PTM MarkerFinder, a software tool to detect and validate spectra from peptides carrying post-translational modifications

ARTICLE *in* PROTEOMICS · AUGUST 2013

Impact Factor: 3.81 · DOI: 10.1002/pmic.201300036 · Source: PubMed

6 AUTHORS, INCLUDING:

Paolo Nanni
University of Zurich
54 PUBLICATIONS **524** CITATIONS
SEE PROFILE

Christian Panse
ETH Zurich
49 PUBLICATIONS **1,213** CITATIONS
SEE PROFILE

Jonas Grossmann
ETH Zurich
66 PUBLICATIONS **1,687** CITATIONS
SEE PROFILE

Ralph Schlapbach
ETH Zurich
76 PUBLICATIONS **2,953** CITATIONS
SEE PROFILE

TECHNICAL BRIEF

PTM MarkerFinder, a software tool to detect and validate spectra from peptides carrying post-translational modifications

Paolo Nanni[1], Christian Panse[1]*, Peter Gehrig[1], Susanne Mueller[2], Jonas Grossmann[1] and Ralph Schlapbach[1]*

[1] Functional Genomics Center Zurich, UZH/ETH Zurich, Zurich, Switzerland
[2] Institute of Microbiology, Department of Biology, ETH Zurich, Zurich, Switzerland

Received: January 24, 2013
Revised: March 29, 2013
Accepted: April 24, 2013

Mass spectrometry (MS) analysis of peptides carrying post-translational modifications is challenging due to the instability of some modifications during MS analysis. However, glycopeptides as well as acetylated, methylated and other modified peptides release specific fragment ions during CID (collision-induced dissociation) and HCD (higher energy collisional dissociation) fragmentation. These fragment ions can be used to validate the presence of the PTM on the peptide. Here, we present PTM MarkerFinder, a software tool that takes advantage of such marker ions. PTM MarkerFinder screens the MS/MS spectra in the output of a database search (i.e., Mascot) for marker ions specific for selected PTMs. Moreover, it reports and annotates the HCD and the corresponding electron transfer dissociation (ETD) spectrum (when present), and summarizes information on the type, number, and ratios of marker ions found in the data set. In the present work, a sample containing enriched *N*-acetylhexosamine (HexNAc) glycopeptides from yeast has been analyzed by liquid chromatography-mass spectrometry on an LTQ Orbitrap Velos using both HCD and ETD fragmentation techniques. The identification result (Mascot *.dat* file) was submitted as input to PTM MarkerFinder and screened for HexNAc oxonium ions. The software output has been used for high-throughput validation of the identification results.

Keywords:
Data mining / High-throughput proteomics / Interpretation of mass spectra / Mass spectrometry / Post-translational modifications

Additional supporting information may be found in the online version of this article at the publisher's web-site

The assignment of stable post-translational modifications (PTMs) by liquid chromatrography-mass spectrometry (LC-MS/MS) analysis is well established. However, the characterization of peptides carrying labile modifications still remains challenging due to the instability of these PTMs during MS fragmentation.

For example, collision-induced dissociation (CID) spectra of phosphorylated peptides are characterized by intense peaks corresponding to the neutral loss of the phosphate group. Acetylated and methylated residues release distinctive immonium ions (m/z 98.0964 for mono-methylated lysine, m/z 126.0913 and 143.1179 for acetylated lysine) and unique neutral fragments (MH^+ 59.0735 for tri-methylated lysine) during CID fragmentation, making it possible to distinguish between acetylated and tri-methylated lysine [1].

Correspondence: Dr. Paolo Nanni, Functional Genomics Center Zurich, UZH/ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
E-mail: paolo.nanni@fgcz.uzh.ch
Fax: +41-44-6353922

Abbreviations: CID, collision-induced dissociation; **ETD**, electron transfer dissociation; **HCD**, higher energy collisional dissociation; **Hex**, hexose; **HexNAc**, N-acetylhexosamine; **NeuAc**, N-acetylneuraminic acid; **PQD**, pulsed Q collision-induced dissociation

*These authors contributed equally to this work.

