12 h at 37 °C. The digests were desalted using a C_{18} Sep Pak cartridge (Waters, Milford, MA) and lyophilized to dryness using a SpeedVac concentrator (Thermo Electron, San Jose, CA)." [28]. Glycopeptides were enriched by lectin-attinity chromatography using wheat germ agglutinin, and further fractionated by high pH reversed phase chromatography. Each fraction was analyzed by LC-MS/MS using ETD as the activation method on an LTQ Orbitrap Velos on-line coupled to a Waters nanoAcquity UPLC. The precursor ions were measured in the Orbitrap, the ETD data were acquired in the linear ion trap. Supplemental activation for the ETD experiments was enabled. These ETD data were interrogated using Protein Prospector v. 5.9.0; *Mus musculus* proteins in UniProt (downloaded 07/06/2011) were searched, a randomized sequence for each entry was added to the database; only tryptic cleavages were considered, one missed cleavage was permitted; mass accuracy required was 10 ppm and 0.6 Da, for the precursor ions and the fragments, respectively; carbamidomethylation of Cys residues was considered as fixed modification; while N-terminal Gln cyclization, Met oxidation, protein N-terminal acetylation and Ser/Thr fucosylation were listed as variable modifications, 3 modifications/peptide were permitted. The results were filtered using the following criteria: minimum score = 22, max E = 0.001, and precursor mass accuracy within 5 ppm [21].

Further details about the synaptosome preparation, and the fractionation of the digest have been published [21, 28].

Human serum sample

Human serum samples were prepared with a slightly modified version of the FASP protocol [2]. One ml human serum was supplemented to give a final concentration of 50 mM DTT, 6M guanidine and 50 mM Tris (pH:7.5), reduction was performed on 95 °C for 30 min, and 100 µl 200 mM iodoacetamide solution was added for the alkylation (30 min in the dark). The reagent excess was removed, the buffer was replaced, the digestion proceeded and the resulting peptides were collected according to the published protocol (for further details see Supplement 1.). Glycopeptides

were enriched by lectin affinity chromatography using Concanavalin A followed by wheat germ agglutinin. LC-MS/MS analyses of the resulting peptide mixtures were performed on an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific LTD) on-line coupled to a Waters nanoAcquity UPLC. Samples were injected onto a Waters Symmetry C18 trapping column (180µm* 20mm, 5µm particle size, 100Å pore size; Waters 186003514) and after trapping with 3% B for 5 minutes (flow rate: 5 µl/min) the peptide mixture was transferred to and separated on a Acclaim PepMap RSLC column (75µm*250mm, 2µm particle size, 100Å pore size; Dionex 164536) developing a linear gradient of 3-40% B in 55 min (flow rate: 300nl/min; solvent A: 0.1% formic acid/water, solvent B: 0.1% formic acid/acetonitrile). Mass measurements were performed in the Orbitrap, and the 3 most abundant multiply charged ions were selected for HCD analysis. ETD data acquisition(performed in the linear ion trap) was triggered by the detection of the HexNAc oxonium ion, m/z 204.0867±10ppm among the top 50 HCD fragments [29] (minimum signal intensity:100). Supplemental activation for the ETD experiments was enabled. Normalized collision energy for HCD experiments was set to 32. MS/MS data of z=2 and z>2 charge states were acquired in separate experiments. Proteome Discoverer (Thermo Scientific, v1.4.0.288) was used to generate peak lists from the raw data. Database searches were performed with the ETD data using Byonic (v2.0-25, Protein Metrics Inc.) with the following parameter set: human subset of the Uniprot database (downloaded 11/13/2013, 39697 entries); tryptic peptides with maximum one missed cleavage, mass accuracies within 5 ppm and 0.6 Da for precursor and fragment ions, respectively. Fixed modification was carbamidomethylation of Cys residues. Variable modifications were Met oxidation (common, maximum 3 modifications/peptide); cyclization of N-terminal Gln residues (rare, maximum 1 modification/peptide); protein N-terminal acetylation (rare, maximum 1 modification/peptide) and N-glycans (rare, maximum 1 modification/peptide using the “N-glycan 57 human plasma” glycan set). Decoy sequences were concatenated to forward entries and common contaminants (using the built-in “common contaminant” database of 69 proteins) were also

considered. Acceptance criteria: 2% false discovery rate on the protein level, minimum peptide score: 200.

The sample preparation protocol for dataset PDX001277 from the PRIDE repository (http://www.ebi.ac.uk/pride/archive/) has been included in Supplement 2, where the appropriate search parameters for the preformed data analysis as well as the acceptance criteria are also listed. Experimental details about testing different alkylating agents with the FASP protocol are presented in Supplement 3.

Results and discussion

First we demonstrate how underalkylation may interfere with glycopeptide analysis. We illustrate this point with a specific type of O-glycosylation. O-linked glycans may feature different core structures, among them there is a deoxyhexose, fucose (Fuc). We have reported that in a mouse synaptosome dataset numerous high scoring O-fucosyl peptides were detected (Table 1) [21]. However, a consensus motif of CXXGGT/SC has been described for O-fucosylation [30], and none of the identified sequences fulfilled this requirement. While one can always assume that the consensus motif definition should be more flexible, the only common denominator among the identified sequences was the presence of a Cys residue in proximity to the glycosylation site. Once underalkylation is considered the explanation becomes obvious, the modifying glycan is HexNAc (GlcNAc or GalNAc) instead of the Fuc. Figure 1 demonstrates the quality of the data as well as the difficulty to identify the misassignments. The ETD spectrum features almost complete c and $z^{\cdot}$ ion series, only the Thr-8's c fragment is missing, while in the C-terminal series the Cys residue, not unexpectedly, produced a w fragment (Peptide fragmentation nomeclature [31]) instead of a $z^{\cdot}$ ion [32] that does not reveal whether the side-chain was alkylated or not. Not only the mass of HexNAc is the same as the two other modifications combined, but even their elemental compositions are identical: HexNAc = $C_8H_{13}NO_5$ = $C_6H_{10}O_4$ + C_2H_3NO = Fuc + carbamidomethyl group (the elemental compositions of the structures added to the peptide are listed). Thus, only two fragments prove our

conclusions: one is the result of the characteristic acetyl loss of N-acetylhexosamines, the other is formed by the loss of the modifying sugar, both from the charge-reduced precursor ion. (Quite a few of the sequences listed in Table 1 were also detected with the HexNAc and properly carbamidomethylated [28]).

