Automated Phosphopeptide Identification Using Multiple MS/MS Fragmentation Modes

Mathias Vandenbogaert,1,2,3,4,5,6,7 Véronique Hourdeu,1,2,3,4,5 Olivia Jardin-Mathe,4,5 Jean Bigeard,6,7 Ludovic Bonhomme,4,5,6,7 Véronique Legros,4,5,6,7 Heribert Hirn,4,6,7 Benno Schwikowski,8,9 and Delphine Pfieger4,5,6

INTRODUCTION

Phosphorylation of proteins plays an important role in signal transduction, protein/protein interaction, subcellular localization, protein activity and stability, and is therefore heavily studied using mass spectrometry-based proteomics. Phosphorylation is classically studied by affinity enrichment of proteolytic phosphopeptides followed by an analysis by the coupling of liquid chromatography and tandem mass spectrometry (LC-MS/MS). It is usually more difficult to identify phosphorylated than unmodified peptides, in part because of the loss of phosphoric acid that frequently occurs to the detriment of the peptide backbone fragmentation during classical CID. Several MS/MS fragmentation methods for the proteomic characterization of phosphopeptide samples have been examined.1−4 In particular, in the widely used linear ion trap, either stand-alone or within the hybrid instruments LTQ-Orbitrap or LTQ-FTMS, classical MS2 or multistage activation (MSA) fragmentation may be performed. The ion trap can be further equipped with an electron transfer dissociation (ETD) module, which allows generating fragmentation spectra mostly made up of c and z ions, instead of the classical b and y fragments, and has the advantage of maintaining labile modifications such as phosphorylations. An additional higher collisional dissociation

ABSTRACT: Phosphopeptide identification is still a challenging task because fragmentation spectra obtained by mass spectrometry do not necessarily contain sufficient fragment ions to establish with certainty the underlying amino acid sequence and the precise phosphosite. To improve upon this, it has been suggested to acquire pairs of spectra from every phosphorylated precursor ion using different fragmentation modes, for example CID, ETD, and/or HCD. The development of automated tools for the interpretation of these paired spectra has however, until now, lagged behind. Using phosphopeptide samples analyzed by an LTQ-Orbitrap instrument, we here assess an approach in which, on each selected precursor, a pair of CID spectra, with or without multistage activation (MSA or MS2, respectively), are acquired in the linear ion trap. We applied this approach on phosphopeptide samples of variable proteomic complexity obtained from Arabidopsis thaliana. We present a straightforward computational approach to reconcile sequence and phosphosite identifications provided by the database search engine Mascot on the spectrum pairs, using two simple filtering rules, at the amino acid sequence and phosphosite localization levels. If multiple sequences and/or phosphosites are likely, they are reported in the consensus sequence. Using our program FragMixer, we could assess that on samples of moderate complexity, it was worth combining the two fragmentation schemes on every precursor ion to help efficiently identify amino acid sequences and precisely localize phosphosites. FragMixer can be flexibly configured, independently of the Mascot search parameters, and can be applied to various spectrum pairs, such as MSA/ETD and ETD/HCD, to automatically compare and combine the information provided by these more differing fragmentation modes. The software is openly accessible and can be downloaded from our Web site at http://proteomics.fr/FragMixer.