N- and *O*-linked glycans are another class of modifications that form characteristic low-mass product ions during CID, higher energy collisional dissociation (HCD), or pulsed Q collision-induced dissociation (PQD) [2]. Some of these ions are oxonium ions corresponding to monosaccharides, for example, N-acetylhexosamine ($HexNAc^+$, m/z 204.0866), hexose (Hex^+, m/z 163.0601), and N-acetylneuraminic acid ($NeuAc^+$, m/z 292.1032). In addition, glycopeptide spectra contain signals resulting from internal fragmentation of HexNAc (m/z 186.0761, 168.0655, 144.0655, 138.0550, 126.0550) or Hex (m/z 145.0495, 127.0390, 115.0390, 109.0284) [3]. MS/MS fragmentation of N- and *O*-linked glycopeptides may also lead to the formation of marker ions containing more than one monosaccharide unit, for example, 325.1135 ($Hex\text{-}Hex^+$) and 366.1400 ($Hex\text{-}HexNAc^+$).

Peptides with labile modifications have been fragmented using different techniques, and their characteristic fragmentation patterns have been used to identify, validate, and localize the modification site. Methods based on additional activation of modified peptides after neutral loss of the modification (MS3, multistage activation) or on alternative fragmentations (electron capture dissociation (ECD) and electron transfer dissociation (ETD)) have been proposed and successfully applied to sequence modified peptides [4].

A common MS-based approach takes advantage of the glycan-specific markers: first, HCD or PQD spectra are acquired and screened for these characteristic oxonium ions to detect potentially modified peptides. The precursors are then fragmented by ETD for peptide sequencing and modification site localization. This approach has been successfully applied to the identification and characterization of glycopeptides [3, 5–7].

HCD and ETD fragmentations can be performed on every peptide in a single run, in a two stage approach (a discovery LC-MS/MS run for the detection of putative glycosylated peptides, followed by a targeted ETD LC-MS/MS analysis [3]), or in a more elegant combination of fragmentation techniques, named HCD product ion-triggered ETD (HCD-PD-ETD) [3, 6, 7]. In the latter, ETD spectra are acquired only when the oxonium ions of interest are detected in the HCD MS/MS spectra, reducing the instrument duty cycle, while improving the sensitivity and simplifying the data analysis. Potential pitfalls of this approach include a faulty triggering of the ETD event, the requirement of a priori knowledge of the triggering ions, and different peptide assignments from HCD and ETD spectra.

In any case, mass spectrometric data from glycoproteomic experiments often contain hundreds of glycopeptide fragmentation spectra, which need to be validated. Although manual validation of all individual spectra in such experiments is obviously not feasible, annotated spectra of the modified peptides are required for publication and reviewing purposes.

Software tools to screen large data sets for potentially modified peptides and validate the assignment of peptide sequences are necessary. PTM MarkerFinder is a tool that selects spectra of interest and greatly facilitates the visualization and quality assessment of the assignment of modifications. Briefly, the PTM MarkerFinder software searches a posteriori for any type of product ion(s) appearing in the HCD or PQD spectra, reports and annotates the corresponding ETD spectrum (when present) and summarizes information on the type, number, and ratios of product ions found in the data set. An overview of the data analysis workflow using PTM MarkerFinder software is shown in Fig. 1.