One can presume that similar misassignments indicating the presence of a Fuc in the glycan and a carbamidomethyl group on the peptide instead of an additional HexNAc in the oligosaccharide also may happen in N-linked glycopeptides if the underivatized Cys is located sufficiently close to the glycosylation site. We believe that this danger actually could be higher for N-glycan determination in any high-throughput studies since there may not be any obvious warning signs such the one we encountered in the O-glycosylation study described above. At the same time, in both cases data obtained by collisional activation (either by ion trap or beam-type CID) easily could confirm the identity of the sequence modified in form of Y_0 and Y_1 fragments, for any O-linked or N-linked glycopeptide, respectively (nomenclature [33]) [34, 35].

On the same basis overalkylation also may lead to incorrect N-glycan composition assignments. Fucosylated structures with an additional carbamidomethyl group on the peptide may be misinterpreted as featuring a HexNAc instead of the Fuc. Overalkylation represents a higher interference potential, since a series of amino acid side-chains may be derivatized [9-13]. However, if the site of modification is the N-terminus, or the side-chain of Lys, His, or Asp, collisional dissociation could come to the rescue, and will reveal the molecular mass of the sequence modified, and thus, also the +57.0216 Da discrepancy.

Here we will present an example, where spotting the problem may not be so trivial. In a recent human glycopeptide study different glycoforms of tryptic peptide ^{74}TVLTPATNHMGN*VTFTIPANR^{94} of Complement 3 protein were identified (Asn-85, the glycosylation site is labeled with an asterisk). Oligomannose structures, $GlcNAc_2Man_{4-7}$ as well as hybrid oligosaccharides, $GlcNAc_3Hex_{3-6}$ were among the glycans assigned from the ETD data. Some of the glycopeptides featured Met-83 oxidized. While evaluating the assignments we noticed a peculiar phenomenon. i) The glycoforms

with the hybrid glycans were only detected with the Met oxidized; ii) in the HCD spectra of these peptides Y_1 was not detected at all, only the corresponding fragment without the oxidized Met side-chain (CH_3SOH loss = -63.9983 Da) was abundant. Appropriate peptide fragment ions if detected were also observed in this 'truncated' form, i. e. with an α-amino-3,4-dehydrobutyric acid instead of the methionine (Figure 2, Supplemental Figure 1).

In this particular sequence the glycosylation site and the Met-sulfoxide are separated by only one residue. Thus, an interaction between the oxidized side-chain and the glycan could not be excluded *per se*. However, we could not find any other Met(O)-containing glycopeptide that displayed similar behavior. The HCD spectra of such peptides always featured the 'normal' Y_1 fragment as well as the product of the neutral loss of CH_3SOH (-64 Da), and the intact Y_1 was usually more abundant (Supplemental Figure 2).

Thus, we started to search for a more plausible explanation. Our hypothesis was that Met-83 was carbamidomethylated, not oxidized. We have demonstrated long time ago that beta-elimination-like partial side-chain loss is the favored fragmentation step for Met-sulfonium ethers [36] and also described in detail the CID characteristics of carbamidomethyl-methionine-containing peptides [37]. The diagnostic fragmentation step there is the loss of CH_3-S-CH_2CONH_2, i.e. -105.0249 Da that will transform the Met residue into a α-amino-3,4-dehydrobutyric acid (see Supplemental Figure 1B). The characteristic CH_3SOH loss (-64 Da) from an oxidized methionine yields the same fragment (Supplemental Figure 1B).

If the Met side-chain is carbamidomethylated instead of oxidized, then the peptide mass is 41 Da higher, thus, the glycan structure has to be so much lighter. The mass difference between a HexNAc and a Hex is 41.0265 Da, and the elemental composition of the new structural combination will be exactly the same as the old one: oxidation of Met + HexNAc in the glycan = O + $C_8H_{13}NO_5$ = $C_8H_{13}NO_6$ = C_2H_3NO + $C_6H_{10}O_5$ = carbamidomethyl on the peptide + Hex in the glycan. Once we scrutinized the ETD data in this light it becomes obvious that the fragmentation of Met-sulfonium ether has not been completely eliminated by the different activation method. On the contrary, the most abundant

fragments in the ETD spectra are the 58 and 105 Da losses from the charge-reduced precursor ion (Figure 3), i.e. either the carbamidomethyl group is removed (CH_2CONH_2 = 58.0293 Da) or, just like in the CID experiments, CH_3-S-CH_2-$CONH_2$ is eliminated (-105.0249 Da). Moreover, c_{11}(2+) at m/z 598.8 indicates carbamidomethylation of the N-terminal part of the peptide sequence. In addition, we also discovered two shorter isoforms of the same sequence stretch, featuring Lys or Arg instead of Thr in position-89. Both of these shorter sequences produced a $z_5^{\cdot}$ fragment, clearly confirming that Asn-85 is modified with a $GlcNAc_2Man_6$ structure and Met-83 is carbamidomethylated (Figure 4). Based on the peak areas of the extracted ion chromatograms of the unmodified, Met-oxidized and Met-carbamidomethylated variants of the Complement C3 [74-94] $GlcNAc_2Man_6$ glycoform (Supplemental Figure 4) we estimated that the extent of Met-carbamidomethylation was approximately 7.6 %.

