

Higher collision energy fragmentations for analysis of multiply phosphorylated peptides

Michaela Scigelova¹, Matthias Biel¹, Sajid Bashir², Kyle Beran³, Martin Hornshaw¹, Anastassios Giannakopoulos¹

¹Thermo Fisher Scientific, Bremen, Germany; ²Texas A&M University-Kingsville, Chemical Biology Research Group, Department of Chemistry, Kingsville, TX, 78363

³The University of Texas of the Permian Basin, Department of Physical Sciences, Odessa, TX 79762

Overview

Purpose: Mass spectrometric analysis of multiply phosphorylated peptides; evaluation of information content of MS/MS spectra acquired with different fragmentation techniques.

Methods: Synthetic standards of peptides with 1-4 phospho-residues per molecule were analyzed using CID and HCD (higher energy collision dissociation) with high mass accuracy/resolution detection. Spectra were manually interpreted, and a software tool written in-house was used.

Results: The manual inspection revealed that optimised HCD spectra of multiply phosphorylated peptides contained enough information to assign unambiguously the location of the phosphorylation site(s). Software tools currently used for database searches, however, show a lack of discerning power for correct phosphosite(s) assignment when multiple phosphorylations occur within the peptide sequence. Improved software tools are needed to efficiently extract the diagnostic information present in HCD spectra to improve the performance and throughput of phosphopeptide analysis.

Introduction

Analysis of phosphorylated peptides by collisional dissociation in a mass spectrometer is more difficult than that of non-phosphorylated peptides because the loss of the phosphate group(s) is the predominant fragmentation pathway. This results in low yield of peptide backbone fragments that are the basis of subsequent identification and phosphorylation site assignment. The fragmentation of phosphorylated peptides can be improved significantly by using phosphorylation-directed multistage tandem MS where an ion originating from a neutral loss signal detected in the MS² stage is selected for a second round of fragmentation (MS³) to obtain more peptide sequence information (1). Unfortunately, this strategy favors the analysis of monophosphorylated peptides because multiply phosphorylated peptides will subsequently lose more phosphoric acid and give mostly ambiguous assignment results. Electron capture/transfer dissociation (ECD/ETD) is another fragmentation technique that can be applied to multiply phosphorylated peptides because this method primarily results in peptide backbone fragmentation without concomitant loss of the phosphate groups (2, 3). Implementing yet another technique, a higher energy collisional dissociation (HCD), on modern hybrid ion trap-Orbitrap mass spectrometers (4) offers such advantages as no low-mass cut off (enabling detection of phospho-Tyr immonium ions, for instance), relatively low intensities of neutral loss peaks within the HCD spectrum, and, last but not least, a fully automated integration with other fragmentation techniques within the analysis method.

In this preliminary study we investigated in detail the information content of HCD fragmentation spectra obtained for multiply phosphorylated peptides. We tried to find out whether HCD provides more useful information than current search engines can utilise. What type of information is it? Can we formulate some suggestions for software improvement?

Methods

Synthetic phosphopeptide solutions (5 pmol/μl; kindly provided by Karl Mechtler, IMP Vienna, Austria) were analyzed by infusion (3 μl/min) in MeOH/water (50:50) containing 1% formic acid, using CID and HCD fragmentation at various energies. Detection at 100,000 resolving power in the LTQ Orbitrap™ mass spectrometer was used. The spectra were searched with Mascot™ against a human database to which the sequences of the synthetic phosphopeptides were added, and the (in)correct assignment of phosphorylated residue within the sequence was noted. HCD fragmentation spectra for the synthetic phosphopeptides were also examined manually, and information contained there was compared to that utilized by the Mascot search engine. Based on this information, a prototype software tool was developed to be used in conjunction with Mascot searches.

Results

Interpreting fragmentation spectra of phosphopeptides, namely, assigning the correct position(s) of the phosphate residue(s), remains a challenging task. The capabilities of commercial software such as SEQUEST and Mascot have been 'extended' with software tools developed by individual research groups (such as Ascore (5) and Modificomb (6)). While most of these handle well the phosphosite assignment for mono-phosphorylated peptides, they have not been specifically designed for multiply phosphorylated peptides. This is an issue, in particular as recently developed phosphopeptide enrichment methods (7) start bringing more multiply phosphorylated peptides to our attention.

The first question we wished to answer was related to the quality of information contained in CID vs. HCD spectra. Figure 1 shows a typical outcome of CID fragmentation – neutral losses of phosphate dominate the MS/MS spectrum over a wide range of energies. These spectra did not yield any identification.