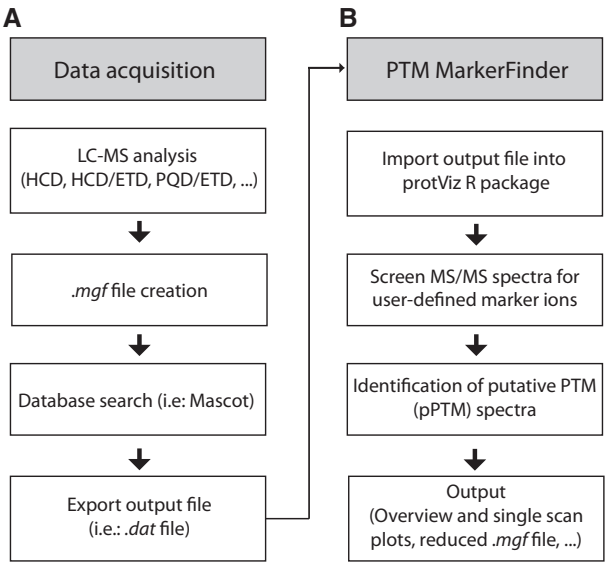

Figure 1. Overview of the PTM MarkerFinder workflow: (A) the samples containing modified peptides are analyzed by LC-MS/MS analysis using HCD/PQD fragmentation technique (with or without additional ETD fragmentation). Spectra are converted into mgf files, the database searches are performed (i.e., Mascot), and the outputs are exported (i.e., as .dat file); (B) PTM MarkerFinder screens the spectra for user defined marker ions, extracts the spectra of *putative PTM* (*pPTM*) peptides, and produces graphical and text outputs.

In the described approach spectra of the sample of interest are acquired by LC-MS/MS on an LTQ-Orbitrap Velos instrument using a combination of HCD/ETD or PQD/ETD fragmentation techniques (Fig. 1A). Since the only prerequisite for PTM MarkerFinder is the presence of marker product ions, the input LC-MS/MS data set could consist only of HCD or PQD. The use of CID fragmentation on LTQ-Orbitrap instruments is not suggested, because of the low mass cutoff limitation inherent to all ion traps.

For every sample, the protein identification is performed using search engines such as Mascot (Matrix Science), one of the leading protein database interrogation software for MS data. In case the expected modification is present in the database, it can be selected for database searches (i.e., HexNAc and Hex).

PTM MarkerFinder (Fig. 1B) then screens the MS/MS spectra in the output file (i.e., Mascot *.dat* file) for marker ions specific for selected PTMs. The software tool is a function of

