Challenges and strategies for targeted phosphorylation site identification and quantification using mass spectrometry analysis

Kevin Blackburn and Michael B. Goshe

Advance Access publication date 24 December 2008

Abstract

Despite its importance, the ‘ultimate’ method to identify and quantify site-specific protein phosphorylation using mass spectrometry (MS) has yet to be established. This is as much a function of the dynamic range of instrumentation as it is the complexities surrounding the isolation and behavior of phosphopeptides. Phosphorylation site analysis using MS can be quite challenging when analyzing just one protein and quickly becomes a daunting task when attempting to perform proteome-wide measurements. Data-dependent tandem MS-based methods which are useful for the discovery and characterization of novel phosphorylation sites often lack the dynamic range and quantitative aspect required for studying the temporal phases of phosphorylation. While targeted methods such as multiple reaction monitoring do provide a highly specific and quantitative methodology for studying phosphorylation changes over time, they are not suited for initial discovery of previously unreported sites of phosphorylation. Data-independent acquisition represents a relatively new approach for simultaneous qualitative and quantitative sample analysis which holds promise for filling this technological gap.

Keywords: phosphorylation site stoichiometry; phosphorylation site identification; data-dependent acquisition; data-independent acquisition; multiple reaction monitoring; LC/MS²

BIOLOGICAL IMPORTANCE OF PROTEIN PHOSPHORYLATION

The reversible phosphorylation of proteins has been, and continues to be, the most studied post-translation modification due to its role of controlling protein function and directing the physiological responses of cells and their environment [1–5]. Phosphorylation of the solvent-accessible hydroxyl of Ser, Thr and Tyr residues by kinases at site-specific motifs convert this polar functional group into a larger, more polar anionic group that is used to modulate protein activity and can be countered by the removal of the phosphate moiety via phosphatases. This chemical alteration changes the functional activity of many cellular proteins such as receptors [6–8], kinases [9, 10], phosphatases [11, 12], translation [13, 14] and transcription factors [15, 16]. The highly regulated control of phosphorylation is important for correctly propagating signal transduction events to maintain cellular function and development in processes such as immune response, cell differentiation, cell-cycle progression, and cellular communication.

Corresponding author. Michael B. Goshe, Department of Molecular and Structural Biochemistry, North Carolina State University, 128 Polk Hall, Campus Box 7622, Raleigh NC 27695-7622, USA, Tel: +1-919-513-7740; Fax: +1-919-515-2047; E-mail: michael_goshe@ncsu.edu

Kevin Blackburn is the Supervisor of the Biochemistry Mass Spectrometry Laboratory at North Carolina State University. He has served as co-instructor for the ASMS Peptides and Proteins Short Course since 2002. His current research interests include the development of high-performance chromatography and mass spectrometry techniques and their application to the study of protein post-translational modifications and proteome dynamics of plant systems.

Mike Goshe is an Assistant Professor of Molecular and Structural Biochemistry and is the Director of the Biochemistry Mass Spectrometry Laboratory at North Carolina State University. His current research interests involve the development and application of mass spectrometry techniques for mapping and quantifying sites of protein phosphorylation directed toward fundamental studies in cell signaling and viral infection and characterizing protein interactions and dynamics.
and apoptosis [15–18]. Aberrant phosphorylation can disrupt this network to promote deleterious effects that can lead to manifestations of disease such as cancer [12, 19–21]. Thus, the study of protein phosphorylation is an active area of research in biology extending to plants, animals, and humans.

**BIOCHEMICAL ANALYSIS OF PHOSPHORYLATION**

From the perspective of a biochemist, a number of classical biochemical approaches may be undertaken to characterize protein phosphorylation from a functional standpoint. These often include the use of cloned and overexpressed proteins, proteins isolated from cells via immunoprecipitation with monoclonal antibodies, or isolation of epitope-tagged proteins from a transgenic cell line or organism using immunoprecipitation or affinity chromatography. In these types of studies, phosphorylation is often monitored by using $[^{32}\text{P}]\text{ATP}$ to characterize kinase activity and potential substrates. The extent of protein phosphorylation can be measured by phosphoimaging based on the incorporation of a $[^{32}\text{P}]$-radiolabel, immunoblotting using antibodies recognizing amino acid residues modified with a phosphate moiety, and phosphorylation specific staining using Pro-Q Diamond (http://www.invitrogen.com).

