Published in final edited form as:

Anal Biochem. 2008 April 1; 375(1): 1–10. doi:10.1016/j.ab.2007.11.023.

Taking Aim at Shotgun Phosphoproteomics

Jason D. Hoffert and Mark A. Knepper

Laboratory of Kidney and Electrolyte Metabolism; National Heart, Lung, and Blood Institute;
Bethesda, MD 20892

Abstract

Shotgun phosphoproteomics employs liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) to analyze phosphopeptides from complex protein mixtures, allowing detection and quantification of phosphorylation events on a global scale. Within the past few years, this powerful technique has been used to uncover novel phosphorylation sites as well as explore changes in protein phosphorylation during cellular signaling in various systems. This review presents a general overview of current phosphoproteomic methodologies, including summaries of various approaches to phosphopeptide enrichment, alternative MS fragmentation strategies, and the latest software for analysis of phosphopeptide data sets.

Keywords

phosphoproteome; phosphoproteomic; phosphorylation; neutral loss; label-free; mass spectrometry

INTRODUCTION

Protein phosphorylation is one of the most common post-translational modifications in eukaryotic cells and plays a critical role in a vast array of cellular processes. Phosphorylation predominantly occurs on the amino acids serine, threonine, and tyrosine [1], with serine being the most frequently modified residue. A recent study of more than 2,000 HeLa cell phosphoproteins demonstrated that the distribution of phospho-serine, -threonine, and -tyrosine sites was 86.4%, 11.8%, and 1.8%, respectively [2]. The relative importance of phosphorylation in biology is reflected in the large number of kinase genes (518) present in the human genome [3]. At the protein level, different cell types may express completely different sets of kinases, with each kinase often capable of phosphorylating multiple substrates. To add to this complexity, a given phosphoprotein is often modified at multiple sites through the actions of distinct kinases.

Within the past few years, global detection and quantification of these complicated phosphorylation events has been made possible through the development of so-called “shotgun” proteomic methodologies (reviewed in [4]). Traditional shotgun proteomics is a “bottom-up” approach that utilizes liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS)¹ to analyze proteolytic peptide fragments derived from highly complex mixtures of proteins from a particular cell or tissue. Historically, analysis of phosphopeptides by mass

Address for correspondence: Jason D. Hoffert, PhD National Institutes of Health Bldg. 10, Room 6N260 10 CENTER DR MSC 1603 BETHESDA, MD 20892-1603 Phone: (301)496-9010, Fax (301)402-1443 e-mail: hoffertj@nhlbi.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

spectrometry has been challenging due to a number of factors including low stoichiometry, ion suppression effects, and reduced peptide backbone fragmentation. However, the availability of more sensitive instrumentation and the development of new chromatography techniques for enriching phosphopeptides have facilitated large-scale phosphoproteomic analysis, allowing researchers to routinely detect hundreds or even thousands of phosphorylation events in a single experiment. Despite many technological advances, shotgun phosphoproteomics remains largely “bottom-up”, providing little information at the phosphoprotein level. Combining this technique with “top-down” MS analysis of intact proteins has the potential to expand the capabilities of this method [5-7].

The main objective of this review is to present a detailed analysis of a typical shotgun phosphoproteomics experiment, including potential caveats, common variations in method, and examples from the current literature. While this review includes references to a number of large-scale phosphoproteomic studies published within the last few years, it is not meant to be comprehensive. In addition, we have chosen to focus specifically on studies that have utilized electrospray ionization (ESI)-based tandem mass spectrometry to identify the overall phosphoproteome of a given system. Therefore, studies dealing with sub-phosphoproteomes (e.g. the phosphotyrosine proteome [8-10], reviewed in [11]) will not be covered.

OVERALL EXPERIMENTAL DESIGN

A general outline for a typical shotgun phosphoproteomics experiment is given in Figure 1. A sample containing both phosphorylated and non-phosphorylated proteins is digested with a site-specific protease, normally trypsin which cleaves the polypeptide chain after arginine and lysine residues. The peptide sample is then subjected to one or more chromatography steps in order to enrich phosphopeptides and to remove a significant portion of non-phosphorylated peptides that could interfere with subsequent analysis. The phosphopeptide sample is then applied to a reversed-phase HPLC column which further stratifies peptides based on relative hydrophobicity. This chromatography step is coupled online to a tandem mass spectrometer through an electrospray source.

The mass spectrometer initially performs a survey scan to obtain the mass-to-charge (m/z) ratios of full-length precursor (“parent”) ions (MS^1). These precursor ions are individually selected and fragmented via a process called collision-induced dissociation (CID) which produces random fragmentation at the peptide bond creating a “daughter” ion spectrum (MS^2). This spectrum is searched against a protein sequence database that has been digested *in silico* with the same protease used for the initial digestion in order to obtain a phosphopeptide identification. For phosphopeptides modified at serine or threonine residues, the MS^2 spectrum is often characterized by a predominant “neutral loss” peak which corresponds to the parent ion that has lost phosphoric acid. This peak can be selected for further fragmentation and analysis at the MS^3 level.

