Published in final edited form as:

Cell Commun Adhes. 2006 ; 13(5-6): 249–262.

Methodologies for Characterizing Phosphoproteins by Mass Spectrometry

Philip R. Gafken¹ and Paul D. Lampe

¹*Proteomics Facility, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109*

Keywords

Phosphorylation; Mass spectrometry; Kinase; Sequencing

Introduction

Here, we review methods for determination of phosphorylation sites in proteins. Phosphorylation is one of the major means of posttranslational regulation of proteins and a large percentage of proteins are phosphorylated at some point during their life cycle. When searched via the Ovid database, “protein phosphorylation” returns over 5,000 reports in the last 2 years and over 25,000 over the last 10 years. Phosphorylation is a very rapid and reversible method of changing the function of proteins through altering the activity of an enzyme, changing the stability of a protein or changing how the protein interacts with other proteins. Determining the functionality of protein phosphorylation events often involves studying the effects of site directed mutations that convert the residue that can be phosphorylated to one that cannot be phosphorylated (i.e., alanine for serine or threonine and phenylalanine for tyrosine) or one that can mimic the charge of the phosphorylation event (i.e., glutamic or aspartic acid). Furthermore, the creation and exploitation of antibodies that are specific for proteins only when phosphorylated at specific residues has revolutionized the study of signaling pathways and cancer biology in general. Creation of phosphospecific antibodies or site-directed mutants requires knowledge of which sites are phosphorylated. Sometimes an educated guess can be sufficient, particularly if the kinase responsible for the phosphorylation site is known and it has a reliable “consensus” sequence that determines substrate specificity. However, some proteins are phosphorylated on more than a dozen residues including proteins such as p53 and the protein our laboratory works on, the gap junction protein connexin43 (Cx43).

Gap junctions are tightly packed clusters of intercellular channels that directly connect the cytoplasm of adjacent cells. They coordinate cell-to-cell communication within tissues and allow for the transfer of molecules less than 1000 Daltons between cells such as ions, simple sugars, amino acids, nucleotides and second messengers (e.g., Ca^{2+} , cAMP, cGMP, IP3) (Goodenough and Paul, 2003; Saez et al., 2003; Willecke et al., 2002). In vertebrates, gap junctions are composed of proteins from the connexin family, which is comprised of 21 members in humans (Goodenough and Paul, 2003; Saez et al., 2003; Willecke et al., 2002). Connexin proteins possess four hydrophobic membrane-spanning domains. The C-terminal region (residues 240–382) of Cx43 appears to be the primary region that becomes phosphorylated. Replacement of Cx43 with a truncated version (Cx43K258Stop) lacking the C-terminal, cytoplasmic tail region yielded mice that died shortly after birth due to an epidermal

barrier defect, not the heart defect that is present in Cx43 deficient mice (Maass et al., 2004). Phosphorylation of Cx43 has been implicated in the regulation of gap junctional communication and gating at several stages of the cell cycle and the subcellular localization during connexin oligomerization, trafficking, assembly/disassembly, and degradation. Thus, considerable evidence indicates that Cx43 is a highly phosphorylated and a highly regulated protein. Several reviews on the functional role of connexin phosphorylation have been published (Laird, 2005; Lampe and Lau, 2004; Moreno, 2005; Solan and Lampe, 2005). Phosphorylation site determination has also been reviewed (Carr et al., 2005; Kobe et al., 2005; Loyet et al., 2005; Zeller and Konig, 2004). This review highlights some of the recent methods utilized to determine phosphorylation sites within proteins and is biased on methods with which we have had experience and those we feel have the most general applicability to characterizing phosphoproteins. It is not meant to provide a thorough summary of all the methods and techniques associated with identification of phosphorylation sites in proteins.

Evolution of Methods for Phosphosite Determination

Determination of kinase specificity or “consensus sequences” often involved the use of purified kinases and [γ - ^{32}P]ATP to phosphorylate purified peptides or proteins followed by proteolytic digestion, purification and automated Edman degradation-based protein sequencing. However, determination of phosphorylation events that occurred in intact cells usually required metabolic labeling of cells with mCi amounts of ^{32}P -inorganic phosphate followed by purification (usually immunoprecipitation) and separation of tryptic fragments via high performance liquid chromatography (HPLC) or 2-D peptide analysis. The disadvantages of these methods are primarily of large losses of signal at each step (primarily low incorporation of ^{32}P into the protein of interest), the difficulty in optimization of the amount of cellular lysate per mCi of radioactivity to yield sufficient signal to detect both radioactivity and protein/peptide/phosphorylation site sequence, and high levels of radioactivity can lead to unanticipated cellular responses. Nonetheless, this methodology has led to identification of many phosphoproteins including many connexin and cell adhesion proteins (Crow et al., 1990; Hertlein et al., 1998; Musil et al., 1990; Paul et al., 1991; Saez et al., 1990) and, in some cases, identification of specific phosphorylation sites. Variations on these methods have also been utilized. For example, after metabolic labeling, we immunoprecipitated Cx43, added recombinant Cx43 and digested the radiolabeled and recombinant Cx43 with trypsin, and separated the peptide mixture by reversed-phase HPLC. The chromatographic elution was monitored for ^{32}P to identify when phosphopeptides eluted and the radioactive fraction and the next eluting fraction are analyzed via mass spectrometry to determine the identity of the phosphorylated peptides (Solan and Lampe, 2005). This method assumes approximate co-elution of phosphorylated and non-phosphorylated peptide and hence would not work well with short peptides (e.g., less than 10 residues) where phosphorylation could radically affect elution time. Also, this methodology allows for the detection of peptides that contain phosphorylation sites present at low stoichiometry as long as enough ^{32}P can be incorporated to allow detection, but it does not provide information about the specific amino acid that is phosphorylated if the peptide contains more than one phosphorylatable residue.