However, these measurements do not provide detailed information regarding the exact site or sites of protein phosphorylation or associated phosphorylation stoichiometries. The use of bioinformatic approaches like interrogating the amino acid residue sequence of a protein for documented phosphorylation motifs (e.g. Scansite [22] http://scansite.mit.edu) may reveal a few putative sites. However, any computational outcome is only as good as the basis set used to establish the searching algorithm, and thus these approaches may miss novel sites of phosphorylation. Site-directed mutagenesis to convert Ser/Thr residues to Ala or Tyr residues to Phe can be used to create mutant proteins that allow an indirect measure of protein phosphorylation and/or protein activity due to the substitution of specific residues. In some cases, phosphorylation may not be totally abolished in the site-directed mutant or the introduction of the mutation could cause a structural perturbation within the protein that prevents the actual residue, as opposed to the mutated one, from being phosphorylated. Even adding $[^{32}\text{P}]\text{ATP}$ exogenously to the sample is of limited use since this method can only ascertain phosphorylation events that occur after introduction of the radiolabel, and thus does not provide any information pertaining to basal level phosphorylation.

**PHOSPHORYLATION IS A COMPLEX POST-TRANSLATIONAL MODIFICATION**

Because of the limitations described above, many biologists have turned to using mass spectrometry (MS) analysis to pinpoint exact sites of phosphorylation and to quantify their extent of phosphorylation in both purified protein and phosphoproteome contexts. However, simply applying conventional MS techniques used for protein identification often fall short for measurements regarding phosphorylation. This is due in part to the low abundance of phosphorylated species in which site-specific phosphorylation is often substoichiometric. Phosphorylation stoichiometry typically spans greater than four orders of magnitude, a dynamic range not usually achievable with commercial MS instruments.

Therefore, what might be achievable with immunoblotting or phosphoimaging where amplification of the signal allows for a much greater dynamic range of detection may not be achievable by MS measurements where there is currently no analogous means of phosphopeptide signal amplification. In addition, phosphorylation events are transient and reversible; thus during a given physiological response, the extent of protein phosphorylation could be relatively high at one time point but below detectable levels at a later time point. Consequently, a set of protein phosphorylation sites could be heavily modified in a biological system (phosphoprotein or phosphoproteome) at one moment while other biologically interesting and important phosphorylation sites may only be marginally phosphorylated. This temporal component in phosphorylation and the difficulties associated with both identifying the sites of modification and quantifying the extent of phosphorylation are analytical challenges which are further complicated when multiple phosphorylation sites are contained within several residues of each other on the same phosphopeptide. During preparation and handling, artificial changes in phosphorylation levels may also be introduced whenever a protein is exposed to kinases and/or phosphatases contained within the system of interest. To minimize these extraneous phosphorylation events while reducing sample-to-sample variability,
sample preparation steps often include the addition of kinase and phosphatase inhibitors to limit these enzymatic activities within each sample, thus promoting the identification of true biological phosphorylation events.

In addition to dynamic stoichiometry, full characterization of protein phosphorylation essentially requires complete sequence coverage for a given protein, a feat that is not often achieved using bottom-up approaches such as liquid chromatography–tandem mass spectrometry (LC/MS/MS) analysis of phosphoprotein or phosphoproteome proteolytic digests. This is due to a variety of reasons. For example, the use of a single protease may generate peptides too large to be adequately fragmented by collision-induced dissociation (CID), a situation which may be alleviated by using another protease or multiple proteases. However, the inherent sample complexity of these peptide mixtures can also preclude the detection of phosphopeptides.

To provide more reliable phosphopeptide detection from simple and complex mixtures of peptides, a variety of enrichment techniques have been used for the specific capture of phosphopeptides/phosphoproteins. These methods target the phosphate moiety for non-covalent capture and subsequent elution and include immobilized metal affinity chromatography [23–28], the use of titanium dioxide supports [29–32], phospho-amino acid motif antibody affinity enrichment [33, 34], and ion exchange chromatography [35–37] among others [38, 39] and their combinations [27, 40–44]. However, the presence of other anionic or basic functional groups can lead to the capture of nonphosphorylated peptides making reproducibility difficult to achieve across different samples. While affinity methods have been demonstrated in the literature to successfully enrich for phosphopeptides, the extent of phosphopeptide identification with any given approach is highly dependent upon the sample and the laboratory performing the analysis. This is due in part to the fact that all binding is based on the law of mass action and each protein or proteome sample will contain different concentrations of certain peptides which affect phosphopeptide binding in a quantitative manner. Thus, in some cases, nonspecific binding of a highly abundant nonphosphorylated peptide becomes competitive over a low abundance phosphopeptide under certain loading conditions. Consequently, there is no clear consensus in the literature as to the ‘best’ affinity isolation approach.

