Mapping and analysis of phosphorylation sites: a quick guide for cell biologists

Noah Dephoure[a], Kathleen L. Gould[b], Steven P. Gygi[a], and Douglas R. Kellogg[c]
[a]Department of Cell Biology, Harvard Medical School, Boston, MA 02115-5701; [b]Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN 37232-2175; [c]Department of Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, Santa Cruz, CA 95064-1077

ABSTRACT A mechanistic understanding of signaling networks requires identification and analysis of phosphorylation sites. Mass spectrometry offers a rapid and highly sensitive approach to mapping phosphorylation sites. However, mass spectrometry has significant limitations that must be considered when planning to carry out phosphorylation-site mapping. Here we provide an overview of key information that should be taken into consideration before beginning phosphorylation-site analysis, as well as a step-by-step guide for carrying out successful experiments.

INTRODUCTION

One of the most difficult and important challenges in cell biology and medicine is to understand how signaling networks integrate and relay signals. An important step in analyzing signaling networks is the identification and characterization of phosphorylation sites, both on individual proteins and more globally on the proteome. When done carefully, phosphorylation-site analysis provides definitive information on functional relationships between signaling proteins. In the case of multisite phosphorylation on a single protein, it is also a necessary step in defining the mechanism of phosphorylation (i.e., processive or distributive), which determines the nature of the signaling response.

Although protein phosphorylation has been studied for decades, recent advances in mass spectrometry (MS) have revolutionized analysis of signaling by allowing rapid identification of phosphorylation sites with precision and sensitivity. Nevertheless, limitations remain that have important implications for the design of mapping experiments and the subsequent interpretation of biological experiments that make use of the MS-derived data. Here we provide an overview of technical considerations for successful analysis of serine, threonine, and tyrosine phosphorylation events by mass spectrometry, as well as a discussion of alternative, "old-fashioned" approaches that can be used to complement and/or bolster mass spectrometry data. We have not aimed to provide comprehensive protocols. Rather, we provide general guidelines and technical information that should be considered before starting a project. The principles and practice of mass spectrometry have been extensively reviewed elsewhere (Steen and Mann, 2004; Gingras et al., 2007; Walther and Mann, 2010).

GENERAL CONSIDERATIONS

The basics of phosphorylation-site mapping

In the simplest mapping experiment, a purified protein is digested with a protease that cuts at defined sites to produce small peptides. Tandem mass spectrometry (MS/MS) is used to measure intact peptide masses and fragment ion masses. Computer algorithms then identify peptides by matching these experimental data to theoretical spectra derived from sequence databases (Figure 1A). Phosphopeptide mapping is accomplished by matching the data to theoretical spectra that consider all possible phosphorylated versions of each peptide (Figure 1B). A number of commercial and open-source search algorithms are widely available (Eng et al., 1994; Perkins et al., 1999; Fenyo and Beavis, 2003; Craig and Beavis, 2004; Geer et al., 2004; Tabb et al., 2007). Although each is subject to its own caveats and hurdles, most known small covalent posttranslational modifications can also be mapped using these principles.
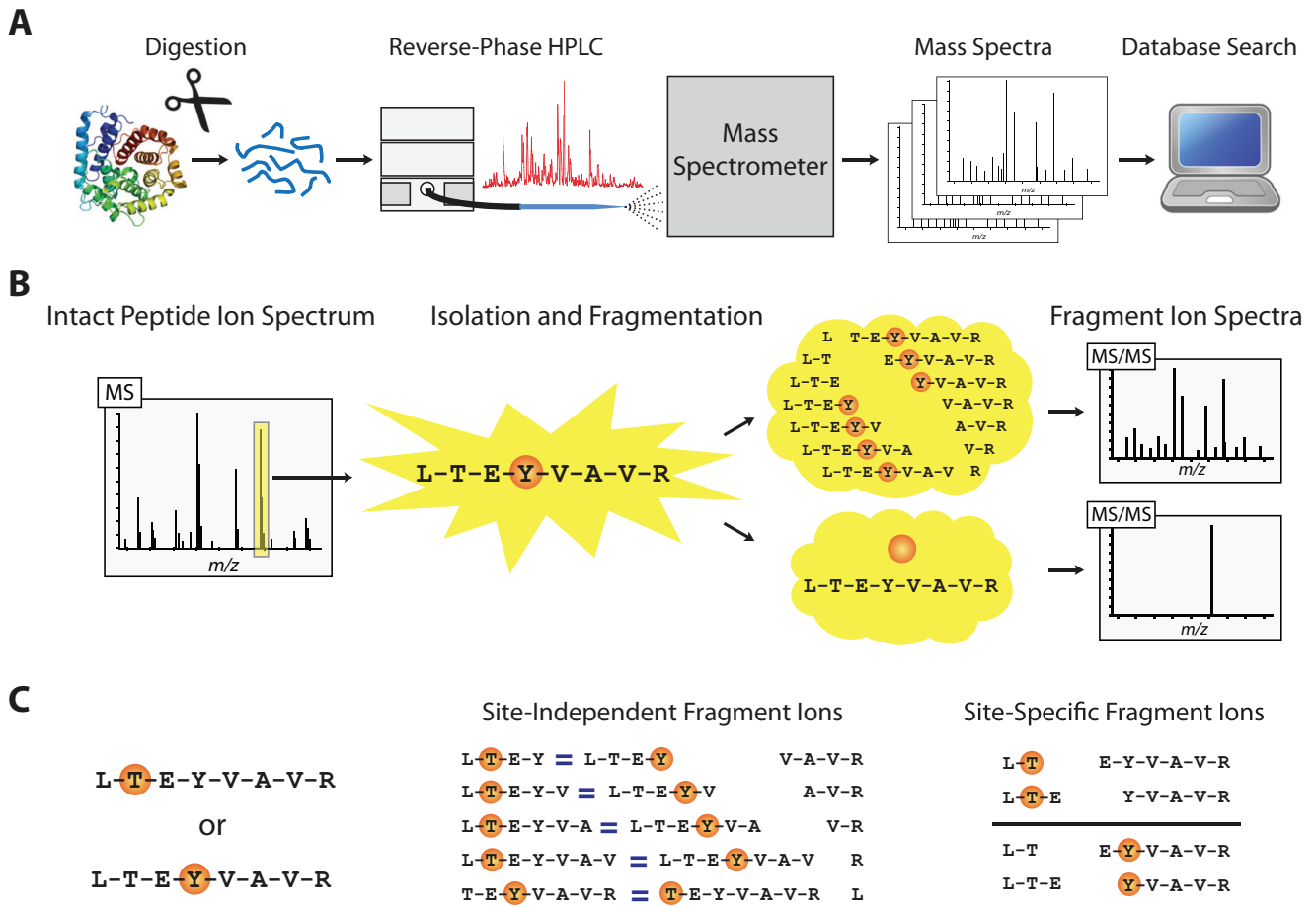
FIGURE 1: Tandem mass spectrometry (MS/MS) analysis of protein phosphorylation. (A) Protein samples are digested with a proteolytic enzyme. The resulting peptides are separated by reverse-phase high-performance liquid chromatography. Peptides enter the mass spectrometer as they elute from the column. Peptide matching is done algorithmically using spectral data and sequence database information. (B) Two basic types of information are generated in the mass spectrometer. The masses of intact peptide ions are determined in a full scan (MS or MS1). Peptides are then isolated one at a time, as depicted for the highlighted peak, and fragmented by colliding them with an inert gas. The resultant fragment ions are detected in a MS/MS (or MS2) scan. With sufficient coverage of fragment ions, the position of the phosphorylated residue, circled in orange, can be identified from the MS/MS (upper path). Fragmentation, however, often liberates the relatively labile phosphate groups at the expense of more informative fragmentation of the peptide backbone (lower path). In extreme cases, MS/MS spectra are dominated by a single ion representing the intact peptide stripped of phosphate. (C) Site localization is dependent on the detection of ions that can distinguish between possible phosphorylatable residues. In the example shown, a threonine and a tyrosine are separated by one amino acid. Fragment ions in the central panel either do not contain a phosphorylatable residue or will have equivalent mass for both possible phosphopeptides. If the threonine is phosphorylated, we would expect to see some or all of the top four site-specific fragment ions shown on the right. If, instead, the phosphate lies on the tyrosine, we would see the bottom four.

