QUANTITATIVE PHOSPHOPROTEOMIC ANALYSIS OF SIGNALING NETWORK DYNAMICS

Forest M. White[1]*

1*Department of Biological Engineering and David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge MA 02139*

Summary

Protein phosphorylation mediated cellular signaling is a highly regulated, dynamic process that controls many aspects of cellular biology. Over the past few years many methods have been developed to quantify temporal dynamics of protein phosphorylation, including mass spectrometry, which can be applied in both an unbiased, discovery mode and in a targeted mode to monitor specific phosphorylation sites. Other methods, such as kinase activity assays and antibody microarrays, have been applied to quantify central nodes in the signaling network, yielding intriguing biological insights. This review provides a concise overview of the latest advances in quantitative analysis of signaling dynamics including a brief commentary on the future of the field.

Introduction

Almost all aspects of cell biology and physiology are regulated by protein post-translational modifications (PTMs), which regulate protein conformation, activation, degradation, sub-cellular localization, and protein-protein interactions. Perhaps the best studied PTM is reversible phosphorylation of serine, threonine and tyrosine residues. Protein phosphorylation mediated signaling networks regulate cellular response to environmental cues including mechanical stress, growth factors and cytokines, cell-cell interactions, and cell-matrix interactions. Moreover, protein phosphorylation plays a key role in regulating most cellular processes including proliferation, migration, apoptosis, gene transcription, including alternative splicing, and protein translation. Given the importance of this PTM, it is not surprising that dysregulation of protein kinases and phosphatases has been linked to a vast number of pathologies, including cancer [1], auto-immune diseases [2], metabolic disorders [3], and pathogenic infections [4].

Over the past decade, many methods have been developed with the ultimate goal of determining signaling pathways and phosphorylation events regulating normal and abnormal cellular processes. The challenges facing these phosphoproteomic methods are similar to those facing many proteomic methods: limited sample amounts, highly complex samples, and huge dynamic range. Adding to these challenges, phosphorylation site stoichiometry is usually less than 100% and can be dynamic and tightly regulated. Quantifying temporal dynamics therefore requires high sensitivity, accurate quantification, analysis of multiple time points, and sample

*Corresponding Author, Forest M. White, Dept. of Biological Engineering, MIT, Bldg. 56-787a, 77 Massachusetts Ave., Cambridge, MA 02139, 617-258-8949, fwhite@mit.edu.

preparation procedures that effectively freeze enzymatic processes to capture the physiological state of the biological sample. Data generated in these efforts will typically be very complex (e.g. many phosphorylation sites quantified at each of multiple time points across multiple biological conditions), requiring computational algorithms to decipher the signaling networks and yield biological insight [5]. Over the past two years, several recently developed methods have successfully addressed these issues and have now provided the first glimpses of phosphorylation-mediated signaling network dynamics.

Mass spectrometry based analysis of ErbB signaling dynamics

The ErbB cellular signaling network plays a central role in many biological processes and has been associated with multiple human cancer types. Over the past four years several manuscripts have attempted to define this network and quantify temporal dynamics in response to receptor stimulation. This series of studies highlights several important issues about the trajectory of discovery-mode phosphoproteomics experiments; here a short description of each method will be followed by a brief commentary on some of these issues. In the first study, stable isotope labeling of amino acids in cell culture (SILAC) was used to quantify changes in the EGFR signaling network at 5 time points following immunoprecipitation of tyrosine phosphorylated proteins [6]. Enrichment at the protein level enabled the quantification of both tyrosine phosphorylated proteins as well as associated proteins that co-precipitated with tyrosine phosphorylated proteins. However, much of the quantification was performed on non-tyrosine phosphorylated peptides, and only a few phosphorylation sites were identified. In the second manuscript, iTRAQ (isobaric tag for relative and absolute quantification)-labeled tyrosine phosphorylated peptides were immunoprecipitated from EGF-stimulated samples at four time points, providing site-specific identification and quantification of temporal dynamics for over one hundred tyrosine sites within the EGFR network [7]. By using a pan-specific antibody to perform unbiased enrichment for tyrosine phosphorylated peptides, this study was now able to quantify temporal dynamics of many novel tyrosine phosphorylation sites. This enrichment technique was subsequently applied to quantify the effects of increased HER2 (ErbB2) expression in the context of EGF and HRG stimulation [8]. To gain functional insight, dynamic phosphoproteomic data was compared to cell phenotypic data to determine tyrosine phosphorylation sites that were most strongly correlated to migration and proliferation. Although putative linkages were provided for many tyrosine phosphorylation sites in this study, serine and threonine sites were absent unless they occurred on a tyrosine phosphorylated peptide. To obtain a more global perspective on ErbB signaling dynamics, a recent study eliminated the immunoprecipitation step altogether and instead employed a combination of strong cation exchange (SCX) and TiO_2 (titanium dioxide) to enrich for phosphorylated peptides. This strategy was applied to SILAC-encoded cells stimulated with EGF at five time points, and resulted in identification of >6000 phosphorylation sites, including over 1000 sites whose phosphorylation was modulated by EGF stimulation [9]. A subcellular fractionation step was included to quantify dynamic changes in phosphorylation in nuclear vs. cytosolic compartments at each time point, since spatial information is critical for many aspects of cellular signaling.