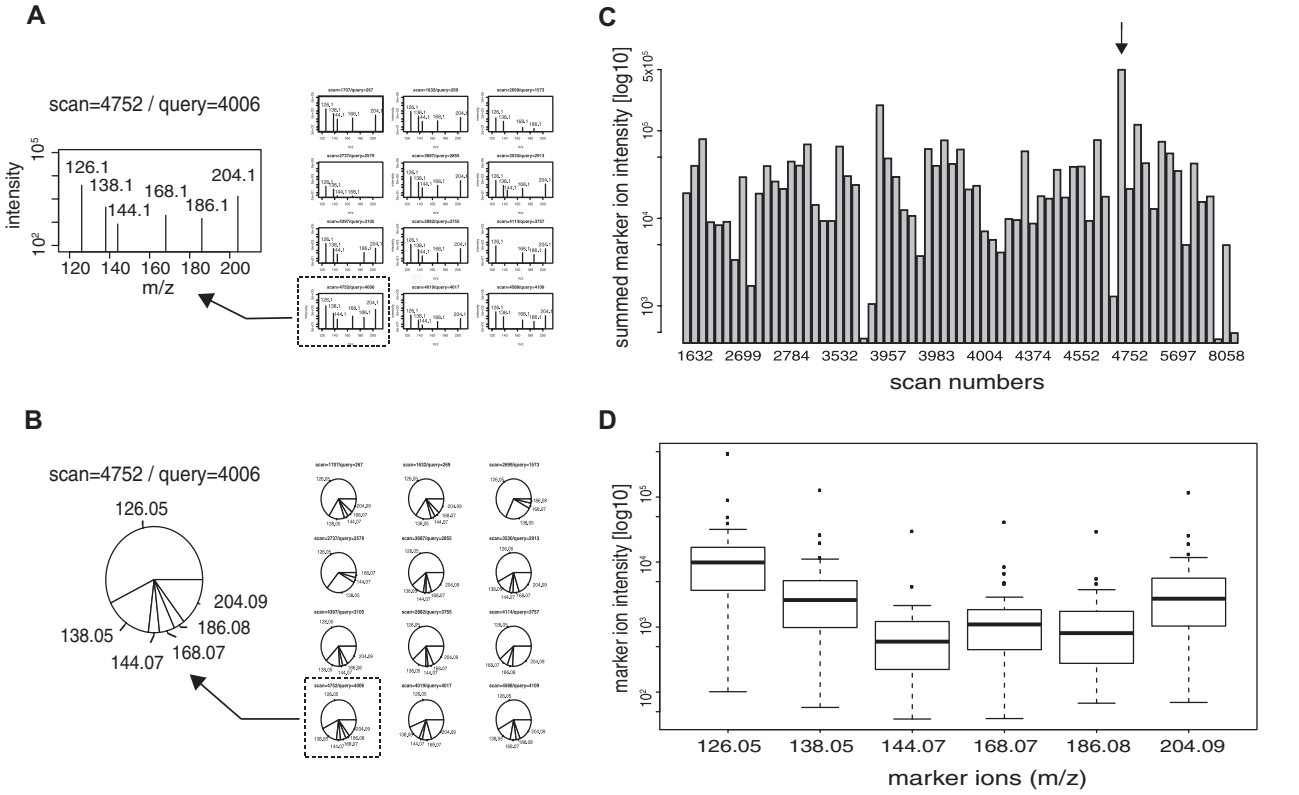

Figure 2. Example of PTM MarkerFinder output from a search for N-HexNAc modifications on a single Mascot output .dat file: (A) overview of the marker ion intensities of the HCD *pPTM spectra*; (B) pie chart for each pPTM spectrum showing the contribution of single marker ions to the total marker ion intensity. In both (A) and (B) only few *pPTM spectra* are reported here, and scan 4752 is highlighted; (C) box plots of marker ion intensities from all *pPTM spectra* (the arrow indicates the scan 4752); (D) summed marker ion intensities for each *pPTM spectrum*.

the R package protViz that requires an R-object containing the mass spectrometric measurement, the database search results, and a list of marker ions (http://cran.r-project.org/web/packages/protViz/index.html, version higher than 0.1.25) [8]. The R-object can be created from any search engine output; however, only the perl script for the conversion of Mascot *.dat* file is included in protViz ("mascotDat2RData.pl"). The PTM MarkerFinder function iterates over each MS/MS spectrum of the mass spectrometric measurement and searches for the marker ions specified by the user (default window is 20 ppm). Since the *.dat* file does not provide information on the type of fragmentation technique applied, PTM MarkerFinder considers the tandem mass spectrum containing the marker ions as HCD scan. Spectra containing at least two product ions and with the sum of the marker ion intensities representing at least 2% of the total ion intensities are annotated with the corresponding ion series (*b, y*) using the protViz:peakplot method and are subsequently called *putative PTM* (*pPTM*) *spectra*. Afterwards, the function screens the succeeding scans for the corresponding ETD scan acquired from the same peptide ion and plots its peptide sequence assignment (*c, z*, and *y* ions). Note that the PTM MarkerFinder expects the ETD scan immediately after the HCD scan. If the MS acquisition changes the method has to be adapted accordingly. The final PTM MarkerFinder output contains information about the presence and the intensity of marker ions in *pPTM spectra* and peptide sequence assignments (Figs. 2 and 3).