In order to investigate the occurrence of overalkylation in other datasets, we selected a project from the PRIDE repository where the only covalent modification intentionally introduced was the carbamidomethylation of Cys residues (PXD001277). In this project the reduction and alkylation was followed by an overnight in-solution digestion without removing the reagent excess (the experimental details provided with the dataset were included in Supplement 2.). Reanalyzing 5 files from this dataset we found that more than 10% of the spectra assigned represented overalkylated sequences (Supplement 2). We also performed experiments with iodoacetamide and N-ethylmaleimide as the alkylating agents using the FASP protocol [2], where the reagent excess is removed *prior to* the digestion. Depending on the experimental conditions, 1-7% of the confidently identified peptides represented overalkylated peptides (Supplement 1 and 3). Overalkylation was detected even when only ~100-fold excess of the alkylating agent was applied (Supplement 1). Based on these data we believe overalkylation regularly occurs in high throughput proteomic experiments.

Summary

Mass spectrometry, a relatively unbiased analytical tool is widely used for post-translational modification (PTM) analysis including site-specific glycosylation analysis of single proteins as well as very complex mixtures. Glycosylation is an 'umbrella PTM'. In each extracellular glycopeptide analysis per definition numerous different modifications have to be considered for each potential site, unless the suite of glycans present has been characterized prior to the glycopeptide analysis. Recently both single glycoprotein characterizations and high-throughput glycosylation studies rely heavily on beam-type CID (HCD) and ETD analysis of glycopeptides in line with automated data interpretation. Although there have been attempts to connect the information content of data acquired by the different activation methods it has not got far enough [38, 39]. Our findings underscore the need for further software development for more reliable glycopeptide interpretation. In addition, we also would like to draw attention to the interference of different covalent modifications that we believe is more common than reported.

Acknowledgement

This work was supported by NIH grant NIGMS 8P41GM103481 (to the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF, Director: A. L. Burlingame),OTKA grant 105611 (to Z. Darula) and NFU GOP1.1.1-11-2012-0452 (to K.F. Medzihradszky). Z. Darula was supported by the Janos Bolyai Fellowship of the Hungarian Academy of Sciences.

References

1. Medzihradszky, K. F. *Methods Enzymol*. **2005,** 405, 50-65.
2. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat Methods.* **2009,** *6,* 359-362.
3. Beck, F.; Burkhart, J. M.; Geiger, J.; Zahedi, R. P.; Sickmann, A. *Methods Mol Biol.* **2012,** *893,* 101-113.
4. Arsene, C. G.; Ohlendorf, R.; Burkitt, W.; Pritchard, C.; Henrion, A.; O'Connor, G.; Bunk, D. M.; Güttler, B. *Anal Chem.* **2008,** *80,* 4154-4160.

5. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J.C. *Anal Chem.* **2003,** *75,* 6023–6028.
6. Gallien, S.; Bourmaud, A.; Domon, B. *J Proteome Res.* **2014,** *13,* 2688-2695.
7. Webster, D. M.; Thompson, E. O. *Aust J Biol Sci.* **1982,** *35,* 125-135.2.
8. Meza, J. E.; Scott, G. K.; Benz, C. C.; Baldwin, M. A. *Anal Biochem.* **2003,** *320,* 21-31.
9. Gundlach, H. G., Moore, S, Stein, WH. *J Biol Chem.* **1959,** *234,* 1761-1764.
10. Jones, M. D.; Merewether, L. A.; Clogston, C. L.; Lu, H. S. *Anal Biochem.* **1994,** *216,* 135-146.
11. Lapko, V. N.; Smith, D. L.; Smith, J. B. *J Mass Spectrom.* **2000,** *35,* 572-575.
12. Boja, E. S.; Fales, H. M. *Anal Chem.* **2001,** *73,* 3576-3582.
13. Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. *Nat Methods.* **2008,** *5,* 459-460.
14. Larsen, M. R.; Trelle, M. B.; Thingholm, T. E.; Jensen, O. N. *Biotechniques* **2006,** *40,* 790-798.
15. North, S. J.; Hitchen, P. G.; Haslam, S. M.; Dell, A. *Curr Opin Struct Biol.* **2009,** *19,* 498-506.
16. Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat Biotechnol.* **2003**, *21,* 660-666.
17. Schjoldager, K. T.; Vester-Christensen, M. B.; Goth, C. K.; Petersen, T. N.; Brunak, S.; Bennett, E. P.; Levery, S. B.; Clausen, H. *J Biol Chem.* **2011,** *286,* 40122-40132.
18. Srinivasan, S.; Romagnoli, M.; Bohm, A.; Sonenshein, G. E. *J Biol Chem.* **2014,** *289,* 33676-33688.
19. Kroslak, T.; Laforge, K. S.; Gianotti, R. J.; Ho, A.; Nielsen, D. A.; Kreek, M. J. *J Neurochem.* **2007,** *103,* 77-87.
20. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc Natl Acad Sci U S A*. **2004,** *101,* 9528-9533.
21. Trinidad, J. C.; Schoepfer, R.; Burlingame, A. L.; Medzihradszky, K. F. *Mol Cell Proteomics.* **2013,** *12,* 3474-3488.
22. Parker, B. L.; Thaysen-Andersen, M.; Solis, N.; Scott, N. E.; Larsen, M. R.; Graham, M. E.; Packer, N. H.; Cordwell, S. J. *J Proteome Res.* **2013,** 12, 5791-5800.
23. Darula, Z.; Chalkley, R. J.; Lynn, A.; Baker, P. R.; Medzihradszky, K. F. *Amino Acids* **2011,** *41,* 321-328.