KEYWORDS: phosphopeptides, phosphosite localization, combination of spectra, LTQ-Orbitrap, Mascot, FragMixer software
(HCD) cell located at the rear of the Orbitrap allows another fragmentation type in which the generated fragments are reliably detected in that analyzer with high mass accuracy and high resolution, without the low-range mass cutoff that applies in ion traps. Several recent studies reported the acquisition of two scans on every selected precursor ion from complex phosphopeptide samples to obtain complementary fragmentation patterns and thus increase the number of confidently identified phosphorylated sequences.\(^9\) Ion trap CID and HCD fragmentations were thus combined on iTRAQ-labeled phosphopeptide samples to get both sensitive identification and reliable quantification from the CID and HCD spectra, respectively.\(^6\) These two types of spectra were also acquired side by side on phosphopeptides to highlight the near absence of gas-phase rearrangement of the phosphate group during HCD fragmentation and the production of a phosphorylation-site-specific \(x\) ion fragment.\(^7\) The fragmentation modes CID and ETD (or ETD with supplemental activation) were also associated to characterize phosphopeptide samples.\(^10\)–\(^14\) Similarly, CID and electron capture dissociation (ECD) fragmentation modes were combined to obtain more confident phosphopeptide sequences and modification sites.\(^15\),\(^16\) In most of these studies, the two lists of phosphopeptide identifications provided by the complementary scan events were compared to assess the respective merits of the fragmentation modes: peptide sequence covered by the fragments, possibility to localize the phosphosite, bias toward certain classes of peptides, etc.\(^19\) Sometimes both lists were simply merged, and duplicates eliminated. Yet, how to handle divergent phosphorylation sites inferred from such a spectrum pair is a rarely addressed issue; it was indeed studied in the case of CID/ETD pairs\(^19\) and CID/ETD pairs,\(^13\) but, to our knowledge, no software solution has been described to automatically provide a consensus phosphorylated sequence from spectrum pairs. Programs previously developed to handle CID/ETD spectrum pairs,\(^17\) or collisionally activated dissociation (CAD)/ECD spectrum pairs on a LTQ-FT instrument,\(^18\) focused on unmodified peptides and were designed for those specific pairs of spectra. Our software FragMixer presented here constitutes a straightforward and flexible solution to interpret phosphopeptide identifications provided by Mascot from any combination of spectrum pairs.

Database search engines usually attribute to every fragmentation spectrum lists of putatively matching (phosphorylated) sequences, of decreasing probability that the match is correct. Simply exporting the list of top ranking matches is not correct. Briefly, to attain a given false localization rate (FLR), cut-offs of MD-score were estimated for different fragmentation schemes, in particular ion trap CID, MSA, ETD and HCD on the LTQ-Orbitrap.

In the present study, we acquired consecutive MS2 and MSA spectra to study in detail the complementary performances of the two fragmentation modes to identify phosphorylated peptides. Because these two scans are produced by very similar fragmentation processes, one might expect to get very redundant identifications. However, we deemed interesting to test the value of combining these two scans because (i) two independent studies reached contradictory conclusions in that MS2 or MSA spectra provided better identification of very complex phosphopeptide samples\(^27\),\(^28\) analyzed on LTQ-FT or LTQ-Orbitrap instruments and (ii) MS2 and MSA can be combined on the hybrid instruments LTQ-Orbitrap and LTQ-FT but also on stand-alone ion traps. We further evaluated the interest of systematically combining ion trap MS2 and MSA spectra within LC−MS/MS analyses of complex samples to improve precise phosphosite localization. Finally, we developed FragMixer, a program to automatically deduce consensus phosphorylated sequences from the pairs of identifications provided by Mascot, primarily by relying on the MD-score. FragMixer also allows processing data consisting of one single scan type (e.g., CID), or of other spectrum pairs (e.g., CID/ETD or CID/HCD), where each scan type requires distinct Mascot search parameters. FragMixer should be useful to obtain more confident identifications from phosphopeptide samples.

## MATERIALS AND METHODS

### Chemicals

Tris (2-carboxyethyl)phosphine (product reference C-4706) and methyl methanethiosulfonate (MMTS) were purchased from Sigma-Aldrich. HPLC gradient grade acetonitrile (ACN), normapur grade formic acid and acetic acid (AA) were purchased from VWR. All buffers and solutions were prepared using ultrapure water (Milli-Q, Millipore). Sequencing-grade modified porcine trypsin (EC 3.4.21.4) was purchased from Promega. C18 ultramicro- and microspin columns were purchased from Harvard Apparatus (product references 74-7226 and 74-4607). The IMAC PHOS-Select iron affinity gel was obtained from Sigma-Aldrich (product reference P9740).

### Plant Material, Plant Protein Extraction, Protein Sample Digestion

*Arabidopsis thaliana* plants (Col-O ecotype) were used to obtain protein samples. The detailed protocol carried out to yield proteolytic peptide samples is described in the Supporting Information. Briefly, several subcellular fractions were obtained, enriched in nuclear (samples A), cytoplasmic (samples B), or membrane proteins (samples C). The different samples Ai, Bi and Ci correspond to technical replicates, obtained by independent digestion procedures of different aliquots of precipitated proteins.