Mass Spectrometry Approaches

The development of ionization techniques during the 1980s and 1990s that allowed for the ionization of proteins and peptides began to make mass spectrometers extremely valuable to biochemists. The two most notable ionization techniques developed were Matrix-Assisted Laser Desorption Ionization (MALDI) and ElectroSpray Ionization (ESI), for which the primary developers of these two techniques were awarded the 2002 Nobel Prize in Chemistry. Today, both MALDI and ESI are coupled with numerous types of mass analyzers, including ion trap, time-of-flight, and Fourier-transform ion cyclotron resonance to create mass

spectrometers that are sold by a variety of analytical instrumentation companies. Mass spectrometers should be evaluated on a project-by-project basis for such characteristics as mass resolution, mass accuracy, sensitivity, and dynamic concentration range. As it is difficult to discuss the advantages and limitations of each type of instrument within the broad field of qualitative and quantitative phosphoprotein analysis, and since the development of mass spectrometers is occurring at a rapid pace, we suggest that experienced biological mass spectrometry labs be consulted during the planning stages of phosphoprotein experiments to determine the most appropriate instrumentation and hence methodologies likely to provide the desired results.

Two general approaches exist for characterizing proteins by mass spectrometry. The first is called the “bottom-up” approach (Figure 1A) where a protein or a mixture of proteins is digested with a protease to produce peptides and then mass spectrometry is performed on the peptide mixture for the qualitative and/or quantitative characterization of the peptides. It is important to note that spectral information from a peptide is used to characterize the protein from which the peptide was derived. The second approach is called the “top-down” approach (Figure 1B) and here whole proteins are separated and then individually analyzed by the mass spectrometer. The vast majority of mass spectrometry-based proteomics experiments, including those related to phosphoprotein analysis, has historically taken place via a bottom-up approach. This has been due to the ability of mass spectrometers to easily provide the necessary spectral information for characterizing peptides and the difficulties associated with characterizing whole proteins by mass spectrometry. Even though bottom-up has been the dominating approach, recent advances in mass spectrometry are making top-down experiments possible (see Future directions). The remainder of this review will summarize techniques and methods that are based upon bottom-up proteomics.

Within the bottom-up approach, one of two schemes is usually followed for the characterization of phosphoproteins. The first is the structural scheme where phosphorylation on a highly enriched protein is characterized with the goal of locating regions, or hopefully specific amino acids, that contain a phosphate group. This is accomplished by isolating the phosphoprotein, digesting the protein with a protease, and then characterizing the resulting peptides by mass spectrometry. Since the mass spectrometry data complexity for this scheme is manageable, data can be analyzed manually or with automated protein database search algorithms (e.g., SEQUEST, MASCOT, X!TANDEM). Peptides that are detected and characterized are then “mapped” to the primary sequence of the protein. The goal is to obtain as much sequence coverage as possible to have thorough coverage of the protein. If coverage of the map is low, due to peptides that are lost during preparation or that are of a size that is not suitable for detection by mass spectrometry, a parallel proteolytic digestion using a different protease may be employed to produce complimentary sequence coverage. The quantity of phosphoprotein needed for a mapping experiment varies greatly, being dependent on such factors as the solubility of the phosphoprotein and the stoichiometry of phosphorylation. To help overcome these potentially limiting factors, phosphoprotein quantities of at least one to ten picomoles are generally recommended to generate extensive peptide maps. This is a stark difference to general protein identification where only a few peptides sequenced back to a protein is necessary, and thus as little as 100 femtomoles of a protein are required.

A second general methodology within the bottom-up approach is the phosphoproteomics scheme, and it is typically used to catalog or detect novel phosphorylation sites within complex mixtures. Instead of analyzing a single protein, as with the structural scheme, a protein mixture is digested with a protease, the resulting peptide mixture can be enriched for phosphopeptides (see below), and mass spectrometry analysis is then performed. The resulting dataset is often very complex and automated database searching for identifying sites of phosphorylation is

required. Using this scheme, hundreds to thousands of phosphorylation sites can be identified in a single experiment.

Separation and Enrichment techniques

The extent of phosphorylation within a cell varies dramatically from protein to protein with phosphorylation stoichiometry often being much less than one. This results in a large number of nonphosphorylated peptides being present within a small number of phosphorylated peptides. In turn, this adds unwanted complexity to the mass spectrometry experiment, making it difficult to find mass spectrometric signals from phosphopeptides within the large background of nonphosphorylated peptides. Therefore, within the workflow of bottom-up proteomics, robust separation and enrichment strategies for phosphopeptides are beneficial.