We tested the PTM MarkerFinder workflow on a sample containing enriched glycopeptides from yeast. A detail description of the procedure can be found in the Supporting Information. In short, membrane proteins were extracted from log phase yeast cells and digested using the filter aided sample preparation protocol by Wisniewski et al. [9]. Peptides were mixed with Concanavalin A (ConA) lectin and loaded onto 30 kDa cutoff filter units according to the protocols by Zielinska et al. [10]. Unglycosylated peptides were washed away while ConA-bound glycopeptides were retained on the filter. Finally, they were released with endoglycosidase H, which cleaves *N*-linked mannose-rich oligosaccharides from glycopeptides, generating a truncated sugar molecule with one HexNAc residue remaining on the asparagine. This procedure resulted in the enrichment of HexNAc glycopeptides. The sample was analyzed by LC-MS/MS on an LTQ-Orbitrap Velos (Thermo Scientific) instrument using a combination of HCD and ETD fragmentation techniques. The raw file was converted to a Mascot Generic Format file (mgf) using Proteowizard (http://proteowizard.sourceforge.net/) and the

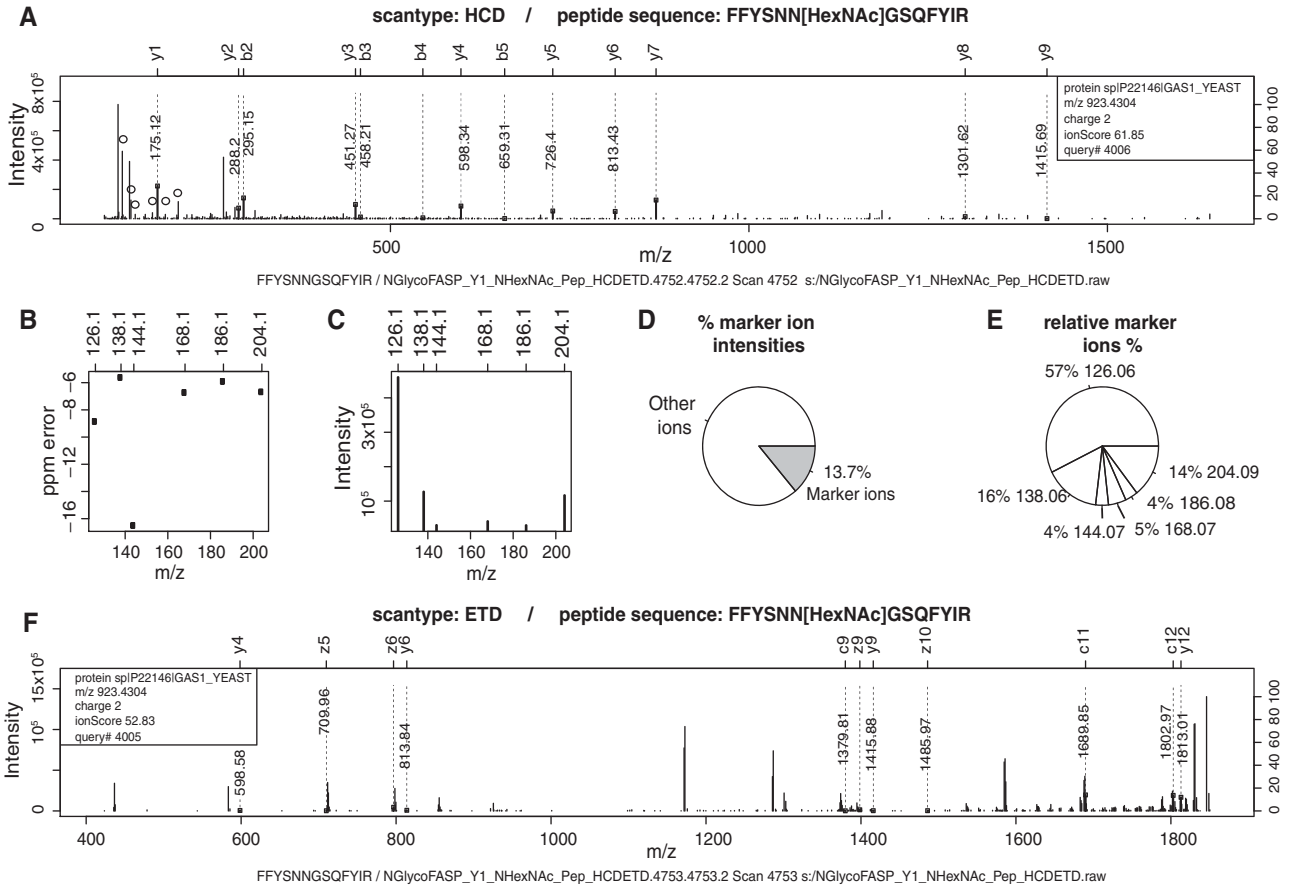

Figure 3. For every *putative PTM spectrum*, the following plots are provided: (A) HCD spectrum, with marker ions highlighted (circles) and peptide sequence assignment; (B) marker ion mass deviation (ppm); (C) marker ion intensities; (D) pie chart showing the contribution of summed marker ion intensities to the total fragment intensity; (E) pie chart showing the contribution of single marker ions to the total marker ion intensity; (F) ETD spectrum (plotted only if present and assigned).

HCD spectra were deconvoluted using H-score [11]. Protein identification was performed using Mascot v2.3 and the resulting *.dat* file was submitted to PTM MarkerFinder to screen for HexNAc specific marker ions (as described earlier).

The output provides the following information: (i) plots summarizing the distribution of marker ions in the *pPTM spectra* during the whole measurement (Fig. 2); (ii) plots reporting the peptide sequence assignment and the marker ion information for each *pPTM spectrum* (Fig. 3); (iii) *.csv* table containing the list of marker ions found in each *pPTM spectrum*; (iv) reduced mgf file, containing the m/z values and the intensities of the *pPTM spectra*.