### C18-IMAC Protocol

Proteolytic peptide samples were acidified with AA to reach a pH of about 3 and cleaned on C18 spin columns (Harvard Apparatus). They were eluted from the C18 phase with 100 μL of 60% ACN, 3% AA to which 100 μL of water were added, to
reached a composition of 30% ACN, 1.5% AA that is compatible with phosphopeptide enrichment on the IMAC resin. Each sample was then incubated on 10 μL of packed IMAC beads for 2 h on a wheel rotating at 15 rpm. After removing the supernatant, IMAC beads were washed three times with 200 μL of 30% ACN, 1.5% AA with 2 min incubation each time and quickly once with 200 μL of water. Peptides retained on the resin were finally eluted with 50 μL of 400 mM NH₄OH. The pH of the IMAC eluates was immediately lowered to about 5 by addition of 10% AA. Phosphopeptide samples were kept at −80 °C until LC−MS/MS analysis.

**LC−MS/MS Analyses**

NanoLC−MS/MS analyses were performed on a Dual Gradient Ultimate 3000 chromatographic system (Dionex). Approximately 1/3 of the IMAC eluates obtained on samples A and 1/10−1/5 of those obtained from samples B and C were injected per LC−MS/MS analysis (this represented about 0.2−0.5 μg of total peptide). Phosphopeptide samples were loaded onto a C18 precolumn (Acclaim PepMap C18, 5 mm length × 75 μm particle size, 100 Å porosity, Dionex). After desalting for 5 min with buffer A (water/ACN/formic acid, 98/2/0.1, v/v/v), peptide separation was carried out on a C18 capillary column (Acclaim PepMap C18, 15 cm length × 75 μm ID, × 3 μm particle size, 100 Å porosity, Dionex) with a gradient starting at 100% solvent A, ramping to 70% solvent B (water/ACN/formic acid, 20/80/0.1, v/v/v) over 70 min, and then to 100% solvent B over 2 min (held 10 min), and finally decreasing to 100% solvent A in 3 min. The column was finally re-equilibrated with 100% solvent A for 20 min. The LC eluent was sprayed into the MS instrument with a glass emitter tip (Pico-tip, FS360-50-15-CE-20-C10.5, New Objective, Woburn, MA, USA). The LTQ-Orbitrap XL mass spectrometer (Thermo-Fisher Scientific) was operated in positive ionization mode. Ion species bearing 1 or ≥2 charges were excluded from fragmentation; dynamic exclusion of already fragmented precursor ions was applied for 90 s with an exclusion mass ≥ m/z capillary column (Acclaim PepMap C18, 15 cm length × 75 μm particle size, 100 Å porosity, Dionex) was applied for 90 s with an exclusion mass ≥ m/z capillary column (Acclaim PepMap C18, 15 cm length × 75 μm particle size, 100 Å porosity, Dionex). During MS/MS scans, fragmentation and detection occurred in the linear ion trap analyzer in centroid mode. Neutral loss masses specified for MSA fragmentation were 32.67 and 49 Da. Precursor selection window was 2 Da for MS2 and MSA. The automatic gain control (AGC) allowed accumulating up to 10⁶ ions for FTMS scans and 10⁴ ions for ITMSⁿ scans. Maximum injection time was set to 500 ms for FTMS scans and 100 ms for ITMSⁿ scans. The MSA-only and MSA/ETD analyses mentioned at the end of the Results section were performed on an LTQ-Orbitrap Velos instrument.

**Database Searches, Resulting Scores**

RAW data files acquired on the LTQ-Orbitrap instrument were converted to Mascot-compatible .MGF files using the software Proteome Discoverer 1.2. All MSⁿ spectra containing at least one peak were exported into the .MGF file. Database searches were performed using the Mascot server v2.2.07 and v2.2.6 with the following parameters: database TAIR10 (release 2010/12/14, 35386 sequences); enzymatic specificity: tryptic with two allowed missed cleavages; fixed modification of cysteine residues (Methylthio(C)); possible phosphorylation of S, T and Y residues; 5 ppm tolerance on precursor masses and 0.6 Da tolerance on fragment ions; fragment types taken into account were those specified in the configuration “ESI-trap”.

For each fragmentation spectrum, the database searches resulted in one or more best (possibly low-scoring) amino acid sequence identification, with an associated ion score, and a set of putative phosphosites. These different spectrum/phosphopeptide matches were ranked by Mascot by decreasing score; the MD-score was then calculated as the score difference between the first match and the Nth match (N > 1), for all those highest-ranked matches that corresponded to the same amino acid sequence. The MD-score has been shown to be an effective criterion for determining phosphosites. All other matches from the same scan, corresponding to other (not-highest-scoring) amino acid sequences, were discarded.