SDS-PAGE

Even though phosphoproteins would be expected to have the same electrophoretic migration on a gel as their nonphosphorylated counterparts, it has often been observed that a phosphoprotein will have an altered electrophoretic migration as compared to the nonphosphorylated form. Uniform binding of SDS to the phosphoprotein may be disrupted due to the negatively-charged phosphate group on the protein, which in turn, could decrease the charge density of the phosphoprotein relative to the nonphosphorylated form. If the charge density difference between the two forms is large enough, a phosphorylated protein would have a retarded migration and appear at a higher molecular weight relative to the nonphosphorylated form (the phosphorylated protein's increased molecular weight as observed on the gel is often much higher than the 80 Dalton mass increase of a phosphate group, which is a mass increase not typically resolved on a gel). Cx43 demonstrates multiple electrophoretic isoforms when analyzed by polyacrylamide gel electrophoresis (SDS-PAGE), including a faster migrating form that includes non-phosphorylated (P0 or NP) Cx43, and at least two slower migrating forms, commonly termed P1 and P2. Both P1 and P2 co-migrate with P0 following alkaline phosphatase treatment, suggesting that phosphorylation is the primary covalent modification detected in SDS-PAGE analysis (Crow et al., 1990; Musil et al., 1990). Due to the widespread, routine use of electrophoresis and the often observed altered electrophoretic mobility of phosphoproteins, SDS-PAGE can be a powerful tool for separating phosphorylated from nonphosphorylated versions of the same protein. An added advantage to SDS-PAGE separation of phosphoproteins is the ability to readily perform proteolytic digestions "in-gel" to convert phosphoproteins into phosphopeptides. Typical proteases used for in-gel digestions are trypsin, endoproteinase Arg-C, endoproteinase Lys-C, and chymotrypsin. The resulting peptides can either be analyzed by mass spectrometry or they can be further processed for the enrichment of phosphopeptides, ultimately leading to the characterization of the phosphoprotein.

Antibody purification

Affinity purification is another method for purifying proteins that can be used in conjunction with SDS-PAGE or by itself. Antibodies raised against a protein can be used to immunoprecipitate the protein and search for phosphorylation sites. Immunoprecipitation allows for the isolation of a protein under a variety of biological conditions to assess changes in phosphorylation on that protein. Similarly, antibodies raised against a specific phosphosite on a protein can be used for immunoprecipitation. Under this scenario, other phosphosites on a protein can be assessed when one phosphosite is known (the epitope of the antibody). However, caution must be exercised when a protein is phosphorylated at multiple serines as certain phosphorylation events might be mutually exclusive and be missed during subsequent analysis. Phosphospecific antibodies can be used to determine the proteins that bind to a phosphoprotein (protein-phosphoprotein interactions) using phosphosite-specific immunoprecipitation followed by analysis of the binding partners. Furthermore, antibodies

specific for phosphotyrosines but not affected by the surrounding amino acids have been successfully used to immunoprecipitate the “phosphotyrosineome” of cells. Since phosphoserine and phosphothreonine are much more abundant in cells and these antibodies seem to have less specificity, phosphoproteome-wide experiments are much more involved.

Metal-assisted enrichment

With bottom-up proteomics focusing on the analysis of peptides from proteolytic digestions, the enrichment of phosphopeptides from proteolytic digests can drastically simplify peptide mixtures. A widespread technique for isolating phosphopeptides is Immobilized Metal Affinity Chromatography, or IMAC (Posewitz and Tempst, 1999). Here, a metal is chelated to a metal binding ligand on a resin allowing for the positive charge on the metal to attract the negative charge of the phosphate group of the phosphopeptides. Non-phosphorylated peptides can be washed away from the bound phosphopeptides and the phosphopeptides can then be eluted. We obtained approximately a 10-fold enrichment of phosphorylated to nonphosphorylated peptides with these methods (TenBroek et al., 2001). Typical metals used for IMAC are Fe^{3+} , Cu^{2+} , and Ga^{3+} . A major drawback to IMAC is that the immobilized metals can retain peptides that are not phosphorylated, but have net negative charges due to the presence of aspartate and glutamate residues. One technique for overcoming this drawback is to convert aspartate and glutamate carboxyls in a peptide to methyl esters prior to IMAC purification, thus removing their negative charges (Ficarro et al., 2002). A newer metal-based affinity technique for isolating phosphopeptides uses titanium oxide-based solid phase material (Pinkse et al., 2004). In-depth optimization of this technique has been described (Larsen et al., 2005) and it has been reported to have higher selectivity for phosphopeptides and to be technically easier than IMAC, making this technique an attractive alternative to IMAC.

SCX

A novel technique for enriching phosphopeptides on a proteomic scale has recently been described that uses strong cation exchange (SCX) chromatography. The logic for using SCX to enrich for phosphopeptides comes from an analysis of the *in silico* trypsin digestion of the National Center for Biotechnology Information (NCBI) human protein database. This analysis showed that approximately 68% of tryptic peptides produced from complete trypsinization of the proteins in the database would have a net charge in solution of +2 at pH 2.7 while approximately 3% of these tryptic peptides would have a net charge in solution of +1 at pH 2.7 (Beausoleil et al., 2004). It should be mentioned that peptides containing “missed” tryptic cleavages (i.e. arginines or lysines immediately flanked by proline to their C-terminal side) would have net solution charges higher than +2, depending on the number of missed cleavages, but the number of peptides containing missed cleavages would be relatively small compared to the number of fully-tryptic peptides. Since a phosphate group at an acidic pH maintains a negative charge, most fully-tryptic peptides that are phosphorylated would have their net solution charge reduced from +2 to +1. SCX separates peptides primarily based on charge making most phosphopeptides that contain a single phosphate group elute in the less complex +1 solution charge region during separation. This results in highly enriched phosphopeptide fractions with the majority of the nonphosphorylated peptide background being removed. These enriched fractions can either be directly analyzed by mass spectrometry or they can be further purified by other techniques, such as IMAC. Using SCX combined with mass spectrometry, over 500 phosphorylation sites in the developing mouse brain (Ballif et al., 2004) and over 2000 phosphorylation sites in HeLa cells (Beausoleil et al., 2004) have been detected.

