Phosphoproteomic approaches to elucidate cellular signaling networks
Katrin Schmelzle and Forest M White

Protein phosphorylation is crucial in the regulation of signaling pathways that control various biological responses. Recent progress in diverse methodologies to investigate protein phosphorylation in complex biological samples has resulted in more rapid, detailed and quantitative analyses of signaling networks. In particular, advances in mass spectrometry (MS) have enabled the identification and quantification of thousands of both known and novel phosphorylation sites. Initial MS-based information can be complemented with a variety of recently developed and improved phosphoproteomic techniques. These include multiplexed microbead or kinase activity assays, flow cytometry based single-cell analysis, protein microarrays and interaction studies. The combination of multiple approaches, coupled with phenotypic response measurements, computational modeling and biochemical manipulations, will ultimately reveal the mechanistic regulation of signaling networks.

Addresses
Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: White, Forest M (fwhite@mit.edu)

Introduction
Phosphorylation of proteins regulates multiple effects including translocation, protein–protein interactions, and activation or inactivation. This post-translational modification is a crucial regulator of cell-signaling pathways; dysregulated phosphorylation has been implicated in various diseases, including cancer [1]. In many cases, a protein can be phosphorylated on multiple sites, which can either act independently or synergistically when phosphorylated simultaneously. It is therefore important to investigate the level of phosphorylation for individual sites on a given protein in addition to looking at the overall level of protein phosphorylation. When considering that a high percentage of the proteome is phosphorylated, a very complex picture of the status of a cell dependent on the level of phosphorylation is revealed. Often, phosphorylation occurs at low stoichiometry or can occur on a protein with low expression level; both conditions represent a challenge for the detection of the phosphorylation site and could be determining factors limiting the number of phosphorylation sites identified and investigated so far. In the past few years, several advances in mass spectrometry (MS) based approaches have enabled the analysis of thousands of phosphorylation sites. Therefore, MS combined with enrichment strategies for phosphorylated proteins and peptides is the tool of choice for the identification of novel phosphorylation sites. Significant developments have occurred not only in MS, however, but also in various other phosphoproteomic technologies. Thus, the phosphoproteomics field is entering a new era. Here, we will discuss a selection of different approaches (Table 1) developed predominantly in the past two years. The goal of this review is to give a concise overview of current approaches in the field of phosphoproteomics and to show how a combination of several approaches can be used to obtain a more comprehensive understanding of a given signaling pathway.

MS-based approaches
Currently, the most powerful tool to interrogate the phosphoproteome is enrichment for phosphopeptides followed by reverse-phase liquid chromatography combined with tandem mass spectrometry (LC-MS/MS). When sample preparation and instrumentation are chosen appropriately, thousands of phosphorylation sites can be identified. In the following section, we will focus on different enrichment approaches that have been successfully coupled to LC-MS/MS for the identification of many phosphorylation sites (Figure 1).

Enrichment strategies
We focus here on the two most commonly used methods of enrichment, immobilized metal affinity chromatography (IMAC) and strong cation exchange (SCX), which have been applied separately or in combination in recent studies. IMAC enrichment is based on phosphate affinity for the immobilized metal chelated to the resin. When used in combination with chemical modification of the carboxylic groups (esterification) to prevent nonspecific binding, this technique is highly selective for phosphorylated peptides [2–4]. As an alternative to IMAC, titanium dioxide based enrichment should require significantly less column preparation time, but this technique is still in development and has not yet been applied to the analysis of biological samples [5]. SCX, when performed at low pH, separates phosphorylated from non-phosphorylated tryptic peptides on the basis of the charge difference associated with the
negatively charged phosphate group. This enrichment technique has been used to fractionate samples either alone or in combination with IMAC prior to LC-MS/MS analysis [6**,7,8**,15]. Although SCX can be used to enrich for phosphorylated peptides, this method tends to be less selective than the IMAC enrichment of esterified phosphorylated peptides.