24. Darula, Z.; Medzihradszky, K. F. *J Am Soc Mass Spectrom.* **2014,** *25,* 977-987.
25. Medzihradszky, K. F.; Kaasik, K.; Chalkley, R. J. *Anal Chem.* **2015,** *87,* 3064-3071.
26. Bern, M.; Kil, Y. J.; Becker, C. *Curr. Protoc. Bioinformatics* **2012**, Chapter *13*: Unit13.20.
27. Bern, M.; Kil, Y. L.; Medzihradszky, K. F.; Tang, W.; Bern, N.; Becker, C. *J. Am. Soc. Mass Spectrom.* **2015**, submitted.
28. Trinidad, J. C.; Barkan, D. T.; Gulledge, B. F.; Thalhammer, A.; Sali, A.; Schoepfer, R.; Burlingame, A. L. *Mol Cell Proteomics.* **2012,** *11,* 215-229.
29. Zhao, P.; Viner, R.; Teo, C.F.; Boons, G.J.; Horn, D.; Wells, L. *J. Proteome Res.* **2011,** *10,* 4088-4104.
30. Harris, R. J.; Spellman, M. W. *Glycobiology* **1993,** *3,* 219-224.
31. Biemann, K. *Methods Enzymol.* **1990,** *193,* 886-887.
32. Chalkley, R. J.; Brinkworth, C. S.; Burlingame, A. L. *J Am Soc Mass Spectrom*. **2006,** *17,* 1271-1274.
33. Domon, B.; Costello, C. E. *Glycoconjugate J*. **1988,** *5,* 397-409.
34. Peter-Katalinić, J. *Methods Enzymol.* **2005,** *405,* 139-171.
35. Medzihradszky, K. F. *Methods Enzymol.* **2005,** *405,* 116-138.
36. Ghiladi, R. A.; Knudsen, G. M.; Medzihradszky, K. F.; Ortiz de Montellano, P. R. *J Biol Chem.* **2005,** *280,* 22651-22663.
37. Medzihradszky, K. F.; Chalkley, R. J. *Mass Spectrom Rev.* **2015,** *34,* 43-63.
38. Wu, S. W.; Pu, T. H.; Viner, R.; Khoo, K. H. *Anal Chem*. **2014,** *86,* 5478-5486.
39. Mayampurath, A.; Yu, C. Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H. *Anal Chem.* **2014,** *86,* 453-63.

Table 1 A few sequences identified as O-fucosylated from ETD data acquired from a mouse synaptosome tryptic digest (full, detailed list in ref. [21]).

Peptide	**Score**[1]	**Expect**[1]
APSC(Carbamidomethyl)S(Fuc)GLSMVSGARPGPGPAATTHK	49.2	6.90E-08
VC(Carbamidomethyl)S(Fuc)AAPPSVLNETGFSLTVPASAK	44.6	3.90E-06
NAEGSTVT(Fuc)C(Carbamidomethyl)TGSIR	53.2	7.40E-06
NQSLPVMMGSFGAPVC(Carbamidomethyl)T(Fuc)TSPK	37.4	3.80E-05
VTPSLNSAPAPAC(Carbamidomethyl)S(Fuc)STSHLK	47.7	2.70E-05
LDFGQGSGS(Fuc)PVC(Carbamidomethyl)LAQVK	55.8	4.60E-07
QKAPFPAT(Fuc)C(Carbamidomethyl)EAPSR	42.5	4.80E-05
FPFGSSC(Carbamidomethyl)T(Fuc)GTFHPAPSAPDK	37.8	6.30E-07
IQTDT(Fuc)C(Carbamidomethyl)HSTVVHSPEVYSVIIR	51.4	1.50E-06
YPATC(Carbamidomethyl)VT(Fuc)DIMLSHK	43.2	3.00E-06
LGPVYC(Carbamidomethyl)QAS(Fuc)FSGTNIIGNK	39.6	5.30E-06
AAT(Fuc)C(Carbamidomethyl)FSTTLTNSVTTSSVPSPR	41.2	1.00E-06

[1]Both the score and the E-values refer to the incorrectly assigned structures. In the correct assignments the sequences are modified with a HexNAc, and the Cys is not alkylated.