Dendrimer capture

A recently developed technique for affinity purification of phosphopeptides is based on the use of a polymeric dendrimer support (Tao et al., 2005). Phosphates on phosphopeptides, generated from proteolytic digestion of protein mixtures, are activated using carbodiimide and imidazole to react with excess amines on a dendrimer to form phosphoramidate bonds. After removal of nonphosphorylated peptides, the phosphorylated peptides are eluted via acid hydrolysis. This technique has been designed as a “one-pot” reaction to minimize purification steps and relative phosphopeptides quantification of two samples can be performed via standard methyl esterification (use of methanol or deuterated methanol) chemistry.

Qualitative Analysis

Generating sequence information on a phosphopeptide provides the identity of the protein from which the peptide was derived, and it provides information about the location of the phosphorylated amino acid. The generation of sequence data by mass spectrometers is performed by a two-stage process called tandem mass spectrometry, also known as MS/MS. The first stage of the process (Figure 2A) is the creation of ions in the ionization source followed by the separation of ions in a mass analyzer followed by detection of the ions. This first stage is called the MS or MS1 analysis and it is used to measure the molecular mass of the peptides present. A second stage (Figure 2B) follows by selecting ions of the same mass-to-charge ratio (m/z) with the first mass analyzer for fragmentation by collision-induced dissociation (CID), where the selected ions collide with inert gas molecules (such as helium or argon) to ultimately undergo fragmentation, mostly along the peptide backbone. The resulting fragment ions are then separated by a second mass analyzer and detected. This second stage is called the MS2 or MS/MS analysis and it is used to provide mass information about the fragments of a peptide after CID and it ultimately provides information about the sequence of the phosphopeptide.

Obtaining peptide sequence information from a fragmentation spectrum can occur via two routes: manual peptide sequencing and/or automated protein database searching. Manual sequencing from mass spectrometry data is exceedingly challenging due to complex tandem mass spectra that are generated. Typically, manual sequencing is attempted when some *a priori* knowledge of the phosphopeptide exists, such as the phosphopeptide's amino acid sequence is suspected. In most cases automated database searching is conducted. Here, the fragmentation spectra produced by the mass spectrometer are compared to theoretical fragmentation spectra generated from the *in silico* digestion of protein sequences in a protein database. The comparison process assigns a score to each tandem mass spectrum that is matched to a theoretical, *in silico* peptide fragmentation spectrum to provide a measure that the matched spectrum is correct. It is important to note that database search algorithms are readily capable of accounting for mass increases to peptides due to phosphorylation (or almost any modification) making them extremely useful for identifying phosphopeptides.

Even though mass spectrometry is very powerful for characterizing phosphoproteins and numerous accounts of successful characterizations are available in the literature, two significant issues exist that make phosphopeptide characterization by mass spectrometry nontrivial. First, the vast majority of mass spectrometry experiments for peptides are set up to detect positively-charge ions. This is done since the N-terminus of peptides along with the amino acids histidine, lysine, and arginine are protonated at low pH. It is generally observed that a phosphorylated peptide has a suppressed response as compared to its nonphosphorylated counterpart during mass spectrometry experiments performed in the positive-ion mode. Second, quite often the major fragmentation product from collision-induced dissociation of phosphopeptides is from the neutral loss of phosphoric acid or a phosphate group and water. This can result in fragmentation spectra that provide limited information about the sequence of the peptide and/or the inability to locate the site of phosphorylation within the peptide.

(DeGnore and Qin, 1998). Additional stages of fragmentation (e.g. MS³) of ions in the tandem mass spectrum can sometimes provide additional information to fully sequence and locate the phosphorylated residue.

Addressing the limitations of qualitative analysis of phosphopeptides by mass spectrometry, a derivatization method was developed that uses cystamine to react with the beta-eliminated phosphoserine and phosphothreonine to produce aminoethylcysteine and beta-methylaminoethylcysteine, respectively (Knight et al., 2003). β -elimination of the phosphoserine and reaction with thiol reagents has been around for many years (Meyer et al., 1986); we used it in conjunction with Edman degradation in the late 80's to determine *in vivo* phosphorylation sites (Lampe and Johnson, 1990). However, this technique is unique since both aminoethylcysteine and beta-methylaminoethylcysteine are lysine analogs sensitive to digestion by lysine-recognizing proteases (e.g., trypsin and Lys-C). This means that digestion of the derivatized proteins results in a new peptide map, as compared to the non-derivatized version, that aids in mapping sites of phosphorylation. Tandem mass spectrometry of the derivatized peptides results in fragmentation spectra that are readily interpretable. However, since this technique and various modifications require careful chemistry and the associated yield losses, it has not been universally utilized.

Quantitative Analysis

Protein phosphorylation is very dynamic, constantly changing throughout the life of a cell. Measuring the changes in phosphorylation is critical to understanding the biology of a phosphorylation event. Three quantitation strategies that rely on mass spectrometry have recently been developed that have direct utility towards measuring changes in protein phosphorylation; namely, SILAC, iTRAQ, and AQUA. Each of these strategies is described below. Other chemical labeling techniques that rely on stable isotope incorporation using ¹⁸O labeled water during trypsin digestions and stable isotope incorporation via methyl esterification also can be considered but will not be described here.