Biological insight from mass spectrometry-based phosphoproteomics

Together, these four manuscripts provide an overview of the progression of the field of phosphoproteomics, and perhaps of the proteomics field as a whole. As expected, each subsequent effort has resulted in a significant increase in the size of the phosphoproteomics data set, a trend that is driving continued methodological and instrument development. However, it is important to consider the advantages and disadvantages, summarized in Table 1, of each of these methods, as there may be a "sweet spot" in which biological insight may be maximized. For instance, a very large amount of mass spectrometric analyses (117 LC-MS/

MS analyses) were required to obtain the massive data set in the Olsen *et al.* manuscript, a level of effort that is not compatible with multiple biological replicates across multiple biological conditions. As an alternative to very large scale data collection for a single perturbation condition, it may be beneficial to target a subset of the phosphoproteome and monitor temporal dynamics across multiple biological conditions, as in the Wolf-Yadlin *et al.* manuscript. Coupling this set of phosphoproteomics data to quantitative phenotypic measurements also insures that the perturbation and corresponding modulation of the signaling network had the desired phenotypic effect, thereby providing a critical check on the biological relevance of the stimulation conditions. This "targeted discovery" approach can be extended to serine/threonine phosphorylation sites by performing a peptide immunoprecipitation with either phospho-specific antibodies [10] or motif-specific antibodies [11], as has been demonstrated recently with ATM/ATR substrates in the DNA damage response.

The need for validation of mass spectrometry data

One concern with the trend toward massive phosphoproteomic data sets is the inability to properly validate MS/MS spectral assignments for each phosphorylated peptide. Performed properly, spectral validation requires the assignment of all abundant ions in the spectrum, necessitating a significant time commitment. To bypass this bottleneck, many mass spectrometry labs have utilized decoy search strategies and statistical methods to estimate false-positive identification rates [12]. Statistical validation is a greater concern for phosphoproteomics as compared to the larger field of proteomics, since each phosphorylation site is typically defined by a single MS/MS spectrum. Unfortunately, using statistical methods it is impossible to tell if any given phosphorylation site is correctly identified, as the MS/MS spectrum remains unvalidated. False positive identifications are particularly dangerous for biologists interested in studying the function of these selected phosphorylation sites, as each phosphorylation site may take 1–2 years to fully investigate. This situation is exacerbated because most biologists do not have the requisite expertise to assess the accuracy of the assignment even if the raw MS/MS spectrum is provided, and it is often mistakenly assumed that all published phosphorylation assignments are correct.