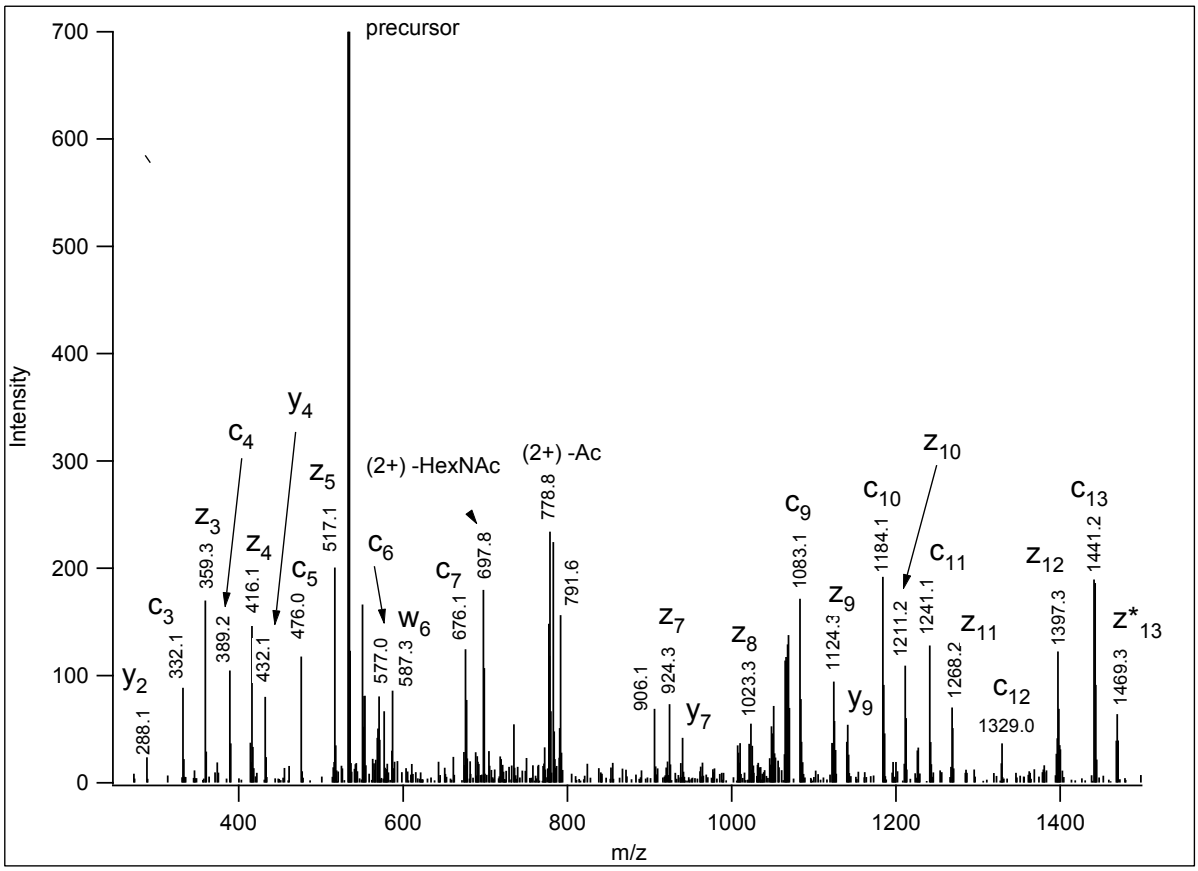
Figure Legends

Figure 1. ETD spectrum of m/z 533.5834(3+) incorrectly identified as NAEGSTVT(Fuc)C(Carbamidomethyl)TGSIR. In the correct assignment the Cys is not alkylated and the peptide is modified with a HexNAc. The detected fragments are identical for both structures as shown in the scheme under the spectrum. Only the characteristic 42 Da (Ac) and 203 Da (HexNAc) losses, both from the charge-reduced precursor ion, help to identify the glycan. Peptide fragments are assigned according to Biemann [31]. The asterisk indicates a z+1 fragment.

Figure 2. HCD data of precursor masses, 913.1652(4+) (upper panel), and 923.4220(4+) (lower panel) that were identified from ETD data as ^{74}TVLTPATNHM(Oxidized)GNVTFTIPANR^{94} glycoforms with an oligomannose ($GlcNAc_2Man_6$) and a hybrid ($GlcNAc_3Hex_5$) structure. Peptide fragments retaining the core GlcNAc (exact additive mass: 203.0794) are indicated with the mass increment: +203. The assignment of the second glycoform is incorrect. When correctly assigned this glycopeptide features the same $GlcNAc_2Man_6$ structure as the first, but the Met residue is carbamidomethylated, not oxidized. The fragmentation patterns for the two different assignments are presented in Supplemental Figure 1.

Figure 3. ETD spectrum of TVLTPATNHM(Carbamidomethyl)GN($GlcNAc_2Man_6$)VTFTIPANR, precursor ion at m/z 923.4220(4+) within 3 ppm of the calculated value. Losses characteristic of the Met-sulfonium ether are labeled in red. Either the carbamidomethyl group is removed (-58 Da) or CH_3-S-CH_2-$CONH_2$ (-105) is eliminated. The fragmentation pattern is shown under the spectrum. Carbohydrate units are labeled following the CFG nomenclature: (http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml)

Figure 4. ETD data of the shorter Complement 3 [74-89] isoforms with Arg (upper panel) and Lys (lower panel) in position 89, precursor ions at m/z 799.3546(4+) and m/z 792.3521(4+), respectively, both within 2 ppm of the calculated values. These data confirmed the $GlcNAc_2Man_6$ glycan structure, and also indicated that Met-83 is carbamidomethylated. Losses characteristic of the Met-sulfonium ether are labeled in red. Either the carbamidomethyl group is removed (-58 Da) or CH_3-S-CH_2-$CONH_2$ (-105) is eliminated. The fragmentation pattern is presented under the spectra, fragments in bold were detected in both spectra, fragments underlined were detected only in the spectrum of the peptide with the C-terminal Lys (lower panel). Carbohydrate units are labeled following the CFG nomenclature: (http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml).


precursor
Intensity
m/z
(2+) -HexNAc
(2+) -Ac
y2
c3
z3
c4
z4
y4
c5
z5
c6
w6
c7
z7
y7
z8
c9
z9
y9
c10
z10
c11
z11
c12
z12
c13
z*13
288.1
332.1
359.3
389.2
416.1
432.1
476.0
517.1
577.0
587.3
676.1
697.8
778.8
791.6
906.1
924.3
1023.3
1083.1
1124.3
1184.1
1211.2
1241.1
1268.2
1329.0
1397.3
1441.2
1469.3

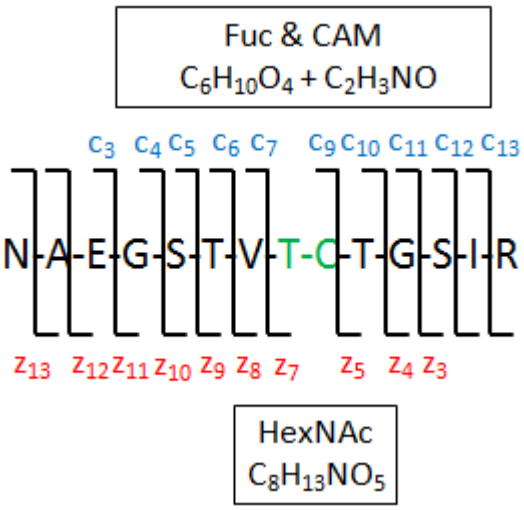

Fuc & CAM
C6H10O4 + C2H3NO
c3 c4 c5 c6 c7 c9 c10 c11 c12 c13
N-A-E-G-S-T-V-T-C-T-G-S-I-R
z13 z12 z11 z10 z9 z8 z7 z5 z4 z3
HexNAc
C8H13NO5

Figure 1.