Although theoretically equivalent to the identification of unphosphorylated peptides, a number of issues complicate phosphopeptide identification. First among these is that the phosphate moiety is relatively labile, and during fragmentation it is often released at the expense of more informative cleavage of the peptide backbone (Figure 1B). Thus phosphopeptide issues can generate reduced-quality MS/MS spectra, leading to fewer high-confidence peptide matches. A second problem can be the scarcity of phosphorylation within the protein of interest. For reasons that are often unclear, many phosphorylation sites appear to have low occupancy; that is, only a small fraction of the peptide molecules are detected as being phosphorylated (Olsen *et al.*, 2010; Wu *et al.*, 2011). Third, researchers should recognize that even when phosphorylation is detected on a peptide, it might not be possible to identify the precise site of modification. Same-sequence peptides phosphorylated on different residues have identical intact masses. When fragmented for MS/MS, only fragments resulting from breakage points located between the two sites can distinguish them (Figure 1C). All other fragment ions are the same for these two distinct phosphopeptides. The closer together the two sites, the lower is the number of possible site-determining fragment ions. Thus peptides with adjacent or multiple phosphorylated residues can be problematic. The probability of observing site-determining ions is further hampered by biases in peptide fragmentation that favor breakage at certain points and disfavor it at others. Even the highest-quality MS/MS spectra rarely contain all possible fragment ions and may not yield sufficient information

to localize the site within the peptide sequence. A number of scoring methods have been developed to assess the confidence of MS/MS-assigned phosphorylation-site localization. One of these, the Ascore algorithm, provides a probabilistic measure of the correctness of site assignment based on the observation or absence of site-determining fragment ions (Beausoleil et al., 2006).

One often underappreciated limitation of phosphorylation-site identification arises from the reliance on a single sequence-specific protease, such as trypsin or lysyl endopeptidase (lysC). Not all phosphorylated residues lie in regions that will generate MS-friendly peptides upon cleavage by a single protease. In the best experiments with a single protease, 80–90% of the protein sequence may be detected, but it is not uncommon to find significantly lower coverage. If any prospective regions or sites are of particular interest or if a comprehensive analysis is desired, it is essential to examine the protein sequence and make a suitable choice of enzyme that will generate peptides of ~8–20 amino acids. Repeating the analysis with a different protease or with a double digest can greatly improve overall sequence coverage and may reveal additional sites of protein modification. Heavily basic regions, which may harbor sites for basophilic kinases, can be difficult to map. Sites that lie in such regions may end up on peptides too short for analysis using trypsin and lysC. Enzymes that cut at nonbasic sites, such as GluC and chymotrypsin, can be useful for overcoming this problem, but the resulting peptides can contain multiple basic residues, producing higher-charge-state ions, which can be difficult to identify with standard fragmentation methods.

