Understanding protein phosphorylation on a systems level

Jimmy Lin, Zhi Xie*, Heng Zhu and Jiang Qian*

Advance Access publication date 7 January 2010

Abstract

Protein kinase phosphorylation is central to the regulation and control of protein and cellular function. Over the past decade, the development of many high-throughput approaches has revolutionized the understanding of protein phosphorylation and allowed rapid and unbiased surveys of phosphoproteins and phosphorylation events. In addition to this technological advancement, there have also been computational improvements; recent studies on network models of protein phosphorylation have provided many insights into the cellular processes and pathways regulated by phosphorylation. This article gives an overview of experimental and computational techniques for identifying and analyzing protein phosphorylation on a systems level.

Keywords: *systems biology; protein phosphorylation; mass spectrometry; protein microarray; kinase; phosphoprotein*

INTRODUCTION

Since the discovery of phosphoserine by Levene and Lipmann in 1932, protein phosphorylation has been shown to be one of the most central regulatory and signaling processes in the cell. Affecting up to 30% of the proteome and essential in the regulation of cellular functions, phosphorylation networks are also essential backbones of the communication system within cells [1]. Phosphorylation can regulate a variety of important protein functions, including subcellular localization, protein degradation and stabilization, as well as biochemical activities [2–4].

Simply defined, phosphorylation is a reversible protein posttranslational modification (PTM), catalyzed by the family of proteins called protein kinases. In eukaryotes, this protein family is highly conserved and is classified into three major subfamilies on the basis of the residues they modify (i.e. Ser/Thr, Tyr and dual specificity kinases). However, in lower eukaryotes (e.g. the budding yeast) and plants, there are no bona fide tyrosine kinases [5]. It is generally believed that Tyr kinases evolved much later than the Ser/Thr kinases. In most fully sequenced eukaryotes, protein kinases are well annotated. In yeast, there are 122 kinases identified [6–8], whereas in *Drosophila* there are 251 [9]. In humans, the kinome consists of 518 protein kinases [10]; in mice, there are 540 members [11]. As with all protein family identification, these numbers are not final; there may be identified kinases that are not catalytically active, as well as other kinases that do not share high sequence homology with known kinases and, thus, are yet to be discovered.

Corresponding author: Jiang Qian, Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. Tel: +443-287-3882; Fax: +410-502-5382; E-mail: jiang.qian@jhmi.edu

Jimmy Lin is an MD/PhD candidate at the Johns Hopkins School of Medicine and an MHS (Bioinformatics) candidate at the Bloomberg School of Public Health. His research interests include computational genomics and proteomics as well as cancer diagnostics and therapeutics.

Zhi Xie is a postdoctoral fellow in the Wilmer Eye Institute at the Johns Hopkins School of Medicine. His research interests include computational analysis of protein phosphorylation, protein–DNA interactions and deterministic and stochastic modeling of genetic regulatory networks.

Heng Zhu, PhD, is an Assistant Professor in the Department of Pharmacology at the Johns Hopkins School of Medicine. His research interests include protein posttranslational modification, transcription factor regulatory networks, host–pathogen interaction and biomarker identification using a proteomics approach.

Jiang Qian, PhD, is an Assistant Professor in the Wilmer Eye Institute at the Johns Hopkins School of Medicine. His research interests include global analysis of gene regulation and signaling pathways.

*These authors contributed equally to this work.

Traditionally, protein phosphorylation has been studied by using several biochemical approaches, including *in vivo* labeling, phosphopeptide mapping, protein purification-associated kinase activity and phosphoamino acid analysis [12]. Although these approaches laid invaluable foundations for understanding protein phosphorylation, they not only are labor intensive but also focus on small numbers of individual proteins. With recent technological advances, protein phosphorylation events can now be studied on a more global and systematic scale; large-scale high-throughput approaches have exponentially increased our understanding of the complexity of the phosphorylation networks.

Here, we review the contemporary high-throughput approaches in protein phosphorylation research (Table 1). In the first part, we examine identification of phosphoproteins and phosphorylated sites in these proteins. In the second part, we focus on identification of kinase–substrate relationships. In the last part, we discuss the construction of phosphorylation networks and their global analysis.

IDENTIFICATION OF PHOSPHOPROTEINS AND PHOSPHORYLATION SITES

The challenge

Only with recent technological developments, researchers have started to identify and study phosphoproteins and phosphorylation sites in a high-throughput manner. Several obstacles make the task difficult. (1) There is no simple central unifying sequence homology among the phosphoproteins. While smaller subsets of phosphoproteins may share some sequence homology, these are still not well annotated and understood. (2) Because protein phosphorylation is a PTM and can occur on a wide variety of proteins, there is an extremely large set of possibly phosphorylated proteins with few unifying characteristics besides phosphorylation. (3) Biologically, phosphoproteins are often present in little amounts and often only a small portion of the same protein is phosphorylated. Thus, high sensitivity is required for detection. (4) Phosphorylation events are often transient and dynamic. Although some proteins fluctuate regularly in their phosphorylation status during the cell cycle, other proteins are phosphorylated only under specific conditions and in response to specific stimuli. Thus, comprehensive understanding would require multiple measurements of the same relationships under different conditions and at different times. (5) Phosphoproteins usually have multiple phosphorylation sites, and additional research is needed to determine which site is physiologically relevant. (6) A phosphoprotein can be phosphorylated by multiple kinases, and the physiological and biological results may differ. Therefore, understanding of phosphorylation is complex and multifaceted. For the reasons above, the