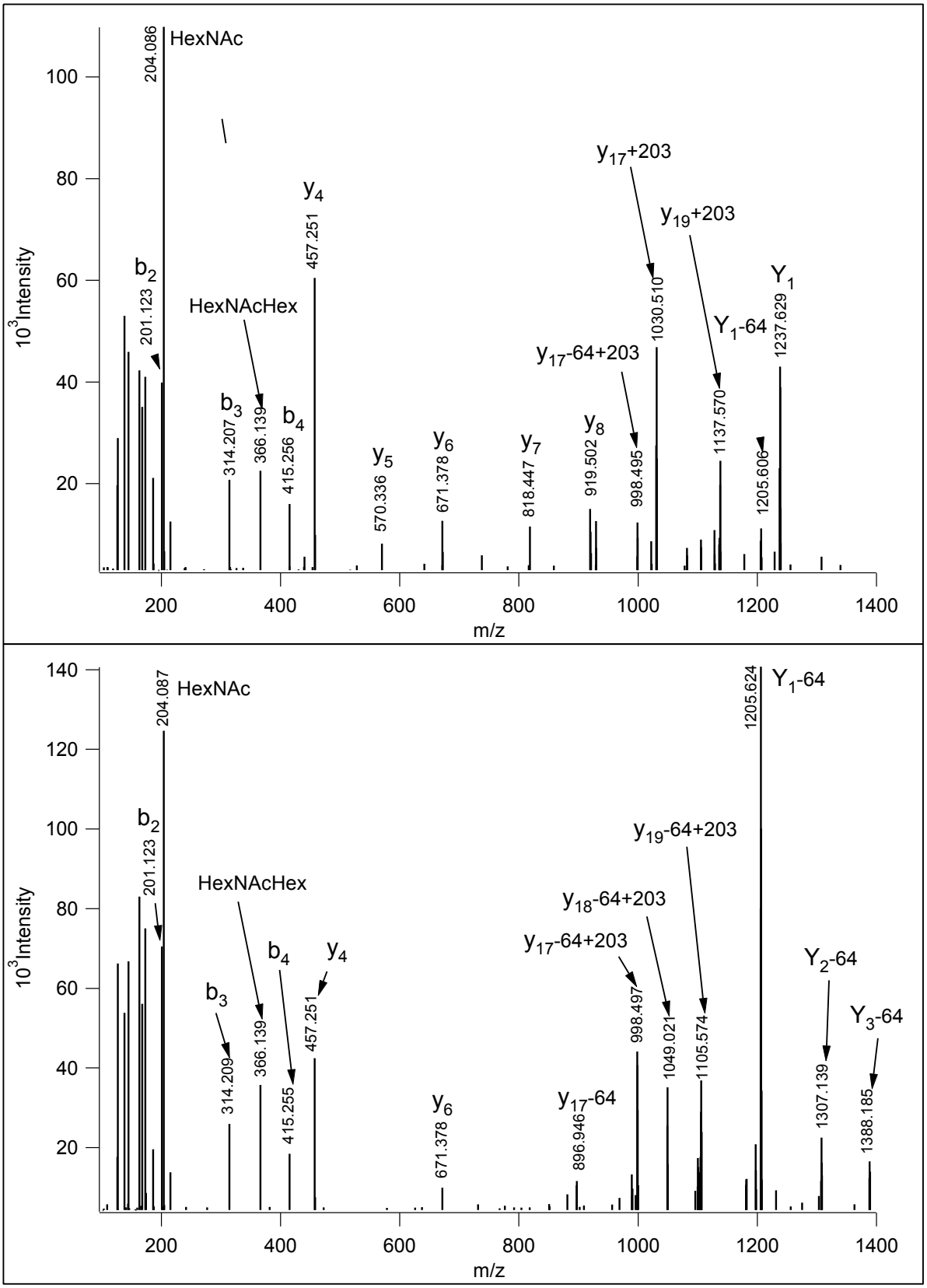
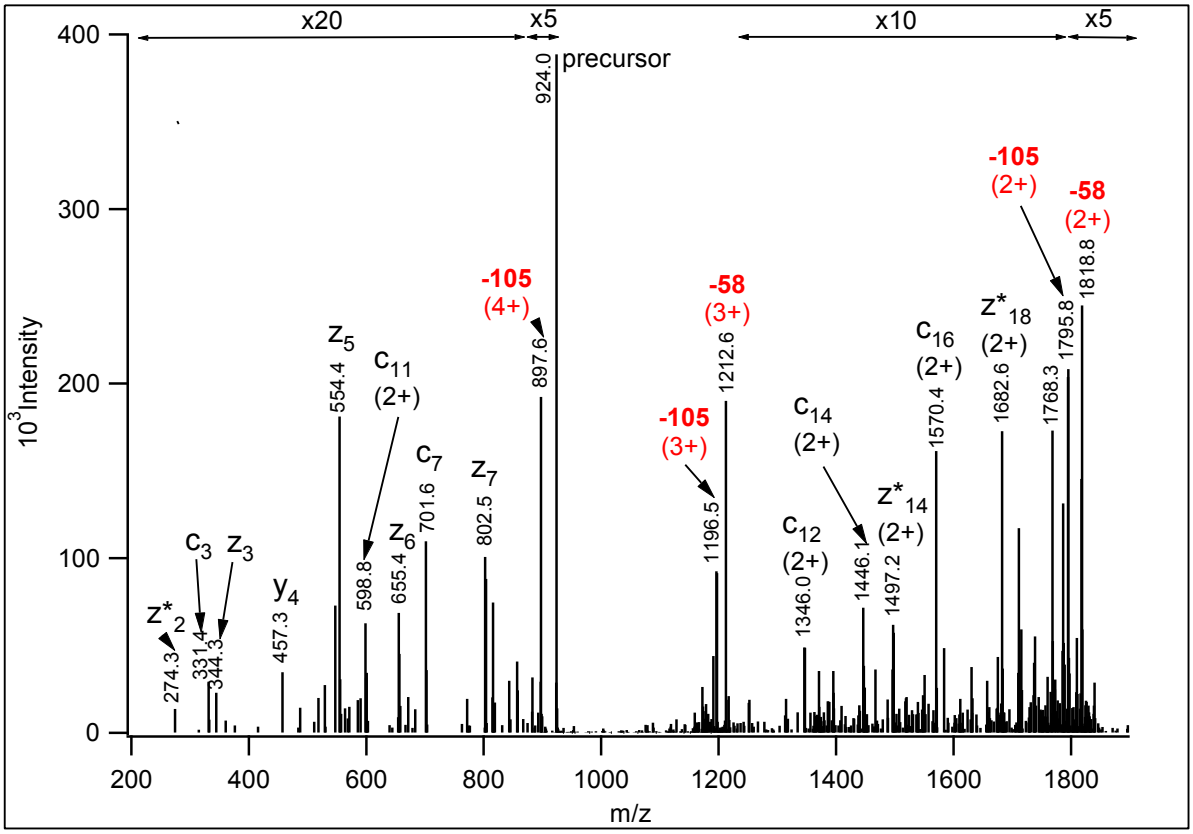