While the above description provides a general summary for a typical phosphoproteomic experiment, there are a number of variations on this theme that require a more detailed exploration. The following descriptions include a number of references to recent shotgun phosphoproteomic experiments that are summarized in Table 1, which also highlights important differences in experimental method.

¹ABBREVIATIONS

CID, collision-induced dissociation; ECD, electron capture dissociation; ESI, electrospray ionization; ETD, electron transfer dissociation; IMAC, immobilized metal affinity chromatography; iTRAQ, isobaric tags for relative and absolute quantification; LC-MS/MS, liquid chromatography-coupled tandem mass spectrometry; MS^2 , MS/MS; MS^3 , MS/MS/MS; PhosphoPIC, phosphopeptide identification and compilation; QUIL, Quantification Using Isotope Labeling; QUOIL, quantification without isotope labeling; SCX, strong cation exchange; SILAC, stable isotope labeling by amino acids in cell culture; TiO_2 , titanium dioxide

SAMPLE PREPARATION

Since phosphoproteins may represent only a small percentage of the total protein present in a particular cell, phosphoproteomic analysis usually begins with a large amount of starting material. Most large-scale studies utilize protein samples in the milligram range, including a recent analysis of HeLa cell nuclear phosphoproteins which utilized 8 mg of protein [12]. However, obtaining this much protein often proves difficult, particularly when analyzing purified cell types from mammalian tissues [13;14]. Phosphoproteomics of subcellular fractions may require less total protein for digestion since one has already eliminated non-relevant phosphorylated and non-phosphorylated proteins which could compete in the subsequent LC-MS/MS analysis [14]. The amount of starting material that is required for a particular experiment can also depend on other factors such as inclusion of multiple chromatography steps as well as the sensitivity of the mass spectrometer used for detection.

Another key component of the sample preparation phase is the choice of homogenization buffer. Traditional non-denaturing buffer solutions such as Tris [15], PBS [16], and sucrose-triethanolamine [14] with added protease and phosphatase inhibitors can be used for the initial homogenization of cells. These buffers have the advantage of keeping organelles and membranes intact, allowing for subsequent subcellular fractionation. The sample is then diluted or buffer-exchanged into a solution containing chaotropic agents such as guanidine or urea in order to denature proteins prior to protease digestion. If a whole-cell protein preparation is desired, the sample can also be homogenized directly in this denaturing buffer [13]. Ionic detergents such as SDS can also be used; however, they are difficult to remove prior to in-solution protease digestion.

A comparison of the effectiveness of various sample preparation methods on the distribution of phosphopeptide identifications is shown in Figure 2. Renal collecting duct membrane protein pellets were resuspended in either 50mM ammonium bicarbonate (NH_4HCO_3), 6M guanidine, or 1% SDS. Prior to digestion with trypsin, guanidine was diluted below 0.5M and SDS was removed via a standard precipitation protocol. Surprisingly, the sample that was digested in ammonium bicarbonate without prior solubilization produced the most phosphopeptide identifications, suggesting that protease digestion does not require the presence of denaturants. A similar result was recently found in an analysis of A431 cell lysates digested in the presence and absence of 1M urea [17]. However, when comparing replicate samples, digestion of solubilized proteins may yield more consistent results.

An alternative to in-solution digestion of phosphoproteins is to run an initial preparative 1-D SDS-PAGE gel, cut out the protein bands, and perform in-gel protease digestion. This method offers the added advantage of fractionating the protein sample and has been utilized effectively for shotgun phosphoproteomics by Steven Gygi's group at Harvard University [12;18-20].

PHOSHOPEPTIDE ENRICHMENT

Methods for enriching phosphopeptides from a complex sample fall into two general categories: affinity-based approaches and chemical derivitization. Affinity-based approaches are more common and include strong cation exchange (SCX), immobilized metal affinity chromatography (IMAC), and metal oxide affinity chromatography (MOAC).

SCX separates proteins based on interaction of positively charged (cationic) functional groups in proteins (i.e. side chains of lysine, arginine, and histidine, as well as amino termini) with a negatively charged (anionic) resin (e.g. polysulfoethylaspartamide-linked silica). At low pH values (< 3.0), tryptic peptides with no missed cleavages normally have a solution charge state of +2 and bind tightly to the SCX column, whereas phosphopeptides, due to the presence of negatively charged phosphate groups, have a reduced charge state and elute in earlier fractions

[12;21]. SCX is effective when utilized in combination with prior separation by 1-D gel or in combination with other forms of chromatography such as IMAC [13;22] or MOAC [2;15].