Table 1: High-throughput approaches in protein phosphorylation research

Technology	Information detected	Advantages	Disadvantages
Gel-based systems	Phosphoproteins and phosphorylation sites	Low cost and large separation possible; specific detection	Sensitivity differs depending on antibody attributes
Mass spectrometry	Phosphoproteins and phosphorylation sites	Large-scale measurement with many sites identified in parallel	Detection limited by enrichment processes; complexity of spectra requires extensive computational effort and existing database
Random peptide library screening	Kinase–substrate relationships	Allows for higher throughput detection of relationships	Dependent on the size of the library, sequences are often too short
Genetic perturbation	Kinase–substrate relationships	Biologically important phenotypic detection	Results may be due to indirect effects
Protein microarray	Kinase–substrate relationships	Allows unbiased screening of large number of relationships	*In vitro* protein may not reflect *in vivo* protein function
Computational approaches	Phosphoproteins and phosphorylation sites; kinase–substrate relationships	Fast, low cost, allows high-order integration	Dependent on existing information, high false-positive rate

understanding of phosphoproteins is non-trivial, and many researchers have contributed to this initiative [13, 14]. Here, we focus on the technologies of the high-throughput approaches in identifying phosphoproteins and phosphorylation sites.

Gel-based systems

Two-dimensional gel electrophoresis, or 2-DE, is a classic proteomics method to separate protein mixtures and detect phosphorylation changes. Its general use allows the separation of proteins in a complex mixture in various dimensions [15]. With the development of phosphospecific stains (e.g. Pro-Q Diamond dye) [16, 17] and the incorporation of modified adenosine triphosphate (ATP) analogues, such as with radio-labeled phosphate or additional bulky R groups [18–20], the separated phosphoproteins can be visualized and detected on the gel. With use of this approach, a number of phosphoproteins have been identified in the extracellular signal-regulated kinase (ERK) pathway [21]. Another gel-based method of phosphoprotein detection is immunoblotting, which takes advantage of phosphospecific antibodies that have been developed against different phosphorylated amino acids [22]. For example, anti-phosphotyrosine antibodies have been generated to detect a wide range of tyrosine-related phosphoproteins [23, 24]. However, antibodies raised against phospho-Ser/Thr generally have low affinity and specificity. Alternatively, phosphorylated peptides on specific Ser, Thr or Tyr residues have been used to create phosphorylation site-specific antibodies. Indeed, more than 300 of such antibodies are currently available [25]. While this method is powerful, different antibodies behave differently and often have different detection ranges and limits. The conclusions must take into consideration the differing sensitivity, specificity and behavior of the different antibodies.

Mass spectrometry

Mass spectrometry (MS) enables simultaneous identification of large numbers of proteins in complex samples [26]. One of the greatest breakthroughs in phosphoproteomics is the application of MS to identify thousands of *in vivo* phosphorylation sites in a single study [27]. Even though MS enables unbiased and large-scale detection of proteins and peptides, due to the aforementioned obstacles of low abundance, low ratio and transiency of the modifications, many adaptations have been made to this technology for phosphoproteomics.

To circumvent the problem of low abundance, many strategies have been developed to enrich phosphoproteins or phosphopeptides prior to MS analysis [15, 28–30]. One strategy is to selectively enrich for phosphoproteins and remove the non-phosphoproteins via immunoprecipitation with antibodies [31, 32]. Although not all antibodies effective for immunoblotting can be directly translated to use with immunoprecipitation, large panels of antibodies still remain that can be used to enrich the protein population [31]. Because of the high affinity and specificity of anti-phosphotyrosine antibodies, they have been applied to obtain a global picture of tyrosine-phosphorylated proteins [31, 33–38]. Another enrichment method is immobilized metal affinity chromatography (IMAC) [39, 40]. Simply put, it is an electrostatic-based method that takes advantage of the negative charge of phos-phosphate groups and the positive charge of metal ions. The interaction is mediated by using nitriloacetic acid (NTA), iminodiacetic acid (IDA) and Tris (carboxymethyl) ethylenediamine (TED) linkers. As expected, since charge is not specific to particular amino acids, other negatively charged acidic amino acids, such as glutamic and aspartic acid, also bind non-specifically to metal ions [3]. Nonetheless, IMAC has been effective in determining the yeast phosphoproteome, plasma membrane phosphoproteins and cell line phosphoproteins [3, 41, 42]. Other similar approaches include strong cation-exchange chromatography (SCX) [43, 44] and enrichment by titanium dioxide and zirconium dioxide [45].

Because of the complexity of the MS spectra, extensive computational analysis is needed before phosphopeptide identification. The first computational step requires the implementation of search algorithms to match the obtained spectrum with known spectra databases [46]. Many software packages perform this step including Mascot [47], SEQUEST [48], OMSSA [49], X! Tandem [50], GutenTag [51], InsPecT [52] and Spectral Networks Analysis [53]. Because of the high false-positive rate of matching the complex MS spectra, a second computational step is to filter out known false positives, using algorithms such as DTASelect [54–56] and PhosphoPIC [57]. Finally, there is a step for further curation and confirmation.