SILAC

Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is based around *in vivo* labeling of proteins in cell culture with amino acids that contain stable (nonradioactive) isotopes (e.g., ¹³C and ¹⁵N)(Ong et al., 2002; Ong et al., 2003). In its simplest form, two separate cell cultures are grown in a pair-wise fashion (Figure 3); for example, culture A might be HeLa cells grown under "normal" conditions while culture B might be HeLa cells grown in the presence of a phosphatase inhibitor. The growth conditions of the cells are identical (except for the presence of the drug), but the growth media of culture B has an essential amino acid (one not synthesized by the cell) replaced with an isotopically "heavy" form of that amino acid (e.g., [¹³C6]-arginine). To date, a number of cell lines have been used in SILAC experiments and the growth and morphology of the cells have not been affected by the isotopically labeled amino acid (Ong et al., 2002). After about five rounds of doubling, cellular proteins are essentially 100% enriched in the labeled amino acid. After culturing, the light and heavy cell populations are combined into one pool and the proteins are isolated. The protein pool is then digested with a protease, typically trypsin, to form a peptide pool that is analyzed by mass spectrometry. Each peptide analyzed will be present in two forms, the light and the heavy form. The two forms have the same chemical properties so they have the same chromatographic retentions, ionization efficiencies, and fragmentation characteristics but they are distinguishable based on the mass difference due to the heavy isotope incorporation. The peak signals produced by the heavy and light forms of a peptide are measured by the mass spectrometer and a relative quantification of that peptide from the two cultures is calculated. Tandem mass spectrometry is also performed in the same experiment on either the heavy or

the light form so the identity of the peptide, and the protein it comes from, is determined. To ensure that the growth of the cells is not affected by the isotope strategy, one can reverse which treatment receives the modified media.

It is important to note that all peptides, both phosphorylated and nonphosphorylated, that contain the isotopically-labeled amino acid are available for relative quantification. To assist with enrichment of phosphoproteins in the SILAC method, immunoprecipitation of a target protein can be performed after cell lysis (Ballif et al., 2005), with mass spectrometry performed on the proteolytically digested proteins for quantification. To assist with the enrichment of phosphopeptides in the SILAC method, SCX chromatography, IMAC, or a combination of both can be employed after proteolytic digestion (Gruhler et al., 2005). This second approach enriches the phosphopeptides and helps remove nonphosphorylated peptides that can act as noise in the quantification experiment.

iTRAQ

A second method for the global quantification of proteins and protein modifications is an *in vitro* chemical labeling procedure called iTRAQ. The iTRAQ reagent consists of four isobaric (same nominal mass) tags that can be used to label up to four separate protein samples (Figure 4A); for example, one sample might be “normal” HeLa cells while the three remaining samples might be HeLa cells grown at three different concentration of phosphatase inhibitors. The tags contain three regions: a peptide reactive region, a reporter region, and a balance region (Ross et al., 2004). The peptide reactive region of the tag consists of an NHS ester and is designed to react with the N-termini and lysines of peptides after protease digestions. The four reporter groups appear in the tandem mass spectrum at m/z 114, 115, 116, and 117. The attached balance groups are designed to make the total mass of the balance and reporter group 145 Da for each tag, resulting in balance groups of 31 Da, 30 Da, 29 Da, and 28 Da, respectively.

Protein samples for quantification are separately isolated, separately digested proteolytically, and each sample is chemically labeled with one of the iTRAQ reagents (Figure 4B). After labeling, the samples are combined and then analyzed by mass spectrometry. As the iTRAQ reagents are isobaric, identical peptides from each sample will have identical masses, so there is no division of the precursor signals in the first stage of mass analysis that could lead to increased spectral complexity by the combination of multiple samples. Additionally, the isobaric nature of the reagent increases the ion population for a given peptide by summing the amount of a peptide from each sample, thus making peptides easier to detect. During tandem mass spectrometry, fragmentation takes place along the peptide backbone allowing for qualitative analysis, while fragmentation also takes place between the reporter and balance region of the tag resulting in intense reporter ions in the tandem mass spectrum (Figure 4C). The relative amounts of these reporter ions correspond to the relative amounts of the peptides present in the four samples. It should be stressed that in contrast to SILAC and AQUA (see below), it is during tandem mass spectrometry experiments, and not the first stage of mass analysis, that relative quantification of peptides takes place.

Phosphoproteins can be analyzed in an identical manner as nonphosphorylated proteins with the iTRAQ methodology. Since the iTRAQ reagent labels phosphopeptides to the same degree as nonphosphorylated peptides and it does not affect the stability of phosphopeptides, no changes to the iTRAQ labeling procedure are necessary. Enrichment strategies, such as IMAC (Sachon et al., 2006; Zhang et al., 2005) or immunoprecipitation with antiphosphotyrosine antibodies (Zhang et al., 2005), can be utilized to remove nonphosphorylated peptides to focus the analysis on site-specific phosphorylation. Also, since iTRAQ is an *in vitro* labeling procedure it can be applied to clinical samples such as tumor tissues and fluids (e.g., serum and urine). Overall, iTRAQ is a very powerful method for quantifying phosphorylation on a proteomic scale.