To address the quantification of phosphorylation, the use of isotope coding strategies [45–48] involving chemical, metabolic or enzymatic incorporation of stable isotopes into phosphopeptides or phosphoproteins have been developed to label two distinct samples. These methods provide a means to quantify abundance changes of phosphorylation events using mass spectrometry, and in some instances, facilitate differential phosphorylation site mapping. Although a number of isotope coding approaches for phosphorylation analysis have been reported in the literature, there is no preferred ‘best’ method.

Overall, the implementation of these advanced fractionation and labeling techniques to promote more efficient MS analysis of phosphoproteins are often hit or miss, depending upon a variety of factors including sample quantity, complexity, and whether the MS acquisition scheme employed will be a targeted hypothesis-driven investigation to study specific, known phosphorylation sites of interest or a discovery-based mode of analysis for the detection and quantification of novel sites of phosphorylation. In this context, the implementation of an appropriate MS acquisition strategy for identifying and quantifying protein phosphorylation, as opposed to a sample-specific enrichment or labeling approach, is actually more critical to achieving biologically meaningful results within a given experimental system.

**ANALYTICAL APPROACHES AND CHALLENGES FOR MASS SPECTROMETRY ANALYSIS OF PHOSPHOPROTEINS**

The current bulk of protein and proteomic characterization heavily relies on LC/MS/MS to interrogate constituent peptides of protein or proteome samples [49, 50]. In the bottom-up approach, proteins are proteolytically digested and the resulting peptides separated by reversed-phase liquid chromatography and analyzed by tandem mass spectrometry. During analysis, peptide masses and masses of the related peptide product ions generated by CID of each intact peptide are measured. Phosphopeptides are recognized by the additional mass of 80 Da due to each phosphorylation event. Based on the fragmentation pattern generated during MS/MS, the site(s) of phosphorylation may be determined by a concomitant 80 Da shift in affected product ions, but detection of the modification can be very difficult...
due to both the biological and chemical nature of protein phosphorylation.

The bottom-up approach in proteomics is typically performed using a data-dependent acquisition (DDA) scheme for LC/MS/MS as illustrated in Figure 1. During peptide elution from the LC column, a survey MS scan is performed which produces a spectrum comprised of all precursors that are eluting at a specific time. The DDA acquisition method utilizes a set of user-defined criteria to select and serially interrogate on-the-fly the most intense ion for subsequent fragmentation using CID to generate a product ion spectrum of the selected precursor for subsequent protein database searching or manual interpretation. Thus for DDA analysis, the generation of a product ion spectrum for a peptide at a specific \( m/z \) is based on its intensity relative to the other peptides at a given retention time. In order to achieve the selection of additional peptides for MS/MS analysis, a more elaborate acquisition scheme is usually implemented for DDA in which additional precursors (typically up to eight) detected in the MS survey scan are individually selected for CID. Although this serial process produces more product ion spectra for additional peptides, it is still limited because the most abundant peptides will be selected for CID first and will usually be interrogated for only a portion of their full chromatographic elution profiles. To circumvent this behavior and avoid acquiring redundant information, once a particular \( m/z \) has been selected for CID, it is not eligible for another round of MS/MS analysis until a particular time has elapsed in order to permit other \( m/z \) values to be selected for CID. This dynamic exclusion approach increases the number of peptides that can be identified during a single LC/MS/MS analysis and increases the likelihood of interrogating lower abundance species such as phosphopeptides. In addition, one can further limit the candidates that may be selected for interrogation through the use of a hypothesis-driven, targeted DDA method where a list of candidate peptide \( m/z \) values may be used to select only specific peptides of interest. For example, if one were to collect data for a set of phosphopeptide masses hypothesized to be present in a sample, targeted DDA may be used to increase the likelihood of MS/MS interrogation of these species of interest while minimizing the probability of interference with the precursor selection from high abundance peptide components which...
are of no interest. Because of the data-dependent, intensity-driven nature of DDA which results in irregular sampling of MS survey scans, it is inherently a qualitative tool for the assignment of phosphorylation sites. Although labeling with isotope-coded tagging reagents may be used in combination with DDA for site-specific quantitation [51–58], the methodology still suffers with duty cycle limitations due to serial MS/MS events and a fundamental bias towards interrogating the highest abundance components in a sample.