Specifically, since phosphoproteins often contain more than one phosphosite, care must be taken to examine how each site performs differently under different conditions. Programs such as MSQuant [58, 59] and Ascore [59, 60] have enabled this curation and confirmation process.

The most powerful application of MS is its ability to determine the cell status in terms of phosphorylation sites on a global scale. Many studies have been performed on a variety of organisms and in different physiological conditions; these organisms include yeast, *Drosophila*, mouse and human [61–65]. For example, by using a strategy based on SCX, more than 2000 phosphopeptides were enriched from 967 proteins in the nuclear fraction of HeLa cell lysate [44]. This accomplishment illustrates that a large-scale comparative phosphorylation analysis of different cell states using MS could simultaneously profile hundreds, even thousands, of phosphorylation sites. Subsequently, the same group conducted a study on cell cycle-regulated phosphorylation in HeLa cells and identified more than 14 000 phosphorylation sites from 3682 proteins [66]. Interestingly, many sites contain the cyclin-dependent motif, a finding that suggests most of the sites may be the direct targets of cyclin-dependent kinase (CDK). In another study, Yang and coworkers performed phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation [67]. They found that irradiation stimulates the signaling pathways, given that the number of phosphopeptides increased in both low- and high-dose irradiation. More importantly, a significant enrichment in the CDK motifs was observed after low-dose irradiation, whereas the motifs for 3-phosphoinositide-dependent protein kinase-1 (PDK1) and AKT/RSK were enriched in high-dose irradiation.

In yeast, an integrated phosphoproteomic technology identified more than 700 phosphopeptides from key regulator proteins of a prototypical G-protein-coupled receptor signaling pathway [68]. Among these identified phosphopeptides, 139 were differentially regulated in response to mating pheromone; interestingly, these regulated proteins were components belonging to the mitogen-activated protein kinase (MAPK) signaling pathway. This example, as well as the ones above, demonstrates the power of the global analysis of phosphorylation in revealing the signaling pathways involved in different biological processes.

IDENTIFICATION OF KINASE–SUBSTRATE RELATIONSHIP

The challenge

The traditional approach to identifying upstream kinase(s) of a substrate of interest has been mainly based on *in vitro* assays. Starting with a single pure recombinant kinase, one can query many of the possible targets [69–73]. However, due to the focused nature of this method, it is not suitable for high-throughput analyses. As reviewed in the previous section, many phosphoproteins and phosphorylation sites have been documented, mainly via the MS approach. However, the connections between substrates/sites and their upstream kinases are still missing. One possible way to identify upstream kinase(s) for a phosphorylation site is to compare with known phosphorylation motifs (see below for computational approaches). If the phosphorylation motif of one kinase matches the identified phosphorylation site, it is likely to be the upstream kinase of the phosphorylation site. However, two main problems hinder this strategy. First, most kinases currently do not have a known phosphorylation motif. Second, many kinases share similar phosphorylation motifs. For example, many CDKs have the p[S/T]P motif, while kinases from calmodulin-dependent protein kinase (CAMK) group often have the Rxxp[S/T] motif [74]. Therefore, new approaches are needed to identify the kinase–substrate relationship.

Random peptide library screening

One large-scale method to identify kinase–substrate relationships is using peptide libraries to query exhaustively possible peptides of a particular kinase in order to characterize the preferred peptides for the kinase in question. The results can then be used to define specificity of the kinase. For example, by scoring an immobilized peptide library, polo-box domain (PBD) was identified as a specific binding domain of the mitotic kinase polo-like kinase 1 (Plk1), and this domain mediates Plk1 localization to substrates [75]. In a large-scale study, a positional scanning peptide library was constructed consisting of 198 distinct peptide mixtures that were used to profile the sequence preferences for phosphorylation by eight protein serine/threonine kinases [76]. Following the derivation of preferred peptides of kinases, one useful application is to predict novel substrates of a kinase by scanning protein databases with the preferred peptides. While this method allows one to start getting a global picture of the

kinase–substrate relationships, the derived peptide preferences of kinases are often too degenerative to pinpoint the full-length protein substrates. This limitation inevitably results in a high rate of false positives of predicted substrates.

Genetic perturbation

Another method for identifying substrates of a particular kinase is to introduce genetic perturbation by either knocking down or knocking out the kinase of interest in cells. The rationale for this approach is as follows: if the perturbation of a kinase results in similar phenotype as the perturbation of a substrate, it is likely that the substrate is a target of the kinase. In systems where gene knockout is straightforward via homologous recombination, such as the budding yeast, comparisons can be made between the wild-type and knockout strains and the affected phosphoproteins should be hypophosphorylated in the mutant [77]. However, in higher eukaryotes where homologous recombination is often not practical, RNA interference (RNAi) technology offers a convenient means of depleting specific proteins.

From small-scale experiments, Sachs and coworkers studied a signaling network of 11 proteins that were perturbed individually and responses of the system were measured in a large number of replicates (700–900 times). A Bayesian network was employed to identify the best network model fitting all perturbation data [78]. In a large-scale study of using genetic perturbation by RNAi, several hundred human kinases were targeted by RNAi and around 11% of the kinome was found essential for promoting cell survival [79]. In addition, many new kinase–substrate pairs were also revealed. While this method is powerful, there are a few potential drawbacks. First, there may be incomplete depletion and some nonspecific effects. Second, it is difficult to distinguish the direct and indirect effects of the depletion of the kinase since the phenotype caused by kinase depletion reflects a consequence of a number of misregulating pathways. Third, just depleting particular kinases or substrates may fail to yield a phenotype due to compensatory pathways.