**Filtering of Results from Different Fragmentation Modes**

The distinct fragmentation modes sometimes lead to conflicting amino acid sequences and modification sites. To reconcile this information, we defined the following sequence of two filter rules, which are depicted in Figure 1.

1. **Sequence Filter (Scan Pair Filter).** Retain an identified amino acid sequence if its ion score is above a given threshold S. Note that, for any given spectrum pair, this rule may result in no sequence, a single sequence, or two distinct sequences; for instance, in the case of one or two misidentifications, but also when two nearly isobaric species are cofragmented.

2. **Phosphosite Filter.** For any amino acid sequence that passes the first filter, if phosphosites with an MD-score greater or equal to a given threshold P exist, retain them, and classify them as “certain”. If no such site exists, retain all phosphosites associated to MD-scores below P and classify them as “uncertain”.

As an example, successive S/T residues are a typical case where multiple phosphosites are retained by the phosphosite filter, as true phosphosites are difficult to unambiguously determine in this case. Note that the phosphosite identifications for any given sequence may originate from one or both spectra...
of a pair, depending whether one or two sequences passed the previous filtering step.

The FragMixer Software

In the course of this study, we created the software tool FragMixer that implements the above selection and filtering sequence, beginning with the Mascot output, using the Java programming language, and the msParser 2.4.0 Toolkit (Matrix Science). Our software can be operated through a graphical user interface or launched from a command line interface for batch processing of larger numbers of Mascot output files.

Input Files. As input, FragMixer requires either a single Mascot DAT file containing identifications from a combined MS2/MSA run, or a pair of DAT files corresponding to two different fragmentation modes, such as those produced when the two types of acquired scans require different search parameters, for example, MSA/ETD or ETD/HCD pairs. FragMixer can also process a single DAT file corresponding to conventional analyses with a single fragmentation mode (such as MS2 spectra acquired in the linear ion trap).

Parameter Settings. Parameters are set in the dialog box shown in Figure 2. The threshold $S$ of the sequence filter is defined in the section “Filtering criteria” using one of four options: (i) the scan-specific Mascot identity threshold, (ii) the scan-specific Mascot homology threshold, (iii) a global, user-specified threshold, or (iv) a threshold corresponding to a user-specified false discovery rate (FDR). In the last case, a database search of the given spectra against a “target–decoy” protein database containing normal (target) and decoy (reverse) sequences must have been performed with Mascot. Briefly, each peptide is then assigned by FragMixer the maximum Mascot ion-score resulting from the two fragmentation modes. Next, peptides are sorted in the order of decreasing ion score, and FDR estimates are calculated for the first $N$ peptides, $N = 1, \ldots$, as the ratio between the sizes of the decoy and target sets of peptides in the first $N$ peptides. The first $N − 1$ peptides are selected, where $N$ is the first integer whose FDR estimate exceeds the user-defined FDR threshold. FragMixer indicates (in the “StatReport”, see below) the ion score threshold used to filter peptides to approach the targeted FDR.

Next, in the section “Peptides to report” the user can indicate whether all or only the post-translationally modified peptides should be exported. FragMixer was specifically developed to characterize phosphopeptides, yet peptides bearing other types of modifications can be filtered from nonmodified ones using this option. Finally, in the section “Phosphosite localization”, the threshold $P$ of the phosphosite filter is specified, as a threshold on either the MD-score, or on the normalized Mascot delta score (which is the MD-score divided by the score of rank 1 peptide). The PeptideReport table provides two best-scoring sequences from each spectrum pair and their associated scores (peptide identification scores and MD-scores for the two scans) and the automatic classification of the putative phosphosites (certain or uncertain).

The “PeptideReport” and “ProteinReport” also contain a consensus sequence for each peptide pair. It is determined as follows and displayed. For any sequence that has phosphosites identified with certainty (by the above phosphosite rule), the sequence is output with putative phosphosites in lowercase and each one enclosed in parentheses, followed by an asterisk. Uncertain putative phosphosites are enclosed in parentheses, followed by the number of asterisks corresponding to the number of modifications (e.g., “(tsGVs)***” represents two phosphorylations within a range of three possible phosphosites). When uncertain putative phosphosites are identified for the same amino acid sequence from the two spectra of a pair, the consensus sequence indicates the union of the two sequence stretches.

