Letter to the Editor

Non-covalent dimer formation in LC-MS analysis

Katalin F. Medzihradszky, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, San Francisco, CA; & BRC, Institute of Biochemistry, Szeged, Hungary

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

Abstract

Evidence will be presented that in the article "Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides." written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo, published in *Anal Chem*. **2014** *86,* 5478-5486, non-covalent homo- and heterodimers were mis-identified as glycopeptides bearing well-defined N-linked structures, where the unexplained mass was attributed to excessive O-glycosylation. Non-covalent dimer formation of abundant components has not previously been considered as a complication in high throughput proteomic analyses.

Letter

My observations are based on information presented in the article "Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides." written by Sz-Wei Wu, Tsung-Hsien Pu, Rosa Viner, and Kay-Hooi Khoo, published in *Anal Chem*. **2014** *86,* 5478-5486, and the raw data associated with it.

Wu *et al*. beautifully presented that the different MS/MS activation methods: ion trap CID, beam-type CID (HCD) and ETD deliver different information about the glycopeptide structure, underscoring the necessity of using these data combined. They also demonstrated the utility of 'Sweet Heart for HCD', a software developed by them for N-linked glycopeptide analysis. This software heavily relies on the identification of a characteristic/diagnostic N-glycopeptide fragment, Y1, i.e. the intact peptide retaining the innermost GlcNAc. They illustrated the diagnostic value of this fragment in N-linked glycopeptide identification by comparing data interpretation delivered by Sweet Heart and Byonic, a commercially available search engine that is marketed for glycosylation analysis (ProteinMetrics/Thermo). The data analyzed derived from a tryptic digest of recombinant human soluble EGF-receptor and were acquired using an Orbitrap Fusion Tribrid mass spectrometer. In the original publication Supplementary Tables 1-

1 and -2 (fully and partially reproduced as Figs S1 and S11, respectively) sum up the cases where Byonic was obviously wrong, since the Y1 fragment detected contradicted the peptide assignment. These tables themselves are excellent illustrations of some of the issues proteomic studies have to deal with constantly. Quite a few of the mis-assignments might be the results of faulty peak-picking, i.e. of incorrect monoisotopic ion assignments. Unfortunately, this is a recurring issue with high mass, high charge-state and low intensity ions that are more common for large mass glycopeptides than unmodified sequences. Byonic now has its own peak-picking feature that eliminates most of these problems. However, overlapping ion clusters and missing monoisotopic peaks (see Table 1 and precursor ion profiles in Figs S2, S4, S8, S13, S17, & S20 versus calculated values in Figs S3, S7, S10, S14, S18, & S21) still may prevent the accurate determination of the molecular mass. However, the Y1 fragment and carbohydrate fragmentation data together may deliver the correct assignment, i.e. the 'Sweet Heart for HCD' approach could succeed.

There are examples for metal- and ammonia-adduct formation and covalent modifications. Such fortuitous modifications during sample preparation/analysis are the downfall of automated data interpretation. Search engines cannot overwrite the search parameters even when it is blatantly obvious for the human observer that something unexpected happened. For example, in this glycosylation study CID fragmentation indicated that mannose residues were randomly partially formylated (+28 Da)(Fig. S2 in the original paper) or acetylated (+42 Da). Whenever glycans are covalently modified or form non-covalent adducts, even if the nature of the modification cannot be deciphered, the same Y1 fragment shows that spectra are from glycoforms of the same peptide, and this was indeed correctly presented in the article.

The more interesting cases are those mis-identifications when, after correcting for the mass of the peptide sequence based on the Y1 fragment detected, the authors could not account for 1500-3500 Da mass differences (see the list in Figs S1 and S11). The authors concluded the HCD and CID data of m/z

1344.6(5+) (Fig S2) represents the peptide NCTSISGDLHILPVAFR bearing $GlcNAc_2Man_7$ at Asn-1, and an O-linked structure is responsible for the 3278 Da mass that is 'missing'. This is a very unlikely explanation. There are only 3 potential O-glycosylation sites in this peptide. If all these sites were occupied with a SAGalGalNAc structure (one of the most common mammalian O-linked carbohydrates), whose presence is suggested in the CID spectrum by the m/z 657 ion (Fig S2/C), this would only account for 1968 Da. There are no fragments indicating larger O-linked structures. Furthermore, there is no sign of sialic acid in the HCD spectrum (Fig S2/B): its abundant oxonium ions should be seen at m/z 274 and 292; and the ion trap CID spectrum (Fig S2/C; Figure 1/B) does not look typical either. In the CID spectrum one would expect to see an ion series of somewhat even abundance corresponding to the loss of terminal sugars or multi-unit parts of an antennae, as well as a dominant Y1 fragment, just like in Figure 1/A. However, the CID spectrum (Fig S2/C & 1/B) is dominated by 2 species: the NCTSISGDLHILPVAFR peptide bearing $GlcNAc_2Man_6$ and $GlcNAc_2Man_7$ structures, abundant 2+ and 3+ ions were detected for both. This observation gave me the idea that the precursor ion picked for the MS/MS analysis must have been the (5+) heterodimer formed from these glycoforms, i.e.