**Global versus targeted MS approaches**

Global approaches aimed at analyzing the largest possible proportion of the phosphorylation sites (phosphoserine, phosphothreonine and phosphotyrosine) using whole-cell lysate or subcellular fractions have been widely applied to different systems, leading to the identification of hundreds or even thousands of phosphorylation sites [2–4,6**,7,8**,9–13]. Even though these numbers are impressive, the phosphorylation sites detected are still primarily abundant sites, as can be seen, for example, in the low numbers of tyrosine phosphorylated peptides identified with such approaches. This result is not surprising given the low relative level of tyrosine phosphorylation, which comprises less than 0.1% of phosphorylation in the cell [14]. To access tyrosine phosphorylation in complex mixtures, targeted approaches using immunoprecipitation (IP) with antiphosphotyrosine antibodies have been employed. These approaches can be applied at the protein level, resulting mainly in information describing total protein phosphorylation levels [15,16–19], or at the peptide level, leading to site-specific information including the identification of hundreds of sites [20**,21]. In order to increase the detection of tyrosine phosphorylation sites following antiphosphotyrosine protein IP, further

---

Table 1

<table>
<thead>
<tr>
<th>Phosphoproteomic aspect</th>
<th>Method</th>
<th>Representative example</th>
<th>Phosphorylation sites studied</th>
<th>Specific information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection and semiquantitative analysis of novel and known phosphorylation sites</td>
<td>MS combined with enrichment steps</td>
<td>(see Figure 1)</td>
<td>Dozens to 1000s</td>
<td>Identification of novel sites, many sites identified in parallel</td>
</tr>
<tr>
<td>Semiquantitative, MS-free targeted assays for phosphorylation sites</td>
<td>Western blotting</td>
<td>Standard technology</td>
<td>1 at a time</td>
<td>Can be applied to any site for which an antibody is available</td>
</tr>
<tr>
<td></td>
<td>Multiplex western blotting</td>
<td>Human breast tissue and breast cancer cell lines</td>
<td>12</td>
<td>Multiplex approach</td>
</tr>
<tr>
<td></td>
<td>Reverse-phase protein microarray</td>
<td>Liver tissue, with and without metastasis [35]</td>
<td>24</td>
<td>Chip format higher throughput than western blotting</td>
</tr>
<tr>
<td></td>
<td>Flow cytometry</td>
<td>Human primary naïve CD4+ cells, various treatments [41**]</td>
<td>9 + 2 phospholipids</td>
<td>Analysis of phosphorylation dynamics in individual cells</td>
</tr>
<tr>
<td></td>
<td>Multiplex microbead suspension array</td>
<td>Jurkat cells treated with anti-CD3 antibody or sodium pervanadate [38*]</td>
<td>10</td>
<td>Solution-phase analysis, higher throughput than western blotting</td>
</tr>
<tr>
<td>Phosphoprotein interactions</td>
<td>Protein microarrays</td>
<td>Interactions of phosphorylated peptides from ErbB receptor with SH2 and PTB domains [43*]</td>
<td>61</td>
<td>In vitro method, binary, monitors direct interactions</td>
</tr>
<tr>
<td></td>
<td>Multiplex microbead suspension array</td>
<td>EGF- and insulin-treated HeLa cells [44*]</td>
<td>Used to assess the direct or indirect interaction of selected proteins with various domains (25 SH2 domains in the example)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Immunoprecipitation or affinity purification and semiquantitative MS for interacting proteins</td>
<td>EGF-treated HeLa cells, Grb2-SH2 domain [45]</td>
<td>Identifies proteins interacting with a specific protein or domain</td>
<td></td>
</tr>
<tr>
<td>Kinase activity</td>
<td>Kinase assay</td>
<td>Insulin- and EGF-treated HT29 cells [47*]</td>
<td>Can be applied to the study of specific kinases (3 kinases in the example)</td>
<td></td>
</tr>
<tr>
<td>Kinase substrates</td>
<td>Proteome chip</td>
<td>Yeast proteome [48*]</td>
<td>Only provides protein-level information rather than information on phosphorylation sites (substances of 82 kinases in the example)</td>
<td></td>
</tr>
</tbody>
</table>