IMAC employs a trivalent metal cation, such as Ga^{+3} or Fe^{+3} , that has been immobilized to a solid support via chelation with either nitriloacetic acid (NTA) or iminodiacetic acid (IDA). Phosphopeptides are coordinated to this positive matrix via electrostatic interactions. The reaction is normally performed at low pH (< 3.0) to prevent non-specific binding of acidic peptides. The majority of non-phosphorylated peptides are removed through a series of low pH washes, followed by elution of phosphopeptides using a high pH solution (e.g. 0.1M NH_4HCO_3 , pH 9.0). A number of recent studies have utilized IMAC as the sole method of phosphopeptide enrichment [13;14], although a significant amount of non-phosphorylated peptides remain in the eluate. Therefore, IMAC is most effective when utilized in combination with a prior chromatography step such as SCX [22]. IMAC alone has a tendency to enrich for multiphosphorylated peptides [23], whereas IMAC combined with prior SCX seems to enrich for singly-phosphorylated peptides [22].

MOAC most commonly consists of beads coated with titanium dioxide (TiO_2) which coordinates the phosphate groups on phosphopeptides through a “bridging bidentate” mechanism distinct from IMAC coordination chemistry [24]. To prevent non-specific binding of acidic peptides, a “quenching agent” such as 2,5-dihydroxybenzoic acid (DHB) is often added to the binding and wash steps [2;24]. A recent study demonstrated that the sensitivity of TiO_2 is compromised for peptides with multiple basic residues [25].

Another method for phosphopeptide enrichment is chemical derivitization via phosphoramidate chemistry (PAC). Unlike the affinity techniques described previously, this method consists of reversible covalent modification of the phosphate group. This technique employs an initial methyl-esterification reaction which caps and protects free carboxyl groups on peptides. Through a complex reaction, phosphate groups are subsequently activated which allows coupling of the phosphopeptides to a solid support (either an amino-derivatized dendrimer [26] or maleimide-derivatized glass beads [16]) via a phosphoramidate linkage. After a series of stringent washes, phosphopeptides are released via acid hydrolysis of the phosphoramidate bond, leaving the phosphate group intact.

The specificity and sensitivity of these enrichment methods can be quite variable depending on the composition of the protein sample as well as the specific experimental conditions employed. A recent study by Bodenmiller *et al.* compared the performance of these different methods [27]. In that study, phosphopeptides from the cytosolic fraction of *D. melanogaster* Kc167 cells were enriched using IMAC (Fe^{+3}), phosphoramidate chemistry (PAC), pTiO_2 (phthalic acid as the quenching agent) and dhhTiO_2 (DHB as the quenching agent). Both IMAC and PAC outperformed TiO_2 in both specificity and sensitivity. Among the two variations of TiO_2 enrichment, samples quenched with phthalic acid possessed more than twice the number of phosphorylation sites and a fewer number of contaminating nonphosphorylated peptides, suggesting that phthalic acid is a superior quenching agent. All methods predominantly isolated singly phosphorylated peptides, although IMAC had a greater tendency of isolating doubly phosphorylated peptides. The most intriguing finding from this study was the low percentage of overlap (33-35%) in identified phosphorylation sites among the different techniques, which demonstrates a complementary relationship among the different phosphopeptide enrichment strategies.

LC-MS/MS ANALYSIS OF PHOSPHOPEPTIDES

Rapid developments in MS instrumentation have made LC-MS/MS an ideal technique for large-scale phosphoproteomic analysis. The latest generation of ion trap (IT) mass analyzers offers high sensitivity detection of phosphopeptides. In addition, versatile “hybrid” instruments

can combine the sensitivity of linear ion trap (LIT) technology with high mass accuracy detection via Orbitrap technology (e.g. LTQ-Orbitrap) or Fourier Transform-Ion Cyclotron Resonance (FT-ICR) (e.g. LTQ-FT) allowing accurate phosphopeptide quantification [2;13;22]. While these instruments can perform traditional MS/MS analysis of phosphopeptides, they can also perform higher level MSⁿ analysis.

Neutral loss scanning and MSⁿ analysis

Neutral loss of phosphoric acid from phosphopeptides occurs predominantly at phosphorylated serine and threonine residues, resulting in conversion of these amino acids to dehydroalanine and dehydroaminobutyric acid, respectively. A detailed mechanism for this phenomenon has been proposed by Tholey *et al.* which includes a six-membered intermediate that explains the paucity of neutral loss at phosphotyrosine residues [28]. When an ion trap mass spectrometer is run in positive ion mode, fragmentation of phosphopeptides by collision-induced dissociation (CID) often results in loss of phosphoric acid (H₃PO₄; 98 Da) from the precursor ion and detection of a predominant neutral loss peak in the MS² spectrum (Figure 1). Neutral loss effectively suppresses peptide backbone fragmentation which is often necessary for confident phosphopeptide identification and phosphorylation site localization. To overcome this drawback, the mass spectrometer can select the neutral loss ion based on the mass difference from its precursor ion (−98 Da) and trigger an additional round of CID fragmentation and detection at the MS³ level. A comparison of the quality between MS² and MS³ spectra for a singly-phosphorylated peptide is shown in Figure 3. The MS² spectrum is dominated by the neutral loss peak (NL) with a few low intensity b and y fragment ions (Figure 3B). In contrast, the MS³ spectrum shows an increased number and intensity of fragment ions, particularly so-called “site-determining” ions (*red peaks*), that will assist in correct phosphorylation site localization (Figure 3C).