### RESULTS

Side-by-Side Comparison of MS2 and MSA Fragmentation Modes for Phosphopeptide Identification

Two reports previously compared the effectiveness of MS2 and MSA fragmentation modes to characterize very complex phosphopeptide samples. Separate LC–MS/MS analyses were carried out on hybrid LTQ-Orbitrap or LTQ-FT instruments using one or the other fragmentation type. In both studies, approximately 1000 phosphopeptides were identified per run. Depending on the report, either MS2 or MSA was determined to be the better fragmentation option. Considering these contradictory results, we wished to assess further the efficiency and the possible complementary nature of MS2 and MSA in providing reliable phosphopeptide identification, in terms of amino acid sequence and phosphosite localization.
compare the two types of scans, we performed LC−MS/MS analyses that triggered the acquisition of both spectra on every peptide selected for fragmentation: MS2/MSA spectrum pairs were acquired on the three most intense ions detected in the preceding FTMS preview scan. In this way, both scans were acquired nearly at the same time in the elution peak of the peptide species, and their performances for peptide identification could be compared without the significant variability that would be entailed by independent technical repeat experiments. Three types of phosphopeptide samples (A, B, C) of variable proteomic complexity, obtained from different subcellular compartments from Arabidopsis thaliana whole plants, were thus analyzed in quadruplicate (A,B), or duplicate (C) in a total of 10 LC−MS/MS analyses.

Overall, we acquired 3252 MS2/MSA spectrum pairs in which at least one spectrum of the pair provided a phosphopeptide identification with a score above the threshold S (chosen as the Mascot identity threshold), such that a sequence was each time reported after our filtering operation. In 19 out of the 3252 pairs, two distinct sequences in amino acids were proposed by Mascot. However, in these cases, only one of the two amino acid sequences possessed a score above the threshold S and thus passed our sequence filter. We determined by manual inspection that, in most of these cases, the lower-scoring identification originated from spectra of very poor quality, from which no convincing sequence information could be determined.

Supporting Information Figure S1 provides a synthetic view of the obtained identifications. When considering the 3233 MS2/MSA pairs that identified the same amino acid sequence (but possibly different phosphorylation sites) as rank 1 matches, MSA yielded identifications with better scores than MS2 in 2008 cases (62.1%). Besides, among these spectrum pairs identifying the same amino acid sequence, an average of 84.4% (2730/3233) of spectrum pairs identified the same phosphosite(s) in the rank 1 sequences. Within this subclass of spectrum pairs, MSA spectra provided higher-scoring identifications than MS2 in 61.8% of cases (1687/2730 cases). The side-by-side comparison of MS2 and MSA spectra acquired on the same precursor ions thus allowed the conclusion that MSA provides better identification of phosphopeptides than MS2 in about 3/5 of cases, which is in agreement with the study of Ulintz et al. who concluded that MSA should be preferred over MS2 to analyze complex phosphopeptide samples.

**Results of the Spectrum Pair Approach**

The previous comparison of the effectiveness of MS2 and MSA spectra in identifying phosphopeptides considered as simple evaluation criterion the Mascot score attributed to each spectrum. It allowed estimating that neither fragmentation mode was radically bested by the other one. This drove us to dig further into the possible complementary nature of the two spectra, and to ask whether it would be interesting to systematically analyze phosphopeptide samples by acquiring MS2/MSA spectrum pairs on each precursor ion. The success in phosphopeptide identification from paired MS2 and MSA spectra was evaluated with the help of our program FragMixer (see Materials and Methods section).

The 10 LC−MS/MS analyses showed that in 2353 out of 3233 MS2/MSA pairs (i.e., 72.8%), the two scans identified the same amino-acid sequence with scores above the identity threshold calculated by Mascot for every MS^n spectrum, whereas in 880 MS2/MSA pairs (i.e., 27.2%), only one scan identified an amino acid sequence with a score above that threshold. For 27.2% of the identified peptide identifications, it then appeared worth combining MS2 and MSA to determine a reliable amino acid sequence from the fragmentation data.