HexNAc
204.086
b2
201.123
HexNAcHex
b3
314.207
366.139
b4
415.256
y4
457.251
y5
570.336
y6
671.378
y7
818.447
y8
919.502
y17-64+203
998.495
y17+203
1030.510
y19+203
Y1-64
1137.570
Y1
1237.629
1205.606
10^3 Intensity
m/z
HexNAc
204.087
b2
201.123
HexNAcHex
b3
314.209
366.139
b4
415.255
y4
457.251
y6
671.378
y17-64
896.946
y17-64+203
998.497
y18-64+203
1049.021
y19-64+203
1105.574
Y1-64
1205.624
Y2-64
1307.139
Y3-64
1388.185
10^3 Intensity
m/z

Figure 2.


x20
x5
x10
x5
precursor
924.0
-105
(4+)
897.6
-58
(3+)
1212.6
-105
(3+)
1196.5
-105
(2+)
-58
(2+)
1818.8
1795.8
1768.3
1682.6
1570.4
1497.2
1446.1
1346.0
1212.6
1196.5
897.6
802.5
701.6
655.4
598.8
554.4
457.3
344.3
331.4
274.3
z5
c11
(2+)
c7
z7
z6
c3
z3
y4
z*2
c16
(2+)
z*18
(2+)
c14
(2+)
z*14
(2+)
c12
(2+)
10^3 Intensity
400
300
200
100
0
200
400
600
800
1000
1200
1400
1600
1800
m/z