AQUA

Unlike SILAC and iTRAQ that provide relative quantitative information, the AQUA strategy provides an absolute quantification of a protein of interest (Kirkpatrick et al., 2005). Here a peptide from the protein of interest is constructed synthetically to contain stable isotopes; the isotopically-labeled synthetic peptide is called an AQUA peptide (Figure 5). The stable isotopes are incorporated into the AQUA peptide by using isotopically heavy amino acids during the peptide synthesis process. The synthetic peptide then has a mass increase, say of 10 Daltons due to the incorporation of a single [$^{13}\text{C}_6$, $^{15}\text{N}_4$]-arginine into the synthetic peptide, relative to the native peptide. Even though the mass difference between the native and the synthetic peptide allows the mass spectrometer to differentiate between the two forms, both forms have the same chemical properties resulting in the same chromatographic retention, ionization efficiency, and fragmentation distribution. Experimentally, a *known* amount of the isotopically labeled peptide is added to a protein mixture, proteolytically digested, and analyzed by mass spectrometry. Since the native peptide and its synthetic counterpart have the same chemical properties, the mass spectrometry signal from the quantified synthetic peptide can be compared to the signal of the native peptide, ultimately allowing for the absolute quantity of the protein to be determined (Gerber et al., 2003). Multiple AQUA peptides can be used to quantify multiple proteins in a single experiment.

The same strategy can be used for quantifying a phosphorylated protein, except the AQUA peptide is synthesized with a phosphoserine, phosphothreonine, or phosphotyrosine in addition to the isotopically labeled amino acid. Additionally, when both the phosphorylated and nonphosphorylated forms of an AQUA peptide are used in a single experiment, both the amount of the total protein and the extent of phosphorylation at that site can be determined simultaneously (Gerber et al., 2003).

Future Directions

Current proteomics strategies mostly employ some form of the bottom-up approach. This approach has been very appealing since most peptides behave similarly to one another, and their chemical characteristics can be predicted making separations and enrichments of peptides straightforward. The bottom-up approach has also been embraced since mass spectrometer's limits of detection for peptides are much lower for peptides than for proteins. The major drawback of the bottom-up approach, however, is the loss of information about the whole protein. Some peptides, including modified peptides, may be lost during sample preparation or the peptide may not be of an optimal size to be detected by the mass spectrometer. Also, the nature of the bottom-up approach makes alternatively-spliced and post-translationally truncated proteins difficult to detect. These drawbacks in the bottom-up approach have resulted in serious research efforts to make top-down analysis of proteins by mass spectrometry routine; however, limitations to this strategy exist. First, commercial mass spectrometers that are suitable for top-down experiments are not widely available. Second, there are limitations in the ability to perform large-scale, or proteomics-scale, separations of proteins that are compatible with mass spectrometry. Third, a routine methodology to perform these experiments in a high throughput manner does not exist. The tremendous amount of research going into the development of top-down strategies will likely solve these limitations and this technology should be widely available within the next five years.

A second area of development is new peptide dissociation techniques. As described above, during CID phosphopeptides are prone to partial or complete loss of phosphoric acid while minimal fragmentation takes place along the peptide backbone, revealing little or no information about the peptide's sequence. The recently developed fragmentation techniques of Electron Capture Dissociation (ECD) and Electron Transfer Dissociation (ETD) have been shown to fragment the peptide backbone while leaving the phosphoserine/phosphothreonine

intact. Not only have ECD and ETD been used to sequence phosphopeptides (Shi et al., 2001; Syka et al., 2004), ECD has also been used to sequence phosphoproteins in top-down experiments (Shi et al., 2001). While both fragmentation techniques are of great utility to fragmenting phosphopeptides, to date ECD has been exclusively coupled to FTICR instruments, the most expensive type of mass spectrometry instrumentation available. Challenges also exist with coupling ECD to chromatography limiting the effectiveness of ECD in LC-MS experiments. ETD, on the other hand, is likely to be more useful since it is easily adapted to more cost-efficient ion trap mass spectrometers that are commonly used for protein and peptide analysis. Also, ETD takes place on time scales that allow it to be more useful in LC-MS experiments. Commercial mass spectrometers are currently available with ECD while instruments with ETD are just starting to become available. Soon, both techniques will be more widely available to make a broad impact in characterizing phosphoproteins.

Acknowledgements

Work on Cx43 phosphorylation was supported by grant GM55632 to PDL.

References

- Ballif BA, Roux PP, Gerber SA, MacKeigan JP, Blenis J, Gygi SP. Quantitative phosphorylation profiling of the erk/p90 ribosomal s6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc Natl Acad Sci U S A* 2005;102:667–72. [PubMed: 15647351]
- Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics* 2004;3:1093–101. [PubMed: 15345747]
- 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;101:12130–5. [PubMed: 15302935]
- Carr SA, Annan RS, Huddleston MJ. Mapping posttranslational modifications of proteins by ms-based selective detection: Application to phosphoproteomics. *Methods Enzymol* 2005;405:82–115. [PubMed: 16413312]
- Crow DS, Beyer EC, Paul DL, Kobe SS, Lau AF. Phosphorylation of connexin43 gap junction protein in uninfected and rous sarcoma virus-transformed mammalian fibroblasts. *Mol Cell Biol* 1990;10:1754–1763. [PubMed: 1690850]
- DeGnove JP, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* 1998;9:1175–88. [PubMed: 9794085]
- 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;20:301–5. [PubMed: 11875433]
- Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem ms. *Proc Natl Acad Sci U S A* 2003;100:6940–5. [PubMed: 12771378]
- Goodenough DA, Paul DL. Beyond the gap: Functions of unpaired connexon channels. *Nat Rev Mol Cell Biol* 2003;4:285–94. [PubMed: 12671651]
- 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;4:310–27. [PubMed: 15665377]
- Hertlein B, Butterweck A, Haubrich S, Willecke K, Traub O. Phosphorylated carboxy terminal serine residues stabilize the mouse gap junction protein connexin45 against degradation. *J Membrane Biol* 1998;162:247–57. [PubMed: 9543497]
- Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications. *Methods* 2005;35:265–73. [PubMed: 15722223]
- Knight ZA, Schilling B, Row RH, Kenski DM, Gibson BW, Shokat KM. Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* 2003;21:1047–54. [PubMed: 12923550]