$[(M_{(Man6)}M_{(Man7)})H_5]^{5+}$

The authors kindly shared the raw data with me. After a thorough investigation I am convinced that all those precursor ions that were reported with the corrected amino acid sequence shortened to NCTSISGDLHILPVAFR, but with a large 'mass defect', represent non-covalent dimers formed in the gas-phase in the mass spectrometer. In addition to the example discussed above, CID and MS data of five additional non-covalent dimers are presented in the Supplement (Figs S4, S8, S12, S16 & S19). The most convincing proof for the existence of heterodimers is shown in Fig. S15, which is the HCD spectrum of a glycopeptide and a unrelated peptide heterodimer. Peptide fragments unambiguously confirm the presence of *both* components. Wu *et al.*, correctly identified the Y1 fragment and thus, the glycosylated

sequence in this 'compound' (as well as for all the other non-covalent assemblies). Nevertheless, Sweet Heart could not deliver the correct interpretation.

Non-covalent gas-phase cluster formation is a known phenomenon. Studies on the fragmentation mechanism of peptide clusters [1] or their utilization for controlled peptide bond formation [2] have been reported. Similarly, non-specific non-covalent gas-phase interaction between peptides and oligosaccharides has been described [3]. More than 20 years ago we reported that abundant analytes could be detected in the high energy CID spectra of any 'background' precursor ion, probably due to non-specific cluster formation between the analytes and the liquid 'matrix'[4]. With this said, I am not aware of any publication when non-covalent dimerization has been considered in the context of high throughput proteomic experiments. One cannot exclude the possibility that non-covalent homo- and heterodimers are formed during nLC/MS experiments from the more abundant components of a complex mixture of high dynamic range (almost all 'real life' samples qualify), and such non-covalent dimers may 'dwarf' the minor components. Ions representing such dimers may be selected for MS/MS analysis, and there is a good chance for mis-interpretation. This danger is obviously enhanced whenever a study is aimed at the characterization of relatively 'rare' large molecules, i.e. whenever precursors of high m/z value, high charge state and relatively low abundance are targeted for MS/MS analysis. Such experiments would include studies of crosslinked peptides from protein complexes, or as in the present example, intact glycopeptide analysis. It is not known how frequently this might occur and thus, how much of a problem this might represent for automated data interpretation. I cannot help but wonder how many times this might have resulted in reporting incorrect structural assignments already.