Obtaining quantitative information can be important

Advances in phosphoproteomics have led to a data explosion. More than 100,000 different phosphorylation sites have been reported in literature-curated databases (PhosphoSitePlus [www.phosphosite.org], Phospho.ELM [phospho.elm.eu.org], and PHOSIDA [www.phosida.com]). It is unclear, however, how many of these sites are physiologically relevant. Indeed, numerous studies have reported mutating mapped sites with little or no effect on function or phenotype, and many more such observations have almost certainly gone unreported. There are two important technical issues that can lead to identification of phosphorylation sites that are not physiologically relevant. First, standard mass spectrometry provides no information on the stoichiometry (also referred to as occupancy) of the phosphorylation site. This is because phosphorylated and unphosphorylated peptides are chemically distinct and therefore behave differently during mass spectrometry. Thus it is not possible to determine the stoichiometry of phosphorylation simply by comparing the amounts of phosphorylated and unphosphorylated peptide. Second, methods for phosphopeptide enrichment, coupled with leaps in instrument sensitivity, have enabled the detection of very low occupancy sites. Together these considerations mean that mass spectrometry can detect peptides that are phosphorylated at such low stoichiometry that they may be meaningless. Fortunately, quantitative mass spectrometry methods, which can measure the stoichiometry of phosphorylation or relative changes in phosphorylation, can help zero in on biologically relevant sites.

Obtaining quantitative information is particularly important when mapping sites phosphorylated by a purified kinase in vitro, where high kinase concentrations, long reaction times, and the absence of cellular phosphatases can potentially result in low-level spurious phosphorylation events that are not biologically relevant. In one example, budding yeast Wee1 was phosphorylated by mitotic Cdk1 in vitro. Standard phosphorylation-site mapping identified >20 phosphorylation sites, but quantitative analysis showed that only eight of the sites were phosphorylated at a measurable stoichiometry (Harvey et al., 2011).

Techniques for obtaining quantitative information

Although qualitative differences can be inferred from the presence or absence of detected sites and crude quantitative assessments made from the relative number of times a site is identified in different samples (spectral counting), neither is truly quantitative. The emergence and refinement of quantitative methods for mass spectrometry have transformed it into a robust experimental platform. The most common quantitative methods use stable-isotope tags incorporated into sample peptides and provide relative quantification. Peptides differing only in isotopic composition are chemically identical; they coelute from the liquid chromatography (LC) column, enter the mass spectrometer together, and have identical ionization efficiencies. Their distinct masses, however, can be observed simultaneously in the mass spectrometer. The relative peak heights of the two species provide a reliable quantitative ratio (Figure 2, A and B). Tags can be incorporated either metabolically, by growing cells in heavy isotope–enriched media, or chemically, after digesting cellular proteins. An alternative, targeted approach, termed AQUA, uses isotope-labeled synthetic peptides spiked into the experimental sample at a known concentration to provide absolute quantification (Gerber et al., 2003). The cost of peptide synthesis prohibits large-scale application of this method.

In the most popular metabolic labeling method, known as stable isotope labeling with amino acids in cell culture (SILAC), cells are grown in media containing heavy isotope–labeled amino acids, usually lysine and/or arginine, as the sole source of these amino acids (Figure 2A; Ong et al., 2002). With an appropriately chosen enzyme that cuts at these amino acids, nearly all resulting peptides will be labeled. Because samples can be combined at an early point in the experiment, usually after either cell harvesting or cell lysis, SILAC experiments are unparalleled for controlling variation in sample handling. However, high media costs can limit the scope of experimentation, and the required media conditions are not compatible with all systems. Reductive dimethylation with heavy (deuterated) or light formaldehyde and sodium cyanoborohydride is a common chemical labeling alternative to SILAC (Figure 2A). It is simple, robust, economical, and can be applied to systems that are not amenable to metabolic labeling (Boersema et al., 2009). However, because labeling must be done after digestion, at the peptide level, it is not as finely controlled as SILAC. In addition, because the heavy label is carried by deuterium, which affects chromatographic retention time, heavy/light peptide pairs no longer perfectly coelute, which can complicate quantification. Isobaric tags are an emerging class of chemical labels that allow multiplexed quantitative analysis (Figure 2C). Unlike the previously described methods, different reagents in each set have identical masses. Thus same sequence peptides labeled with them appear as a single species in a MS scan. However, upon fragmentation, each label releases a unique reporter ion that can be measured in a MS/MS scan (Figure 2D). Two different commercial products—isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMTs)—are available in sets of up to eight unique labels for iTRAQ and six for TMT (Thompson et al., 2003; Ross et al., 2004). The multiplexed analysis enabled by isobaric tags can be extremely powerful; however, they are a somewhat costly and less mature platform.