EGF, epidermal growth factor; Grb2, growth factor receptor-bound protein 2; PTB, phosphotyrosine binding; SH2, Src homology 2.
affinity-based enrichment for phosphorylated peptides [20**,22,23], dendrimer conjugation chemistry [24] or the use of specific marker ions (e.g. phosphotyrosine immonium ion with mass-to-charge ratio 216.04) during MS analysis [25] have been used. As a complementary, even more targeted, approach individual proteins of interest have been isolated and phosphorylation sites mapped by MS [26–28]).
Phosphoproteomics in cellular signaling

Cataloging versus relative quantification

Most efforts in MS-based phosphoproteomic analyses have focused on the identification of phosphorylation sites (cataloging), and only recently has the focus started to switch to more quantitative studies [6**,15**,19,20**,22,23]. In brief, relative quantification is typically performed by stable-isotope labeling of different samples with different isotopes, followed by analysis of the mixture and comparison of intensities. The most common labeling strategies for phosphoproteomics are stable isotope labeling with amino acids in cell culture (SILAC), eferistation with isotopically enriched alcohols (e.g. methanol or ethanol), and use of the isobaric reagent iTRAQ from Applied Biosystems. SILAC is restricted to cultured cells, provides analysis of up to three samples in parallel, and generates quantitative information from the full scan mass spectrum. Quantification occurs at the MS level with eferistation, but labeling is introduced post-lysis and is typically limited to two samples analyzed in parallel. Because the iTRAQ reagent is an amine-reactive probe with four isobaric isoforms, labeling is introduced post-lysis, up to four samples can be analyzed in parallel, and quantification occurs in the MS/MS spectrum, through comparison of marker fragment ions. As a more detailed analysis of quantification strategies is beyond the scope of this review, interested readers are referred to a recent review by Ong and Mann [29].

Which MS-based method should be chosen?

Not surprisingly, the MS method of choice depends on the sample and the proteins of interest. As mentioned above, IMAC or SCX enrichment could be suitable for global analysis of the most abundant phosphorylation sites, whereas for a specific question a more targeted approach will often lead to more informative results. In general, sample fractionation followed by analysis of the individual fractions results in the identification of more sites than a single, unfractinilated analysis. Unfortunately, sample fractionation also typically leads to decreased sample yield and increased analysis time. The required depth of coverage should be considered in choosing the method.

In addition to sample preparation and enrichment, MS instrumentation also plays an essential role. Following collision-activated dissociation (CAD), phosphoserine and phosphothreonine residues tend to exhibit neutral loss of phosphoric acid, often resulting in poor quality MS/MS spectra that can inhibit identification of the peptide sequence. In the past few years, the best results have been achieved either with a quadrupole time-of-flight instrument to minimize intensity of the ion corresponding to neutral loss of phosphoric acid or with the linear ion trap instrument, performing MS/MS/MS on the neutral loss peak to increase the number of identifications [6**,8*]. Most recently, electron transfer dissociation (ETD) technology [30,31] has been demonstrated as a very promising alternative to CAD because of higher sequence coverage combined with phosphate retention in MS/MS mode, enabling more facile identification of peptide sequences and phosphorylation sites.

Antibody-based approaches (non-MS)

To monitor previously identified phosphorylation sites, the combination of phosphospecific antibodies and western blotting has been the gold standard. However, until recently the limited throughput of this approach, with only one phosphorylation site investigated at a time, has driven the development of other, high-throughput approaches. Arrays using phosphospecific antibodies to investigate phosphorylation sites have been evaluated [32,33] and used to interrogate dozens of phosphorylation sites simultaneously [34*,35]. As this technology requires antibodies with high-affinity and specificity, currently only a limited number of phosphorylation sites can be analyzed [36]. However, further development might lead to an even broader application of microarray technology for phosphoprotein studies. In a slight twist on this general concept, Gembitsky et al. [37] developed a tyrosine phosphorylation microarray to investigate the overall tyrosine phosphorylation of dozens of proteins rather than individual sites. An alternate approach, multiplexed microbead suspension array (developed by LumineX), was employed to investigate the kinetics of ten phosphorylation events in parallel and could, theoretically, be extended to simultaneous analysis of 100 phosphorylation sites, if highly specific, high-affinity antibodies were available [38*]. In addition, the combination of phosphospecific antibodies and western blotting has been recently multiplexed and used, for example, to investigate twelve phosphorylation sites in parallel in different breast and breast cancer cell lines [39].