Numerous large-scale phosphoproteomic studies have utilized neutral loss scanning and MS³ analysis [2;13;15;22]. Some of these studies have documented substantial increases in overall phosphopeptide identification rates through this approach. In an analysis of the yeast pheromone signaling pathway, Gruhler *et al.* reported the identification of 724 unique phosphopeptides through a combination of MS² and MS³ analysis as opposed to only 417 unique phosphopeptides from MS² analysis alone [22]. In addition, a recent study of kidney collecting duct membrane fractions [14] identified only 72 phosphoproteins from an examination of MS² spectra, in comparison to 151 from MS³ analysis. In contrast to these reports, a recent analysis of nocodazole-arrested HeLa cell lysates by Beausoleil *et al.* found that only 8.7% of all non-redundant phosphorylation sites were contributed solely by MS³ spectra and that analysis of MS³ spectra provided very little additional information with respect to proper phosphorylation site assignment [18]. The relative contribution of MS³ spectra to the final phosphopeptide data set may depend on a number of factors including the type of sample being analyzed, differences in techniques for phosphopeptide purification, variations in MS instrumentation parameters such as CID energy, and the overall quality of MS² spectra. However, inclusion of MS³ data lends an added level confidence to phosphopeptide identifications [29;30], particularly in light of a recent report demonstrating a variety of sequence motifs and chemical modifications that can produce false-positive phosphopeptide identifications when analyzed solely at the MS² level [31].

Electron Transfer Dissociation (ETD)

While traditional CID fragmentation/neutral loss scanning has been a staple of large-scale phosphoproteomics analysis for a number of years, an alternative strategy has recently emerged called ETD that offers a number of distinct advantages. ETD, as well as its precursor ECD (electron capture dissociation), operate on the principle of bombarding multiply-protonated peptide ions with electrons which results in extensive peptide backbone fragmentation and

generation of c and z type ions [5]. As opposed to the b and y ions of traditional CID analysis that are formed through fragmentation at the peptide bond, c and z ions are produced through cleavage of the bond between the amide nitrogen and alpha carbon. One of the main advantages of ETD is that labile post-translational modifications (e.g. phosphorylation, glycosylation, acylation, ubiquitylation and sumoylation) remain intact during MS/MS analysis (reviewed in [32]). ETD fragmentation has also been used for “top-down” MS analysis of intact proteins [33] [32]. Another advantage of ETD is that it appears to be more sensitive than CID fragmentation for analysis of phosphopeptides. This technique has demonstrated its robustness in a recent analysis of yeast in which only 30- μ g of total protein yielded 1,252 phosphorylation sites on 629 proteins [34]. In another study employing ETD, Molina *et al.* identified 1,435 phosphorylation sites from human embryonic kidney 293T cells, of which 1,141 (~ 80%) were not previously described [35]. In this study, ETD was able to identify more phosphopeptides than CID in samples digested with either Lys-C (244 vs. 132) or trypsin (243 vs. 171). However, the majority of these identifications did not overlap, suggesting that ETD and CID fragmentation are complementary techniques.

Post-MS acquisition data analysis

Following the initial collection of MS spectra, there are a number of steps that are critical to generating high quality phosphopeptide data sets: proper database searching, exclusion of nonphosphorylated peptide identifications, estimation of data set false discovery rate (FDR), application of appropriate filtering parameters, and correct phosphorylation site assignment (Figure 4).

A popular method for searching phosphopeptide data sets is the target-decoy approach which was first described in an LC-MS/MS analysis of the yeast proteome [36] and later applied to analysis of phosphoproteomes [14;18]. This approach involves searching a composite database that contains protein sequences in both forward (“target”) and reversed (“decoy”) orientations. The number of false-positive hits found in the decoy database is then used to predict the number of false positives found in the forward database in order to calculate the false discovery rate (reviewed in [37]). Popular search algorithms that are compatible with the target-decoy strategy include Mascot [38], SEQUEST [39], OMSSA [40], and X! Tandem [41]. A recent comparison of these four programs indicated that when run on high mass accuracy data sets, Mascot and SEQUEST offered higher sensitivity for phosphopeptide identifications [42]. This study also demonstrated that replicate runs of the same sample significantly increased the total number of identifications. Although a particular search algorithm may perform better under certain conditions, investigators should not limit their searches by using just one algorithm since each has been shown to identify distinct sets of peptides [43]. In addition, recent alternative strategies to traditional database searching based on peptide sequence tags (GutenTag [44], InsPecT [45], and Spectral Networks Analysis [46]) may improve identification of peptides with multiple post-translational modifications.