Changes in phosphorylation can also be monitored by determining the stoichiometry of a particular site. Measuring stoichiometry, however, is more challenging than measuring relative levels between samples. For a single site, labeled peptides can be
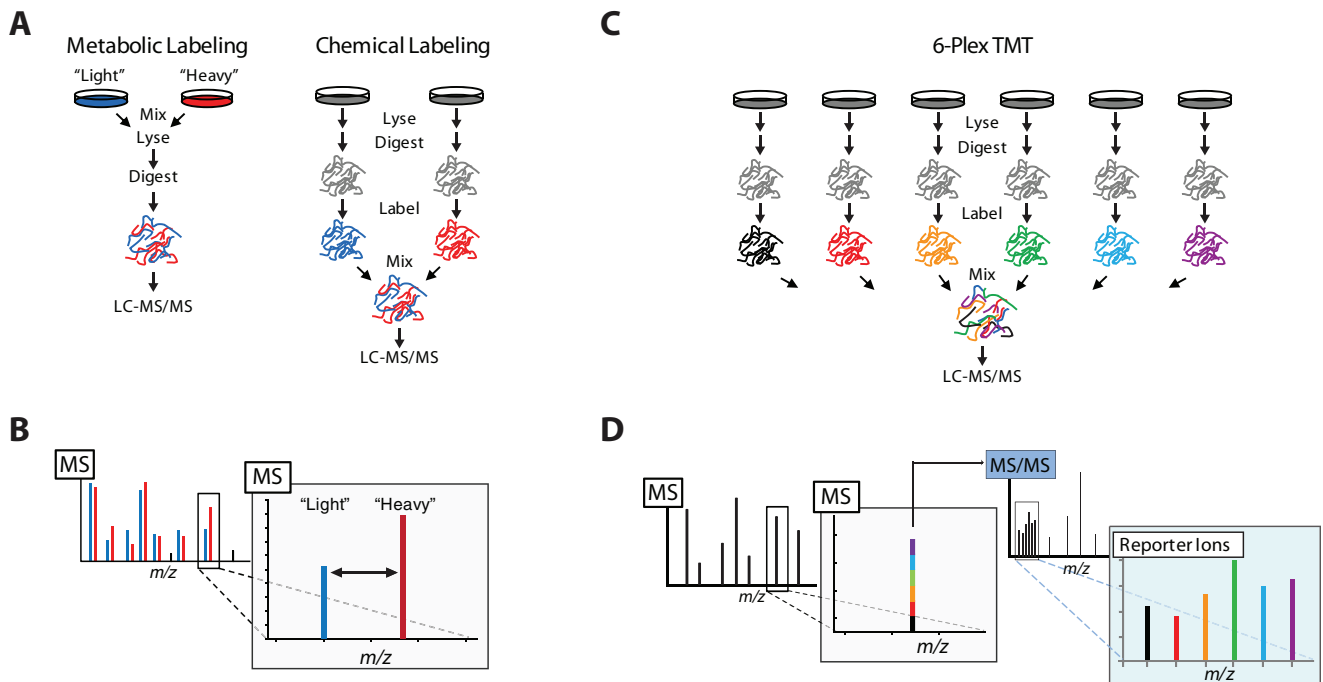
FIGURE 2: Stable-isotope labeling methods for quantitative mass spectrometry. (A) Cells can be labeled either metabolically by growing them in media containing heavy isotope–enriched nutrients such as amino acids in SILAC or chemically after lysis and digestion. Once labeled, samples can be combined for LC-MS/MS analysis. (B) Same sequence peptides from samples labeled with heavy and light isotopes are chemically identical. They coelute from the reverse-phase HPLC column and enter the mass spectrometer together. In a full scan of intact peptides they appear as doublets, separated by the characteristic added mass of the isotope label. Peak heights provide relative quantification. (C) Isobaric labels, such as the depicted 6-plex TMTs, allow multiplexed quantitative analysis. They are chemically incorporated after peptide digestion, before mixing and analysis. (D) The different labels in each set of isobaric labels have identical masses, and thus in the full scan, each peak is actually a composite of peptides from each sample. However, upon fragmentation, each label releases a unique reporter ion that can be detected in a MS/MS scan. Peak heights provide relative quantification of all six samples.

synthesized and spiked into samples to determine absolute levels of both the phospho and nonphospho forms of a peptide from which occupancy can be determined (Stemmann et al., 2001). As previously mentioned, the cost of labeled peptides is limiting. A more general method relies on phosphatase treatment of a labeled phosphoprotein and quantitative comparison with a mock-treated sample. The resulting ratio of the unphosphorylated peptide before and after removal of the phosphate indicates the fractional phosphorylation occupancy (Zhang et al., 2002b; Harvey et al., 2011). Note that this method cannot resolve occupancies of multiple sites on a single peptide. A few schemes based on this approach have been reported, and it has even been modified to work in large-scale experiments (Wu et al., 2011). However, such analyses are still not routine.

Mapping sites on isolated proteins versus large-scale proteomic analysis of phosphorylation

The development of efficient methods for phosphopeptide enrichment, along with improvements in sensitivity and the maturation of quantitative methods, has made it possible to identify and quantify >10,000 sites from thousands of proteins in a day or two of mass spectrometric analysis (Walther and Mann, 2010). In this kind of experiment, cells are subjected to an experimental manipulation, and proteome-wide changes in phosphorylation are detected by comparison to a control sample. This approach has proven to be extraordinarily powerful for mapping signaling networks. However, it

provides minimal sequence coverage of individual proteins, and many proteins, especially those of lower abundance, will be missed completely. Therefore it cannot provide detailed phosphorylation-site information for individual proteins.

A STEP-BY-STEP GUIDE TO THE IDENTIFICATION AND ANALYSIS OF PHOSPHORYLATION SITES

The following summarizes key steps and considerations for successful analysis of phosphorylation sites on individual proteins. We assume here that the kinase of interest phosphorylates its target independently of other kinases, which is usually the case. However, some kinases can only phosphorylate proteins that have previously been "primed" by another kinase, which can significantly complicate in vitro reconstitution of phosphorylation and analysis of phosphorylation site mutants.

Step 1: Optimization of assays for detection of phosphorylation events in vivo and in vitro

An important first step is to optimize methods for detecting phosphorylation events. The most convenient means of detecting protein phosphorylation is via electrophoretic mobility shift. Detection is rapid and simple, works on proteins modified in vivo and in vitro, and allows one to determine the stoichiometry of phosphorylation simply by assessing the fraction of shifted protein. A limitation of this approach is that many phosphorylation events do not cause a shift in electrophoretic mobility. Phosphorylation-induced shifts in

electrophoretic mobility are likely due to local context-dependent effects on the flexibility of the peptide chain rather than to changes in molecular weight or overall charge. As a result, the effects of phosphorylation on electrophoretic mobility are unpredictable. Empirical manipulation of the ratio of acrylamide to bisacrylamide can improve resolution of differently phosphorylated forms of a protein (Nishiwaki *et al.*, 2007). In addition, a recently developed method for incorporating a phosphate-binding molecule (termed Phos-Tag) into polyacrylamide gels can allow high-resolution detection of multiple phosphorylation events (Kinoshita *et al.*, 2006, 2012). Isoelectric focusing can also provide high resolution of multiple sites. Although it is more labor intensive, isoelectric focusing works on most proteins and can provide information on how many phosphates are attached for each charge isoform.