Protein microarray

Protein microarrays, also known as protein chips, are miniaturized and parallel assay systems that contain small amounts of purified proteins in a high-density format [8, 80]. This approach allows the simultaneous determination of a variety of analytes from small amounts of samples within a single experiment. The development of protein microarray technology has revolutionized the determining of the relationship between enzymes and their substrates. Functional protein microarrays are typically prepared by immobilizing individually purified proteins onto a microscope slide using a standard contact chip writer or non-contact microarrayer. A protein microarray can be viewed as a substrate array when an enzymatic reaction is performed on it to identify potential downstream targets [81]. Many types of enzymatic reactions have been developed for various types of PTMs, such as phosphorylation [5, 8, 82], ubiquitylation [83], acetylation [84] and deoxyribonucleic acid (DNA) crosslinking [85]. Protein chips offer many advantages for studying protein phosphorylation. Thousands of proteins can be rapidly screened for enzyme–substrate relationships in an unbiased fashion with very small amounts of reagents and under a variety of test conditions. In addition, closely related kinases with known redundant functions can be readily differentiated at the molecular level on the basis of their substrate profiles.

Zhu and coworkers first analyzed the substrate specificity of 119 yeast kinases on 17 different substrates using nanowell protein chips [80]. Tiny wells with 10 μl capacity were created in a silicon elastmer, polydimethylsiloxane (PDMS), to carry out parallel kinase reactions. Substrates were first covalently crosslinked to the bottom of the wells, and individually purified yeast kinases were added in kinase buffer in the presence of radio-labeled ATP. After the reactions, the nanowell chips were washed extensively, dried and exposed to X-ray film to visualize the phosphorylation signals. By doing so, phosphorylation profiles of almost all of the yeast kinases were obtained. Another example of the same approach involved identifying potential upstream kinases of a yeast protein Crz1p [86]. The identified kinase, Hrr25p, was not only identified but was also validated with *in vitro* and *in vivo* methods.

Later, the same group managed to accomplish the so-called 'Phosphorylome Project' using the yeast proteome microarrays [5]. The goal was to identify all the potential protein substrates of each yeast kinase. *In vitro* kinase reactions were carried out on the yeast proteome chips using 87 individually purified kinase(s) complexes in the presence of ^{33}P-ATP. In all, 4129 phosphorylation events, involving 1325 different proteins, were identified. To ensure that the signals resulted from

phosphorylation events, 5% sodium dodecyl sulfate (SDS) was used to denature proteins on chips to remove signals from binding of kinase proteins or ^{33}P-ATP. Those phosphorylation results have been assembled into a first-generation, global kinase signaling network in yeast. While this experiment is set up from a kinase perspective, one can also examine the data from a complementary view. From a substrate perspective, more than a quarter of the substrates could be recognized by three or more kinases, thus showing the multivalent behavior of some substrates; however, the vast majority of substrates are recognized by only a few kinases.

The use of protein microarrays has now been extended to different organisms. Using the first Arabidopsis protein chip containing 1690 proteins, 48 and 39 potential substrates have been identified for the *Arabidopsis thaliana* MAPKs, MPK3 and MPK6, respectively [87]. Nearly all of the 48 potential MPK3 substrates were confirmed by other *in vitro* methods, and one of the identified MPK6 substrates was also shown to be the MAPK substrate *in vivo* by another independent study. Recently, a comprehensive MAPK target network in Arabidopsis was constructed using protein microarrays [82]. In total, 570 MPK phosphorylation substrates have been identified for 10 different MAPKs and some of the substrates were validated by reconstitution experiments. Because of the high-throughput nature of these studies with small amounts of protein, this powerful method will start to revolutionize our understanding of kinase–substrate relationships on a global scale.

Computational approaches

Recently, a number of computational approaches have been proposed to predict kinase–substrate relationships [88]. Most of these prediction algorithms rely on phosphorylation motifs; if a kinase has a known phosphorylation motif, one can scan the protein sequences and the proteins with matched sites are predicted to be potential substrates of the kinase [89]. This method has been shown effective in yeast to identify and predict protein kinase A (PKA) substrates [90]. While such approaches are powerful and useful, a drawback is the high rate of false positives due to the fact that phosphorylation motifs often degenerate and are not the only determinant for the specificity for kinase–substrate interaction.

A more sophisticated computational approach would be to build networks based on diverse information [91]. In the system named NetworKIN, phosphorylation networks are created based on phosphomotif and protein–protein interaction networks [74]. Predictions of kinase–substrate interactions based solely on phosphomotif have low accuracy. However, predictions of relationships are greatly improved by taking into consideration the network context. Many of these predicted relationships were verified using experimental methods [74]. In total, this study shows that the topological position of kinases, phosphatases and their substrates provides rich information that will be useful for understanding and prediction of the network.