After the database search, data sets must be filtered both to remove non-phosphorylated peptide identifications and to estimate false positive rates. Two publicly-available software programs, DTASelect (<http://fields.scripps.edu/DTASelect/>) [47-49] and PhosphoPIC [14], can filter phosphopeptide data sets according to predetermined false discovery rates. DTASelect offers a useful “Contrast” tool that can compare peptide similarities and differences among multiple samples. PhosphoPIC includes an option for merging phosphopeptide data sets and is useful for determining common and unique phosphopeptide identifications from multiple MS levels (i.e. MS² and MS³). Both programs allow the user to remove nonphosphorylated peptides and both programs are compatible with SEQUEST.

Once the phosphopeptide data set is filtered, the site(s) of phosphorylation for each sequence must be confirmed. If there are multiple serine, threonine, or tyrosine residues present in the

peptide sequence, there is a distinct possibility that the “top-ranked” peptide from the database search has an incorrect assignment. Traditionally this process was performed by manual examination of spectra; however, the use of software programs has automated this task. One program called MSQuant (<http://msquant.sourceforge.net>) [2;29] is compatible with MS² and MS³ searches performed using Mascot. The scoring is probability based and takes into account the number of matched b and y ions within an observed mass range for all possible combinations of phosphorylation sites within a peptide sequence. Another program, Ascore (<http://ascor.med.harvard.edu/ascor.php>), also utilizes a probability-based scoring algorithm which takes into account both the presence and intensity of site-determining ions in MS/MS spectra that have been searched using SEQUEST [14;18].

MS-based Quantification of Phosphopeptides

Quantification of phosphopeptides is often the ultimate goal of large-scale phosphoproteomics experiments. Measuring relative or absolute phosphopeptide (phosphoprotein) abundances under different experimental conditions (e.g. hormone treatment, response to stress) or disease states, across multiple time points, is critical to understanding the dynamic regulatory networks involved in these processes. A number of MS-based quantification techniques are currently employed for measuring relative changes in phosphopeptide abundance.

Stable isotope labeling by amino acids in cell culture (SILAC) [50] involves metabolically labeling distinct populations of cells through incubation in media containing either standard (“light”) or stable isotope-labeled (“heavy”) versions of an essential amino acid (e.g. lysine, arginine, leucine). The two protein samples are mixed together prior to trypsinization, followed by phosphopeptide enrichment and analysis by LC-MS/MS. Phosphopeptide quantification occurs by comparing the relative abundances of the full length heavy and light parent ions throughout a specific chromatographic time period. The area under the curve of each “extracted ion chromatogram” (XIC) is directly proportional to the abundance of that peptide in the sample. Programs such as MSQuant [2;29] and QUIL [51], have automated this process for large data sets. Using the SILAC method, Gruhler *et al.* identified 139 phosphopeptides from yeast that changed in abundance at least two-fold in response to mating pheromone [22]. In a recent study of EGF-treated HeLa cells, Olsen *et al.* employed multiple combinations of heavy and light amino acids permitting quantification of phosphopeptides from 5 different timepoints [2]. In that study, the authors found that 14% of the 6,600 phosphorylation sites identified were modulated at least 2-fold by EGF treatment. Although SILAC is a popular approach in cultured cells, it is not practical for whole organism labeling given the associated costs and is not feasible for short-term studies using isolated mammalian tissues.

Another labeling strategy that can be used for phosphopeptide quantification is called iTRAQ (isobaric Tags for Relative and Absolute Quantification) [52]. This method involves labeling peptides from different samples with amine-reactive isobaric tags, combining the samples, and measuring the intensity of distinct reporter ions that are generated following fragmentation in the mass spectrometer. Each tag consists of: (1) a reporter group of a specific mass (114-117 Da) which allows simultaneous quantification of a peptide from up to 4 samples (although an 8-plex version will be available soon), (2) a “balancer” region to make each tag isobaric, and (3) a peptide reactive group. Studies using iTRAQ for large-scale phosphopeptide quantification have focused on phosphotyrosine-containing proteins [10;53]. Due to the small size of the reporter ions, iTRAQ is only compatible with wider mass range instruments such as triple quadrupole and Orbitrap mass spectrometers, not with traditional ion traps.

An alternative to the labeling methods described above is the label-free approach. As the name suggests, this technique does not require labeling of proteins or peptides prior to mass spectrometry and involves a direct comparison of a particular peptide's extracted ion chromatogram among separate LC-MS runs. Since no expensive labeling reagents are

necessary, there are tremendous cost benefits to using this strategy, not to mention the amount of time saved during sample preparation. The reproducibility and linearity of the label-free approach was recently demonstrated for protein samples from rat kidney and human plasma [54]. In that study, nearly ideal Pearson's correlation coefficients were found for peptide ion peak areas and retention times when comparing identical samples among multiple LC-MS runs. In addition, the study also introduced a user-friendly, BioWorks/SEQUEST-compatible software program for label-free quantification, QUOIL, which has been employed for quantitative and statistical analysis of phosphopeptides from multiple data sets [13;14]. QUOIL incorporates algorithms for outlier-resistant mean estimation and for adjusting statistical significance threshold in multiplicity of testing (step-down p-value) to minimize the rate of false positives. For a current synopsis of alternative programs that are available for label-free quantification, see the review by Wong *et al.* [55].