Incorporation of ^{32}P either in vitro or in vivo is highly sensitive but provides no information on the stoichiometry of phosphorylation of individual sites. However, an advantage of in vivo labeling with ^{32}P is that it can provide definitive results in less than a week as to whether a protein is phosphorylated in vivo (methods available in Cooper *et al.*, 1984; Den Haese *et al.*, 1995). Immunoblotting with phospho-specific antibodies is another common method for monitoring phosphorylation. However, not all sites can be detected with this approach. Effective antibodies that recognize phosphotyrosine have been available for >20 years. However, no comparably reliable antibodies exist for detecting phosphoserine or phosphothreonine. A number of antibodies that recognize phosphorylated residues within specific short linear motifs are available. These phospho-motif antibodies can be used to track phosphorylation due to certain classes of kinases (Zhang *et al.*, 2002a).

To preserve phosphorylation, it is essential that extract buffers and SDS–PAGE sample buffers contain high concentrations of phosphatase inhibitors. Sodium fluoride and β-glycerol phosphate, used together at 50 and 100 mM, respectively, are good inexpensive options.

Once a site has been identified, it may be possible to generate phosphospecific antibodies that recognize the site and its surrounding context. These powerful reagents allow rapid detection of the site in vivo and in vitro, but their production is expensive and not guaranteed. In our collective experience, only 50% of phosphopeptides have yielded useful phosphospecific antibodies. In addition, such antibodies provide no information on the stoichiometry of phosphorylation.

Step 2: Identification of the kinase or kinases that act directly on the protein of interest

Many proteins are phosphorylated by multiple kinases. Therefore it is not productive to simply mutate all phosphorylation sites identified on a protein isolated from cells, since this approach could blur the contributions of individual kinases. Instead, it is best to define which kinase or kinases act directly on a protein and then analyze the contribution of each kinase independently. The following criteria should be used to define whether a kinase acts directly on a protein of interest:

- Consider the consensus recognition site of the kinase of interest. Although far from perfect, there are many clear examples of site preferences, and there are good computational tools for examining consensus sites (e.g., http://elm.eu.org/, http://netphorest .info/, http://scansite.mit.edu/; Songyang *et al.*, 1994; Yaffe *et al.*, 2001; Obenauer *et al.*, 2003; Miller *et al.*, 2008; Dinkel *et al.*, 2012). Matches between mapped phosphorylation sites and the minimal consensus recognition site for the kinase under

study increase confidence that the relevant kinase has been matched to its target sites.
- Loss-of-function or gain-of-function mutations or small interfering/short hairpin RNA knockdown of the kinase should have the predicted effects on phosphorylation of the putative target protein. Of course, one must be aware of potential redundancies between kinases, which can complicate the analysis.
- Kinase catalytic domains generally do not show a significant affinity for their target proteins; however, many kinases associate with their targets via secondary docking sites (for examples, see Choi *et al.*, 1994; Mortensen *et al.*, 2002; Harvey *et al.*, 2005; Dard and Peter, 2006). Thus detecting binding interactions between the kinase and a putative target increases confidence that the protein is a direct target.
- Purified kinase should be capable of efficiently and quantitatively phosphorylating the target protein on a biologically relevant time scale in vitro. If quantitative phosphorylation requires a 10-fold excess of kinase and a 2-h reaction time but signaling in vivo occurs within minutes, the protein may not be a direct target or additional factors are required.

Fulfilling these criteria is relatively straightforward in yeast, which is why yeast are indispensable organisms for signaling analysis. It can be more difficult to fulfill all of the criteria in animal cells, but one should aim to fulfill as many as possible.

Step 3: Mapping of sites phosphorylated in vitro

The best mapping results come from an approach that combines data obtained in vivo and in vitro. Thus it is preferable to develop an in vitro reconstituted system to generate phosphorylated protein for analysis. To obtain good sequence coverage and quality spectra that yield high-confidence phosphopeptide matches, it is best to obtain as much phosphorylated protein as possible. The amount of starting material required for successful analysis will vary widely, but a "more is better" rule should apply. High-occupancy sites on MS-friendly peptides may be detectable from as little as 2 pmol of total protein (~100 ng of a 50-kDa protein); however, much higher levels (≥10 pmol) are often required.

Ideally, controls should be carried out when establishing the in vitro system to ensure that the kinase of interest is directly responsible for phosphorylation of the protein. Use of a kinase-dead control, specific kinase inhibitors, or analogue-sensitive kinase mutants can help rule out the possibility that phosphorylation is due to a copurifying kinase.

One should also consider whether the isolated substrate protein is already phosphorylated, potentially by multiple kinases. If so, the protein should be treated with phosphatase during purification. If the protein is isolated by affinity chromatography, this can be achieved by treating the protein with lambda or calf intestinal phosphatase while it is still bound to the affinity beads. Even proteins produced in bacteria sometimes have undergone nonspecific phosphorylation and may need to be treated with phosphatase before being used as a substrate.

Two reactions should be carried out: one containing kinase, and a reference reaction containing no kinase or a kinase-dead mutant. If one is using protein labeled by SILAC, the reference and experimental reactions are combined after the reactions and should contain equal amounts of protein labeled with heavy or light isotope. The combined reactions are analyzed directly by mass spectrometry after protease treatment to identify sites phosphorylated by the kinase. Alternatively, the combined reactions can be resolved by electrophoresis, and the band representing the protein of interest

can be excised and analyzed by mass spectrometry to compare levels of phosphorylation, which can improve detection in some cases. If one is not using SILAC, the kinase and control reactions can be independently labeled with mass tags, combined, and analyzed.