Protein vs. peptide analysis

From these same set of four manuscripts another interesting trend can be observed. Over the past decade, mass spectrometry-based phosphoproteomics has shifted from a protein-centric focus, in which spots on 2D-gels were isolated and characterized or thin-layer chromatography was used to detect phosphorylated peptides resulting from single protein digests, to a peptide-centric focus, driven by the desire to obtain larger data sets with more phosphorylation sites. This shift in the field is highlighted by comparing the Blagoev *et al.* and Zhang *et al.* manuscripts [6,7], where the cost of obtaining site-specific phosphorylation information was the failure to detect interacting proteins. In order to understand dynamic regulation of phosphorylation mediated signaling networks it will be necessary to shift our attention back to a more protein centric strategy, to uncover the regulatory pattern on individual proteins. For instance, of the approximately 20 potential phosphorylation sites on EGFR, how many can be phosphorylated simultaneously? Which sites exhibit positive or negative cooperativity [13], and how does the combination of different sites then affect the downstream signaling network? In the near future, one of the most significant challenges will be to understand the dynamics of combinatorial signaling for many proteins across the network. Although difficult, especially given the size of the proteins and number of modifications, this task should be feasible, as indicated in Figure 1, by combining lessons learned from top-down analysis of protein post-translational modifications (e.g. histone modifications [14]) with mass spectrometry based phosphoproteomics techniques, including electron transfer dissociation (ETD) [15,16].

Targeted analysis of cellular signaling dynamics

Although mass spectrometry-based phosphoproteomics has the potential to uncover novel components and provide unprecedented coverage of signaling networks, much of the success in quantifying phosphorylation-mediated cellular signaling dynamics over the past few years has come from targeted approaches, including flow cytometry, kinase activity assays, western blots, antibody microarrays, targeted mass spectrometry, and ELISAs. These techniques typically have much greater throughput and improved sensitivity relative to discovery-mode mass spectrometry, and therefore enable analysis of much smaller sample sizes at many more time points across multiple biological conditions. Of course, the major limitation of these approaches is the inability to detect novel phosphorylation sites, and quantification is therefore limited to pre-selected components within the signaling network. However, by properly choosing the nodes to be analyzed, very interesting insights can be achieved through quantification of a relatively small number of proteins within the network.

Dynamics in the apoptotic signaling network

To assess the role of differential cues in activating apoptosis, phospho-specific western blots were combined with antibody microarrays and kinase activity assays to quantify the state of 11 central nodes in the apoptosis signaling network at 13 time points in response to multiple individual or combined doses of EGF, Insulin, or TNF [17]. This quantitative, dynamic phosphoproteomics dataset was then complemented with multiple quantitative phenotypic readouts for apoptotic response under these same stimulation conditions. Computational analysis of these data sets highlighted the role of select kinases (e.g. IKK, JNK1) at early vs. late response times in regulating the cellular apoptotic response. Interestingly, even though only 11 nodes were quantified, most of the dynamic measurements could be discarded without adversely affecting the predictive power of the computational model, indicating that measuring even fewer nodes may still suffice to give relevant biological insight. This concept has recently been extended to quantify the role of five key kinases in regulating apoptosis in response to adenovirus [18]. Intriguingly, quantification of any single kinase activation state could not predict response across multiple cell types, since each kinase may be differentially activated in each cell type. However, by quantifying the activity of all five kinases together, it was possible to generate a model in one cell line and apply it to correctly predict response in two other cell lines, indicating the presence of common effectors to process the information and program the proper cell response.

Targeted analysis by mass spectrometry

Most of the mass spectrometry based phosphoproteomics studies to date have taken an unbiased, "discovery-mode" approach in order to identify, and in some cases quantify, protein phosphorylation sites within particular signaling networks. Although these studies have often discovered novel phosphorylation sites, the reproducibility of these analyses has been poor. In fact, replicate analyses typically yield only 50–70% reproducibility at the peptide or protein level [19]. To address this concern, a targeted multiple reaction monitoring (MRM)-based analysis has recently been applied to quantify temporal dynamics in the EGFR signaling network [20]. In this method, cells were stimulated with EGF for 7 different time points and analyzed by anti-phosphotyrosine peptide immunoprecipitation followed by IMAC-LC-MS/MS. The initial analyses were performed in discovery mode to identify nodes within the network, but subsequent, targeted analyses were performed using MRM to specifically quantify 226 phosphorylation sites within the network, with high reproducibility (~90% across 4 replicate analyses). This MRM method can now be applied to quantify changes in these specific phosphorylation sites across many biological perturbations, providing a higher-throughput approach to network-level analysis of cellular signaling.