The plots (i) summarize the number and distribution of putative modified peptides in the analyzed LC-MS run and facilitate a prescreening of the data quality. In Fig. 2A, an overview of the marker ion intensities in the *pPTM* HCD *spectra* is given. This allows the detection of scans containing several marker ions (i.e., scan 4752, with 6 HexNAc oxonium ions). Figure 2B and C display a pie chart for each *pPTM spectrum* representing the contribution of single marker ions to the total marker ion intensity, and the summed marker ion intensities for each *pPTM spectrum*, respectively. From Fig. 2C, we understand that scan 4752 shows higher marker ion intensities compared to other *pPTM spectra*. The box plot in Fig. 2D shows the marker ion intensities from all *pPTM spectra*. The reproducibility in the intensity of the HexNAc signature ions can be observed. By using this output, 74 *pPTM* HCD *spectra* have been extracted from the 4083 HCD spectra present in the *.dat* file. The efficiency of PTM MarkerFinder in extracting spectra of putative glycopeptides and in discarding false-positives has been validated by manual comparison to the XIC (extracted ion chromatogram) of the frequently observed ion 204.0866 (Supporting Information Fig. 1). Screening with less stringent settings (intensity and number of marker ions) increased the number of detected *pPTM spectra* (Supporting Information Table 1), but the added *pPTM spectra* did not contain any additional Mascot glycopeptide identifications. This shows that the filter settings are crucial to discard false-positives, that the number of HexNAc glycopeptides in the sample was rather low and that with our initial settings we were able to pinpoint all glycopeptide spectra.

In addition to the annotated spectra, PTM MarkerFinder reports the identified peptide sequence, the type and localization of the modification, the Mascot identification score, the scan, and the query number. When no significant sequence assignment of the *pPTM* HCD *spectrum* is achieved, the uninterpreted spectrum is shown. Corresponding unassigned ETD spectra are not plotted. An example of PTM MarkerFinder output (ii) for a HexNAc modified peptide is reported in Fig. 3. The annotated HCD and ETD spectra are displayed in Fig. 3A and F, respectively (scan 4752 and 4753).

PTM MarkerFinder highlights the identified marker ions (Fig. 3A (circles) and Fig. 3C) and shows their mass deviation (ppm) (Fig. 3B). Moreover, it calculates the relative intensity ratios of such ions and represents them as pie charts (Fig. 3D and E), allowing a quick visual validation of the Mascot results. As an example: if Mascot identifies a HexNAc(N) modification, but the HCD spectrum does not contain any oxonium ion characteristic for this modification, the Mascot assignment can be considered as false. On the other hand, spectra showing a clear HexNAc(N) fragment ion pattern that do not give any significant Mascot output could be considered for manual interpretation. In the example shown, we can see at a glance that both HCD and ETD identified the same peptide sequence and modification site.