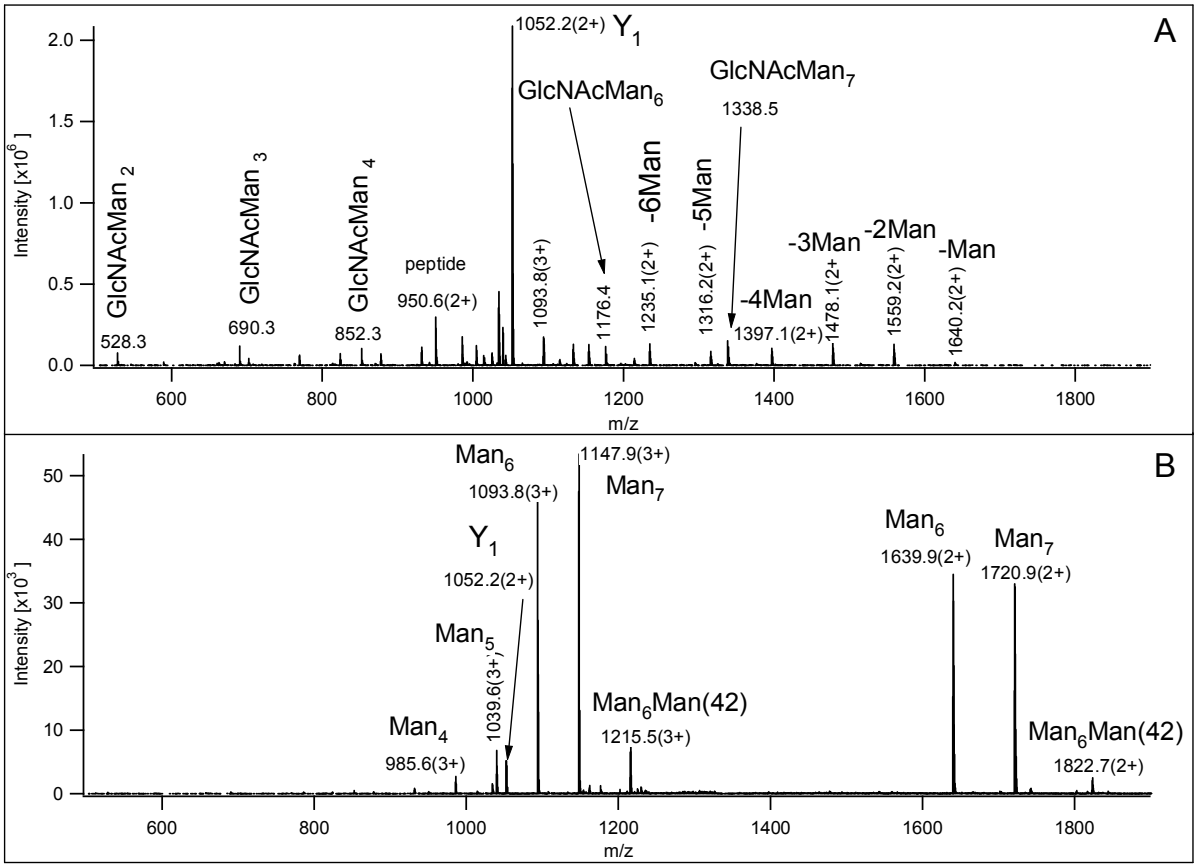

Figure 1. *Panel A* shows the CID data of the NC(Carbamidomethyl)TSISGDLHILPVAFR peptide modified at Asn-1 with a $GlcNAc_2Man_7$ structure, precursor at m/z 1147.5(3+). Oxonium ions are indicated with their sugar compositions, while 'Y' fragments are labeled with the sugar losses. Obviously the identity of the sugar units cannot be determined from these data, so the assignments are based on common knowledge of mammalian N-linked glycosylation. *Panel B* displays the CID of m/z 1344.6(5+). These data suggest that a non-covalent heterodimer of the Man_6 and Man_7 glycoforms was fragmented. Both spectra were acquired in the linear ion trap, so are low resolution and low mass accuracy data.