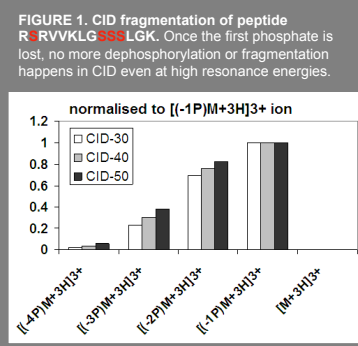
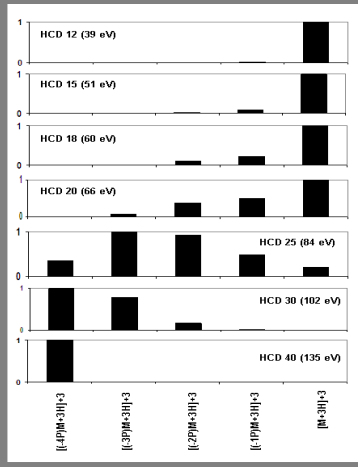


FIGURE 2. HCD fragmentation of peptide RFRVVKLGSSSLGK. Increasing collision energy settings promotes both progressive dephosphorylation of the precursor peptide and its fragmentation.

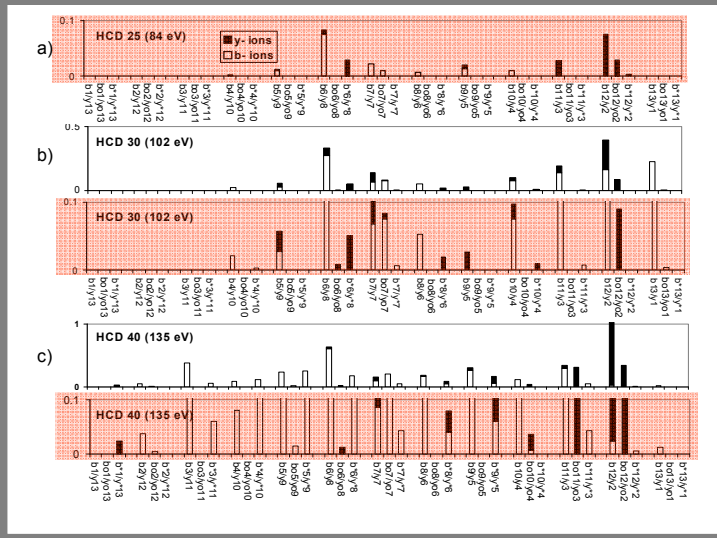


HCD fragmentation, on the other hand, promotes both progressive dephosphorylation (Figure 2) and fragmentation of the peptide. Manual interpretation of 548 peaks (using cut off S/N = 2) assigned 63 fragments as belonging to the analyzed phosphopeptides, while a standard Mascot search assigned 28 fragments of 130 used after spectrum pre-processing. The main issue encountered with our multiply phosphorylated peptide data was that Mascot identification scores were below the confidence threshold, meaning that in a large dataset such phosphopeptide assignments would escape detection altogether.

Mascot settings did not consider a-ions (11 fragments found by manual interpretation) or 3+ fragments (5 fragments found by manual interpretation). Moreover, a lot of information about phospho-site localisation could be coming from the fragmentation series of a partially dephosphorylated precursor peptide. At present, though, only fully dephosphorylated fragment series for peptides with multiple phosphorylation sites are considered in a standard Mascot search for practical reasons.

Based on the manual interpretation of the spectra for multiply phosphorylated peptides (see example in Figure 3) we conclude that there is a lot more information present in HCD spectra which is potentially useful for the correct assignment of the phosphosite(s) but is currently not used by the search engine.

FIGURE 3. Progressive fragmentation of peptide bonds within peptide RFRVVKLGSSSLGK with HCD at different energy settings. Fragments b_n-y_n⁻, b_n⁻-y_n⁰ (loss of H₂O), and b_n⁻-y_n⁻ (loss of NH₃) are shown. Highlighted boxes show enlarged region of less than 0.1% base peak intensity (data normalized to base peak).



To proceed further with our investigations it was clear we needed a specific software tool which would enable extraction of relevant information from the fragmentation spectra of multiply phosphorylated peptides in an automated fashion to avoid tedious manual interpretation. Figure 4 shows a flow chart diagram for the prototype software tool.

First, .mgf format is generated from each spectrum to be searched and sent to Mascot with a relevant set of search parameters (including phosphorylation as a variable modification). If the candidate sequence is a phosphopeptide, data is further processed as follows: for each possible positional isomer/permutation of a phosphorylation site on a given candidate sequence a table of fully and partially dephosphorylated fragments is calculated. The values in the table are matched to the peaks existing in the acquired spectrum (min S/N threshold of 2 required) within a set tolerance (+/- 2 ppm).

The outcome of the 'match' is consolidated for each particular peptide bond fragmentation by adding together the intensities of related fragments. For instance, for a particular fragment b_n, the intensities of observed monoisotopic peaks of a_n⁻, b_n⁻, b_n⁰- and b_n⁻ were added together. Similarly, for a particular fragment y_n⁻ the intensities of observed monoisotopic peaks of y_n⁻, y_n⁰- and y_n⁻ were added together. Moreover, the fragment intensities of partially dephosphorylated precursors were used to boost the overall intensity of the relevant fully phosphorylated species, thus enabling Mascot to make use of this information.

The resulting list of b- and y- fragments with their boosted intensities was then transcribed into an .mgf format and used for the second round of Mascot search.

The results from the first and second Mascot search were compared (manually) and if in agreement with respect to the peptide amino acid sequence and number of phosphorylation sites, the final hit of the second search was reported as a final hit.