Next, we counted the number of MS2/MSA pairs in which both scans identified the same amino-acid sequence with scores above the identity threshold and precise phosphosite assignment was determined from (i) both scans; (ii) only one scan; (iii) neither scan. Category (i) groups the peptides for which it was worth combining MS2/MSA to get confident phosphosite assignment. Referring to the report of Savitski et al. we considered that precise phosphosite localization was obtained from a MS2 or a MSA scan when the MD-score was above 4 (which should lead to a FLR below 5%). The three categories (i) to (iii) represented 1606, 408, and 339 spectrum pairs, respectively; this indicated that in 17.3% of cases (i.e., 408/2353), it was worth combining MS2 and MSA to obtain a precise phosphosite localization and not only determine a reliable amino acid sequence. Our program FragMixer handles automatically spectrum pairs corresponding to cases (i) to (iii) and provides a consensus phosphopeptide sequence reflecting the union of the information contained in both spectra, as long as the scores (identification scores and MD-scores) pass the filtering criteria (see Materials and Methods and Supporting Information Table S1). Within the above category (ii), 57 ± 6% (229 out of 408 spectrum pairs) of precise phosphosite assignments were based on MSA spectra. This slight superiority of MSA over MS2 seems to be coherent with the study of Savitski et al., who reported that MSA was more successful than MS2 in determining the correct phosphorylation localization.

Finally, to visualize further the interest of combining the information provided by the two spectra of a pair, we considered one given LC−MS/MS analysis combining MS2/MSA pairs and plotted the number of phosphopeptides identified with a precisely determined phosphosite in a range of FDR values, when considering individually the MS2 spectra or the MSA spectra or the consensus sequences determined by FragMixer (Figure 3). Besides the fact that MSA spectra were overall slightly better able to discriminate the precise phosphosites as compared to MS2 spectra, this figure clearly shows the gain provided by the consensus sequences regardless of the FDR threshold. At a 1% FDR, about 460 unique phosphopeptides with a precisely determined phosphosite were identified from an MS2/MSA analysis. The three curves show the numbers of identifications determined when selecting MS2 or MSA spectra from the spectrum pairs, or by combining the information provided by both spectra of a pair using FragMixer (FM) (consensus sequences).
identifications were obtained from consensus sequences, to be compared to 360–390 identifications when considering independently the MS2 or MSA spectra.

Precise phosphosite assignment is particularly challenging when the peptide sequence contains a continuous stretch of residues (S, T, Y) that are all theoretical candidates for bearing the modification. We could determine from all of our 10 analyses combining MS2/MSA spectra that 187, 50, and 4 consensus phosphopeptides respectively contained a stretch of 2, 3, or 4 neighboring S/T residues in which the position of the phosphorylation site could not be resolved when choosing a MD-score threshold of 4 (sequence stretches such as (ss)*, (sss)* or (ssss)*, with s exchangeable with t or y). When considering only the MS2 data from the spectrum pairs, these numbers were 345, 43 and 8; when considering only the MSA data, they were 344, 32, and 9. These three sets of figures show another advantage of acquiring the two spectra to reduce uncertainty of phosphosite localization and of using FragMixer to combine the two resulting pieces of information.

We have shown on the previous analyses that combining MS2 and MSA on every precursor helps to identify the amino acid sequences and to precisely localize the phosphosites, whereas one scan type alone might fail in providing one or both pieces of information. However, acquiring MS2 and MSA on every selected precursor approximately doubles the analysis time spent on every ion, which is possibly done at the expense of the total number of different phosphopeptides identified. To address this issue, we analyzed five different sample preparations (A1, B1, B2, C1, C2) using three different acquisition methods: 6 MS2 spectra acquired on the six most intense ions, or 6 MSA spectra on the six most intense ions, or 3 MS2/MSA spectrum pairs on the three most intense peptide ions detected in the FTMS preview scan. The numbers of peptide identifications obtained from the different acquisition methods are illustrated in Figure 4. It is apparent from these analyses that for samples of moderate complexity, such as A1, B1 and C2, the analysis combining MS2 and MSA scans on every precursor can be advantageous or at least compete with MS2-only or MSA-only analyses in terms of total numbers of different (phospho)peptides identified (Figure 4 a and b). For more complex phosphopeptide samples, the decreased MS/MS cycle time was accompanied with a diminished total number of unique identifications (samples B2 and C1). Besides, from all the (pairs of) spectra matched to an identified phosphopeptide in MS2-only, MSA-only and MS2/MSA analyses, we calculated the fractions of phosphorylated sequences exhibiting an ambiguous phosphosite assignment (see Figure 4c). Whatever the sample complexity, an advantage of acquiring MS2 and MSA spectra on each precursor could usually be seen, with an average decrease in the fraction of peptides having phosphosite localization uncertainty from 25% for MSn-only analyses to 20% for the MS2/MSA option. The interest of the “dual spectra” approach would probably be more significant when considering more divergent fragmentation modes, such as MSA/ETD or MSA/HCD. A detailed study using these fragmentation modes is beyond the scope of the present study. But briefly, a phosphopeptide sample obtained from A. thaliana cytosolic fraction was submitted to MSA-only and MSA/ETCID (ETD with supplemental activation) analyses. Then, the fractions of phosphopeptides identified with unclear phosphosites dropped from 17 to 10% when switching from the single- to the dual-spectrum option.