The combination of all the information reported in Figs. 2 and 3 greatly helped the validation of spectra from glycopeptides and the identification of additional spectra from glycopeptides that remained unassigned during protein database searches. In total, PTM MarkerFinder allowed the validation of 24 peptide sequences by both HCD and ETD spectra. The localization of an additional 13 glycosites was validated only by ETD (9) or HCD (4). The remaining *pPTMs* HCD *spectra* could not be confidently assigned to any peptide sequence neither by HCD nor by ETD fragmentation, but since they contain 3–6 marker ions, they most probably represent glycopeptides.

In addition to the plots (i) and (ii), PTM MarkerFinder returns a summary table (*.csv*) containing the following column attributes: "scans," "*m*/*z*," "markerIonMZ," "markerIonIntensity," "markerIonMzError," "markerIonPpmError," and "query" which can be used for statistics (iii). The reduced mgf file (iv) contains only the *pPTM* HCD *spectra* and their corresponding ETD scans, allowing to rerun Mascot searches using different parameters while decreasing dramatically the required computational time. The complete data set is available in the Supporting Information.

Since PTM MarkerFinder is able to search a posteriori for any type of product ion(s), it can be used on data sets containing any type of modification producing marker ions upon HCD or PQD MS/MS fragmentation. To test this, a tryptic digest of yeast proteins was analyzed by LC-MS/MS using HCD fragmentation. The spectra were searched for glycan signatures (such as HexNAc (N) and Hex (S,T)), and methyl (K,R), dimethyl (K,R) and acetyl (K) modifications. Despite the low specificity of some marker ions, PTM MarkerFinder allowed the validation of 45 modified peptides. All the results are reported in the Supporting Information (Table 2). As a conclusion, our software can be applied to validate not only various types of glycosylations, but also completely unrelated protein modifications.

This work was supported by ETH Zurich and University Zurich.

The authors have declared no conflict of interest.

References

[1] Matthiesen, R., Trelle, M. B., Højrup, P., Bunkenborg, J., Jensen, O. N., VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J. Proteome Res.* 2005, *4*, 2338–2347.

[2] Huddleston, M. J., Bean, M. F., Carr, S. A., Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* 1993, *65*, 877–884.

[3] Zhao, P., Viner, R., Teo, C. F., Boons, G.-J. et al., Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* 2011, *10*, 4088–4104.

[4] Boersema, P. J., Mohammed, S., Heck, A. J., Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.* 2009, *44*, 861–878.

[5] Hahne, H., Kuster, B., A novel two-stage tandem mass spectrometry approach and scoring scheme for the identification of O-GlcNAc modified peptides. *J. Am. Soc. Mass Spectrom.* 2011, *22*, 931–942.

[6] Singh, C., Zampronio, C., Creese, A., Cooper, H. J., Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J. Proteome Res.* 2012, *11*, 4517–4525.

[7] Mayampurath, A. M., Wu, Y., Segu, Z. M., Mechref, Y., Tang, H., Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. *Rapid Commun. Mass Spectrom.* 2011, *25*, 2007–2019.

[8] Panse, C., Grossmann, J., protViz: Visualizing and Analyzing Mass Spectrometry Related Data in Proteomics. 2012, R package version 0.1.26

[9] Wiśniewski, J. R., Zougman, A., Mann, M., Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* 2009, *8*, 5674–5678.

[10] Zielinska, D. F., Gnad, F., Wiśniewski, J. R., Mann, M., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 2010, *141*, 897–907.

[11] Savitski, M. M., Mathieson, T., Becher, I., Bantscheff, M., H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. *J. Proteome Res.* 2010, *9*, 5511–5516.