CID has proven to be a quite effective approach for phosphorylation site identification and assignment. However, unequivocal identification of the phosphorylated residue by MS/MS is often difficult since the phosphoester bond is labile under the CID conditions typically employed for peptide fragmentation, resulting in the facile loss of either H₂PO₄ (98 Da) or HPO₃ (80 Da) from phosphoseryl (pSer) and phosphothreonyl (pThr) residues [59]. Although phosphopeptides produce b- and y-type product ions like their unmodified counterparts, the neutral loss of phosphate-related fragments represents a favored fragmentation pathway which results in decreased abundance of sequence specific, peptide backbone product ions. These neutral losses are dependent upon the amino acid sequence of the peptide and the location of the phosphorylation site. While neutral losses sometimes impede peptide identification during database searching, the unique neutral losses that do occur for phosphopeptides can be used as an additional feature for specific detection of phosphopeptides. When analyzed in negative ion mode, phosphopeptides can be specifically detected via the generation of the phosphate-specific product ions at m/z 79 (PO₃⁻) and m/z 97 (H₂PO₄⁻) [60, 61]. In positive ion mode, the preferred dissociation channel for pSer and pThr containing peptides during CID often generates a highly abundant product ion corresponding to the neutral loss of H₂PO₄ from the peptide. With the use of ion traps, these neutral losses can be used to trigger another round of CID of the neutral loss ‘precursor’ using an MS³ event, which produces peptide backbone sequence specific ions which may be used for a subsequent database searching and phosphopeptide site assignment [31, 62, 63]. However this MS²/MS³ or multi-stage activation approach suffers from a loss of ion abundance at each successive stage of CID analysis, thus the MS³ spectra often elicit lower database searching scores [62] but are often preferable over the corresponding MS² derived spectra which are dominated by the neutral loss of phosphate. Although the peptide may be tentatively identified as phosphorylated based on a characteristic neutral loss and quantified based on ion intensity, a sequence and phosphorylation site assignment may not be obtained.

In addition to their low abundance and varying stoichiometries, the partial loss of the phosphate moiety during CID can prevent unequivocal identification of the phosphorylated residue and can prevent phosphopeptides from being quantified with the precision and accuracy necessary to study protein phosphorylation at the residue level. However, the use of alternative dissociation techniques has provided yet another tool for phosphopeptide characterization. The use of electron capture dissociation (ECD) [64–66] and electron transfer dissociation (ETD) [67,68] seems to be ideally suited to the analysis of phosphoproteins because the phosphate moiety remains intact during fragmentation while sequence specific c- and z-type ions are also generated which can be used for sequence determination. At the present time ETD is available on the LTQ ion trap and orbitrap systems from Thermo Scientific (http://www.thermo.com). Although the use of ETD is relatively new and details of the fragmentation mechanisms are not as well characterized and understood as those involving CID, it holds much promise for improved characterization of phosphopeptides and is currently being applied in phosphoproteomic studies [69].

To provide a more selective and sensitive approach for phosphopeptide characterization, targeted MS/MS may be used in a hypothesis-driven context. The neutral loss of the phosphate moiety and the generation of sequence specific product ions for a given phosphopeptide may be exploited by using selected ion monitoring (SIM) or MRM. In SIM or MRM, only a select number of precursor-to-product transitions are monitored during the analysis, resulting in a more specific and sensitive phosphopeptide assay for the targeted species than open (i.e. non-targeted) approaches. While an excellent approach for specific, targeted phosphopeptide detection and quantification, these monitoring techniques are not readily amenable to the identification of novel phosphopeptides in a discovery mode of operation. To partially address this aspect, novel acquisition modes incorporating...
MRM triggering of full scan product ion acquisitions provide at least another level of qualitative data above and beyond those of traditional SIM or MRM methods. Nevertheless, methods utilizing MRM-based acquisitions still remain the gold standard assay for high throughput, targeted, quantitative detection of specific analytes such as phosphopeptides.