Figure 4. Comparison of (phospho)peptide identifications provided by LC–MS/MS analyses containing MS2 or MSA scans only, triggered on the top 6 precursors, or 3 pairs of MS2/MSA scans triggered on the top 3 precursors. All LC–MS/MS analyses were filtered with FragMixer to reach an estimated FDR of 1%. (a) Numbers of unique AA sequences identified: number of different peptide sequences identified; both phosphorylated and unmodified peptides are taken into account; identifications differ in terms of sequence in amino acids and/or number of phosphorylations. (b) Numbers of unique phosphopeptides identified: number of different phosphorylated sequences; they differ in sequence in amino acids and/or in the number of phosphorylation moieties. (c) Proportions of (pairs of) spectra leading to phosphopeptide identifications with unclear phosphosites. Samples A, B and C correspond to phosphopeptide mixtures enriched from different A. thaliana subcellular protein fractions (see Materials and Methods section). Samples C1 and C2 provided largely differing results because they corresponded to very different protein amounts used from the digestion step.
It is finally worthwhile to note that the observed variations in unique phosphopeptide identifications between MS2-only, MSA-only and MS2/MSA arise from the combined effects of the different acquisition methods and of inter-run variability. Figure 5 shows Venn diagrams representing the overlaps of the lists of unique phosphorylated sequences identified from the three analyses (MS2, MSA or MS2/MSA) of each sample. As it has already been reported, replicate analyses provide complementary identifications, yet we observed that usually the majority of phosphopeptide identifications were shared between the three analyses. Finally, the analyses performed here on different samples do not allow us to draw robust conclusions as to the maximum level of complexity and peptide amount justifying the MS2/MSA method under the analytical conditions used here (MS instrument, targeted resolution in FTMS which can impact the cycle time, etc.). Yet they illustrate the fact that combining the two scans on every precursor may be worth testing when a few hundreds of phosphorylated peptides are expected to be identified in a single LC–MS/MS analysis. Finally, FragMixer is perfectly suited to assess on a given sample the value of acquiring two spectra on each selected precursor and to automatically combine the Mascot identification results obtained on both scans.

**DISCUSSION**

The nearly simultaneously acquired MS2 and MSA scans on every precursor showed that MSA tended to yield better phosphopeptide identification than MS2 fragmentation because in approximately 60% of all cases, the MSA spectrum of a given MS2/MSA pair matched a theoretical phosphorylated sequence with a better Mascot score than the corresponding MS2 spectrum. Our observations are in agreement with the report of U tintz et al., who also used Mascot for data interpretation. In contrast to this, Villen et al., who used Sequest, reached the conclusion that MS2 was more suitable for analyzing complex phosphopeptide samples on high-mass-accuracy MS instruments. More generally, it should be noted that the score associated by a search engine to a given MS spectrum significantly depends on the fragment ion types that are taken into account for its interpretation. In the case of a peptide sequence phosphorylated at a S/T residue, Mascot searches among the higher-intensity fragment peaks for fragments in a dehydrated form (loss of H$_3$PO$_4$) and for fragments still bearing the phosphorylation moiety. Two series of y and of b ions are then considered in the ESI-trap configuration. Finally, MS2 or MSA fragmentation may be found to provide better phosphopeptide identifications, depending on the theoretical fragment types that are considered by the search engine.