Single cell phosphorylation profiling by FACS

Targeted analysis of signaling networks has also been demonstrated for signaling studies of single primary human lymphoma and leukemia cells [21,22].In this approach, cells were fixed and permeabilized prior to incubation with fluorophore-labeled phospho-specific antibodies. Multi-color fluorescence detection by FACS (fluorescence-activated cell sorting) enabled quantification of multiple phosphorylation sites simultaneously in thousands of single cells. This technique has now been used to stratify acute myeloid leukemia (AML) patient response to chemotherapeutic agents [22], and when combined with statistical modeling, has enabled mapping of signaling network connectivity [23].

Validation of targeted methods

As with MS-based phosphoproteomics, these targeted approaches also require extensive validation, often prior to application of the methodology. For instance, the multiplex kinase activity platform used in the above experiments was validated for dynamic range, linear response, and specificity prior to implementation in the apoptotic response studies [24]. For antibody-based approaches, extensive validation of antibodies is required prior to using them to quantify phosphorylation state of given proteins, since the vast majority of phospho-specific antibodies may have a significant amount of non-specific binding [10,25]. Proving the need for extensive validation, in a recent comparison of western blots and antibody microarrays, only 4 of 63 phospho-specific antibodies gave a single band on the western blot and gave data similar in trend and magnitude between the two methods [25]. Empirical evidence has demonstrated that many commercially available phospho-specific antibodies do not give a single band on a western blot, automatically invalidating them for use in antibody microarray or phospho-FACS formats. Luminex and ELISA approaches may be less vulnerable to non-specific antibody binding if there is a sandwich of 2 antibodies that both recognize the protein or peptide.

Conclusions

With either unbiased, discovery mode or targeted methods, it has now become possible to generate a systems-level view of quantitative protein phosphorylation dynamics within selected signaling networks. The combination of these techniques should provide unprecedented breadth and depth of coverage in signaling networks, while enabling higher throughput analyses of selected nodes under a multitude of conditions. However, obtaining biological insight from this data is not obvious, and phosphoproteomics data by itself is not sufficient to understand the mechanisms by which cells interpret multiple environmental perturbations to achieve the proper cellular response. Minimally, as indicated in Figure 2, it is necessary to collect phosphoproteomics data under multiple conditions and correlate this information to cellular phenotypic data quantified under the same conditions. In the near future, it will be necessary to extend these studies to obtain higher resolution data, both temporally and spatially, as phosphorylation changes may occur very rapidly [26] and spatial localization can change functional consequence. Although spatial localization is a critical component in regulating cellular signaling, it is exceedingly difficult to perform subcellular fractionation while eliminating enzymatic activity, including kinases, phosphatases, and proteases; improved methodology therefore needs to be developed.

In the more distant future, it will be necessary to query for additional protein modifications (e.g. glycosylation, acetylation, ubiquitination, nitrosylation) in addition to protein phosphorylation, as the signaling networks are not solely regulated by phosphorylation. Of course, the complexity of the network will increase greatly when these additional modifications are considered, but a thorough understanding of the biological system will require this level of complexity. Ideally, the methods that have been recently developed for dynamic protein

phosphorylation analysis will be applicable to other PTMs, enabling rapid progression toward these future goals.

Acknowledgements

References

• 1. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature 2001;411:355–365. [PubMed: 11357143]

2. Gatzka M, Walsh CM. Apoptotic signal transduction and T cell tolerance. Autoimmunity 2007;40:442–452. [PubMed: 17729038]

• 3. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signaling pathways: insights into insulin action. Nat Rev Mol Cell Biol 2006;7:85–96. [PubMed: 16493415]

4. Sirard JC, Vignal C, Dessein R, Chamaillard M. Nod-like receptors: cytosolic watchdogs for immunity against pathogens. PLoS Pathog 2007;3:e152. [PubMed: 18166077]

• 5. Janes KA, Lauffenburger DA. A biological approach to computational models of proteomic networks. Curr Opin Chem Biol 2006;10:73–80. [PubMed: 16406679]