**DATA-INDEPENDENT ACQUISITION APPROACHES FOR TARGETED PHOSPHOPEPTIDE DETECTION AND QUANTIFICATION**

Over the past 15 years, LC/MS/MS-based approaches using DDA and transition-based MS/MS methods such as MRM have been used quite effectively to assign phosphorylation sites as well as quantify phosphorylation site stoichiometry with the implementation of isotope labeling strategies or with the use of stable isotope-labeled internal standard peptides (e.g. AQUA [62]). Although quite powerful, both of these strategies have limitations. Because of intensity-derived switching criteria used with data-dependent methods, open tandem MS approaches such as DDA are by nature biased towards detection of the highest abundance components within a given sample and may never interrogate low abundance components such as low stoichiometry phosphopeptides. Although ‘targeted’ DDA methods allow one to only interrogate specific ions of interest such as a list of potential phosphopeptide m/z values regardless of intensity, other factors may preclude interrogation or detection of these selected phosphopeptides such as the presence of isobaric, unrelated peptides. In addition, targeted DDA methods are hypothesis-driven, requiring prior knowledge of the presence of a particular phosphopeptide in a given sample and is therefore not amenable to a discovery-based mode of operation. Likewise, MRM-based strategies are hypothesis-driven methods which nonetheless provide the ultimate in sensitivity for high-throughput, targeted quantitative phosphopeptide analysis.

Over the past 5 years, data-independent, parallel, multiplex fragmentation approaches have been reported for the analysis of simple and complex mixtures of proteins and hold promise in the area of phosphopeptide characterization [70–75] (Figure 2). Data-independent acquisition (DIA) approaches
allow for maximum qualitative coverage of a given sample with the added advantage of rigorous quantitative information within the same experiment. The enhanced qualitative coverage afforded by DIA analysis provides enhanced detection of low abundance components, including phosphopeptides. While DIA approaches are completely open and not targeted or hypothesis-driven at the data acquisition stage, the comprehensive nature of the resulting datasets enables ‘targeted data interrogation’ for detection of specific components of interest such as phosphopeptides. For example, neutral losses of phosphate from phosphopeptide precursors represent a phosphopeptide signature which may be easily queried post-acquisition from a DIA dataset.

Unlike MS/MS-based DDA strategies where individual ions of interest are sequentially selected and fragmented, DIA approaches do not use a precursor selection step prior to directing ions to the collision cell, thus resulting in fragmentation of all ions present at any given time in a chromatographic separation. In terms of peptide analysis, fragments will be generated for all charge states of each peptide precursor which, in many cases, enhances the resulting product ion intensities for any given peptide compared to the DDA analysis of individual selection and fragmentation of the peptide precursor at just one charge state. Consequently, the level of DIA enhancement will vary from peptide to peptide and is dependent upon the initial charge state distribution. In addition, unrelated peptide precursors which co-elute will also be simultaneously fragmented during DIA. As one might expect, this produces a very complex composite spectrum of product ions from all precursors without product-to-precursor ion relationships which are normally known on the basis of data-dependent precursor selection during MS/MS. Although product-to-precursor relationships are not explicitly captured in the spectral data from a multiplex fragmentation experiment, these relationships may be reconstructed post-acquisition from an LC/MS experiment via retention time alignment and chromatographic profiling of all precursor and product ions. The foundation of this approach is the fundamental relationship that a product ion derived from a particular precursor ion must exactly co-elute with its unique precursor ion [75], and thus the chromatographic profile defined by the retention time apex and peak shape must match between precursor and associated product ions. Once these relationships are established, data-independent LC/MS data can be processed to generate a number of highly reproducible ‘pseudo-MS/MS’ or reconstructed product ion spectra for each precursor ion detected. However, for this approach to work two additional criteria are required for the successful deconvolution of data-independent datasets: excellent mass measurement accuracy and relatively high mass resolution. Because of the complexity of product ion spectra derived without precursor ion selection in data-independent mode, mass measurement accuracy on the order of 10 ppm with a mass resolution of at least 10 000 are required to provide the specificity within the reconstructed product ion spectra in order to make unambiguous, statistically rigorous database search matches.