As the next step, we focused on the optimisation of collision energy used for HCD in order to gain maximum information for phosphosite assignment. We chose a set of phosphopeptides having the same amino acid sequence (WWGSGPSGGSGGGK); potential sites of phosphorylation highlighted) but differing in number and location of the phosphorylated residues (1-3). Fragmentation was carried out at energy range 30-70 (normalized collision energy values). The new software tool was used to process the spectra and the results for each analyzed spectrum were evaluated. The results indicated that a 'normalised collision energy' setting of 50 provided the most informative spectra leading to the correct sequence identification for all synthetic phosphopeptides used.

Processing the spectra from optimised HCD acquisition conditions (i.e. normalized collision energy 50) with our software tool resulted in the improvement in expectancy values which brought the correct hit to the top for all phosphopeptides analyzed. The expectancy values improved by at least 10 orders of magnitude (Table 1) while no such remarkable improvement was noted for 'incorrect' phospho-candidate sequences.

FIGURE 4. Data processing flow-chart with the new software tool. *Fragment tables contain calculated m/z of all possible fragments (a-, b-, y-, b_n⁻, y_n⁰-, b_n⁻-, y_n⁻-series) from fully or partially dephosphorylated precursor phosphopeptides. †Consolidation here denotes reduction of the number of related peaks while at the same time adding together their absolute intensities.

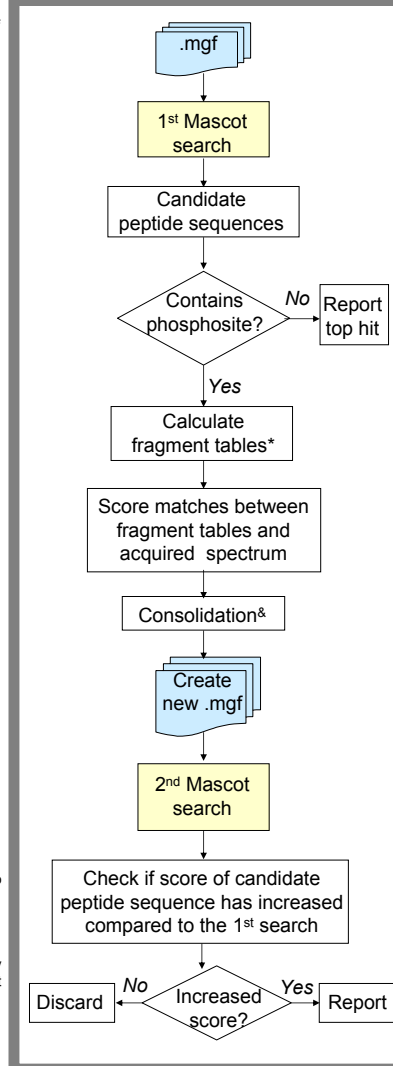


Table 1. Analyzed phosphopeptides with expectancy values from the Mascot search before (Exp value 1) and after (Exp value 2) processing with our software tool. In all cases, a normalized collision energy setting of 50 was employed for HCD fragmentation.

Sequence	Exp value 1	Exp value 2
WWGSGPSGGSGGGK	1.6 e5	2.4 e-5
WWGSGPSGGSGGGK	9.5 e4	1.0 e-6
WWGSGPSGGSGGGK	4.8 e4	1.3 e-9

The improvement in the expectancy values reflects several aspects of data processing: 1) removal of noise in the MS/MS spectrum (min. S/N=2); 2) considering fragments from partially dephosphorylated precursor which are not considered in normal Mascot search; 3) boosting ion intensities of relevant fragments by adding together intensities of related fragments coming from different fragment series, charge states, etc.

Conclusions

- We verified by manual interpretation that HCD fragmentation provides phosphopeptide spectra with high information content relevant for both sequence identification and phosphosite assignment.
- We obtained only low significance hits when using the Mascot search engine for the identification of our multiply phosphorylated peptides from HCD spectra.
- Therefore, we have developed a prototype software tool to pre-process the HCD spectra of multiply phosphorylated peptides to be used in conjunction with the Mascot search engine.
- By applying this software routine, we obtained the correct amino acid sequence, location of phosphorylation(s) and identification confidence above significance threshold for search in a large protein database.

References

- Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J. X., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12130–12135.
- Chalmers, M. J., Hakansson, K., Johnson, R., Smith, R., Shen, J. W., Emmett, M. R., and Marshall, A. G. (2004) Protein kinase A phosphorylation characterized by tandem Fourier transform ion cyclotron resonance mass spectrometry. *Proteomics* **4**, 970–981.
- Schroeder, M. J., Webb, D. J., Shabanowitz, J., Horwitz, A. F., and Hunt, D. F. (2005) Methods for the detection of paxillin post-translational modifications and interacting proteins by mass spectrometry. *J. Proteome Res.* **4**, 1832–1841.
- Olsen, J. V., Macek, B., Lange, O., Makarov, A., Hornung, S. and Mann, M. (2007) Higher energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**, 709–712.
- Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. <