As a complement to mass spectrometry analysis, phosphoamino acid analysis can be applied to the protein of interest isolated from ^{32}P-labeled cells or after in vitro phosphorylation by specific kinases (Kamps and Sefton, 1989). For example, if phosphoamino acid analysis reveals that the protein of interest is phosphorylated on both serine and threonine residues in vivo and in vitro, but mass spectrometry analysis identifies only serine phosphorylation sites, it might be that potentially important modifications were missed because the corresponding threonine phosphopeptides could not be detected by mass spectrometry. Similarly, two-dimensional phosphopeptide mapping can provide a sense of the complexity of phosphorylation, as well as a means of testing whether all important sites of phosphorylation have been identified after mutagenesis, without the need for complex mass spectrometry experiments (Boyle et al., 1991). Classic phosphopeptide mapping also has the advantage that all the phosphopeptides are detected, in contrast to mass spectrometry.

Step 4: Mapping of sites phosphorylated in vivo

To determine whether phosphorylation sites identified in vitro are relevant, it is important to show that they are also phosphorylated in vivo. A number of databases catalogue sites that have been found to be phosphorylated in vivo in large-scale surveys, so an easy first step is to determine whether sites identified in vitro have already been identified in vivo, while keeping in mind that large scale surveys provide low-sequence coverage and miss many sites (for databases see PhosphoSitePlus, Phospho.ELM, and PHOSIDA).

To map sites phosphorylated in vivo, one should first define physiological conditions under which a significant fraction of the protein is phosphorylated. This could involve treating cells with a stimulus that activates signaling or synchronizing cells in the cell cycle. Another approach is to manipulate the signaling pathway genetically such that the kinase of interest is hyperactivated. An ideal situation is when one can purify the substrate protein from control cells, cells in which the relevant kinase is hyperactivated, and cells in which the relevant kinase has been inactivated, which allows one to determine which sites depend on the kinase of interest in vivo.

Once the appropriate conditions have been defined, the substrate protein must be purified under conditions that preserve phosphorylation and yield sufficient amounts of protein. For best results, one should aim for ≥10 pmol of protein. Affinity purification methods that allow specific release of the purified protein will produce the best results. For example, proteins tagged with hemagglutinin or FLAG can be purified using antibody beads and eluted with an excess of peptide (Ho et al., 2002; Harvey et al., 2011). Tandem affinity purification (TAP), multifunctional TAP, or other affinity-based tags can also be used (Rigaut et al., 1999; Ma et al., 2012). Purification of proteins using antibodies raised against the protein of interest can be problematic because elution from the antibody requires harsh conditions that also release large amounts of antibody from the beads, complicating the analysis, although this problem can be circumvented by cross-linking the antibodies to beads. Nonspecific elutions also generate a higher background of contaminant proteins that can interfere with target peptide detection.

To preserve phosphorylation, purification should be carried out in buffers that contain high concentrations of salt and phosphatase inhibitors. High salt concentrations help inhibit phosphatases and reduce nonspecific binding, whereas phosphatase inhibitors minimize dephosphorylation during purification.

Step 5: Interpretation of phosphorylation-site mapping data

Mass spectrometry analysis will yield a list of identified sites. It should include a tally of peptide spectral matches (PSMs), each representing a unique MS/MS spectrum that matches a peptide containing that site, and scoring parameters for peptide identification and site localization. Peptide identification data are routinely filtered to ~1% false-discovery rate using the target-decoy strategy (Elias and Gygi, 2007). Researchers should be aware that the data will contain incorrect matches. The observation of multiple PSMs for a given site either through multiple observations of the same peptide or the detection of different peptide sequences harboring the same site bolsters the confidence of correct site assignment. As discussed earlier, in many cases site assignments cannot always be resolved to a single residue. In some cases, for example, where the study is focused on a kinase with a known consensus motif, local sequence can be used to guide the choice of sites to pursue for further validation. However, in most cases, all possible sites on each peptide must be considered. Sites that are phosphorylated both in vivo and in vitro have high confidence of being relevant sites. Sites that change in occupancy in response to changes in relevant upstream signals are also high-confidence sites. Another consideration that can enhance confidence that correct identification has been made is the conservation of the phosphorylation site(s) throughout evolution. Although proteins phosphorylated at multiple sites within unstructured regions may not show evolutionary conservation of phosphorylation sites, within folded domains, phosphorylation sites are often conserved (Landry et al., 2009; Niu et al., 2012).

Phosphorylation-site mapping screens are almost never saturating. Failure to observe a site is insufficient evidence to conclude that the site is not phosphorylated in the cell. Many factors, including length, hydrophobicity, and charge, affect the chromatographic properties and ionization efficiencies and thus the ease of detection of different peptides. Phospho and nonphospho forms of the same peptide can have very different signal intensities. Thus even the observation of an unphosphorylated peptide is no guarantee that the correlate phosphopeptide is easily detectable. Despite these caveats, the unphosphorylated peptide sequence coverage still provides some indication of the depth of analysis. With sufficient amounts of protein, and barring long stretches of intractable sequence, it should be possible to achieve ≥80% amino acid coverage. Lower coverage decreases confidence that all sites have been identified and often indicates that more protein or an additional protease is needed to generate peptides for the analysis.

For quantitative experiments, the data will include abundance ratios for each phosphopeptide along with signal intensities, often recorded as a signal-to-noise ratio. Unlike protein-level analysis, in which multiple quantified peptides are often observed, phosphopeptides are more frequently detected and quantified only once. There is a strong correlation between signal strength and reproducibility. When selecting sites for further study, investigators should pay close attention to the number of PSMs and the signal strength for peptides harboring each site. It should also be noted that compiling ratios at the site level from multiple peptide measurements is not always trivial. Simply calculating averages or medians of all peptides containing a given site might not reveal the full complexity of cellular phosphorylation patterns. Singly and doubly phosphorylated forms of a peptide might be present at different

levels. One must also not forget that changes in total protein level are not reflected in the phosphopeptide ratios. Wherever possible, separate protein-level measurements made from unmodified peptides should be performed and used to normalize phosphopeptide ratios.