The only commercially available data-independent platform for protein and proteome analysis currently on the market is the LC/MS E platform by Waters Corporation (www.waters.com). In their DIA approach, an LC/MS analysis is conducted where data is captured from alternating scans of low (normal) collision energy and ‘elevated’ collision energy (MS E). Data in the low energy scan provides intact precursor ion m/z and intensity data whereas the elevated energy scan provides product ion data. The current data analysis package ProteinLynx Global Server 2.3 (PLGS 2.3) processes this data to generate reconstructed product ion spectra which are then searchable via any of the typical protein database search algorithms (e.g. Mascot, SEQUEST). Although any database search algorithm may be used, an optimized search algorithm known as IDENTIFY E was specifically designed for MS E data and incorporated into PLGS 2.3. Figure 3 shows product ion spectra obtained for the same bovine serum albumin (BSA) tryptic peptide from both DDA LC/MS/MS analysis (Figure 3A) and DIA LC/MS E analysis (Figure 3B) of a BSA tryptic digest. It is readily observed that the two spectra are qualitatively identical in terms of product ion coverage and relative product ion abundances indicating equivalent information content. Because alternating low and elevated energy scans are acquired over the full elution profile of every precursor during LC/MS E, the quality of LC/MS E derived product ion spectra often exceeds that of the corresponding LC/MS/MS datasets.

In addition to the qualitative data obtained during LC/MS E, quantitative data is acquired for each precursor ion during the low energy scan cycle. Because the sampling of low and elevated energy
scans occurs at regularly spaced intervals across the full elution profile, peak areas extracted for each precursor may be used as a measure of relative abundance in a given sample, thus allowing for cross sample comparisons for relative quantification without the use of isotope tagging reagents or stable isotope-labeled peptide standards. One could, however, spike AQUA-type stable isotope-labeled internal peptide standards into a sample and derive absolute quantities for any number of target phosphopeptide species using LC/MS\textsuperscript{E}. In addition to comparing the intensities of phosphopeptides to determine the relative phosphorylation change across samples, the integrated peak areas of the nonphosphorylated peptides and their corresponding phosphorylated versions may be used to calculate an apparent percentage of phosphorylation for each modified site within each sample which, in turn, can be relatively compared across samples to provide a measure of site-specific stoichiometry.

**APPLICATION OF DATA-INDEPENDENT LC/MS\textsuperscript{E} FOR IDENTIFYING AND QUANTIFYING PHOSPHORYLATION SITES OF THE MEMBRANE PROTEIN RECEPTOR BRII**

We recently reported an application of LC/MS\textsuperscript{E} in the characterization of site-specific stoichiometry of three phosphorylation sites of the BRASSINOSTEROID INSENSITIVE 1 (BRII) leucine-rich repeat receptor-like kinase (LRR RLK) in...
Arabidopsis thaliana [7]. Upon oligomerization with its LRR RLK co-receptor BRI1-ASSOCIATED KINASE 1 (BAK1), the extent of phosphorylation of BRI1 using $[^\gamma-32P]ATP$ was found to increase, thus indicating transphosphorylation of BRI1 by BAK1 [7]. The fundamental question from this result was which BRI1 site or sites displayed enhanced levels of phosphorylation upon BAK1 binding and transphosphorylation. DDA analysis followed by database searching of in-gel tryptic digests of BRI1 samples incubated with or without BAK1 suggested phosphorylation of the tryptic peptide corresponding to BRI1 residues 1157–1171. Unfortunately, the DDA-derived database search matches for this peptide provided an ambiguous assignment of modified sites, as there were three Ser residues and one Thr residue contained within this peptide, all of which could be phosphorylated. Specific product ions present in the DDA spectral match suggested potential phosphorylation of S1166, S1168 and T1169 in both the plus and minus BAK1 samples. Using LC/MS$^E$ we were able to elucidate the exact sites of phosphorylation and provide quantitative information on site-specific stoichiometry as illustrated in Figure 4.