• 6. Blagoev B, Ong SE, Kratchmarova I, Mann M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 2004;22:1139–1145. [PubMed: 15314609]

• 7. 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;4:1240–1250. [PubMed: 15951569]

•• 8. Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim HD, Grantcharova V, Lauffenburger DA, White FM. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Mol Syst Biol 2006;2:54. [PubMed: 17016520]The only manuscript thus far to correlate quantitative mass-spectrometry-based phosphoproteomics with quantitative phenotypic profiling. ErbB signaling is quantified at 4 time points in each of 4 different stimulation conditions

•• 9. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006;127:635–648. [PubMed: 17081983]Currently the most extensive quantitative phosphoproteomic study by MS. This paper uses SCX/TiO$_2$/LC-MS/MS to quantify temporal dynamics at 5 timepoints in the ErbB signaling network the global yeast pheromone response, identifying more than 700 phosphopeptides

•• 10. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 2007;316:1160–1166. [PubMed: 17525332]Uses non-specific binding of phospho-specific antibodies to immunoprecipitate phosphorylated peptides in the DNA damage network.

11. Stokes MP, Rush J, Macneill J, Ren JM, Sprott K, Nardone J, Yang V, Beausoleil SA, Gygi SP, Livingstone M, et al. Profiling of UV-induced ATM/ATR signaling pathways. Proc Natl Acad Sci U S A 2007;104:19855–19860. [PubMed: 18077418]

12. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods 2005;2:667–675. [PubMed: 16118637]

• 13. Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, Furnari FB, White FM. Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. Proc Natl Acad Sci U S A 2007;104:12867–12872. [PubMed: 17646646]

14. Pesavento JJ, Bullock CR, Leduc RD, Mizzen CA, Kelleher NL. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. J Biol Chem. 2008

15. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A 2007;104:2193–2198. [PubMed: 17287358]

• 16. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 2004;101:9528–9533. [PubMed: 15210983]

•• 17. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science 2005;310:1646–1653. [PubMed: 16339439]Targeted phosphoproteomics at multiple timepoints and multiple stimulation conditions to uncover key regulatory nodes in the apoptotic signaling network. Combination of large-scale data acquisition and computational modeling.

• 18. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. Common effector processing mediates cell-specific responses to stimuli. Nature 2007;448:604–608. [PubMed: 17637676]

19. Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM. Temporal dynamics of tyrosine phosphorylation in insulin signaling. Diabetes 2006;55:2171–2179. [PubMed: 16873679]

• 20. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proc Natl Acad Sci U S A 2007;104:5860–5865. [PubMed: 17389395]

21. Irish JM, Czerwinski DK, Nolan GP, Levy R. Altered B-cell receptor signaling kinetics distinguish human follicular lymphoma B cells from tumor-infiltrating nonmalignant B cells. Blood 2006;108:3135–3142. [PubMed: 16835385]

•• 22. Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud O, Gjertsen BT, Nolan GP. Single cell profiling of potentiated phospho-protein networks in cancer cells. Cell 2004;118:217–228. [PubMed: 15260991]Application of phospho-FACS to analyze cytokine stimulation response in cells obtained from AML patients.

• 23. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science 2005;308:523–529. [PubMed: 15845847]

24. Janes KA, Albeck JG, Peng LX, Sorger PK, Lauffenburger DA, Yaffe MB. A High-throughput Quantitative Multiplex Kinase Assay for Monitoring Information Flow in Signaling Networks: Application to Sepsis-Apoptosis. Mol Cell Proteomics 2003;2:463–473. [PubMed: 12832460]

• 25. Sevecka M, MacBeath G. State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling. Nat Methods 2006;3:825–831. [PubMed: 16990815]

26. Dengjel J, Akimov V, Olsen JV, Bunkenborg J, Mann M, Blagoev B, Andersen JS. Quantitative proteomic assessment of very early cellular signaling events. Nat Biotechnol 2007;25:566–568. [PubMed: 17450129]