Step 6: Analysis and interpretation of phosphorylation-site mutants

After identification of all phosphorylation sites possible and their assignment to likely protein kinases, the next step is to mutate the sites so that their biological significance can be ascertained. Typically, serine and threonine phosphosites are mutated to alanine (or valine for threonine), and tyrosine phosphorylation sites are mutated to phenylalanine. Because mass spectrometry can miss sites, it is important to verify that most or all key sites have been identified and mutated. If the mutant protein loses its SDS–PAGE shift, it is likely that most sites have been identified. However, this does not exclude the possibility that some sites have been missed that do not cause a shift. Thus a more rigorous approach is to show that the mutant protein fails to incorporate ^{32}P in a reconstituted in vitro system or that the relevant phosphopeptides identified by in vivo, ^{32}P-labeled, two-dimensional phosphopeptide mapping disappear. It can sometimes also be informative to switch serine for threonine residues or vice versa. Many protein kinases do not distinguish serine from threonine, and if the site is targeted in vivo, the protein's gel shift should not be lost with this switch, and yet a change in phosphoamino acid content of the corresponding peptide can be readily identified, which allows one to directly verify that the relevant phosphorylation site has been identified.

If a phosphorylation site mutant causes a loss of function, there can be the concern that it causes nonspecific damage to the protein. The vast majority of phosphorylation sites occur in regions of proteins that are predicted to be disordered, so it is unlikely that phosphorylation-site mutants disrupt protein structure (Iakoucheva et al., 2004; Gsponer et al., 2008). In addition, a number of criteria can be used to help rule out this possibility. For example, if normal levels of the protein are expressed in vivo, it is likely that the protein undergoes normal folding, since proteins that cannot fold correctly are destroyed. Another helpful test is to determine whether the phosphorylation-site mutant retains a subset of normal functions, which would indicate that the mutants affect specific functions of the protein. If the protein shows normal localization, it clearly retains key functions.

It has become common in protein phosphoregulation studies to mutate phosphorylation sites to "phosphomimetic" residues in an attempt to study the constitutively phosphorylated state. In this approach, serine and threonine are typically mutated to aspartic or glutamic acid residues, whereas tyrosine is substituted with glutamic acid. This approach has two significant shortcomings. First, if the phosphorylation site serves as a recognition signal for an adaptor protein (i.e., 14-3-3, FHA-domain, PTB-domain, and SH2-domain proteins), phosphomimetic mutants will not bind to the adapter protein (Durocher et al., 1999; Zisch et al., 2000; Roberts-Galbraith et al., 2010) because they do not fit into the binding pocket (van der Geer and Pawson, 1995; Yaffe et al., 1997; Durocher et al., 1999). Second, the negative charge introduced by aspartate or glutamate substitutions (−1) does not match that of the phosphorylated residue (generally −1.5) at physiological pH. Neighboring pairs of aspartic or glutamatic acid side chains can overcome the charge differential and may act as better phosphomimetics (Strickfaden et al., 2007; Pearlman et al., 2011). However, the size of

the ionic shell produced by a phosphate group is also different, and so the overall chemical environment created by phosphorylation is very different from that of negatively charged amino acids (Hunter, 2012). It is therefore not surprising that phosphomimetic mutations often fail to reproduce the changes to a protein caused by phosphorylation. As a result of these two limitations, the behavior of phosphomimetic mutations can be uninterpretable. Of course, there are examples in which phosphomimetic substitutions have been highly informative. The constitutive activation of MEK kinases by a phosphomimetic mutation is an excellent example (McKay and Morrison, 2007).

MAIN POINTS TO CONSIDER WHEN PLANNING TO MAP AND ANALYZE PHOSPHORYLATION SITES

- Rigorous, high-confidence mapping results come from an approach that combines information obtained by mapping sites on protein phosphorylated in vivo and protein phosphorylated in vitro by purified kinase.

- Because many proteins are phosphorylated by multiple kinases, care must be taken to ensure that sites identified by mapping can be unambiguously assigned to the relevant kinase.

- Obtaining high sequence coverage requires a minimum of 2–10 pmol of purified phosphorylated protein, which corresponds to 100–500 ng of a 50-kDa protein. More protein is better.

- Even under the best conditions, mass spectrometry can miss phosphorylation sites because the corresponding peptides are lost during sample preparation or are not detected by mass spectrometry.

- Assignment of phosphorylation sites by mass spectrometry cannot always be done with 100% confidence.

- "Old-fashioned" techniques, such as phosphopeptide mapping, can provide an important complement to modern mass spectrometry techniques.

- Before beginning a mapping project, it is important to consider whether good in vivo and in vitro assays are available to determine whether all key sites have been mutated. One should also consider whether good assays are available to analyze the functions or phenotypes of phosphorylation-site mutants.

REFERENCES

Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 24, 1285–1292.

Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4, 484–494.

Boyle WJ, van der Geer P, Hunter T (1991). Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. Methods Enzymol 201, 110–149.

Choi KY, Satterberg B, Lyons DM, Elion EA (1994). Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. cerevisiae. Cell 78, 499–512.

Cooper JA, Sefton BM, Hunter T (1984). Diverse mitogenic agents induce the phosphorylation of two related 42,000-dalton proteins on tyrosine in quiescent chick cells. Mol Cell Biol 4, 30–37.

Craig R, Beavis RC (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466–1467.

Dard N, Peter M (2006). Scaffold proteins in MAP kinase signaling: more than simple passive activating platforms. BioEssays 28, 146–156.

Den Haese GJ, Walworth N, Carr AM, Gould KL (1995). The Wee1 protein kinase regulates T14 phosphorylation of fission yeast Cdc2. Mol Biol Cell 6, 371–385.

Dinkel H et al. (2012). ELM—the database of eukaryotic linear motifs. Nucl Acids Res 40, D242–D251.

Durocher D, Henckel J, Fersht AR, Jackson SP (1999). The FHA domain is a modular phosphopeptide recognition motif. Mol Cell 4, 387–394.