UNDERSTANDING PHOSPHORYLATION NETWORKS

With most proteins having multiple phosphosites, most kinases having numerous substrates and most substrates being mediated by many kinases and substrates, phosphorylation cannot be understood simply as pair-wise interactions or even linear pathways. Comprehensive and accurate understanding of protein phosphorylation must take into account all the intricate and complex interrelationships between all the different components. Phosphorylation networks, as an important part of cellular networks, are defined by the components, that is, kinase and substrates, and the phosphorylation events among those components. One of the most powerful ways of understanding phosphorylation as a network instead of a group of disparate pair-wise relationships is that the overall signal pathways can be discerned [92, 93].

Specificity of kinase–substrate relationship

Identification of consensus sequences (motifs) that the kinases recognize can help us understand the specificity of kinase–substrate interactions. Some well-studied kinases have known phosphorylation motifs identified as the results of traditional experimental studies. In addition, global studies of phosphorylation networks accumulate large amounts of data on phosphorylation sites. All of these motifs and sites are now consolidated into different databases, including PhosphoELM [94, 95], PhosphoSite [96] and Human Protein Reference Database [97].

Computational tools are extremely helpful in determining and generating phosphorylation motifs from known phosphoproteins [76, 98, 99].

Currently available motif-generating tools include Scansite [100], Phosida [101] or NetPhos [102, 103] and motif-x [104]. In particular, motif-x algorithm discovered overrepresented motifs in a number of phosphorylation studies. A more recent study was aimed at creating an atlas of consensus sequence motifs derived from kinase-specific data covering 179 kinases and 104 phosphorylation-dependent binding domains [98].

However, it is worth noting that phosphorylation motif is not the sole determinant for the specificity of the kinase–substrate interactions. Physical protein–protein interactions can help facilitate the phosphorylation reaction. Third, proteins (scaffold proteins) can also contribute to the specificity of kinase–substrate interactions [105].

From *in vitro* to *in vivo* networks

Despite the many aforementioned advances in identification of kinase–substrate interactions by protein microarray assays or computational predictions, these interactions may be more indicative of *in vitro* than *in vivo* relationships of kinases and substrates, and the identified substrates may not be physiologically relevant targets since special environments within cells are not present in *in vitro* assays; differences include presence of regulatory subunits, protein subcellular compartmentalization, regulation in gene expression dynamics and tissue specificity. Therefore, *in vitro* false positives must be separated from true *in vivo* relationships. Many different bioinformatics methodologies, alone or in combination, have been proposed and applied to this problem, including artificial neural networks based on experimentally verified phosphorylation sites [102] and evolutionary proteomics approach [90]. Linding and coworkers, for example, recently proposed a computational approach to map known phosphorylation sites to their corresponding kinases using motif-based predictions with the network context of kinases and phosphoproteins, including pathway information and coexpression data [91]. The authors found that network context improved the success rate of identification of physiologically relevant kinase–substrate interactions, and they further validated several new kinase *in vivo* targets associated with a DNA damage signaling pathway.

Global properties of networks

For a more comprehensive understanding of the functions of phosphorylation in the cellular context, it is important to integrate phosphorylation networks with other well-studied cellular networks. In a recent study involving a global investigation of protein phosphorylation in yeast, a phosphorylation network was constructed through integration with the large data sets covering protein–protein interactions and transcription factor (TF) binding sites. It is notable that the largest group of substrates is TFs, suggesting the importance of phosphorylation in the regulation of cellular activity. Interestingly, in the search for regulatory connections, these seemingly complex networks can be simplified into several types of regulatory modules: interacting substrates, scaffold, kinase cascade, TF-regulated kinase, kinase regulation, kinase feedback loops I and II and heterosubstrate regulation module [5]. This finding reveals some of the basic network building blocks within the more extensive network, allowing researchers to start seeing and decoding these complex interrelationships that exist within the cellular regulatory network.

Dynamics of phosphorylation networks

Phosphorylation is not only a reversible but also a dynamic process that fluctuates in response to different stimuli and conditions. A complete picture of global phosphorylation can only be seen when one considers different cell states, responses to stimuli or different time courses. Therefore, quantifying temporal dynamics requires high sensitivity, accurate quantification and analysis of multiple time points. One approach is to perturb components of a phosphorylation network and observe the behavior of other components. MS has been used to explore the dynamics of a phosphorylation network. The general paradigm is to differentially label samples prior to MS to allow for eventual comparison, which will reveal dynamic changes of components in the phosphorylation network during a time course of a particular treatment or during differential treatments [58, 106]. A common strategy of sample labeling is stable isotope labeling by amino acids in cell culture (SILAC) [107]. With this method, cells are grown differently with isotopically labeled amino acids. This then allows for comparative quantization in MS. In a recent study, SILAC-encoded cells were used after being stimulated with epidermal growth factor (EGF) to determine temporal dynamics of phosphorylation proteins and their sites as a function of stimulus, time and subcellular location. In total, more than 6000 phosphorylation sites and over 1000 sites

whose phosphorylation was modulated by EGF stimulation were identified [58]. The dynamic phosphorylation network provides a global and integrative view of cellular regulation.

SUMMARY

In the past decade, significant progress has been made in understanding kinases, substrates and their interacting relationship on a systems level. While these developments are still in their early stages, the preliminary results and the importance of comprehensive global pictures of phosphorylation have already been shown to be important and effective. However, many more advances need to be made to better elucidate the functions of protein phosphorylation: (i) a deeper coverage of phosphoproteins and a complete mapping of kinase–substrate pairs; (ii) a comprehensive understanding of the dynamic relationships as a network; (iii) the integration of information from other networks, such as protein–protein interaction, transcription and other PTM networks. We expect that these upcoming studies will provide a representative and complete view of protein phosphorylation and its role in controlling cellular processes.