- Kobe B, Kampmann T, Forwood JK, Listwan P, Brinkworth RI. Substrate specificity of protein kinases and computational prediction of substrates. *Biochim Biophys Acta* 2005;1754:200–9. [PubMed: 16172032]
- Laird DW. Connexin phosphorylation as a regulatory event linked to gap junction internalization and degradation. *Biochim Biophys Acta* 2005;1711:172–82. [PubMed: 15955302]
- Lampe PD, Johnson RG. Amino acid sequence of *in vivo* phosphorylation sites in the main intrinsic protein (mip) of lens membranes. *Eur J Biochem* 1990;194:541–547. [PubMed: 2176601]
- Lampe PD, Lau AF. The effects of connexin phosphorylation on gap junctional communication. *Int J of Biochem Cell Biol* 2004;36
- 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;4:873–86. [PubMed: 15858219]
- Loyet KM, Stults JT, Arnott D. Mass spectrometric contributions to the practice of phosphorylation site mapping through 2003: A literature review. *Mol Cell Proteomics* 2005;4:235–45. [PubMed: 15640519]
- Maass K, Ghanem A, Kim JS, Saathoff M, Urschel S, Kirfel G, Grummer R, Kretz M, Lewalter T, Tiemann K, Winterhager E, Herzog V, Willecke K. Defective epidermal barrier in neonatal mice lacking the c-terminal region of connexin43. *Mol Biol Cell* 2004;15:4597–608. [PubMed: 15282340]
- Meyer HE, Hoffmann-Posorske E, Korte H, Meilmeyer LMG. Sequence analysis of phosphoserine-containing peptides. *FEBS Lett* 1986;204:61–66. [PubMed: 3091399]
- Moreno AP. Connexin phosphorylation as a regulatory event linked to channel gating. *Biochim Biophys Acta* 2005;1711:164–71. [PubMed: 15955301]
- Musil LS, Cunningham BA, Edelman GM, Goodenough DA. Differential phosphorylation of gap junction protein connexin43 in junctional communication-competent and deficient cell lines. *J Cell Biol* 1990;111:2077–2088. [PubMed: 2172261]
- 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;1:376–86. [PubMed: 12118079]
- Ong SE, Kratchmarova I, Mann M. Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (silac). *J Proteome Res* 2003;2:173–81. [PubMed: 12716131]
- Paul DL, Ebihara L, Takemoto LJ, Swenson KI, Goodenough DA. Connexin46, a novel lens gap junction protein, induces voltage-gated currents in nonjunctional plasma membrane of *xenopus* oocytes. *J Cell Biol* 1991;115:1077–1089. [PubMed: 1659572]
- Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2d-nanolc-esi-ms/ms and titanium oxide precolumns. *Anal Chem* 2004;76:3935–43. [PubMed: 15253627]
- Posewitz MC, Tempst P. Immobilized gallium(iii) affinity chromatography of phosphopeptides. *Anal Chem* 1999;71:2883–92. [PubMed: 10424175]
- 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;3:1154–69. [PubMed: 15385600]
- Sachon E, Mohammed S, Bache N, Jensen ON. Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: Application to proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom* 2006;20:1127–34. [PubMed: 16521170]
- Saez JC, Berthoud VM, Branes MC, Martinez AD, Beyer EC. Plasma membrane channels formed by connexins: Their regulation and functions. *Physiol Rev* 2003;83:1359–400. [PubMed: 14506308]
- Saez JC, Nairn AC, Czernik AJ, Spray DC, Hertzberg EL, Greengard P, Bennett MVL. Phosphorylation of connexin 32, a hepatocyte gap-junction protein, by camp-dependent protein kinase, protein kinase c and calcium/calmodulin-dependent protein kinase. *Eur J Biochem* 1990;192:263–273. [PubMed: 2170122]
- Shi SD, Hemling ME, Carr SA, Horn DM, Lindh I, McLafferty FW. Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem* 2001;73:19–22. [PubMed: 11195502]

- Solan JL, Lampe PD. Connexin phosphorylation as a regulatory event linked to gap junction channel assembly. *Biochim Biophys Acta* 2005;1711:154–63. [PubMed: 15955300]
- 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;101:9528–33. [PubMed: 15210983]
- 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;2:591–8. [PubMed: 16094384]
- TenBroek EM, Lampe PD, Solan JL, Reynhout JK, Johnson RG. Ser364 of connexin43 and the upregulation of gap junction assembly by camp. *J Cell Biol* 2001;155:1307–18. [PubMed: 11756479]
- Willecke K, Eiberger J, Degen J, Eckardt D, Romualdi A, Guldenagel M, Deutsch U, Sohl G. Structural and functional diversity of connexin genes in the mouse and human genome. *Biol Chem* 2002;383:725–737. [PubMed: 12108537]
- Zeller M, Konig S. The impact of chromatography and mass spectrometry on the analysis of protein phosphorylation sites. *Anal Bioanal Chem* 2004;378:898–909. [PubMed: 14673564]
- Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 2005;4:1240–50. [PubMed: 15951569]

Figure 1. Approaches for characterizing proteins by mass spectrometry

The bottom-up approach (A) relies on breaking down proteins via proteolytic digestion to form peptides. The resulting peptides are then characterized, including sites of phosphorylation, to identify the proteins from which the peptides originated. This approach can be performed on a single protein (as shown in the figure) or it can be performed on a mixture of proteins. The top-down approach (B) subjects a whole protein to analysis by the mass spectrometer where the mass of the protein is measured and mass of the protein's fragments generated in the mass spectrometer are measured.