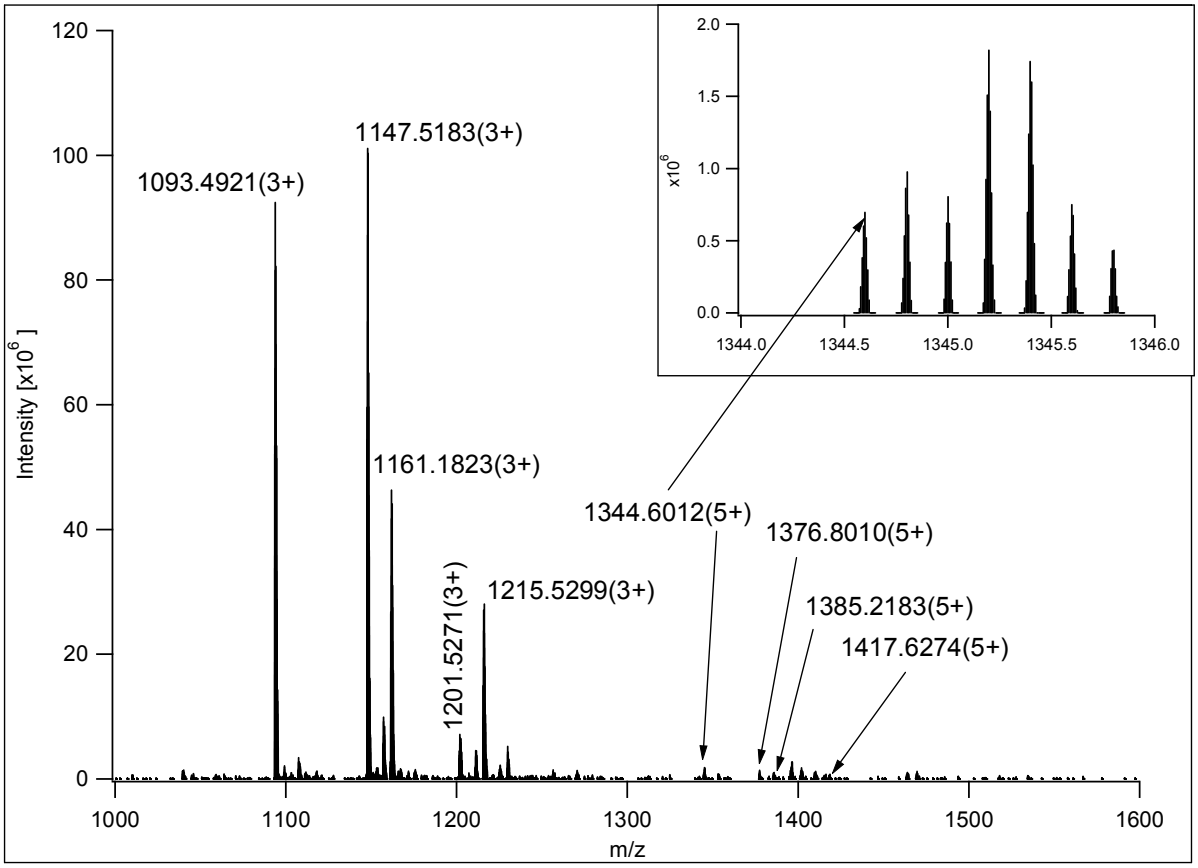

Figure 2. MS survey scan from which the precursor ion for the CID spectrum in Figure 1/B was selected. The inset shows the precursor ion cluster. Table 1 lists the different glycoforms present in this spectrum.

Table 1. Glycoforms of NC*TSISGDLHILPVAFR detected in the spectrum shown in Figure 2

MH^+_{Xtract}	$m/z_{manually}$	z	MH^+_{manual}	Δ ppm	MH^+_{calc}	Relative Intensity	structure, $GlcNAc_2$+
3278.4616	1093.4921	3	3278.4617	3	3278.4502	31.0	Man_6
3294.4471	1098.8206	3	3294.4472	1	3294.4451	0.7	Man_6; Cys(O)
3440.5404	1147.5183	3	3440.5403	11	3440.5031	100.0	Man_7
3456.4944	1152.8363	3	3456.4943	-1	3456.4980	1.8	Man_7; Cys(O)
3468.4984	1156.8376	3	3468.4982	0	3468.4980[a]	4.9	Man_6Man(28)
3482.5202	1161.5116	3	3482.5202	2	3482.5136[c]	19.6	Man_6Man(42)
3493.4240	1165.1427	3	3493.4135	0	3493.4130	3.8	Man_7(Fe)
3498.5193	1166.8466[b]	3	3498.5252	5	3498.5086	1.0	Man_6Man(42); Cys(O)
3602.5667	1201.5271	3	3602.5667	3	3602.5559	14.7	Man_8
3618.5551	1206.8566	3	3618.5552	1	3618.5508	0.6	Man_8; Cys(O)
3630.5606	1210.8584	3	3630.5606	3	3630.5508[a]	4.4	Man_7Man(28)
3644.5752	1215.5299	3	3644.5751	2	3644.5664[c]	24.2	Man_7Man(42)
6717.9590	1344.6012[d]	5	6718.9767	153	6717.9455	0.6	Man_6 & Man_7 dimer
6879.0080	1376.8010	5	6879.9757	-3	6879.9984	1.3	Man_7 dimer
6922.0476	1385.2183[b]	5	6922.0623	8	6922.0089[c]	0.4	Man_7 & Man_6Man(42) Man_6 & Man_7Man(42) dimers
7083.0778	1417.6274	5	7084.1077	6	7084.0617[c]	0.7	Man_7 & Man_7Man(42) dimer

[a] assumed formylation of a Man residue

[b] overlapping ion clusters, ambiguous monoisotopic peak assignment

[c] assumed acetylation of a Man residue

[d] missing monoisotopic peak, probably due to thresholding in data acquisition

C* corresponds to carbamidomethyl Cys; Cys(O) indicates the oxidation of alkyl Cys

The first column represents automated deconvolution values using Xtract, a feature of Thermo's Xcalibur software. The correct value determined for the Man_6 & Man_7 dimer indicates that it can compensate for the missing monoisotopic peak.

Acknowledgement

KFM is supported by NIH grant NIGMS 8P41GM103481, and by the Howard Hughes Medical Institute (to the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF, Director: A.L. Burlingame).

Supporting Information Available. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) Jurchen, J.C.; Garcia, D. E., Williams, E. R. *J Am Soc Mass Spectrom.* **2003** *14,*1373-1386.

(2) Lee, S.; Valentine, S. J.; Reilly, J. P.; Clemmer, D. E. *J Am Chem Soc.* **2011** *133,* 15834-15837.

(3) Comeau, A. N.; Renaud, J. B.; Mironov, G. G.; Berezovski, M. V.; Mayer, P. M. *Int. J. Mass Spectrom.* **2012** *316-318,* 31-39.

(4) Falick, A. M.; Medzihradszky, K. F.; Walls, F. C. *Rapid Commun Mass Spectrom.* **1990** *4,* 318-322.