The utilization of LC/MS$^E$ data to quantify the sites of phosphorylation is dependent upon chromatographic resolution of different phosphopeptide isoforms. The elution profiles for the nonphosphorylated tryptic BRI1 peptide corresponding to residues 1157–1171 along with its monophosphorylated counterparts can be easily distinguished as

**Figure 4:** LC/MS$^E$ analysis of a tryptic digest for BAK1-treated BRII kinase. Extracted ion chromatograms are shown for the (A) nonphosphorylated and (B) monophosphorylated isoforms of the doubly charged BRII tryptic peptide EIAGSGDSQSTIR corresponding to residues 1157–1171. Also shown is the (C) LC/MS$^E$ base peak chromatogram for the same retention time window. Labeled as DDA in panel B is the point at which in a separate, parallel LC/MS/MS analysis, a data-dependent MS/MS triggering event captured a product ion spectrum from the eluting monophosphorylated peptides.
shown in Figures 4A and 4B, respectively, in reference to the base peak chromatogram shown in Figure 4C. This particular peptide contains four potential sites of Ser/Thr phosphorylation (S1162, S1166, S1168 and T1169), two of which are located on adjacent residues. Clearly defined in Figure 4B for the extracted ion chromatogram of \( m/z \) 821.3, which corresponds to the doubly charged monophosphorylated peptide, are two distinctive chromatographic peaks indicative of at least two different phosphopeptides of the same primary sequence, each having a unique phosphorylated residue. Because of the high degree of structural similarity between the different phosphopeptide isoforms, they elute within 30 s of one another. Database searching and analysis of the LC/MS\(^E\) dataset unambiguously indicated the presence of three distinct sites of monophosphorylation on this peptide at S1166, S1168 and T1169 \[7\].

This was proven by the presence of unique, site-specific product ions observed for each of the two chromatographic peaks shown in Figure 4B. For the peptide eluting at 23.96 min, the LC/MS\(^E\) derived product ion spectrum presented in Figure 5 permitted the unequivocal assignment of phosphorylation at S1166. However, the peak generated at 23.42 min was deduced to contain two monophosphorylated peptides, each containing phosphorylation at either S1168 or T1169. The information rich LC/MS\(^E\) spectra indicated that this chromatographic peak shared product ions unique to phosphorylation on both S1168 and T1169, thus confirming coelution of both peptides at 23.42 min. Because S1168 and T1169 are adjacent to one another, phosphorylation at either site would be expected to produce phosphopeptides that possess similar reversed-phase chromatographic properties which was indeed the case as determined by LC/MS\(^E\).

The example of the monophosphorylated BRI1 peptide highlights one of the challenges faced by DDA-based approaches for the unequivocal assignment of specific sites of phosphorylation. In our DDA analysis of the BRI1 peptide digest, the retention time at which an LC/MS/MS spectrum for the precursor at \( m/z \) 821.4 was obtained corresponded to the valley between the two chromatographic peaks representing the three monophosphorylated forms as shown in Figure 4B. This resulted in the generation of a product ion spectrum representing the composite of all three phosphopeptide isoforms. Interrogation behavior such as this is common and contributes to ambiguous site assignments from database searching of DDA results and is consistent with the role serendipity has been reported to play in DDA-based acquisition schemes \[76\]. With three phosphorylation sites, this particular BRI1 phosphopeptide example is typical of the dynamic and complex nature of protein phosphorylation and the analytical challenges which are encountered using LC/MS/MS analysis.

**Figure 5:** MS\(^E\) spectrum of the BRI1 monophosphorylated tryptic peptide EIQAAGSGIDpSQSTIR. This particular phosphopeptide was resolved chromatographically from two other phosphopeptide isomers (Figure 4B) and unequivocal assignment of the phosphorylation site on S1166 was achieved by LC/MS\(^E\). Marked with an arrow is the mass difference between \( y_5 \) and \( y_7 \), corresponding to the site of phosphorylation (pSD).
Table 1: Phosphorylation stoichiometry of identified phosphorylation sites of Arabidopsis BRI1 in the presence or absence of BAK1

<table>
<thead>
<tr>
<th>BRI1 phosphorylation site</th>
<th>Percent phosphorylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+BAK1</td>
<td>-BAK1</td>
</tr>
<tr>
<td>S1166</td>
<td>30</td>
</tr>
<tr>
<td>S1168/T1169</td>
<td>23</td>
</tr>
</tbody>
</table>

Because of the limitations associated with unequivocal site assignment by DDA in this particular example, it is not difficult to understand that quantitative, site-specific stoichiometry assignment to each of these sites would be precluded with an open (i.e. non-targeted) DDA-derived scheme, which is in direct contrast to the open DIA approach used by LC/MS\textsuperscript{E} analysis.