One of the fundamental limitations of MS or the above-mentioned antibody-based approaches is the need to look at cell lysate obtained from populations of cells and the corresponding lack of cellular resolution. To overcome this issue, a recently developed flow cytometry based methodology has enabled the analysis of phosphorylation dynamics in individual cells and can distinguish signaling changes between subpopulations of cells present in numbers as low as 10,000. This approach has been used to investigate up to 11 phosphorylation events in parallel [40,41**,42]. Currently, the main limitation to this methodology is the development of phosphospecific, fluorescence-labeled antibodies compatible with flow cytometry.

Interactions of phosphoproteins and phosphorylated sites

The identification or quantification of protein phosphorylation can inform as to the activity of protein kinases and phosphatases, but cellular signaling networks are also regulated through the recruitment of selected proteins to specific

---

phosphorylation sites. A recent study by Jones et al. [43*] used protein microarrays to quantify interaction affinities between 61 fluorescence-labeled, tyrosine-phosphorylated peptides from ErbB receptors with approximately 150 SH2 (Src homology 2) and PTB (phosphotyrosine binding) domains. By quantifying fluorescence levels at different titrations, \( K_D \) values were determined experimentally for each peptide–domain interaction. Although only binary interactions were investigated and subcellular localization was not considered, this method can indicate many novel potential interactions of phosphorylated proteins.

Alternate approaches to identifying protein–protein interactions in the cell-signaling network have employed SH2 domains to extract interacting proteins and phosphoproteins from complex mixtures. Studies have been performed either in multiplex format, with 25 SH2 domains bound to Luminex beads [44*], or individually, through MS-based analysis of proteins interacting with the SH2 domain of Grb2 [45]. Regardless of format, each of these techniques has provided novel insight into the connectivity of the cellular signaling network.

**Kinase-related approaches**

*In vitro* kinase assays have long been used to establish potential kinase–substrate affinities, but have traditionally been performed on a single kinase–substrate pair at a time. Recently, kinase activity assays have been multiplexed and used to analyze the activities of up to five kinases extracted from cells stimulated under different conditions for different times. Activity was monitored using specific optimal substrates and \([\gamma-^{32}\text{P}]\text{ATP} \) or fluorescence [47*]. In addition to multiplexing the number of kinases, it is also possible to multiplex the number of substrates being investigated. In an effort to determine all possible substrates for selected kinases, each kinase was added to a yeast proteome chip and over 4000 *in vitro* phosphorylation events were detected [48*]. Although the majority of these interactions appear to be

![Simplified scheme showing phosphorylation-related key events in a receptor tyrosine kinase signaling network.](current-opinion-in-biotechnology-figure-2.png)
novel, further validation through in vivo characterization of each kinase–substrate interaction is required.

Combined approaches for more complete insights into signaling pathways

It is already clear from the literature that novel biological insights can be obtained by combining multiple phosphoproteomic analysis methods (i.e. kinase assays, western blotting and protein microarrays) with computational modeling and phenotypic characterization such as cell death assays [49,50]. Additional mechanistic understanding of the biological regulation of cellular signaling networks might be achieved through more comprehensive characterization of a specific signaling pathway in combination with quantitative biological response data.

In designing a strategy to investigate cellular signaling networks, it is important to bear in mind that a signaling pathway normally combines many different phosphorylation, dephosphorylation and interaction events, as shown in a simplified scheme for a receptor tyrosine kinase pathway in Figure 2. It is also worth noting that, although we present a phosphorylation-centric view, other post-translational modifications will also occur and have important roles.

For a more detailed description of signaling pathways, the reader is referred to a review by Hunter [1].