CONCLUSION

Shotgun LC-MS/MS has become an attractive method for identifying and quantifying protein phosphorylation and the recent surge in large-scale phosphoproteomics publications reflects this trend. Comparative studies of different phosphoproteomic methodologies have made it clear that a comprehensive analysis should not involve one strategy, but should employ multiple strategies, particularly when it comes to phosphopeptide enrichment, MS fragmentation, and searching MS spectra. With the growing popularity of this approach, it becomes necessary for both new and seasoned investigators to focus on the quality of the data in terms of limiting both false positives as well as incorrect phosphorylation site assignments. The numerous software applications that were presented in this review will assist in this process. Improvements in the size and quality of phosphopeptide data sets should also occur as developing technologies in the field of mass spectrometry become more widely available.

ACKNOWLEDGEMENTS

We thank Dr. Trairak Pisitkun for helpful insights and review of this manuscript. This work is funded by the Intramural Budget of the National Heart, Lung, and Blood Institute (National Institutes of Health, project no. ZO1-HL001285 to M. A. Knepper)

Reference List

1. Reinders J, Sickmann A. State-of-the-art in phosphoproteomics. *Proteomics* 2005;5:4052–4061. [PubMed: 16196093]
2. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;127:635–648. [PubMed: 17081983]
3. Johnson SA, Hunter T. Kinomics: methods for deciphering the kinome. *Nat.Methods* 2005;2:17–25. [PubMed: 15789031]
4. Wu L, Han DK. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert.Rev.Proteomics* 2006;3:611–619. [PubMed: 17181475]
5. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:9528–9533. [PubMed: 15210983]
6. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal.Chem* 2007;79:3525–3534. [PubMed: 17441688]
7. Millea KM, Krull IS, Cohen SA, Gebler JC, Berger SJ. Integration of multidimensional chromatographic protein separations with a combined “top-down” and “bottom-up” proteomic strategy. *J.Proteome.Res* 2006;5:135–146. [PubMed: 16396504]

8. Pandey A, Podtelejnikov AV, Blagoev B, Bustelo XR, Mann M, Lodish HF. Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc.Natl.Acad.Sci.U.S.A* 2000;97:179–184. [PubMed: 10618391]
9. Steen H, Kuster B, Fernandez M, Pandey A, Mann M. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J.Biol.Chem* 2002;277:1031–1039. [PubMed: 11687594]
10. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:5860–5865. [PubMed: 17389395]
11. Machida K, Mayer BJ, Nollau P. Profiling the global tyrosine phosphorylation state. *Mol.Cell Proteomics* 2003;2:215–233. [PubMed: 12754303]
12. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:12130–12135. [PubMed: 15302935]
13. Hoffert JD, Pisitkun T, Wang G, Shen RF, Knepper MA. Quantitative phosphoproteomics of vasopressin-sensitive renal cells: regulation of aquaporin-2 phosphorylation at two sites. *Proc.Natl.Acad.Sci.U.S.A* 2006;103:7159–7164. [PubMed: 16641100]
14. Hoffert JD, Wang G, Pisitkun T, Shen RF, Knepper MA. An automated platform for analysis of phosphoproteomic datasets: application to kidney collecting duct phosphoproteins. *J.Proteome.Res* 2007;6:3501–3508. [PubMed: 17683130]
15. Macek B, Mijakovic I, Olsen JV, Gnäd F, Kumar C, Jensen PR, Mann M. The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol.Cell Proteomics* 2007;6:697–707. [PubMed: 17218307]
16. Bodenmiller B, Mueller LN, Pedrioli PG, Pflieger D, Junger MA, Eng JK, Aebersold R, Tao WA. An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: application to *Drosophila melanogaster* Kc167 cells. *Mol.Biosyst* 2007;3:275–286. [PubMed: 17372656]
17. Kim SC, Chen Y, Mirza S, Xu Y, Lee J, Liu P, Zhao Y. A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea. *J.Proteome.Res* 2006;5:3446–3452. [PubMed: 17137347]
18. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat.Biotechnol* 2006;24:1285–1292. [PubMed: 16964243]
19. Li X, Gerber SA, Rudner AD, Beausoleil SA, Haas W, Villen J, Elias JE, Gygi SP. Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. *J.Proteome.Res* 2007;6:1190–1197. [PubMed: 17330950]
20. Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. *Mol.Cell Proteomics* 2004;3:1093–1101. [PubMed: 15345747]
21. Lim KB, Kassel DB. Phosphopeptides enrichment using on-line two-dimensional strong cation exchange followed by reversed-phase liquid chromatography/mass spectrometry. *Anal.Biochem* 2006;354:213–219. [PubMed: 16750159]
22. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol.Cell Proteomics* 2005;4:310–327. [PubMed: 15665377]
23. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat.Biotechnol* 2002;20:301–305. [PubMed: 11875433]
24. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol.Cell Proteomics* 2005;4:873–886. [PubMed: 15858219]
25. Klemm C, Otto S, Wolf C, Haseloff RF, Beyermann M, Krause E. Evaluation of the titanium dioxide approach for MS analysis of phosphopeptides. *J.Mass Spectrom* 2006;41:1623–1632. [PubMed: 17089331]