FUNDING

The National Institutes of Health grant No. EY017589 and GM076102.

Key Points

- Protein phosphorylation is critical in regulating many cellular processes, and understanding of these processes has been expanded by the development of technologies enabling large-scale unbiased queries.
- Mass spectrometry is effective in both identifying large numbers of phosphoproteins and monitoring global phosphorylation levels.
- Protein microarray is a complementary technology that enables the discovery of kinase–substrate relationships on a global scale.
- Network models combining different sources of experimental data greatly enable the deciphering of complex protein phosphorylation networks.

References

1. Zolnierowicz S, Bollen M. Protein phosphorylation and protein phosphatases. De Panne, Belgium, September 19–24, 1999. *EMBO J* 2000;**19**:483–8.
2. Manning G, Plowman GD, Hunter T, *et al*. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 2002;**27**:514–20.
3. Ficarro SB, McCleland ML, Stukenberg PT, *et al*. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* 2002;**20**:301–5.
4. Cohen P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem Sci* 2000; **25**:596–601.
5. Ptacek J, Devgan G, Michaud G, *et al*. Global analysis of protein phosphorylation in yeast. *Nature* 2005;**438**:679–84.
6. Brinkworth RI, Munn AL, Kobe B. Protein kinases associated with the yeast phosphoproteome. *BMC Bioinformatics* 2006;**7**:47.
7. Hunter T, Plowman GD. The protein kinases of budding yeast: six score and more. *Trends Biochem Sci* 1997;**22**:18–22.
8. Zhu H, Klemic JF, Chang S, *et al*. Analysis of yeast protein kinases using protein chips. *Nat Genet* 2000;**26**: 283–289.
9. Morrison DK, Murakami MS, Cleghon V. Protein kinases and phosphatases in the Drosophila genome. *J Cell Biol* 2000;**150**:F57–62.
10. Manning G, Whyte DB, Martinez R, *et al*. The protein kinase complement of the human genome. *Science* 2002; **298**:1912–34.
11. Caenepeel S, Charydczak G, Sudarsanam S, *et al*. The mouse kinome: discovery and comparative genomics of all mouse protein kinases. *Proc Natl Acad Sci USA* 2004;**101**: 11707–12.
12. Johnson SA, Hunter T. Kinomics: methods for deciphering the kinome. *Nat Methods* 2005;**2**:17–25.
13. Morandell S, Stasyk T, Grosstessner-Hain K, *et al*. Phosphoproteomics strategies for the functional analysis of signal transduction. *Proteomics* 2006;**6**:4047–56.
14. Stern DF. Phosphoproteomics for oncology discovery and treatment. *Expert Opin Ther Targets* 2005;**9**:851–60.
15. Gorg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 2004; **4**:3665–85.
16. Levine A, Vannier F, Absalon C, *et al*. Analysis of the dynamic Bacillus subtilis Ser/Thr/Tyr phosphoproteome implicated in a wide variety of cellular processes. *Proteomics* 2006;**6**:2157–73.
17. Hopper RK, Carroll S, Aponte AM, *et al*. Mitochondrial matrix phosphoproteome: effect of extra mitochondrial calcium. *Biochemistry* 2006;**45**:2524–36.
18. Stasyk T, Dubrovska A, Lomnytska M, *et al*. Phosphoproteome profiling of transforming growth factor (TGF)-beta signaling: abrogation of TGFbeta1-dependent phosphorylation of transcription factor-II-I (TFII-I) enhances cooperation of TFII-I and Smad3 in transcription. *Mol Biol Cell* 2005;**16**:4765–80.
19. Villarino A, Duran R, Wehenkel A, *et al*. Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. *J Mol Biol* 2005;**350**: 953–63.
20. Shah K, Liu Y, Deirmengian C, *et al*. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci USA* 1997;**94**:3565–70.
21. Lovric J, Dammeier S, Kieser A, *et al*. Activated raf induces the hyperphosphorylation of stathmin and the

reorganization of the microtubule network. *J Biol Chem* 1998;**273**:22848–55.