Current peptide-centric method:
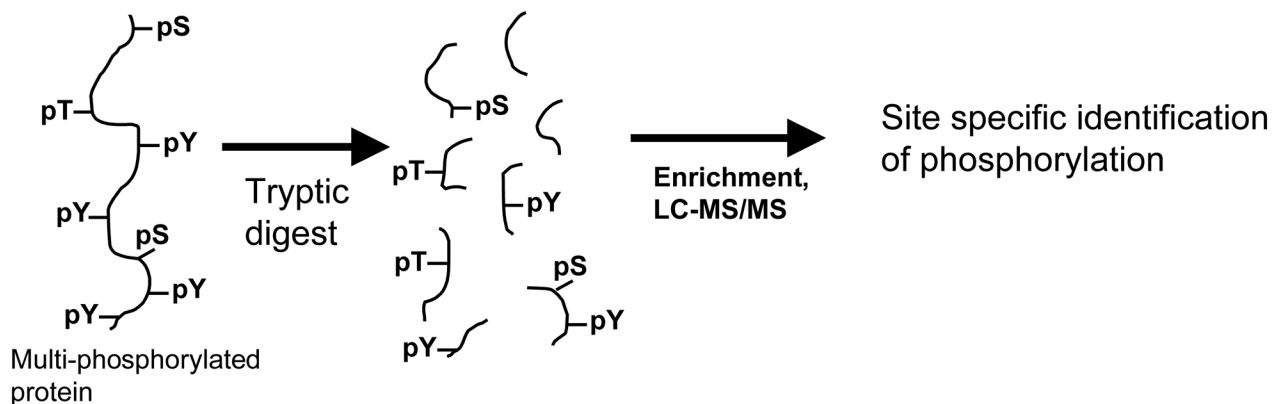
Potential protein-centric method:
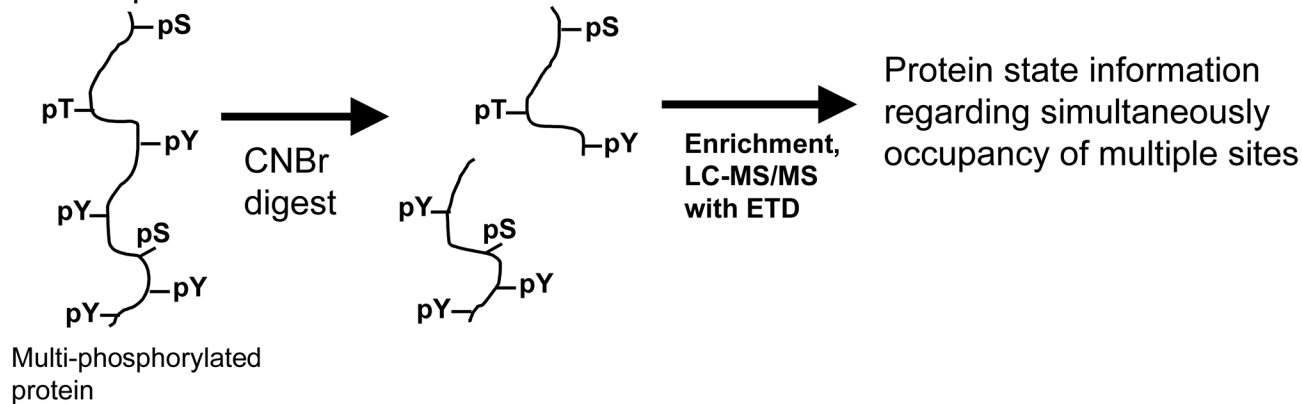
Figure 1.
Current peptide-centric mass spectrometry-based phosphoproteomics methods digest proteins to peptides, resulting in loss of protein state information, although phosphorylation-site specific quantification is possible. In the future, in order to understand how coordinated phosphorylation affects protein function and signaling, it will be necessary to analyze either intact proteins or very large peptides (similar to those that could be generated with a CNBr cleavage). These proteins would then be analyzed by LC-MS/MS with ETD to provide sequence and phosphorylation site occupancy.
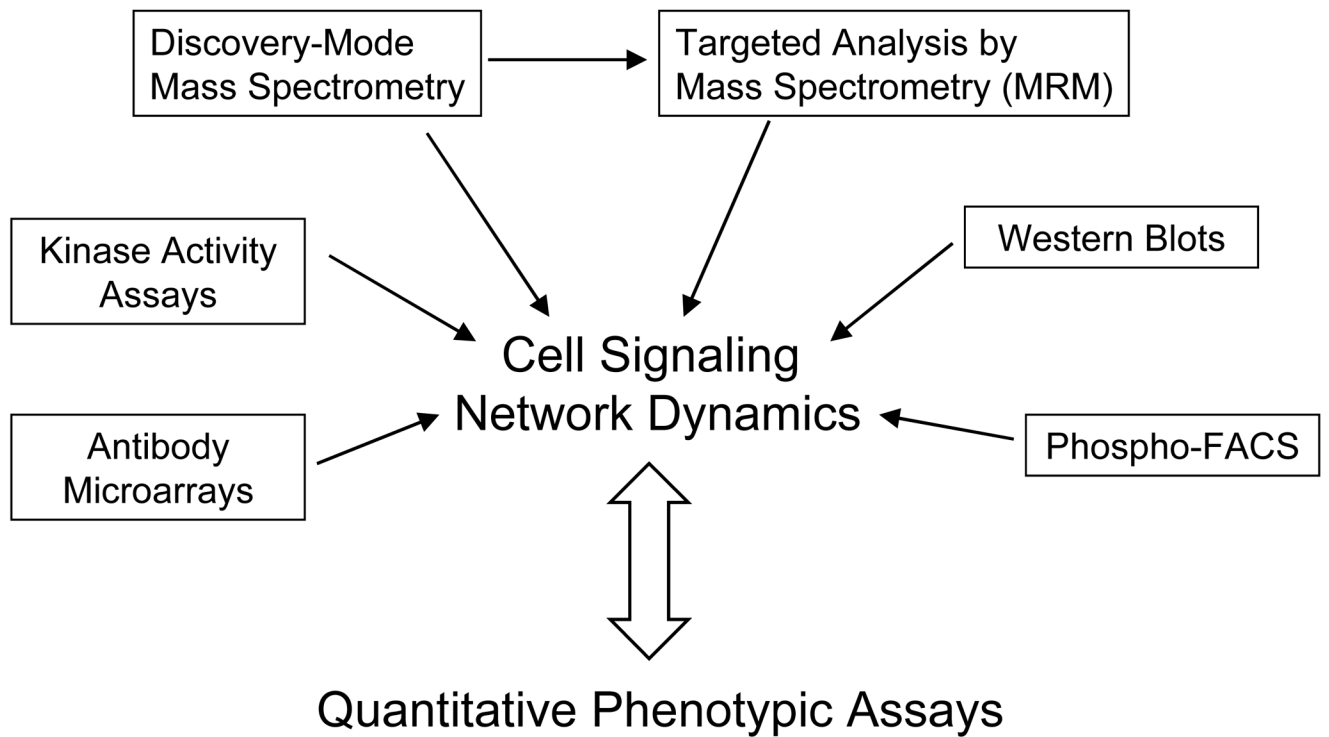
Figure 2.
By combining discovery-mode mass spectrometry with multiple targeted phosphoproteomic methods, it will be possible to obtain a much better systems-wide view of dynamic cellular signaling networks. Combining this information with quantitative phenotypic assays will provide biological insight into the key network components responsible for regulating selected biological responses to cellular perturbations.

Table 1

A comparison of the mass spectrometry-based phosphoproteomics methods that have been applied to quantify temporal dynamics in the ErbB signaling network.

Mass Spectrometry Method	# of Sites	# of MS Analyses	Spatial Information	Biological Insight
Protein pTyr Immunoprecipitation [6]	10's	2 per 5-point timecourse	No	Interacting proteins
Peptide pTyr Immunoprecipitation [7,8,20]	100's	1 per 4-point timecourse, multiple biological conditions	No	Site specificity, Correlation with phenotype
SCX-TiO2	1000's	>100 per 5 point timecourse	Yes cytosolic vs.nuclear	Site specificity, Spatial-temporal regulation