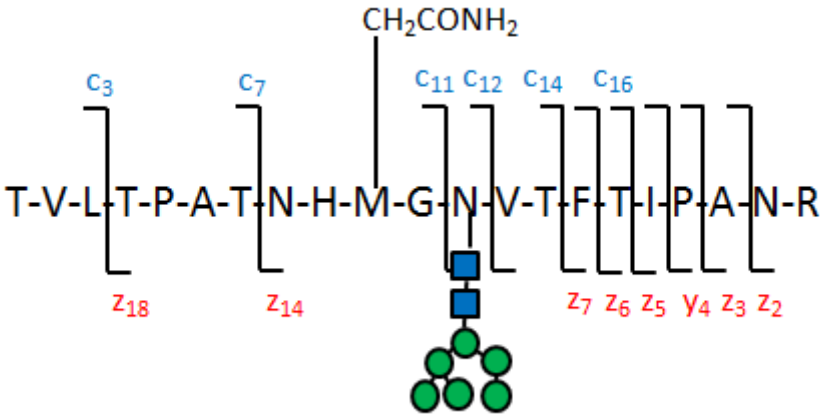
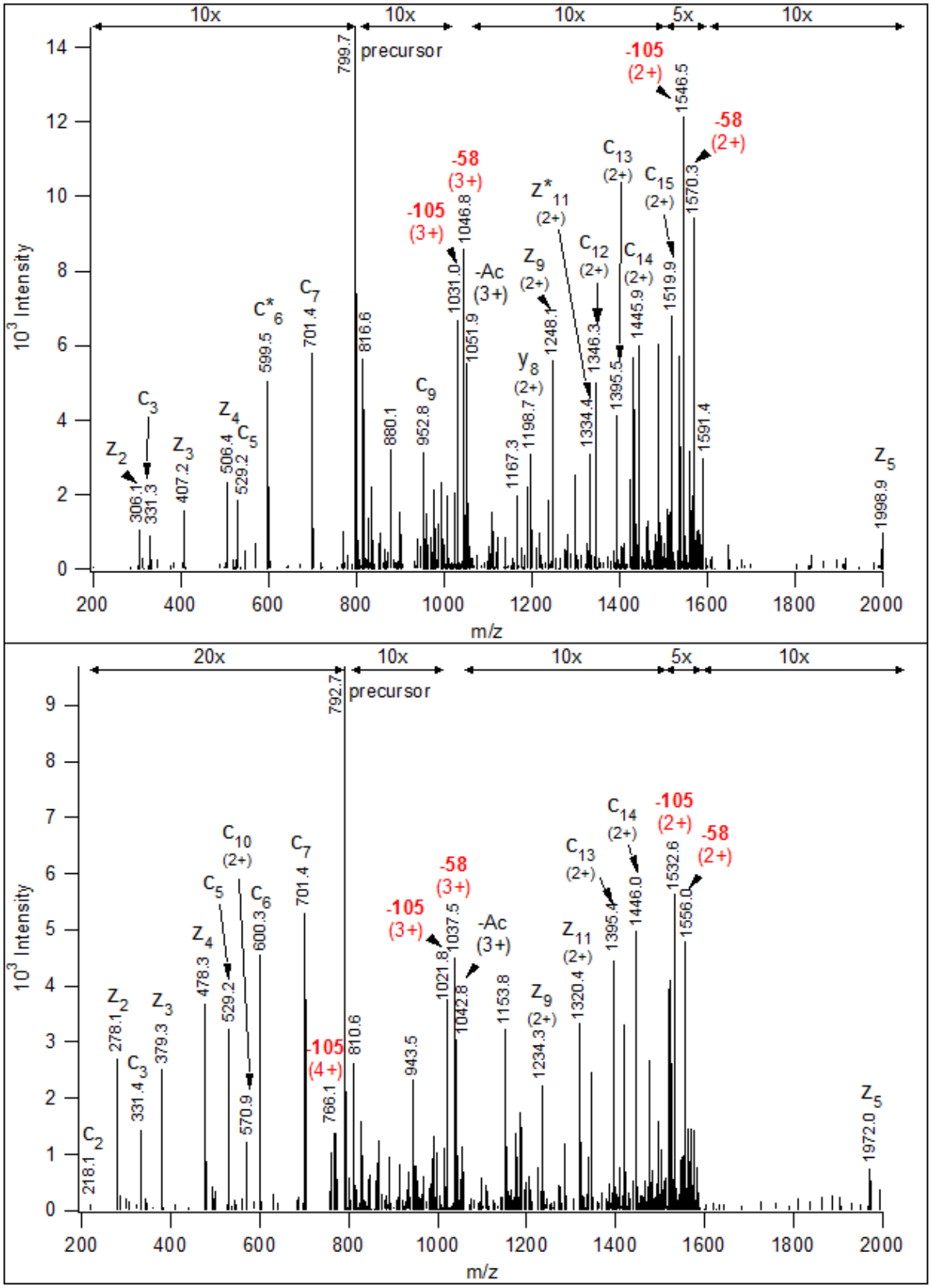

CH2CONH2
c3
c7
c11 c12
c14
c16
T-V-L-T-P-A-T-N-H-M-G-N-V-T-F-T-I-P-A-N-R
z18
z14
z7 z6 z5 y4 z3 z2

Figure 3.


10x
10x
10x
5x
10x
precursor
799.7
-105
(2+)
1546.5
-58
(2+)
1570.3
-58
(3+)
1046.8
-105
(3+)
1031.0
C13
(2+)
C15
(2+)
Z*11
(2+)
C12
(2+)
C14
(2+)
Z9
(2+)
-Ac
(3+)
1051.9
C*6
599.5
C7
701.4
816.6
1248.1
1346.3
1445.9
1519.9
y8
(2+)
1198.7
1395.5
1334.4
C9
952.8
880.1
1167.3
C3
306.1
331.3
Z2
Z3
407.2
Z4
506.4
C5
529.2
1591.4
Z5
1998.9
10^3 Intensity
m/z
20x
10x
10x
5x
10x
precursor
792.7
C10
(2+)
C7
701.4
C5
C6
600.3
-58
(3+)
-105
(3+)
1037.5
-Ac
(3+)
C14
(2+)
-105
(2+)
C13
(2+)
-58
(2+)
1532.6
1446.0
1395.4
1556.0
Z11
(2+)
1320.4
Z4
478.3
Z2
278.1
Z3
379.3
C3
331.4
529.2
570.9
1021.8
1042.8
1153.8
Z9
(2+)
1234.3
-105
(4+)
766.1
810.6
943.5
C2
218.1
Z5
1972.0
10^3 Intensity
m/z

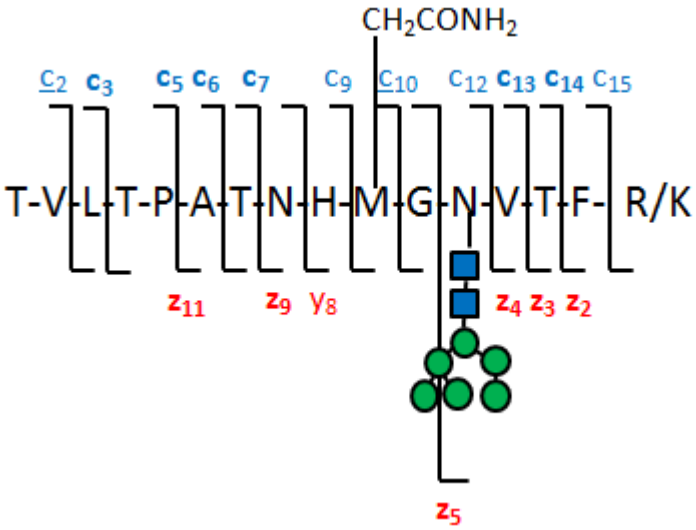

CH2CONH2
c2 c3 c5 c6 c7 c9 c10 c12 c13 c14 c15
T-V-L-T-P-A-T-N-H-M-G-N-V-T-F- R/K
z11 z9 y8 z4 z3 z2
z5

Figure 4.