26. Tao WA, Wollscheid B, O'Brien R, Eng JK, Li XJ, Bodenmiller B, Watts JD, Hood L, Aebersold R. Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat.Methods* 2005;2:591–598. [PubMed: 16094384]
27. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat.Methods* 2007;4:231–237. [PubMed: 17293869]
28. Tholey A, Reed J, Lehmann WD. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J.Mass Spectrom* 1999;34:117–123. [PubMed: 12440389]
29. Olsen JV, Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc.Natl.Acad.Sci.U.S.A* 2004;101:13417–13422. [PubMed: 15347803]
30. Ulintz PJ, Bodenmiller B, Andrews PC, Aebersold R, Nesvizhskii AI. Investigating MS2-MS3 matching statistics: A model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol.Cell Proteomics*. 2007
31. Lehmann WD, Kruger R, Salek M, Hung CW, Wolschin F, Weckwerth W. Neutral loss-based phosphopeptide recognition: a collection of caveats. *J.Proteome.Res* 2007;6:2866–2873. [PubMed: 17569551]
32. Mikesch LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. *Biochim.Biophys.Acta* 2006;1764:1811–1822. [PubMed: 17118725]
33. Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 2005;102:9463–9468. [PubMed: 15983376]
34. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:2193–2198. [PubMed: 17287358]
35. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:2199–2204. [PubMed: 17287340]
36. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J.Proteome.Res* 2003;2:43–50. [PubMed: 12643542]
37. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat.Methods* 2007;4:207–214. [PubMed: 17327847]
38. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–3567. [PubMed: 10612281]
39. Yates JR III, Eng JK, McCormack AL, Schieltz D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal.Chem* 1995;67:1426–1436. [PubMed: 7741214]
40. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J.Proteome.Res* 2004;3:958–964. [PubMed: 15473683]
41. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004;20:1466–1467. [PubMed: 14976030]
42. Bakalarski CE, Haas W, Dephoure NE, Gygi SP. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal.Bioanal.Chem.* 2007
43. Balgley BM, Laudeman T, Yang L, Song T, Lee CS. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol.Cell Proteomics* 2007;6:1599–1608. [PubMed: 17533222]
44. Tabb DL, Saraf A, Yates JR III. GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal.Chem* 2003;75:6415–6421. [PubMed: 14640709]

45. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal.Chem* 2005;77:4626–4639. [PubMed: 16013882]
46. Bandeira N, Tsur D, Frank A, Pevzner PA. Protein identification by spectral networks analysis. *Proc.Natl.Acad.Sci.U.S.A* 2007;104:6140–6145. [PubMed: 17404225]
47. Lu B, Ruse C, Xu T, Park SK, Iii JY. Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal.Chem* 2007;79:1301–1310. [PubMed: 17297928]
48. Tabb DL, McDonald WH, Yates JR III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J.Proteome.Res* 2002;1:21–26. [PubMed: 12643522]
49. Cantin GT, Venable JD, Cociorva D, Yates JR III. Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway. *J.Proteome.Res* 2006;5:127–134. [PubMed: 16396503]
50. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol.Cell Proteomics* 2002;1:376–386. [PubMed: 12118079]
51. Wang G, Wu WW, Pisitkun T, Hoffert JD, Knepper MA, Shen RF. Automated quantification tool for high-throughput proteomics using stable isotope labeling and LC-MSn. *Anal.Chem* 2006;78:5752–5761. [PubMed: 16906720]
52. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol.Cell Proteomics* 2004;3:1154–1169. [PubMed: 15385600]
53. Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM. Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 2006;55:2171–2179. [PubMed: 16873679]
54. Wang G, Wu WW, Zeng W, Chou CL, Shen RF. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J.Proteome.Res* 2006;5:1214–1223. [PubMed: 16674111]
55. Wong JW, Sullivan MJ, Cagney G. Computational methods for the comparative quantification of proteins in label-free LCn-MS experiments. *Brief.Bioinform.* 2007

Figure 1.