Elias JE, Gygi SP (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4, 207–214.

Eng JK, McCormack AL, Yates J 3rd (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Spectrom 5, 976–989.

Fenyo D, Beavis RC (2003). A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem 75, 768–774.

Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004). Open mass spectrometry search algorithm. J Proteome Res 3, 958–964.

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA 100, 6940–6945.

Gingras AC, Gstaiger M, Raught B, Aebersold R (2007). Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 8, 645–654.

Gsponer J, Futschik ME, Teichmann SA, Babu MM (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science 322, 1365–1368.

Harvey SL, Charlet A, Haas W, Gygi SP, Kellogg DR (2005). Cdk1-dependent regulation of the mitotic inhibitor Wee1. Cell 122, 407–420.

Harvey SL, Enciso G, Dephoure NE, Gygi SP, Gunawardena J, Kellogg DR (2011). A phosphatase threshold sets the level of Cdk1 activity in early mitosis. Mol Biol Cell 22, 3595–3608.

Ho Y et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180–183.

Hunter T (2012). Why nature chose phosphate to modify proteins. Phil Trans R Soc Lond B Biol Sci 367, 2513–2516.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32, 1037–1049.

Kamps MP, Sefton BM (1989). Acid and base hydrolysis of phosphoproteins bound to immobilon facilitates analysis of phosphoamino acids in gel-fractionated proteins. Anal Biochem 176, 22–27.

Kinoshita E, Kinoshita-Kikuta E, Koike T (2012). Phos-tag SDS–PAGE systems for phosphorylation profiling of proteins with a wide range of molecular masses under neutral pH conditions. Proteomics 12, 192–202.

Kinoshita E, Kinoshita-Kikuta E, Takiyama K, Koike T (2006). Phosphate-binding tag, a new tool to visualize phosphorylated proteins. Mol Cell Proteomics 5, 749–757.

Landry CR, Levy ED, Michnick SW (2009). Weak functional constraints on phosphoproteomes. Trends Genet 25, 193–197.

Ma H, McLean JR, Chao LF, Mana-Capelli S, Paramasivam M, Hagstrom KA, Gould KL, McCollum D (2012). A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms. Mol Cell Proteomics 11, 501–511.

McKay MM, Morrison DK (2007). Integrating signals from RTKs to ERK/MAPK. Oncogene 26, 3113–3121.

Miller ML et al. (2008). Linear motif atlas for phosphorylation-dependent signaling. Science Signal 1, ra2.

Mortensen E, McDonald H, Yates J, Kellogg DR (2002). Cell cycle-dependent assembly of a Gin4-septin complex. Mol Biol Cell 13, 2091–2105.

Nishiwaki T, Satomi Y, Kitayama Y, Terauchi K, Kiyohara R, Takao T, Kondo T (2007). A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. EMBO J 26, 4029–4037.

Niu S, Wang Z, Ge D, Zhang G, Li Y (2012). Prediction of functional phosphorylation sites by incorporating evolutionary information. Protein Cell 3, 675–690.

Obenauer JC, Cantley LC, Yaffe MB (2003). Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 31, 3635–3641.

Olsen JV et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3, ra3.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376–386.

Pearlman SM, Serber Z, Ferrell JE Jr (2011). A mechanism for the evolution of phosphorylation sites. Cell 147, 934–946.

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 17, 1030–1032.

Roberts-Galbraith RH, Ohi MD, Ballif BA, Chen JS, McLeod I, McDonald WH, Gygi SP, Yates JR 3rd, Gould KL (2010). Dephosphorylation of F-BAR protein Cdc15 modulates its conformation and stimulates its scaffolding activity at the cell division site. Mol Cell 39, 86–99.

Ross PL et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3, 1154–1169.

Songyang Z, Blechner S, Hoagland N, Hoekstra MF, Piwnica-Worms H, Cantley LC (1994). Use of an oriented peptide library to determine the optimal substrates of protein kinases. Curr Biol 4, 973–982.

Steen H, Mann M (2004). The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5, 699–711.

Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW (2001). Dual inhibition of sister chromatid separation at metaphase. Cell 107, 715–726.

Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pryciak PM (2007). A mechanism for cell-cycle regulation of MAP kinase signaling in a yeast differentiation pathway. Cell 128, 519–531.

Tabb DL, Fernando CG, Chambers MC (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res 6, 654–661.

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75, 1895–1904.

van der Geer P, Pawson T (1995). The PTB domain: a new protein module implicated in signal transduction. Trends Biochem Sci 20, 277–280.

Walther TC, Mann M (2010). Mass spectrometry-based proteomics in cell biology. J Cell Biol 190, 491–500.

Wu R, Haas W, Dephoure N, Huttlin EL, Zhai B, Sowa ME, Gygi SP (2011). A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat Methods 8, 677–683.

Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, Cantley LC (2001). A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat Biotechnol 19, 348–353.

Yaffe MB, Rittinger K, Volinia S, Caron PR, Aitken A, Leffers H, Gamblin SJ, Smerdon SJ, Cantley LC (1997). The structural basis for 14-3-3:phosphopeptide binding specificity. Cell 91, 961–971.

Zhang H, Zha X, Tan Y, Hornbeck PV, Mastrangelo AJ, Alessi DR, Polakiewicz RD, Comb MJ (2002a). Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. J Biol Chem 277, 39379–39387.

Zhang X, Jin QK, Carr SA, Annan RS (2002b). N-terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun Mass Spectrom 16, 2325–2332.

Zisch AH, Pazzagli C, Freeman AL, Schneller M, Hadman M, Smith JW, Ruoslahti E, Pasquale EB (2000). Replacing two conserved tyrosines of the EphB2 receptor with glutamic acid prevents binding of SH2 domains without abrogating kinase activity and biological responses. Oncogene 19, 177–187.