22. Kaufmann H, Bailey JE, Fussenegger M. Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* 2001;**1**:194–9.
23. Huang D, Liu X, Plymate SR, *et al*. Proteomic identification of 14-3-3 sigma as a common component of the androgen receptor and the epidermal growth factor receptor signaling pathways of the human prostate epithelial cell line M12. *Oncogene* 2004;**23**:6881–9.
24. Kim SY, Chudapongse N, Lee SM, *et al*. Proteomic analysis of phosphotyrosyl proteins in morphine-dependent rat brains. *Brain Res Mol Brain Res* 2005;**133**:58–70.
25. Mandell JW. Phosphorylation state-specific antibodies: applications in investigative and diagnostic pathology. *Am J Pathol* 2003;**163**:1687–98.
26. Shevchenko A, Jensen ON, Podtelejnikov AV, *et al*. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci USA* 1996;**93**:14440–5.
27. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;**422**:198–207.
28. Paradela A, Albar JP. Advances in the analysis of protein phosphorylation. *J Proteome Res* 2008;**7**:1809–18.
29. Macek B, Mann M, Olsen JV. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* 2009;**49**:199–221.
30. Huber LA. Is proteomics heading in the wrong direction? *Nat Rev Mol Cell Biol* 2003;**4**:74–80.
31. Gronborg M, Kristiansen TZ, Stensballe A, *et al*. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* 2002;**1**:517–27.
32. Witze ES, Old WM, Resing KA, *et al*. Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* 2007;**4**:798–806.
33. Pandey A, Podtelejnikov AV, Blagoev B, *et al*. Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci USA* 2000;**97**:179–84.
34. Pandey A, Fernandez MM, Steen H, *et al*. Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. *J Biol Chem* 2000;**275**:38633–9.
35. Steen H, Kuster B, Fernandez M, *et al*. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 2002;**277**:1031–9.
36. Yeung YG, Wang Y, Einstein DB, *et al*. Colony-stimulating factor-1 stimulates the formation of multimeric cytosolic complexes of signaling proteins and cytoskeletal components in macrophages. *J Biol Chem* 1998;**273**:17128–37.
37. Caratu G, Allegra D, Bimonte M, *et al*. Identification of the ligands of protein interaction domains through a functional approach. *Mol Cell Proteomics* 2007;**6**:333–45.
38. Machida K, Thompson CM, Dierck K, *et al*. High-throughput phosphotyrosine profiling using SH2 domains. *Mol Cell* 2007;**26**:899–915.
39. Andersson L, Porath J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem* 1986;**154**:250–4.
40. Gaberc-Porekar V, Menart V. Perspectives of immobilized-metal affinity chromatography. *J Biochem Biophys Methods* 2001;**49**:335–60.
41. Shu H, Chen S, Bi Q, *et al*. Identification of phosphoproteins and their phosphorylation sites in the WEHI-231 B lymphoma cell line. *Mol Cell Proteomics* 2004;**3**:279–86.
42. Kim JE, Tannenbaum SR, White FM. Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J Proteome Res* 2005;**4**:1339–46.
43. Ballif BA, Villen J, Beausoleil SA, *et al*. Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics* 2004;**3**:1093–101.
44. Beausoleil SA, Jedrychowski M, Schwartz D, *et al*. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* 2004;**101**:12130–5.
45. Pinkse MW, Uitto PM, Hilhorst MJ, *et al*. 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.
46. Bakalarski CE, Haas W, Dephoure NE, *et al*. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal Bioanal Chem* 2007;**389**:1409–19.
47. Perkins DN, Pappin DJ, Creasy DM, *et al*. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;**20**:3551–67.
48. Yates JR, III, Eng JK, McCormack AL, *et al*. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 1995;**67**:1426–36.
49. Geer LY, Markey SP, Kowalak JA, *et al*. Open mass spectrometry search algorithm. *J Proteome Res* 2004;**3**:958–64.
50. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004;**20**:1466–7.
51. Tabb DL, Saraf A, Yates JR, III. GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 2003;**75**:6415–21.
52. Tanner S, Shu H, Frank A, *et al*. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* 2005;**77**:4626–39.
53. Bandeira N, Tsur D, Frank A, *et al*. Protein identification by spectral networks analysis. *Proc Natl Acad Sci USA* 2007;**104**:6140–5.
54. Lu B, Ruse C, Xu T, *et al*. Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal Chem* 2007;**79**:1301–10.
55. Tabb DL, McDonald WH, Yates JR, III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 2002;**1**:21–6.
56. Cantin GT, Venable JD, Cociorva D, *et al*. Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway. *J Proteome Res* 2006;**5**:127–34.
57. Hoffert JD, Wang G, Pisitkun T, *et al*. An automated platform for analysis of phosphoproteomic datasets: application to kidney collecting duct phosphoproteins. *J Proteome Res* 2007;**6**:3501–8.