In addition to the qualitative assignment of three BRI1 phosphorylation sites by LC/MS\textsuperscript{E}, quantitative data for these sites was obtained from the low energy scans. Table 1 summarizes the extent of BRI1 phosphorylation observed for the three sites in the presence or absence of BAK1 based upon the quantitative LC/MS\textsuperscript{E} data. From these results, we may conclude that S1166 is clearly a BAK1 dependent phosphorylation site within BRI1, since the extent of phosphorylation nearly doubles in the presence of BAK1 whereas phosphorylation of S1168/T1169 appears to be BAK1 independent. Although we know that the phosphopeptide isoforms containing either phosphorylated S1168 or T1169 are both present in the sample, we were unable to quantify the two sites independently because of their close structural similarity and resulting chromatographic co-elution. Nevertheless, the LC/MS\textsuperscript{E} data did provide biologically informative site-specific assignment and stoichiometry above and beyond that attainable from a similar, open DDA analysis of the same sample. Because our analysis of BRI1 was performed as a discovery mode experiment in which the primary goal was identification and assignment of novel phosphorylation sites, targeted DDA or MRM-based methods were not appropriate. This BRI1/BAK1 study illustrates the power of the data-independent LC/MS\textsuperscript{E} approach for discovery mode proteomic analysis since it permits the ability to determine phosphorylation site identification and determination of site-specific stoichiometry in a manner that is readily applicable to other post-translational modifications.

**PERSP\textsuperscript{E}\textsuperscript{E}**

Characterization of phosphoproteins and phosphoproteomes continues to progress as novel technologies are developed and integrated for phosphopeptide characterization. Current MS/MS-based analytical strategies only partially meet the demands in this area. Data-dependent MS/MS-based methodologies are plagued by problems with duty cycle, dynamic range, ambiguous site assignments and an inherent lack of quantitative information. MS/MS-based MRM strategies remain state-of-the-art in terms of sensitivity and selectivity for the quantification of specific phosphopeptides but lack the ability to operate in a discovery mode and provide essentially no qualitative data for the assignment of phosphorylation to a specific site within a peptide. Data-independent approaches such as LC/MS\textsuperscript{EI} are beginning to fill the analytical niche for discovery mode phosphorylation analysis where both the identification and quantitative characterization of novel phosphorylation sites may be achieved. Similar to the current trend in the area of biomarker discovery and validation, results from data-independent analyses can be used in a phosphoproteome context to provide the requisite precursor ion and optimal transition information to feed into an MRM-based targeted quantitative analysis pipeline for the high-throughput interrogation of phosphopeptides of interest in complex biological samples.

**Key Points**

- Characterization of protein or proteome phosphorylation is an extremely complex problem which must include unambiguous site assignment along with quantification of site-specific stoichiometry in order to provide the level of detail necessary for interpretation in a meaningful biological context.
- MS methods using data-dependent acquisition schemes are biased towards detecting the highest abundance components in a mixture and sporadically acquire MS and MS/MS data over the LC separation, thus precluding comprehensive targeted phosphopeptide detection and quantification.
- While providing the gold standard for quantification over a wide dynamic range, targeted methods for phosphopeptide detection based on SIM or MRM suffer in discovery mode experiments because of their hypothesis-driven nature.
- Compared to data-dependent or targeted-MRM-based approaches, data-independent methods of acquisition such as LC/MS\textsuperscript{E} provide the most comprehensive analysis of protein and proteome samples, enabling post-acquisition, targeted data interrogation for phosphopeptide species both qualitatively and quantitatively.
- A data-independent approach using LC/MS\textsuperscript{E} was shown to effectively allow unequivocal identification and determination of site-specific stoichiometry of three phosphorylation sites within a single peptide of the cytoplasmic domain of the membrane protein receptor BRI1 subjected to transphosphorylation by BAK1.
Acknowledgements
We thank the research agencies of North Carolina State University and the North Carolina Agricultural Research Service for continued support of our biological mass spectrometry research. The authors also thank Scott Geromans of Waters Corporation for many helpful discussions.

FUNDING
National Science Foundation (MCB-0419819 and DBI-0619250); United States Department of Agriculture (NRI 2006-35604-16739).

References