Figure 2. General scheme for tandem mass spectrometry experiments to characterize phosphopeptides

(A) In the bottom-up proteomics approach, a single stage of mass analysis, or MS, is performed first. Peptides from a protease digestion are converted to ions (depicted as arrows) in the ionization source, separated in a mass analyzer, and detected (note: the size of the arrows is meant to only reflect the size of the ions; the size of the arrows is not meant to indicate the rate at which the ions are traveling). A mass difference of ~80 Da could imply phosphorylated and nonphosphorylated versions of a peptide (peaks 4 and 5 in this example). **(B)** To gain information about the sequence of the peptide and the site of phosphorylation, a second stage of mass analysis is performed, or MS/MS, in which a single ion called the precursor ion is selected by the first mass analyzer and subjected to collision-induced dissociation (CID). The CID process consists of the precursor ion being exposed to gas molecules, such as helium or argon, in a collision cell resulting in collisions that cause the peptide ion to fragment. The resulting fragment ions are separated in a second mass analyzer and detected. The MS/MS fragmentation spectrum for the putative nonphosphorylated peptide ion species 4 (**B-i**) is compared to phosphopeptide ion species 5 (**B-ii**). The peaks in the spectra correspond to fragment masses arising from CID, with differences between the peaks corresponding to the mass of amino acids. The mass differences in both ion species are identical up through the fourth amino acid from the N-terminus producing the same amino sequences (same position on the m/z -axis). The mass differences between the two spectra change at the fifth amino acid due to the presence of phosphoserine (pS) in peptide ion species 5, but not peptide ion species 4. The remainder of the peaks for both peptide ion species has identical mass differences indicating the same amino acid sequence between the two species.

Figure 3. Protein quantification scheme using SILAC

Cell cultures A and B are grown under identical conditions, but cell culture B is fed an isotopically-labeled (heavy) amino acid in place of a naturally occurring, essential amino acid found in culture A. After multiple rounds of cell doublings to ensure complete incorporation of the heavy amino acid, the cells are combined and the proteins are harvested. The proteins are digested with a protease and mass spectrometry is performed on the peptide mixture. Peptides produced from proteins present in each cell population will appear in the MS spectrum as two peaks, separated by the mass of the isotope label used in the heavy amino acid (an increase of six Daltons when $^{13}\text{C}_6$ -arginine is used). The peak heights of in the MS spectrum provide a relative abundance of the proteins present between the two samples. Ions in the MS spectra can be further analyzed by MS/MS to identify the proteins from which the proteins originated. Ultimately, protein identifications can be linked to the quantitative information.

Figure 4. Protein quantification scheme using iTRAQ

(A) Four isobaric (identical nominal masses) iTRAQ reagents are available and each is designed to contain three regions: a peptide reactive region, a balance region, and a reporter region. (B) Experimentally, up to four samples can be prepared in parallel. First the protein mixtures are individually reduced to break disulfide bonds, alkylated to block free sulfhydryls on cysteines, and then digested with a protease to create peptides. Following digestion, each sample is separately labeled with one of the four iTRAQ reagents, termed 114, 115, 116, and 117. The peptide reactive group (see **panel A**) of the reagent will react with free amines on the peptides (amine N-termini and the amine on lysine side chains). After labeling, the four samples are combined into one mixture that is then analyzed by mass spectrometry. Since the iTRAQ reagent is isobaric, a labeled peptide present in each of the four samples appears as one ion species during MS measurements (not shown in the figure). (C) During MS/MS, labeled peptides produce fragmentation spectra that allow for the identification of the protein (region

not circled in the mass spectrum). Also, MS/MS produces fragmentation at sites engineered into the iTRAQ reagent (see **panel A**) yielding reporter ions (m/z 114, 115, 116, 117) that are used for the relative quantification of the proteins from each sample.

Figure 5. Absolute protein quantification scheme using AQUA reference peptides

Selection of the AQUA peptide (**A**) begins with selecting a peptide region from a protein of interest (a protein that has been previously studied by bottom-up proteomics and for which quantitative information is desired), in this case a region that contains a known phosphorylation site. The phosphopeptide is then synthesized to contain an isotopically-labeled amino acid (in this case the C-terminal arginine) and the synthetic peptide is then characterized by liquid chromatography coupled to mass spectrometry (LC MS) to determine its chromatographic retention time and MS/MS fragmentation characteristics. The AQUA peptide is then used experimentally (**B**) by adding a known amount of it to a protein lysate, proteolytically digesting the mixture, and characterizing the resulting peptide mixture by LC MS. With the AQUA peptide and the analyte peptide having the same chemical properties, but differing in mass, both peptides co-elute and have the same fragmentation characteristics. The chromatographic elution profile of the AQUA peptide is compared to the profile of the analyte to determine the absolute quantity of the analyte peptide.