A standard protocol for LC-MS/MS-based phosphoproteomic analysis. Proteins are digested with a site-specific protease, usually trypsin. Phosphorylated peptides are enriched from non-phosphorylated peptides using one or more of the approaches listed. Phosphopeptide samples are then separated on a reversed-phase liquid chromatography (RP-HPLC) column which is connected on-line to the mass spectrometer. Full-length peptide (parent) ions are measured in the initial scan (MS¹) and selected individually for fragmentation by collision-induced dissociation (CID). This process often results in “neutral loss” of phosphoric acid (98 Da; H₃PO₄) from the parent ion and very little backbone fragmentation is detectable in the MS² spectrum. The neutral loss peak can be selected for further CID fragmentation and detection

in the MS³ spectrum to increase the overall phosphopeptide identification rate or to confirm poor quality MS² identifications.

Figure 2.

Effect of sample preparation on the distribution of phosphoprotein identifications. Renal collecting duct membrane protein pellets were resuspended in either 50mM ammonium bicarbonate (NH_4HCO_3), 6M guanidine, or 1% SDS prior to trypsinization and analysis by LC-MS/MS.

Figure 3.

Comparison of the quality of MS² and MS³ spectra for a typical phosphopeptide. A. Amino acid sequence of a phosphopeptide identified as water channel aquaporin-2 (AQP2) with a single phosphorylated residue, Ser-256 (*). The region that is highlighted in red includes 5 b and 5 y fragment ions that represent so-called “site-determining ions” that could distinguish the actual site of phosphorylation (i.e. Ser-256) from another potential site, Ser-261. B. MS² spectrum of the peptide shown in A. The blue neutral loss peak (NL) is predominant while the number and abundance of b and y ions, particularly the site-determining ions (red arrows) is low. Neutral loss peak with loss of 1 water (NL - H₂O). C. Higher quality MS³ spectrum

generated by fragmentation of the neutral loss peak shown in B. There is an increase in the number and abundance of b and y ions relative to the MS² spectrum.

Figure 4.

Analyzing phosphopeptide data sets in the post-acquisition phase. After MS spectra are acquired, the appropriate protein sequence database must be searched using specific criteria in order to generate a list of potential peptide matches. By using various filtering parameters based on a predetermined false discovery rate (FDR), only the highest quality spectra corresponding to predicted phosphopeptides are selected from this list. At this point both MS² and MS³ data sets can be combined and analyzed to determine the total number of phosphoproteins identified as well as the exact site(s) of phosphorylation on each corresponding peptide. Assigning phosphorylation sites can be accomplished using computer programs (automated) or by personal examination of spectra (manual).

Table 1

Summary of recent shotgun phosphoproteomics publications

Primary Author	Year	Organ/Cell Type	Condition	Protein Fractionation	Purification	Mass Spectrometer	Fragmentation
Ballif <i>et al.</i> [20]	2004	mouse brain	whole cell	1-D PAGE	SCX	LCQ DECA XP	CID
Beausoleil <i>et al.</i> [12]	2004	HeLa	nuclear prep	1-D PAGE	SCX	LCQ DECA XP	CID/MS ³
Beausoleil <i>et al.</i> [18]	2006	HeLa	nocodazole treatment	1-D PAGE	SCX	LTQ-FT	CID/MS ³
Bodenmiller <i>et al.</i> [27]	2007	<i>Drosophila</i> Kc167	cytosolic fraction	no	PAC, IMAC, TiO ₂	LTQ-FT	CID
Cantin <i>et al.</i> [49]	2006	HeLa	TNF- α treatment	no	IEF/Ga ⁺ -IMAC	LTQ	CID/MS ³
Chi <i>et al.</i> [34]	2007	yeast	whole cell	no	Fe ³⁺ -IMAC	LTQ	ETD
Gruhler <i>et al.</i> [22]	2005	yeast	pheromone signaling	no	SCX/Fe ³⁺ -IMAC	LTQ-FT	CID/MS ³
Hoffert <i>et al.</i> [13]	2006	kidney collecting duct	vasopressin signaling	no	IMAC	LTQ-FT	CID/MS ³
Hoffert <i>et al.</i> [14]	2007	kidney collecting duct	vasopressin signaling - membrane fraction	no	SCX/Ga ⁺ -IMAC	LTQ-FT	CID/MS ³
Li <i>et al.</i> [19]	2007	yeast	α -factor-arrested	1-D PAGE	IMAC	LTQ-Orbitrap	CID
Macek <i>et al.</i> [15]	2007	<i>B. subtilis</i>	whole cell	no	SCX/Fe ³⁺ -IMAC	LTQ-FT, LTQ-Orbitrap	CID/MS ³
Molina <i>et al.</i> [35]	2007	HEK 293T	pervanadate/calyculin A	no	RP-HPLC/TiO ₂	Agilent 6340 3-D ion trap	ETD
Olsen <i>et al.</i> [2]	2006	HeLa	EGF treatment	no	SCX/TiO ₂	LTQ-FT	CID/MS ³