A general strategy for the analysis of cellular signaling networks is schematically represented in Figure 3. As a starting point in investigating the phosphorylation within a given signaling network, a combination of MS-based approaches (e.g. global and targeted, tyrosine phosphorylation analyses) is of greatest utility, as they enable the detection and relative quantification of both known and novel sites. This information will provide a basis from which to address more in-depth analyses. For instance, the activity of certain kinases (e.g. Akt and Erk) under different conditions can be measured by kinase assays or the regulation of specific phosphorylation sites in different subpopulations of cells can be investigated by flow cytometry with a suitable antibody. Alternately, following site identification, the interaction of phosphorylation sites of interest with other specific proteins can be determined, for example, using a microbead suspension array or in a more general manner by pulling down a specific protein and identifying binding partners by MS. Known phosphorylation sites not covered by the MS approaches, but for which antibodies are available, can be investigated.
using a standard western blot approach as well as higher throughput assays including the multiplex microbead suspension array or reverse-phase protein microarrays. In addition, for specific proteins of interest, phosphorylation sites can be mapped by MS after isolation of the protein. To complement phosphoproteomic data, functional assays (e.g. measuring migration, proliferation and apoptosis) provide downstream biological response data that can be used to help define phosphorylation site function in a given biological system.

These approaches will generate a large amount of data that can subsequently be used for mathematical or computational analysis and model development [51*]. Application of computational models will generate novel hypotheses regarding signaling cascades and specific phosphorylation sites, which might be crucial for certain processes or represent potential drug targets for dysregulated pathways in diseases. These hypotheses can then be tested with biochemical approaches, including the use of kinase inhibitors, RNA interference, antibodies or site-directed mutagenesis.

Conclusions
In the past few years, new methods covering different aspects in the field of phosphoproteomics have been developed. Some are still in an early stage and further improvement as well as broader application by the research community is expected. We propose a possible combination of different approaches including MS, flow cytometry, different types of immunoreaction (e.g. western blot, microbead suspension array and protein microarrays), protein interaction and kinase activity studies. These approaches paired with phenotypic measurements, mathematical or computational modeling and biochemical manipulations might lead to a more comprehensive understanding of signaling pathways and, ultimately, to the identification of potential drug targets in the case of dysregulated pathways in disease.

Acknowledgements
We would like to thank members of the White laboratory for helpful discussions and comments on the manuscript. This work was supported by National Institutes of Health Grants DK42816, R21-AI065354-2, and P50-GM68762.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

16. Analysis of phosphotyrosine-based signaling events in the epidermal growth factor receptor (EGFR) pathway. By immunoprecipitating tyrosine-phosphorylated proteins and closely associated binders by phosphotyrosine protein IP/GE/LC-MS/MS, the temporal dynamics of more than 80 proteins in HeLa cells were determined. This is the first study to look at the temporal dynamics of tyrosine phosphorylated proteins and associated proteins in the EGFR pathway.
Phosphoproteomics in cellular signaling

Schmelzle and White


Analysis of phosphoproteins-based signaling events in the EGFR pathway. Using phosphoprotein peptide IP/IMAC/LC-MS/MS the temporal dynamics of more than 100 phosphorylation sites on more than 70 proteins were quantified. Data analysis using self-organizing maps (SOM) revealed co-regulated phosphorylation sites. This is the first study to directly monitor the temporal dynamics of individual phosphorylation sites in the EGFR pathway.


36. Shults MD, James KA, Lauffenburger DA, Imperiali B: A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. Nat Methods 2005, 2:277-283. The authors have developed and applied a fluorescence-based, sensitive kinase assay, which was applied to whole-cell lysates to investigate the activity of three kinases in parallel.

Reports a global approach to investigate the yeast phosphoproteome using proteome chip technology incubated with kinases and [γ-32P]ATP. This study resulted in more than 4000 in vitro phosphorylation events.


A recent example of a study that used multiple approaches to investigate cell-signaling networks. Multiplexed kinase activity assays, quantitative western blotting, antibody microarrays, and apoptotic response measurements were combined and analyzed using partial least squares regression models.


A recent review of computational models used to analyze data gained from proteomic experiments with regard to proteomic networks.