58. Olsen JV, Blagoev B, Gnad F, *et al*. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;**127**:635–48.
59. Olsen JV, Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA* 2004;**101**:13417–22.
60. Beausoleil SA, Villen J, Gerber SA, *et al*. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 2006;**24**:1285–92.
61. Smith JC, Duchesne MA, Tozzi P, *et al*. A differential phosphoproteomic analysis of retinoic acid-treated P19 cells. *J Proteome Res* 2007;**6**:3174–86.
62. Molina H, Horn DM, Tang N, *et al*. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA* 2007;**104**:2199–204.
63. Wang Y, Ding SJ, Wang W, *et al*. Profiling signaling polarity in chemotactic cells. *Proc Natl Acad Sci USA* 2007;**104**:8328–33.
64. Wilson-Grady JT, Villen J, Gygi SP. Phosphoproteome analysis of fission yeast. *J Proteome Res* 2008;**7**:1088–97.
65. Zhai B, Villen J, Beausoleil SA, *et al*. Phosphoproteome analysis of Drosophila melanogaster embryos. *J Proteome Res* 2008;**7**:1675–82.
66. Dephoure N, Zhou C, Villen J, *et al*. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci USA* 2008;**105**:10762–7.
67. Yang F, Stenoien DL, Strittmatter EF, *et al*. Phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation. *J Proteome Res* 2006;**5**:1252–60.
68. Gruhler A, Olsen JV, Mohammed S, *et al*. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005;**4**:310–27.
69. Manning BD, Cantley LC. Hitting the target: emerging technologies in the search for kinase substrates. *Sci STKE* 2002;**2002**:PE49.
70. Schutkowski M, Reineke U, Reimer U. Peptide arrays for kinase profiling. *Chembiochem* 2005;**6**:513–21.
71. Dammeier S, Lovric J, Eulitz M, *et al*. Identification of the smooth muscle-specific protein, sm22, as a novel protein kinase C substrate using two-dimensional gel electrophoresis and mass spectrometry. *Electrophoresis* 2000;**21**:2443–53.
72. Knebel A, Morrice N, Cohen P. A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *EMBO J* 2001;**20**:4360–9.
73. Preisinger C, Short B, De Corte V, *et al*. YSK1 is activated by the Golgi matrix protein GM130 and plays a role in cell migration through its substrate 14-3-3zeta. *J Cell Biol* 2004;**164**:1009–20.
74. Linding R, Jensen LJ, Pasculescu A, *et al*. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* 2008;**36**:D695–9.
75. Elia AE, Cantley LC, Yaffe MB. Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* 2003;**299**:1228–31.
76. Hutti JE, Jarrell ET, Chang JD, *et al*. A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods* 2004;**1**:27–9.
77. Grant SG, Karl KA, Kiebler MA, *et al*. Focal adhesion kinase in the brain: novel subcellular localization and specific regulation by Fyn tyrosine kinase in mutant mice. *Genes Dev* 1995;**9**:1909–21.
78. Sachs K, Perez O, Pe'er D, *et al*. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005;**308**:523–9.
79. MacKeigan JP, Murphy LO, Blenis J. Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nat Cell Biol* 2005;**7**:591–600.
80. Zhu H, Bilgin M, Bangham R, *et al*. Global analysis of protein activities using proteome chips. *Science* 2001;**293**:2101–5.
81. Tao SC, Chen CS, Zhu H. Applications of protein microarray technology. *Comb Chem High Throughput Screen* 2007;**10**:706–18.
82. Popescu SC, Popescu GV, Bachan S, *et al*. MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays. *Genes Dev* 2009;**23**:80–92.
83. Lu JY, Lin YY, Qian J, *et al*. Functional dissection of a HECT ubiquitin E3 ligase. *Mol Cell Proteomics* 2008;**7**:35–45.
84. Lin YY, Lu JY, Zhang J, *et al*. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell* 2009;**136**:1073–84.
85. Chen CS, Korobkova E, Chen H, *et al*. A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nat Methods* 2008;**5**:69–74.
86. Kafadar KA, Zhu H, Snyder M, *et al*. Negative regulation of calcineurin signaling by Hrr25p, a yeast homolog of casein kinase I. *Genes Dev* 2003;**17**:2698–708.
87. Feilner T, Hultschig C, Lee J, *et al*. High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Mol Cell Proteomics* 2005;**4**:1558–68.
88. Xue Y, Ren J, Gao X, *et al*. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics* 2008;**7**:1598–608.
89. Schwartz D, Chou MF, Church GM. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol Cell Proteomics* 2009;**8**:365–79.
90. Budovskaya YV, Stephan JS, Deminoff SJ, *et al*. An evolutionary proteomics approach identifies substrates of the cAMP-dependent protein kinase. *Proc Natl Acad Sci USA* 2005;**102**:13933–8.
91. Linding R, Jensen LJ, Ostheimer GJ, *et al*. Systematic discovery of in vivo phosphorylation networks. *Cell* 2007;**129**:1415–26.
92. Pawson T, Scott JD. Protein phosphorylation in signaling—50 years and counting. *Trends Biochem Sci* 2005;**30**:286–90.
93. Cui Q, Ma Y, Jaramillo M, *et al*. A map of human cancer signaling. *Mol Syst Biol* 2007;**3**:152.
94. Diella F, Cameron S, Gemund C, *et al*. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* 2004;**5**:79.
95. Diella F, Gould CM, Chica C, *et al*. Phospho.ELM: a database of phosphorylation sites—update 2008. *Nucleic Acids Res* 2008;**36**:D240–4.
96. Hornbeck PV, Chabra I, Kornhauser JM, *et al*. PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 2004;**4**:1551–61.

97. Keshava Prasad TS, Goel R, Kandasamy K, *et al*. Human Protein Reference Database—2009 update. *Nucleic Acids Res* 2009;**37**:D767–72.

98. Miller ML, Jensen LJ, Diella F, *et al*. Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* 2008;**1**:ra2.

99. Yaffe MB, Leparc GG, Lai J, *et al*. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* 2001;**19**:348–53.

100. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* 2003;**31**:3635–41.

101. Gnad F, Ren S, Cox J, *et al*. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* 2007;**8**:R250.

102. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999;**294**:1351–62.

103. Blom N, Sicheritz-Ponten T, Gupta R, *et al*. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 2004;**4**: 1633–49.

104. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005;**23**:1391–8.

105. Ubersax JA, Ferrell JE, Jr. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* 2007;**8**: 530–41.

106. Munton RP, Tweedie-Cullen R, Livingstone-Zatchej M, *et al*. Qualitative and quantitative analyses of protein phosphorylation in naive and stimulated mouse synaptosomal preparations. *Mol Cell Proteomics* 2007;**6**:283–93.

107. Ong SE, Blagoev B, Kratchmarova I, *et al*. 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.