Our analyses acquiring MS2/MSA spectra pairs on every precursor further indicated that the two spectra provided partially overlapping lists of confidently identified phosphopeptides. These lists mostly differed in terms of phosphosite localization but also in terms of amino acid sequence assignments. As a result, when optimizing the cycle time is not a critical issue, it appears to be an interesting option to combine both scans on every fragmented species. FragMixer, which was developed and used in this study to compare identification results yielded by MS2 and MSA spectra, allows direct and straightforward comparison of the strengths and weaknesses of two fragmentation modes. The respective ability of two fragmentation schemes to successfully identify peptides, depending on their size, charge state, basicity, modification state, etc., can be very easily assessed. When dealing with phosphorylated peptides, one can evaluate the efficiency of two fragmentation modes in terms of amino acid sequence determination on the one hand and of phosphosite localization on the other hand. One fragmentation mode may indeed be superior to the other in one or both aspects. Such study may also be performed on other types of modification (such as sulfonation, for example), with the limitation however that MD-score thresholds for validating modification sites have not been described yet. Finally, FragMixer automatically merges the information contained in paired spectra to derive a consensus phosphopeptide sequence.

It is possible to extend the pair scanning strategy by acquiring two more distinct, complementary, fragmentation pairs, such as MSA/HCD or MS2/ETD. Whereas we tried to assess the possible loss of unique phosphopeptide identifications due to the increased cycle time taken by the MS2/MSA spectrum pairs, two independent studies similarly evaluated the impact of combining CID and ETD spectra. Kim et al. reported that the alternating CID/ETD method provided a similar number of unique phosphopeptide identifications as the CID-only method from a complex phosphopeptide sample (about 1200 unique identifications each time). Besides, Molina et al. observed a decrease of 7–17% of unique peptide identifications when analyzing a mixture of 48 standard proteins by acquiring three pairs of CID/ETD scans instead of six CID spectra. With the advent of new-generation hybrid instruments offering faster acquisition speed, such “spectrum pair methods” may become more systematically applicable even to very complex samples. Combining ion trap MS/MS and HCD can thus be better considered on the new-generation LTQ-Orbitrap Velos instrument, because of its increased sensitivity and diminished cycle time in HCD. The complementary information provided by ion trap CID and ETD fragmentation modes in terms of
Application or not of a number of parameters that are likely to determine the MD-score was demonstrated to be quite insensitive to some interlaboratory variability. It is worth mentioning that the MD-score threshold parameter should be either taken directly from the study of Savitski et al. or estimated de novo in FragMixer rather than calculated downstream of Mascot should thus constitute an easy to implement, flexible and robust tool for phosphoproteomics studies.

CONCLUSION

We have compared the benefits of MS2 and MSA fragmentation modes for phosphopeptide sample characterization and reached the conclusion that combining both scans can be a valid option to increase the number of unique phosphopeptides identified and the confidence in phosphosite assignment. Acquiring two more different spectra on every precursor, such as MSA/ETD or ETD/HCD, during an LC–MS/MS analysis is likely to be an even more fruitful acquisition method. To deal with the pairs of spectra thus obtained on every fragmented ion, we have developed the software FragMixer. Using the Mascot Delta Score to estimate the reliability of the phosphosite assignments determined by Mascot, FragMixer processes the information collected on the two spectra to export the most probable phosphorylated sequence and automatically maintain uncertainty in phosphosite assignment when relevant. This is in particular of prime importance when phosphoproteomic data are the basis for further validation experiments, such as introducing point mutations followed by functional in vivo analyses of a protein of interest.

ASSOCIATED CONTENT

Supporting Information

Supplemental figures, tables, and methods. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

E-mail: benno.schwikowski@pasteur.fr; delphine.pflieger@univ-evry.fr.

Present Address

Plate-forme Génotypage des Pathogènes et Santé Publique, Institut Pasteur, 28, rue du Docteur Roux, 75724 Paris Cedex, France.

Author Contributions

These two authors contributed equally to this work.

ACKNOWLEDGMENTS

This work was supported by the CNRS, Genopole-France, Institut National de la Recherche Agronomique, Université d’Evry Val d’Essonne and Région Ile-de-France. DP thanks the Agence Nationale pour la Recherche (ANR) for the funding ANR-2010-JCJC-1608-01. We are also grateful to Abdelkader Namane for administrative support.

REFERENCES


(3) Nagaraj, N.; D’Souza, R. C.; Cox, J.; Olsen, J. V.; Mann, M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J. Proteome Res.* 2010, 9, 6786–6794.


ASSOCIATED CONTENT


